Mathematical Statistics and Stochastic Processes for IT Students

István Szalkai
SZALKAI@ALMOS.UNI-PANNON.HU
Pannon University, Veszprém, Hungary

11.11.2019.
Mathematical Statistics and Stochastic Processes for IT Students

István Szalkai
SZALKAI@ALMOS.UNI-PANNON.HU
Pannon University, Veszprém, Hungary

11.11.2019.
Contents

I Vector valued random variables

1 Two - dimensional random variables and independence
   1.1 General definitions
   1.2 The discrete case
   1.3 Summary and an example
   1.4 The continuous case
   1.5 Conditional probability

2 Higher dimensional random variables
   2.1 Covariance and independence
   2.2 The normal (Gauss-) distributions
      2.2.1 2-dimensional
      2.2.2 n-dimensional
   2.3 The binomial/multinomial (Bernoulli-) distributions
      2.3.1 1-dim = 2-dim
      2.3.2 n-dim (2 ≤ n)
   2.4 The poli-hypergeometric distributions
      2.4.1 1-dim = 2-dim
      2.4.2 n-dim (2 ≤ n)

II Mathematical Statistics

3 Elementary notions
CONTENTS

4 Confidence intervals .............................................. 31
4.1 Interval for the probability .................................. 31
4.2 Interval for the mean when \( \sigma \) is known ................. 33
4.3 Interval for the mean when \( \sigma \) is unknown ............... 34
4.4 Interval for the dispersion .................................... 35

5 Point estimations and hypothesis testing ........................ 37
5.1 General notions .................................................. 37
5.2 Parametric tests ................................................ 39
5.2.1 \( u \)-test for the mean of one sample when \( \sigma \) is known 39
5.2.2 \( t \)-test for the mean of one sample when \( \sigma \) is unknown 41
5.2.3 \( k \)-test for the dispersion of one sample ............... 42
5.2.4 \( u \)-test for the means of two samples .................. 43
5.2.5 \( t \)-test for the means of two samples when \( \sigma_1 = \sigma_2 \) 45
5.2.6 \( F \)-test for the dispersions of two samples .......... 46
5.3 Nonparametric tests ........................................... 47
5.3.1 Goodness of fit ............................................. 47
5.3.2 Homogeneity ............................................... 49
5.3.3 Independence ............................................. 50
5.3.4 Test for correlation ....................................... 51
5.3.5 Normality testing ......................................... 51

6 Regression and the least square method ......................... 55
6.1 The general case ................................................ 56
6.2 Linear regression .............................................. 57
6.3 Estimating the correlation coefficient ....................... 62
6.4 Regression and covariance ................................... 63
6.5 Nonlinear regressions - linearizing methods ............... 67
6.5.1 The Ruler Method ......................................... 68
6.5.2 Exponential regression ................................... 68
6.5.3 Logarithmic regression ................................... 70
6.5.4 Power regression ......................................... 71
6.5.5 Hyperbolic regression ................................... 72
6.5.6 Logit-probit regression .................................. 74
6.6 Nonlinear regressions - direct methods ..................... 76
6.6.1 Quadratic regression ...................................... 77

7 Mathematical background ......................................... 79
7.1 The Student- or \( t \)-distribution ............................... 80
7.2 The \( \chi^2 \) distribution ...................................... 82
III Stochastic Processes 85

8 Introduction 87
  8.1 Elementary notions ........................................ 87
  8.2 Examples ...................................................... 88
    8.2.1 The Brownian motion .................................... 88
    8.2.2 The Poisson process ..................................... 89

9 General stochastic processes 93
  9.1 The state space .............................................. 93
  9.2 The index (parameter-) set .................................. 94
  9.3 The mean-, dispersion- and autocovariance functions .... 94

10 Classical types of stochastic processes 95
  10.1 Processes with stationary independent increments ....... 95
  10.2 Martingales .................................................. 97
  10.3 Markov processes ............................................ 98
  10.4 Stationary processes ......................................... 106
  10.5 Renewal processes ........................................... 107
  10.6 Point processes .............................................. 108
  10.7 Moving average processes .................................. 109
  10.8 Autoregressive processes .................................... 111
  10.9 White noise processes ....................................... 113

References 115

Tables 117

Index 123
Introduction

Mathematical Statistics and Stochastic Processes became extremely important in modern engineering and computer technology. The present book is for engineers and IT experts, so it focuses on applications, illustrations and mainly on computing formulas, serving as few mathematics as necessary. For basic Probability Theory we refer to our short and illustrative summary [SzI1]. (Letters and numbers in square brackets [...] refer to further reading in the section "References".) Not only for curiosity we mention the Hungarian terms as well in brackets and in quotation marks ("...").

We highly acknowledge the funding of the grant EFOP-3.4.3.-A.2.3.

This book contains of 125 pages, 17 Figures and 5 Tables.
Preliminaries: different basic notations

Since many different notations are in use in Probability Theory, let us collect and identify them first. Through this book we also give the Hungarian terms as well in brackets and in quotation marks ("...").

□ = end of a definition / theorem / proof / remark,

[...] = literature reference (see last section),

\( A \cup B \) = disjoint union of sets, that is \( A \cap B = \emptyset \),

\( \mathbb{R}, \mathbb{N} \) = set of real and natural numbers,

\( \mathbb{R}^+, \mathbb{R}_{\geq 0} \) = set of nonnegative numbers,

\( \mathbf{a}, \mathbf{a}' \) = vectors,

\( \exp (x) = e^x \), \( \exp_a (x) = a^x \) are the exponential functions \((a > 0)\),

\( \text{lg}(x), \text{ln}(x), \log(x) \) and \( \log_a (x) \) are the logarithm functions of different bases (see the Remark below),

\( \Omega, T, H \) = sample set (in Hungarian: "eseménytér"),

\( P(A), \Pr(A) \) = the probability of \( A \subseteq \Omega \),

\( \xi, \zeta, X, Y : \Omega \to \mathbb{R} \) = random variables (1-dimensional or real valued or scalar, "valós vagy skalár értékű valószínűségi változó")

r.v. = random variable (v.v.)

\( \xi, \zeta, \mathbf{X}, \mathbf{Y} : \Omega \to \mathbb{R}^n \) = random variables (n-dimensional or vector valued, "többdimenziós vagy vektor értékű valószínűségi változó")

r.v.v. = random vector variable (v.v.v.)

\( f_{\xi}, f, G, H : \mathbb{R} \to \mathbb{R} \) = distribution functions ("eloszlásfüggvények"),

\( f_{\xi}, f, g, h : \mathbb{R} \to \mathbb{R} \) = density functions ("sűrűségfüggvények"),

\( f', \frac{df}{dx}, \frac{d}{dx} f \) = derivatives of \( f \),

\( M(\xi), E(\xi), E\{\xi\}, m_{\xi}, m, \mu(\xi) \) = mean of \( \xi \) = expected value ("átlag, várható érték"),

\( D(\xi), \sigma(\xi), \sigma_{\xi} \) = dispersion of \( \xi \) ("\( \xi \) szórása"),

\( D^2(\xi), \sigma^2(\xi), \sigma^2_{\xi}, \text{var}(\xi) \) = variance of \( \xi \) ("\( \xi \) szórásnégyzete"),

\( \xi^* := \frac{\xi - M(\xi)}{D(\xi)} \) is the standardized version of \( \xi \).
Remark 1 \( \ln(x) \), \( \log(x) \) usually denote the natural logarithm (base \( e \)) and \( \log_{10}(x) \) the \( \log_{10}(x) \), but different books, programs and users can use other choices, please check it in each situation. However, in most applications there is no substantial difference among different bases, since \( \log_b(x) = \log_b(a) \cdot \log_a(x) \) where \( \log_b(a) \) is a constant multiplier, i.e. the Reader may choose his/her favourite.
Part I

Vector valued random variables
We usually make two or more measurements at an experiment, so it is better to consider the r.v. vector of data $\mathbf{\zeta} = (\zeta_1, ..., \zeta_n)$ instead of a set or separate r.v. $\{\zeta_1, ..., \zeta_n\}$. 
Chapter 1

Two-dimensional random variables and independence

Definition I.1 \( \vec{\zeta} : \Omega \to \mathbb{R}^2 \) is a 2 dimensional r.v. or a vector-r.v.

Explanations: \( \vec{\zeta} = (\xi, \eta) = \left( \begin{array}{c} \xi \\ \eta \end{array} \right) \), i.e. \( \vec{\zeta} (\omega) = (\xi(\omega), \eta(\omega)) \) for \( \omega \in \Omega \), so \( \xi \) and \( \eta \) are the coordinate (function)s of \( \vec{\zeta} \).

In fact, \( \xi \) and \( \eta \) are any two r.v. as you like: \( \xi, \eta : \Omega \to \mathbb{R} \).

Sometimes \( \zeta \) or simply \( \zeta \) is written instead of \( \vec{\zeta} \), moreover the (worst) notation \( \zeta = (\zeta_1, \zeta_2) \) is often used.

1.1 General definitions

Definition I.2 The distribution function of \( \vec{\zeta} = (\xi, \eta) \), or the common / joint distr. func. of \( \xi \) and \( \eta \) ("együttes eloszlásfüggvény") is

\[
F_{\vec{\zeta}} : \mathbb{R}^2 \to \mathbb{R}, \quad F_{\vec{\zeta}} (x, y) := P (\xi < x, \eta < y).
\]

(1.1)

In what follows, we simply write \( \zeta \) and \( F_{\zeta} \) instead of \( \vec{\zeta} \) and \( F_{\vec{\zeta}} \).

Theorem I.3 \( F_{\xi} (x) = \lim_{y \to \infty} F_{\vec{\zeta}} (x, y) \) and \( F_{\eta} (y) = \lim_{x \to \infty} F_{\vec{\zeta}} (x, y) \) for any \( x, y \in \mathbb{R} \).

Definition I.4 By the theorem above \( \xi \) and \( \eta \) are called the marginal (or border) distributions of \( \vec{\zeta} \), ("határeloszlás" or "peremeloszlás").
Definition I.5 \( \xi \) and \( \eta \) are **independent** (of each-other) if
\[
\forall x, y \in \mathbb{R} \quad F_{\xi}(x, y) = F_{\xi}(x) \cdot F_{\eta}(y) .
\] (1.2)

(See also [Sz1], (1.10) and (1.15)-(1.17).)

For the following notions \( \xi \) and \( \eta \) do **not** need to have a common distribution function.

Definition I.6 The **covariance** (in Hungarian: "kovariencia") of \( \xi \) and \( \eta \) is:
\[
\text{cov} (\xi, \eta) := M( (\xi - m_{\xi}) \cdot (\eta - m_{\eta}) )
\] (1.3)
where \( m_{\xi} = M(\xi) \) and \( m_{\eta} = M(\eta) \), or, without abbreviations
\[
\text{cov} (\xi, \eta) := M( (\xi - M(\xi)) \cdot (\eta - M(\eta)) ) .
\]
cov\((\xi, \eta)\) is also denoted by \( \sigma_{\xi, \eta} \).

Remark I.7 "co-variance" literally means varying together ("együtt változás").
cov\((\xi, \eta)\) really detects the changing measure of \( \xi \) and \( \eta \). Look: \( \xi - M(\xi) \) and \( \eta - M(\eta) \) are the differences of \( \xi \) and \( \eta \) from their means (movements "up" or "down") in the same time, and (1.3) measures (in some way) the relation of these movements to a single real number.

Especially positive cov\((\xi, \eta)\) means that \( \xi > M(\xi) \) or \( \xi < M(\xi) \) occur "exactly when" \( \eta > M(\eta) \) or \( \eta < M(\eta) \), in one word "\( \xi \) and \( \eta \) move in the same direction" (concerning to their means), i.e. \( \xi \) and \( \eta \) help and strengthen each other. Similarly, negative cov\((\xi, \eta)\) means that \( \xi > M(\xi) \) or \( \xi < M(\xi) \) occur "exactly when not" \( \eta > M(\eta) \) or \( \eta < M(\eta) \), in one word "\( \xi \) and \( \eta \) move in other directions", i.e. \( \xi \) and \( \eta \) impede or weaken each other.

Let us highlight again that the above implications are "not sure" (as in mathematics usually), only "with some probability" (as in mathematical statistics, as usual), or less: concerning the mean (average) of the formulae!
(See also the below theorems and remarks.)

Theorem I.8 For any r.v. \( \xi, \eta \) and \( a, b, c, d \in \mathbb{R} \) real numbers (constant r.v.) we have
\[
(o) \quad \text{cov} (\xi, \eta) = M(\xi \cdot \eta) - M(\xi) \cdot M(\eta) ,
\]
\[
(i) \quad \text{if } M(\xi) = 0 \text{ then } \text{cov} (\xi, \eta) = M(\xi \cdot \eta) ,
\]
\[
(ii) \quad \text{if } \xi \text{ and } \eta \text{ are independent, then } \text{cov} (\xi, \eta) = 0 .
\]
1.1. GENERAL DEFINITIONS

(iii) but the reverse implication is not true in general, however it is true for normal distributions,
(iv) \( D^2 (\xi + \eta) = D^2 (\xi) + D^2 (\eta) + 2 \cdot \text{cov} (\xi, \eta) \) for any two r.v. \( \xi \) and \( \eta \),
(v) \( \text{cov} (\xi, \xi) = D^2 (\xi) \) (auto/self covariance, "saját/ön- kovariancia"),
(vi) \( \text{cov} (\xi, \eta) = \text{cov} (\eta, \xi) \) (symmetry, "szimmetrikusság"),
(vii) \( \text{cov} (a \xi + b, c \eta + d) = ac \cdot \text{cov} (\xi, \eta) \),
(viii) \( \text{cov} (\xi, \eta) = \text{cov} (\xi - M (\xi), \eta - M (\eta)) \),
(ix) \( \text{cov} (a \xi + b, a \xi + b) = a^2 D^2 (\xi) \),
(x) \( \text{cov} (a, \eta) = 0 \),
(xi) \( \text{cov} (a_1 \xi_1 + a_2 \xi_2, b_1 \eta_1 + b_2 \eta_2) = \\
= a_1 b_1 \text{cov} (\xi_1, \eta_1) + a_1 b_2 \text{cov} (\xi_1, \eta_2) + a_2 b_1 \text{cov} (\xi_2, \eta_1) + a_2 b_2 \text{cov} (\xi_2, \eta_2) . \)

Proof. (o) by definition \( \text{cov} (\xi, \eta) = \\
= M ((\xi - m_\xi) \cdot (\eta - m_\eta)) = M (\xi \eta) - M (\xi m_\eta) - M (\eta m_\xi) + M (m_\xi m_\eta) \\
= M (\xi \eta) - m_\eta \cdot M (\xi) - m_\xi \cdot M (\eta) + m_\xi m_\eta \\
= M (\xi \eta) - m_\eta \cdot m_\xi - m_\xi \cdot m_\eta + m_\xi m_\eta \\
= M (\xi \eta) - m_\xi m_\eta = M (\xi \cdot \eta) - M (\xi) \cdot M (\eta) . \)

(i) follows from (o).
(ii) if \( \xi \) and \( \eta \) are independent then \( M (\xi \cdot \eta) = M (\xi) \cdot M (\eta) \) (see [SzI1]).
(iii) we do not prove it here.
(iv) \( D^2 (\xi + \eta) = M ([\xi + \eta - m_\xi - m_\eta]^2) = \\
= M ([\xi - m_\xi]^2) + M ([\eta - m_\eta]^2) + 2 \cdot M ((\xi - m_\xi) \cdot (\eta - m_\eta)) \\
= D^2 (\xi) + D^2 (\eta) + 2 \cdot \text{cov} (\xi, \eta) . \)
(v) by definition \( \text{cov} (\xi, \xi) := M ((\xi - m_\xi)^2) = D^2 (\xi) . \)
(vi) obvious.
(vii) since \( a \xi + b - M (a \xi + b) = a (\xi - M (\xi)) \)
and \( c \eta + d - M (c \eta + d) = c (\eta - M (\eta)) \),
we have
CHAPTER 1. TWO-DIMENSIONAL RANDOM VARIABLES AND INDEPENDENCE

\[
\text{cov}(a\xi + b, c\eta + d) = M\left(ac\,(\xi - m_\xi)\,(\eta - m_\eta)\right) \\
= ac \cdot M\left((\xi - m_\xi)\,(\eta - m_\eta)\right) = ac \cdot \text{cov}(\xi, \eta).
\]

(viii) take \(a = c = 1\), \(b = -M(\xi)\) and \(d = -M(\eta)\) in (vii).
(ix) use (vii), with \(a = c\) and \(b = d\), and (v).
(x) by (o) \(\text{cov}(a, \eta) = M(a \cdot \eta) - M(a) \cdot M(\eta) = a \cdot M(\eta) - a \cdot M(\eta) = 0\).

Remark I.9 (o) Clearly \((\xi \cdot \eta)(\omega) = \xi(\omega) \cdot \eta(\omega)\) for \(\omega \in \Omega\).
(ii) and (iii) say that calculating \(\text{cov}(\xi, \eta)\) can not decide the independence of \(\xi\) and \(\eta\), in the case \(\text{cov}(\xi, \eta) = 0\) we can only say that \(\xi\) and \(\eta\) are uncorrelated ("korrelálatlanok"). See Example I.10 below for details and examples.
(iv) is the generalization of the "Pithagorean Theorem"
\[
D^2(\xi + \eta) = D^2(\xi) + D^2(\eta)
\]
for independent r.v. \(\xi, \eta\), since (iv) is valid for any r.v. \(\xi\) and \(\eta\) (see also [SzI1]).
(vii) Clearly \(\text{cov}(\xi, \eta)\) changes when we change measure units (cm or km), since such a change zooms (in or out) the fluctuations of \(\xi\) and \(\eta\). For this reason \(\text{cov}(\xi, \eta)\) differs from \(\text{cov}(\xi^*, \eta^*)\) where \(\xi^* = \frac{\xi - M(\xi)}{D(\xi)}\) and \(\eta^* = \frac{\eta - M(\eta)}{D(\eta)}\) are the standard versions of \(\xi\) and \(\eta\). This phenomenon is called "\(\text{cov}(\xi, \eta)\) is not normed" or "depends upon the scales" ("skálafüggő"). The normed version of \(\text{cov}(\xi, \eta)\) is the correlation coefficient (see below).
(viii) must be clear by everyday thinking: the covariance ("varying together") must not depend on "where is the zero on our scale" (e.g. measuring temperature in centigrade or Kelvin). See also Remark II.7 at the beginning of Part Statistics.
(x) is also clear: neither a constant \(a\) "varies together" with \(\xi\), nor \(\xi\) with \(a\).

Example I.10 Here we give some examples for r.v. which are uncorrelated but not independent.

First example: Let \(\xi\) be a uniform (continuous) r.v. on the interval \([-1,1]\) and let \(\eta = \xi^2\), clearly \(\xi\) and \(\eta\) are not independent (please check). However, by (o) \(\text{cov}(\xi, \eta) = M(\xi \cdot \xi^2) - M(\xi) \cdot M(\xi^2) = M(\xi^3) - M(\xi) \cdot M(\xi^2) = 0 - 0 = 0\)
since \(M(\xi^3) = M(\xi) = 0\). Similarly \(\text{cov}(\xi, \xi^2) = 0\) for any r.v. symmetric to the origin (i.e. \(M(\xi) = 0\)).

Second example: let \(X\) and \(Y\) be discrete finite r.v. such that \(\text{Im}(X) = \{0, 2\}, \text{Im}(Y) = \{0, 1, 2\}\), \(P(X = 0, Y = 1) = \frac{1}{2}, P(X = 2, Y = 0) = P(X = 2, Y = 2) = \frac{1}{4}\) and the other possibilities are zero:
1.1. GENERAL DEFINITIONS

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1/4</td>
<td>1/2</td>
<td>1/4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
<td>1</td>
</tr>
<tr>
<td>Σ</td>
<td>1/4</td>
<td>1/2</td>
<td>1/4</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

So \( P(X = 0) = P(X = 2) = \frac{1}{2}, \) \( P(Y = 0) = P(Y = 2) = \frac{1}{2} \) and \( P(Y = 1) = \frac{1}{2} \).

Further \( M(X) = M(Y) = 1 \) and \( M(X \cdot Y) = 0 + 0 + 2 \cdot \frac{1}{2} = 1 \) so \( \text{cov}(X, Y) = 0 \), i.e. \( X \) and \( Y \) are uncorrelated. On the other hand \( X \) and \( Y \) are not independent, since

\[
P(X = 0, Y = 1) = \frac{1}{2} \neq P(X = 0) \cdot P(Y = 1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.
\]

(There are many similar examples, e.g. if \((X, Y)\) has the values \((-1,0), (0,1), (1,0), (0,-1)\) with probabilities \(1/4\).)

**Third example:** Let \( \xi = X + Y \) and \( \eta = X - Y \) where \( X \) and \( Y \) are independent Bernoulli (discrete) r.v. with the same parameter \( p \).

\( \xi \) and \( \eta \) are uncorrelated since

\[
\text{cov}(\xi, \eta) = \text{cov}(X + Y, X - Y) = \text{cov}(X, X) - \text{cov}(X, Y) + \text{cov}(Y, X) - \text{cov}(Y, Y)
\]

\[= D^2(X) - D^2(Y) = 0.\]

However \( \xi \) and \( \eta \) are not independent since, for e.g.

\[
P(\xi = 0, \eta = 1) = P(X + Y = 0, X - Y = 1) = 0
\]

(the only solution \( X = \frac{1}{2} \) and \( Y = -\frac{1}{2} \) are impossible), while

\[
P(\xi = 0) \cdot P(\eta = 1) = P(X + Y = 0) \cdot P(X - Y = 1) = p \cdot (1 - p)^3. \quad \square
\]


**Remark I.11** The main disadvantage of \( \text{cov} \) is property (viii): depends on the scales (measure units) \( a \) and \( c \) of \( \xi \) and \( \eta \). The modification \((1.4)\) below handles this problem: \( R(a \xi + b, c \eta + d) = R(\xi, \eta) \).

**Definition I.12** The **Pearson correlation coefficient** or **normed covariance** ("korrelációs együttható, normált kovariancia") is

\[
R(\xi, \eta) := \frac{\text{cov}(\xi, \eta)}{D(\xi) \cdot D(\eta)}.
\]

Other notations are \( r(\xi, \eta) \) and \( \rho \{\xi, \eta\} \). \quad \square
CHAPTER 1. TWO-DIMENSIONAL RANDOM VARIABLES AND INDEPENDENCE

Remark I.13 (i) "co-relation" literary means (common) relation between two objects ("összefüggés").
(ii) This version of the correlation coefficient is named after **Pearson**\(^1\).

Theorem I.14 (i) \(-1 \leq R(\xi, \eta) \leq +1\),
(ii) if \(\xi\) and \(\eta\) are independent (or uncorrelated) then \(R(\xi, \eta) = 0\),
(iii) but the reverse implication is not true (see Theorem I.8),
(iv) for Gaussian distributions:
\(\xi\) and \(\eta\) are independent \(\iff\) \(R(\xi, \eta) = 0\),
(v) \(|R(\xi, \eta)| = 1\) if and only if \(\xi\) and \(\eta\) are "the same":
\[
\eta = a \cdot \xi + b \quad \text{for some } a, b \in \mathbb{R}, a \neq 0 .
\]

Proof. (i) can be deduced from the **Cauchy-Schwarz-Bunyakovszkij** (CSB) inequality\(^2\).
(ii)-(iv) follow from the corresponding parts of Theorem I.8.
(v) For the backward direction let \(\eta = a \xi + b\). Now, by
\[
m_\eta = M(\eta) = M(a \xi + b) = a M(\xi) + b = am_\xi + b
\]
and the definition the enumerator is
\[
cov(\xi, \eta) = M((a \xi - m_\xi) (\eta - m_\eta)) = M((\xi - m_\xi)(a \xi + b - (am_\xi + b)))
\]
\[
= M((\xi - m_\xi)(a (\xi - m_\xi))) = M(a (\xi - m_\xi)^2) = a \cdot D^2(\xi)
\]
and using
\[
D(\eta) = D(a \xi + b) = |a| \cdot D(\xi)
\]

---

\(^1\) Karl Pearson (1857-1936) an English mathematician and bio-statistician.
\(^2\) The **Cauchy - Schwarz - Bunyakovszkij** (CSB) inequality has (at least) three different forms:
- (C) \(\left( \sum_{i=1}^{n} x_i y_i \right)^2 \leq \left( \sum_{i=1}^{n} x_i^2 \right) \cdot \left( \sum_{i=1}^{n} y_i^2 \right)\) for any \(x_1, y_1, ..., x_n, y_n \in \mathbb{R}\) real numbers and \(n \in \mathbb{N}\),
- (C) \(\sum_{i=1}^{\infty} x_i y_i \leq \left( \sum_{i=1}^{\infty} x_i^2 \right) \cdot \left( \sum_{i=1}^{\infty} y_i^2 \right)\) for any \(x_1, y_1, ..., x_n, y_n, ... \in \mathbb{R}\) sequences,
if the sums are finite,
- (BS) \(\left( \int_{a}^{b} f(x) g(x) \, dx \right)^2 \leq \left( \int_{a}^{b} f^2(x) \, dx \right) \cdot \left( \int_{a}^{b} g^2(x) \, dx \right)\) for any functions \(f, g : \mathbb{R} \rightarrow \mathbb{R}\), if the integrals are finite.
In general: \(\langle x, y \rangle^2 \leq \langle x, x \rangle \cdot \langle y, y \rangle\) for any scalar product \(\langle ., . \rangle\). □
1.2. THE DISCRETE CASE

we have \( R(\xi, \eta) = a \cdot D^2(\xi) / |a| \cdot D(\xi) = \pm 1 \).

The other direction is more difficult.

\[ \text{Remark I.15} \]

The main significance of (i) are the limits (bounds) of \( R \), we can estimate and compare the magnitude of \( R \) to the absolute limits. Though the conclusions like "\( R = 0.5 \) means 50% connection between \( \xi \) and \( \eta \)" has no mathematical background or meaning, we feel and say similar sentences.

\[ \text{Remark I.16} \]

However, the cases \( R(\xi, \eta) = \pm 1 \) really mean strict connections: using connection (1.5) we can compute exactly the values of \( \eta \) from \( \xi \) (and back, of \( \xi \) from \( \eta \)) since \( a, b \in \mathbb{R} \) are (fixed) real numbers! We can think that the measuring quantities (devices) are really joined firmly, only the scales are changed (linear transformation), like Celsius and Fahrenheit: \( Y[^\circ F] = 1.8 \cdot X[^\circ C] + 32 \) and \( X[^\circ C] = \frac{1}{1.8} Y[^\circ F] - \frac{32}{1.8} \approx 0.5556 \cdot Y[^\circ F] - 17.7778 \).

The quantities \( \text{cov}(\xi, \eta) \) and \( R(\xi, \eta) \) have many applications in Regression theory in Statistics. More detailed investigation can be found in Section 6.4 "Regression and covariance".

See also Remark II.103 after Theorem II.102.

1.2 The discrete case

\[ \text{Definition I.17} \]

If \( \text{Im}(\xi) = \{x_1, x_2, \ldots, x_n, \ldots\} \) and \( \text{Im}(\eta) = \{y_1, y_2, \ldots, y_m, \ldots\} \) then the distribution of \( \vec{\zeta} = (\xi, \eta) \) (or: the common/joint distribution of \( \xi \) and \( \eta \)) is the set of probabilities: \( \{p_{i,j} : 1 \leq i, j \leq \infty\} \) where

\[ p_{i,j} := P(\xi = x_i, \eta = y_j) . \quad (1.6) \]

Clearly

\[ 0 \leq p_{i,j} \leq 1 \quad \text{and} \quad \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{i,j} = 1 . \quad (1.7) \]

(Any set of real numbers, satisfying (1.7) can be a joint discrete distribution.)

\[ \text{Definition I.18} \]

\[ q_i^{(\xi)} := \sum_{j=1}^{\infty} p_{i,j} = P(\xi = x_i) \quad \text{and} \quad q_j^{(\eta)} := \sum_{i=1}^{\infty} p_{i,j} = P(\eta = y_j) \quad (1.8) \]

are the marginal (or border) distributions ("peremeloszlások") of \( \vec{\zeta} \). \( \square \)
CHAPTER 1. TWO-DIMENSIONAL RANDOM VARIABLES AND INDEPENDENCE

Theorem I.19 In fact, the sets of probabilities
\[ \left\{ q_i^{(\xi)} : 1 \leq i \leq \infty \right\} \quad \text{and} \quad \left\{ q_j^{(\eta)} : 1 \leq j \leq \infty \right\} \] (1.9)
are the distributions of \( \xi \) and \( \eta \).

Proof of Theorem I.19

Theorem I.20 The discrete r.v. \( \xi \) and \( \eta \) are independent if and only if for every \( i, j \in \mathbb{N} \) we have
\[ P (\xi = x_i, \eta = y_j) = P (\xi = x_i) \cdot P (\eta = y_j) \] (1.10)
i.e. \( p_{i,j} = q_i^{(\xi)} \cdot q_j^{(\eta)} \).

(See also [SzI1], (1.2) and (1.15)-(1.17).)

Remark I.21 In other words: (1.2) and (1.10) are equivalent.

Theorem I.22
\[ F_\xi (x, y) = \sum_{x_i < x} \sum_{y_j < y} p_{i,j} \quad \text{for any } x, y \in \mathbb{R} , \]
\[ F_\xi (x) = \sum_{x_i < x} q_i^{(\xi)} \quad \text{and} \quad F_\eta (y) = \sum_{y_j < y} q_j^{(\eta)} . \]

Theorem I.23
\[ M (\xi \cdot \eta) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{i,j} \cdot x_i \cdot y_j , \]
\[ M (\xi) = \sum_{i=1}^{\infty} q_i^{(\xi)} \cdot x_i \quad \text{and} \quad M (\eta) = \sum_{j=1}^{\infty} q_j^{(\eta)} \cdot y_j . \]

1.3 Summary and an example

In case \( \text{Im}(\xi) \) and \( \text{Im}(\eta) \) are finite, then we can arrange all the data in a table as seen below.

<table>
<thead>
<tr>
<th>( \xi ) ( \eta )</th>
<th>( y_1 )</th>
<th>( y_2 )</th>
<th>...</th>
<th>( y_j )</th>
<th>...</th>
<th>( y_m )</th>
<th>( \xi_{\text{marg}} )</th>
<th>( \eta_{\text{marg}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>( p_{1,1} )</td>
<td>( p_{1,2} )</td>
<td>...</td>
<td>( p_{1,j} )</td>
<td>...</td>
<td>( p_{1,m} )</td>
<td>( q_1^{(\xi)} )</td>
<td>( q_1^{(\eta)} )</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>( p_{2,1} )</td>
<td>( p_{2,2} )</td>
<td>...</td>
<td>( p_{2,j} )</td>
<td>...</td>
<td>( p_{2,m} )</td>
<td>( q_2^{(\xi)} )</td>
<td>( q_2^{(\eta)} )</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( x_i )</td>
<td>( p_{i,1} )</td>
<td>( p_{i,2} )</td>
<td>...</td>
<td>( p_{i,j} )</td>
<td>...</td>
<td>( p_{i,m} )</td>
<td>( q_i^{(\xi)} )</td>
<td>( q_i^{(\eta)} )</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( x_n )</td>
<td>( p_{n,1} )</td>
<td>( p_{n,2} )</td>
<td>...</td>
<td>( p_{n,j} )</td>
<td>...</td>
<td>( p_{n,m} )</td>
<td>( q_n^{(\xi)} )</td>
<td>( q_n^{(\eta)} )</td>
</tr>
<tr>
<td>( \eta_{\text{marg}} )</td>
<td>( q_1^{(\eta)} )</td>
<td>( q_2^{(\eta)} )</td>
<td>...</td>
<td>( q_j^{(\eta)} )</td>
<td>...</td>
<td>( q_n^{(\eta)} )</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Two-dimensional finite discrete distribution
As in the previous section, \( \{x_1, x_2, \ldots, x_n\} \) and \( \{y_1, y_2, \ldots, y_m\} \) are the values of \( \xi \) and \( \eta \). The joint distribution of \( \xi \) and \( \eta \) can be seen in the middle of the table: \( p_{i,j} \) was defined in (1.6). The marginal distributions are in the margins of the table: \( q_i^{(\xi)} \) is the sum of the \( i \)-th row, and \( q_j^{(\eta)} \) is the sum of the \( i \)-th column of the table, according to (1.8). Only the middle of the table (the set \( \{p_{i,j}\} \)) is usually given, we ourselves have to compute \( q_i^{(\xi)} \) and \( q_j^{(\eta)} \) by summarizing the rows and columns. For checking, the sums of both marginal distributions (the last row and the last column) must give 1, see the right bottom entry.

Independence can be checked by (1.10): each \( p_{i,j} \) must be equal to the product of (the corresponding) \( q_i^{(\xi)} \) and \( q_j^{(\eta)} \) (in the same row and column). Observe, that if (at least) one \( p_{i,j} \) does not fulfill this equality, \( \xi \) and \( \eta \) are not independent.

Considering only the first and last column/row, we can find the distributions of the (one variable) r.v. \( \xi/\eta \) respectively, i.e. not considering the other, so \( M(\xi) \), \( M(\eta) \), \( D(\xi) \) and \( D(\eta) \) can be computed easily from these columns/rows, as in ordinary (one dimensional) probability theory, or see the second line of Theorem [I.23].

The mean \( M(\xi \cdot \eta) \) can be computed also by Theorem [I.23], the picked \( p_{i,j} \) must be multiplied by \( x_i \) and \( y_j \) (in the same row and column) and summed for all \( p_{i,j} \). Finally use the formulae \( \text{cov}(\xi, \eta) = M(\xi \cdot \eta) - M(\xi) \cdot M(\eta) \) and \( R(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{D(\xi) \cdot D(\eta)} \).

**Example I.24** The price \( (X) \) and quality \( (Y) \) were investigated for a certain product, the numbers in the table show how many products were found for each category in a shop\(^\text{3)}\). Calculate \( \text{cov}(X,Y) \), \( R(X,Y) \) and estimate the measure of dependence of \( X \) and \( Y \).

<table>
<thead>
<tr>
<th>( X ) ( \backslash ) ( Y )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>41</td>
<td>53</td>
<td>72</td>
<td>33</td>
</tr>
<tr>
<td>30</td>
<td>12</td>
<td>10</td>
<td>11</td>
<td>18</td>
</tr>
</tbody>
</table>

**Solution I.25** The given dataset contains the number of products in each category, not probabilities. So, we have to calculate relative frequencies for approximating the probabilities. The sum is \( 2+6+6+4+41+\ldots+12+10+11+18=268 \), so the common- and the marginal distributions are the following:

\(^{3)}\) \( \xi \) and \( \eta \) were replaced to \( X \) and \( Y \) for technical reasons only.
CHAPTER 1. TWO-DIMENSIONAL RANDOM VARIABLES AND INDEPENDENCE

<table>
<thead>
<tr>
<th>X  \ Y</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>( q_i^{(X)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2/268</td>
<td>6/268</td>
<td>6/268</td>
<td>4/268</td>
<td>18/268</td>
</tr>
<tr>
<td>20</td>
<td>41/268</td>
<td>53/268</td>
<td>72/268</td>
<td>33/268</td>
<td>199/268</td>
</tr>
<tr>
<td>30</td>
<td>12/268</td>
<td>10/268</td>
<td>11/268</td>
<td>18/268</td>
<td>51/268</td>
</tr>
<tr>
<td>( q_i^{(Y)} )</td>
<td>55/268</td>
<td>69/268</td>
<td>89/268</td>
<td>55/268</td>
<td>268/268</td>
</tr>
</tbody>
</table>

Independence checking, e.g. 2. row 4. column: \( 199/268 \times 55/268 \neq 33/268 \)
so \( X \) and \( Y \) are not independent.

Means (expected values):

\[
\]

\[
M(Y) = 1*(55/268) + 2*(69/268) + 3*(89/268) + 4*(55/268) = 680/268 \approx 2.5373,
\]

\[
M(X) = 10*(18/268) + 20*(199/268) + 30*(51/268) = 5690/268 \approx 21.2313,
\]

\[
cov(X,Y) = M(XY) - M(X)M(Y) = 14120/268 \approx 0.1966.
\]

Since \( \text{cov}(X,Y) > 0 \), \( X \) and \( Y \) strengthen each other, the move "in the same" direction.

Dispersions and \( R(X,Y) \):

\[
M(Y^2) = (1^2)*(55/268) + (2^2)*(69/268) + (3^2)*(89/268) + (4^2)*(55/268) = 2012/268 \approx 7.5075,
\]

\[
M(X^2) = (10^2)*(18/268) + (20^2)*(199/268) + (30^2)*(51/268) \approx 475.0000,
\]

\[
D(Y) = \sqrt{M(Y^2) - M^2(Y)} = \sqrt{7.5075 - 2.5373^2} \approx 1.0342,
\]

\[
D(X) = \sqrt{M(X^2) - M^2(X)} = \sqrt{475.0000 - 21.2313^2} \approx 4.9224,
\]

\[
R(X,Y) = \frac{\text{cov}(X,Y)}{D(Y)D(X)} = \frac{0.1966}{1.0342 \times 4.9224} \approx 0.0386.
\]

Since \( R(X,Y) \) is small \((\approx 4\%)\), the connections between \( X \) and \( Y \) is weak.

End of the solution.
1.4 The continuous case

It is very similar to the discrete case.

Definition I.26 The density function of $\zeta$ is the common/joint density function of $(\xi, \eta)$, i.e. the function $h : \mathbb{R}^2 \rightarrow \mathbb{R}_+^*$ such that for any $a, b, c, d \in \mathbb{R} \cup \{-\infty, +\infty\}$, $a \leq b$ and $c \leq d$ we have

$$P(a \leq \xi \leq b, \ c \leq \eta \leq d) = \int_{a}^{b} \int_{c}^{d} h(x, y) \, dy \, dx.$$ (1.11)
**Remark I.27** Any function \( h : \mathbb{R}^2 \to \mathbb{R} \) is suitable if \( 0 \leq h(x,y) \) and

\[
\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y) \, dy \, dx = 1 . \tag{1.12}
\]

Clearly

\[
\int_{-\infty}^{+\infty} h(x,y) \, dy = f_\xi(x) \quad \text{and} \quad \int_{-\infty}^{+\infty} h(x,y) \, dx = f_\eta(y) \tag{1.13}
\]

are the marginal density functions of \( \xi \) and \( \eta \). Further (by (1.11))

\[
F_\xi(b,d) = \int_{-\infty}^{b} \int_{-\infty}^{d} h(x,y) \, dy \, dx . \tag{1.14}
\]

**Theorem I.28** The continuous r.v. \( \xi \) and \( \eta \) are independent if and only if for every \( x, y \in \mathbb{R} \) we have

\[
h(x,y) = f_\xi(x) \cdot f_\eta(y) , \tag{1.15}
\]

and, if and only if for any \( a, b, c, d \in \mathbb{R} \cup \{-\infty, +\infty\} \), \( a \leq b \) and \( c \leq d \) we have

\[
P(a \leq \xi \leq b, c \leq \eta \leq d) = P(a \leq \xi \leq b) \cdot P(c \leq \eta \leq d) \tag{1.16}
\]

i.e.

\[
\int_{a}^{b} \int_{c}^{d} h(x,y) \, dy \, dx = \left( \int_{a}^{b} f_\xi(x) \, dx \right) \cdot \left( \int_{c}^{d} f_\eta(y) \, dy \right) . \tag{1.17}
\]

(See also [SzI1], (1.2) and (1.10).)

**Theorem I.29** \( M(\xi, \eta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \cdot y \cdot h(x,y) \, dy \, dx \),

\[
M(\xi) = \int_{-\infty}^{+\infty} x \cdot f_\xi(x) \, dx \quad \text{and} \quad M(\eta) = \int_{-\infty}^{+\infty} y \cdot f_\eta(y) \, dy . \tag{1.18}
\]
1.5 Conditional probability

Considering two (dimensional) r.v. questions like \( P(\xi = x \mid \eta = y) \), \( P(\xi < x \mid \eta < y) \) naturally occur. By elementary probability theory we clearly have

\[
P(\xi = x \mid \eta = y) = \frac{P(\xi = x \land \eta = y)}{P(\eta = y)}, \quad (1.18)
\]

\[
P(\eta = y \mid \xi = x) = \frac{P(\xi = x \land \eta = y)}{P(\xi = x)} \quad (1.19)
\]

and

\[
P(\xi < x \mid \eta < y) = \frac{P(\xi < x \land \eta < y)}{P(\eta < y)}. \quad (1.20)
\]

Using the notations of the previous sections we can write for discrete r.v.

\[
P(\xi = x_i \mid \eta = y_j) = \frac{p_{i,j}}{q_j}, \quad P(\eta = y_j \mid \xi = x_i) = \frac{p_{i,j}}{q_i}. \quad (1.21)
\]

\[
P(\xi = x_i \mid \eta \leq y_j) = \frac{\sum_{\ell=1}^{j} p_{i,\ell}}{\sum_{\ell=1}^{j} q^{(\ell)}}, \quad \text{and} \quad P(\xi \leq x_i \mid \eta \leq y_j) = \frac{\sum_{\ell=1}^{i} \sum_{j=1}^{j} p_{s,\ell}}{\sum_{\ell=1}^{j} q^{(\ell)}}, \quad (1.22)
\]

for continuous r.v.

\[
P(\xi < b \mid \eta < d) = \frac{\int_{-\infty}^{b} \int_{-\infty}^{d} h(x,y) \, dy \, dx}{\int_{-\infty}^{+\infty} \int_{-\infty}^{d} h(x,y) \, dy \, dx}. \quad (1.23)
\]

**Definition I.30** The conditional distribution functions (clearly) are

\[
F_{\xi}(x|y) = P(\xi < x \mid \eta = y) \quad \text{and} \quad F_{\eta}(y|x) = P(\eta < y \mid \xi = x). \quad (1.24)
\]

For continuous r.v. the conditional density functions are

\[
f_{\xi}(x|y) = \frac{h(x,y)}{f_{\eta}(y)} \quad \text{and} \quad f_{\eta}(y|x) = \frac{h(x,y)}{f_{\xi}(x)} \quad (1.25)
\]

for the conditions "\( \eta = y \)" and "\( \xi = x \)", respectively. \( \square \)
Chapter 1. Two-Dimensional Random Variables and Independence

Theorem I.31 For continuous r.v.

\[ f_{\xi}(x|y) = \frac{\partial F_{\xi}(x|y)}{\partial x} \quad \text{and} \quad f_{\eta}(y|x) = \frac{\partial F_{\eta}(y|x)}{\partial y} , \quad (1.26) \]

further

\[ F_{\xi}(x|y) = \frac{1}{f_{\eta}(y)} \cdot \frac{\partial H(x,y)}{\partial y} \quad \text{and} \quad F_{\eta}(y|x) = \frac{1}{f_{\xi}(x)} \cdot \frac{\partial H(x,y)}{\partial x} . \quad (1.27) \]

Definition I.32 The conditional means (of \( \xi \), assuming \( \eta = y \), and of \( \eta \)
assuming \( \xi = x \)) are, for discrete r.v.:

\[ M(\xi | \eta = y_j) = \sum_{i=1}^{\infty} x_i \cdot P(\xi = x_i | \eta = y_j) = \frac{1}{q_j(x)} \sum_{i=1}^{\infty} x_i \cdot p_{i,j} \quad (1.28) \]

and

\[ M(\eta | \xi = x_i) = \sum_{j=1}^{\infty} y_j \cdot P(\eta = y_j | \xi = x_i) = \frac{1}{q_i(\xi)} \sum_{j=1}^{\infty} y_j \cdot p_{i,j} , \quad (1.29) \]

for continuous r.v.:

\[ M(\xi | \eta = y) = \int_{-\infty}^{+\infty} x \cdot f(x|y) \, dx \ , \]

\[ M(\eta | \xi = x) = \int_{-\infty}^{+\infty} y \cdot g(y|x) \, dy \ , \quad (1.30) \]

which can also be written as

\[ M(\xi | \eta = y) = \frac{1}{f_{\eta}(y)} \cdot \int_{-\infty}^{+\infty} x \cdot h(x,y) \, dx \quad (1.31) \]

and

\[ M(\eta | \xi = x) = \frac{1}{f_{\xi}(x)} \cdot \int_{-\infty}^{+\infty} y \cdot h(x,y) \, dy . \quad (1.32) \]
Chapter 2

Higher dimensional random variables

In practice, a random variable is a physical (or other) quantity we measure during our experiment. However, in most cases, more than one quantity are measured for one experiment. Further, the connection among these quantities, in general, is not known (complicated, or even, the connection itself we want to reveal), so we must consider these quantities to be distinct random variables, and investigate the connection among them later.

2.1 Covariance and independence

Definition I.33 \( \vec{\xi} : \Omega \rightarrow \mathbb{R}^n \) is an \( n \)-dimensional r.v. or a vector-r.v.

Explanations: \( \vec{\xi} = (\xi_1, \xi_2, ..., \xi_n) = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} \), i.e. \( \vec{\xi} (\omega) = (\xi_1 (\omega), ..., \xi_n (\omega)) \) for \( \omega \in \Omega \), so \( \xi_1, ..., \xi_n \) are the coordinate (function)s of \( \vec{\xi} \).

In fact, \( \xi_1, ..., \xi_n \) are any \( n \) r.v. as you like.

Sometimes \( \xi \) or simply \( \xi \) is written instead of \( \vec{\xi} \), moreover the (worst) notation \( \xi = (\xi_1, ..., \xi_n) \) is often used.

The dimension \( n \) can also be denoted by \( \mu \) and by any other letter.

Definition I.34 \( M \left( \vec{\xi} \right) := (M (\xi_1), ..., M (\xi_n)) \in \mathbb{R}^n \) is an \( n \)-dimensional vector.
**Definition I.35** For $\xi : \Omega \rightarrow \mathbb{R}^n$ and $\eta : \Omega \rightarrow \mathbb{R}^m$ the **covariance matrix** ("kovariencia mátrix") is

$$\text{cov} \left( \xi, \eta \right) := \left[ \text{cov} \left( \xi_i, \eta_j \right) \right] \in \mathbb{R}^{n \times m}. \quad (2.1)$$

In case $\xi = \eta$ the matrix $C = \text{cov} \left( \xi, \xi \right)$ is called **auto/self covariance matrix** ("auto/saját- kovariencia mátrix").

**Theorem I.36** If the elements of $C$ (auto cov.matrix) are denoted by $c_{i,j}$, then

(i) \( c_{i,j} = c_{j,i} \), that is $C$ is symmetric,

(ii) \( c_{i,i} = D^2 (\xi_i) \) (the diagonal of $C$),

(iii) $C$ is positive semidefinite\(^1\),

(iv) if $\eta = A \cdot \xi + m$ for some real $A \in \mathbb{R}^{m \times n}$ and $m \in \mathbb{R}^m$,

then \( \text{cov} \left( \eta, \eta \right) = A \cdot \text{cov} \left( \xi, \xi \right) \cdot A^T \). \( \square \)

In the next Sections we briefly introduce the most important higher dimensional distributions.

### 2.2 The normal (Gauss-) distributions

#### 2.2.1 2-dimensional

**Definition I.37** The **2-dimensional normal** (Gauss-) r.v.-s are determined by the distribution functions

$$f \left( x_1, x_2 \right) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-r^2}} \cdot e^{-\frac{1}{2\sigma_1^2 \sigma_2^2 \left( 1-r^2 \right)} \left( \frac{(x_1-m_1)^2}{\sigma_1^2} - 2r \frac{(x_1-m_1)(x_2-m_2)}{\sigma_1 \sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2} \right)} \quad (2.2)$$

or, in modern notation

---

\(^1\) **Definition:** The real quadratic matrix $A = [a_{i,j}] \in \mathbb{R}^{n \times n}$ is **positive definite** if $x^T A x > 0$ for each $x \in \mathbb{R}^n$ where $x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} x_i x_j \quad \Box$

**Theorem:** A symmetric matrix is positive-definite if and only if all its eigenvalues are positive, that is, the matrix is positive-semidefinite and it is invertible. \( \Box \)
2.2. THE NORMAL (GAUSS-) DISTRIBUTIONS

\[ f(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - r^2}} \cdot \exp \left( \frac{-1}{2(1 - r^2)} \left( \frac{(x_1 - m_1)^2}{\sigma_1^2} - 2r \frac{(x_1 - m_1) \cdot (x_2 - m_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} \right) \right) \]

where \( m_1, m_2 \in \mathbb{R}, \sigma_1, \sigma_2, r \in \mathbb{R}^{+0}, -1 < r < 1 \) are any real numbers.

**Theorem I.38** The marginal distributions \( \xi \) and \( \eta \) are also normal, and \( M(\xi) = m_1, M(\eta) = m_2, D(\xi) = \sigma_1, D(\eta) = \sigma_2 \) and \( R(\xi, \eta) = r \).

2.2.2 n-dimensional

**Definition I.39** For any \( k \)-dimensional r.v. \( \vec{\xi} = (\xi_1, ..., \xi_k) \) where \( \xi_1, ..., \xi_k \) are standard normal r.v. (i.e. \( M(\xi_i) = 0 \) and \( D(\xi_i) = 1 \) for \( i = 1, ..., k \)) and real matrix \( A \in \mathbb{R}^{n \times k} \) and \( m \in \mathbb{R}^n \) the following \( n \)-dimensional r.v. \( \vec{\eta} := A \cdot \vec{\xi} + m \) is called \textit{n-dimensional normal r.v.} □

**Remark I.40** Be careful with the dimensions \( n \) and \( k \)!

An alternative definition is the following:

**Definition I.41** Let \( A = [a_{i,j}] \in \mathbb{R}^{n \times n} \) a symmetric\footnote{Definition: The real quadratic matrix \( A = [a_{i,j}] \in \mathbb{R}^{n \times n} \) is symmetric if \( A^T = A \), i.e. \( [a_{i,j}] = [a_{j,i}] \) for each \( i, j = 1, ..., n \). The symmetric matrix \( A \) is positive definite if \( \mathbf{z}^T A \mathbf{z} > 0 \) for each \( \mathbf{z} \in \mathbb{R}^n \) where \( \mathbf{z}^T A \mathbf{z} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} x_i x_j \). □} positive definite quadratic matrix and let \( B = [b_{i,j}] := A^{-1} \) the inverse matrix and let \( d_B := \det(B) \) the determinant of \( B \). Let further \( m_1, ..., m_n \in \mathbb{R} \) be any real numbers. Then \( \vec{\xi} = (\xi_1, ..., \xi_n) \) is an \( n \)-dimensional normal (Gaussian) r.v. if the joint density function is

\[ f_{\vec{\xi}}(x_1, ..., x_n) = \frac{\sqrt{d_B}}{(2\pi)^{n/2}} \cdot \exp \left( -\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - m_i) b_{i,j} (x_j - m_j)}{2} \right) \] (2.3)

□

Theorem: A symmetric matrix is positive-definite if and only if all its eigenvalues are positive, that is, the matrix is positive-semidefinite and it is invertible. □
2.3 The binomial/multinomial (Bernoulli-) distributions

2.3.1 1-dim = 2-dim

Recall the well known (1-dimensional) Bernoulli- or binomial distribution: given \( A \subseteq \Omega \), \( p = P(A) \), fix an \( m \in \mathbb{N} \), repeat the experiment \( m \)-many times (independently and with the same conditions) and let

\[
\xi := \text{the number of occurrences of } A.
\]

Then we have, taking \( q = 1 - p \)

\[
P(\xi = k) = \binom{m}{k} p^k q^{m-k} \quad \text{for } 0 \leq k \leq m. \tag{2.4}
\]

Observe now first, that in fact, we have a partition of \( \Omega \) to \( \{ A, \overline{A} \} \) since \( A \cup \overline{A} = \Omega \) and \( A \cap \overline{A} = \emptyset \). Second, together with \( \xi \) we also know the number of occurrences of \( \overline{A} \), i.e. we can let

\[
\xi_2 := \text{the number of occurrences of } \overline{A}
\]

and have

\[
P(\xi_2 = \ell) = \binom{m}{\ell} p^{m-\ell} q^\ell \quad \text{for } 0 \leq \ell \leq m \tag{2.5}
\]

and, of course \( p + q = 1 \) and \( k + \ell = m \).

This observation will be generalized for larger partitions in the next section.

2.3.2 n-dim (2 \( \leq n \))

**Definition I.42** Let \( A_1 \cup^* A_2 \cup^* \ldots \cup^* A_n = \Omega \), \( P(A_i) = p_i \), \( \sum_{i=1}^n p_i = 1 \), repeat the experiment \( m \)-many times, independently and with the same conditions, \( m \in \mathbb{N} \) is fixed, and let

\[
\xi_i := X_i := \text{number of } A_i \text{ occurring for } i = 1, \ldots, n.
\]

Then \( \overrightarrow{\xi} = (\xi_1, \ldots, \xi_n) \) is called \( n \)-dimensional binomial / multinomial / Bernoulli r.v. \( \square \)
Remark: If your experiment is choosing (sampling) \( m \) many elements from a set \( H \), which contains \( n \)-kind of objects, then the above term "independently and with the same conditions" means, that you must put back ("visszatenni") the chosen element before the next choosing. This method is called sampling with repetitions / putting back ("visszatevéses mintavétel").

**Theorem I.43** The distribution is: for any nonnegative integers \( k_1, \ldots, k_n \in \mathbb{N} \)

\[
P(\xi_1 = k_1, \ldots, \xi_n = k_n) = \begin{cases} 
\frac{m!}{k_1! \cdots k_n!} \cdot p_1^{k_1} \cdots p_n^{k_n} & \text{if } k_1 + \ldots + k_n = m \\
0 & \text{otherwise}
\end{cases}
\]

where \( p_i = P(A_i) \) for \( i = 1, \ldots, n \).

**Warning:** \( n \in \mathbb{N} \) is the size of the partition of \( \Omega \) and \( m \in \mathbb{N} \) is the number of experiments (repetitions).

**Remark I.44** The fraction \( \frac{m!}{k_1! \cdots k_n!} \) above is called polinomial or multinomial coefficient and usually is denoted as

\[
\binom{m}{k_1, \ldots, k_n} = \frac{m!}{k_1! \cdots k_n!}.
\]

(2.6)

### 2.4 The poli-hypergeometric distributions

It is the same as the binomial distribution, but without repetitions/putting back ("visszatevés / ismétlés / ismétlődés nélkül").

#### 2.4.1 1-dim = 2-dim

The well known Hypergeometric distribution is the following. Let \( A_1 \cup^* A_2 = H \), \( |H| = N \), \( |A_1| = M_1 \), \( |A_2| = M_2 = N - M_1 \), repeat the drawings from the set \( H \) for \( m \)-many times (\( m \in \mathbb{N} \) is fixed) without repetitions/putting back, and let

\[
\xi := \text{the number of occurrences of } A = A_1.
\]
Then we have
\[ P(\xi = k) = \binom{M_1}{k} \binom{N-M_1}{m-k} \binom{N}{m} \] for \( 0 \leq k \leq m \). \hspace{1cm} (2.7)

As in the Bernoulli distribution, we have a 2-element partition of \( H = A_1 \cup^* A_2 \), so the above is, in fact, 2-dimensional. The generalization is easy, go to next subsection.

2.4.2 n-dim \((2 \leq n)\)

**Definition I.45** Let \( A_1 \cup^* A_2 \cup^* \ldots \cup^* A_n = H \), \( |A_i| = M_i \), \( \sum_{i=1}^{n} M_i = N = |H| \) and choose without repetitions/putting back ("visszatevés / ismétlés / ismétlődés nélkül") from the set \( H \) for \( m \) many times (\( m \in \mathbb{N} \) is fixed), and let
\[ \xi_i := X_i := \text{number of } A_i \text{ occurring, without repetitions/putting back for } i = 1,\ldots,n . \]
Then \( \overrightarrow{\xi} = (\xi_1,\ldots,\xi_n) \) is called \( n \)-dimensional binomial / multinomial / Bernoulli r.v. \hspace{1cm} \Box

**Theorem I.46** The distribution is: for any nonnegative integers \( k_1,\ldots,k_n \in \mathbb{N} \)
\[ P(\xi_1 = k_1,\ldots,\xi_n = k_n) = \begin{cases} \binom{M_1}{k_1} \cdots \binom{M_n}{k_n} \binom{N}{m} & \text{if } k_1 + \ldots + k_n = m \hspace{1cm} \Box \\ 0 & \text{otherwise} \end{cases} \]

**Warning:** \( N = |H| \in \mathbb{N} \) is the size of the set \( H \), \( n \in \mathbb{N} \) is the size of the partition of \( H \) and \( m \in \mathbb{N} \) is the number of experiments (drawings) from the set \( H \).
Part II

Mathematical Statistics
Chapter 3

Elementary notions

Definition II.1  i) The result of a measuring is \( n \) many real numbers \( x_1, \ldots, x_n \).
ii) A statistical sample ("minta") is \( n \) many r.v. \( (\xi_1, \ldots, \xi_n) \) OR \( (X_1, \ldots, X_n) \).
iii) The degree of freedom ("szabadsági fok") is \( s = n - 1 \) in the above case. In other cases it often has another formula, where we always describe them.

Definition II.2  i)
\[
\bar{\xi} = \bar{\xi} := \frac{\xi_1 + \ldots + \xi_n}{n}
\]
(3.1)
is the empirical (greek)/ practical ("tapasztalati") average/ mean/ expected value.

ii) \( \bar{(\xi^2)} = \bar{(\xi^2)} := \frac{\xi_1^2 + \ldots + \xi_n^2}{n} \) is the empirical squared mean.

iii) The empirical variance and dispersion are
\[
\sigma^2 := \frac{1}{n} \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2 = \frac{(\xi_1 - \bar{\xi})^2 + \ldots + (\xi_n - \bar{\xi})^2}{n}
\]
(3.2)
and
\[
\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2}
\]
(3.3)

iv) The corrected ("korrigált, javított") empirical variance and dispersion are
\[
(\sigma^*)^2 := \frac{n}{n - 1} \cdot \sigma^2 = \frac{(\xi_1 - \bar{\xi})^2 + \ldots + (\xi_n - \bar{\xi})^2}{n - 1}
\]
(3.4)
and
\[
\sigma^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2} = \sqrt{\frac{n}{n-1}} \cdot \sigma . \tag{3.5}
\]

\[\square\]

**Remark II.3** The empirical and the corrected dispersions are often denoted by \(s\) and \(s^*\) to distinguish from the theoretical dispersion \(\sigma\).

The empirical and corrected variances and dispersions can be calculated easier:

**Theorem II.4**
\[
\sigma^2 = \left(\frac{\bar{\xi}^2}{n}\right) - \left(\frac{\bar{\xi}}{n}\right)^2 = \frac{\xi_1^2 + \ldots + \xi_n^2}{n} - \left(\frac{\xi_1 + \ldots + \xi_n}{n}\right)^2 , \tag{3.6}
\]
\[
\sigma = \sqrt{\left(\frac{\bar{\xi}^2}{n}\right) - \left(\frac{\bar{\xi}}{n}\right)^2} \quad \text{and so} \quad \sigma^* = \sqrt{\frac{n}{n-1} \left(\frac{\bar{\xi}^2}{n}\right) - \left(\frac{\bar{\xi}}{n}\right)^2} . \tag{3.7}
\]
\[\square\]

**Example II.5** Let \(\{\xi_1, \ldots, \xi_n\} = \{20.0, 20.2, 20.4, 20.7, 20.7, 21.0, 21.1, 21.3, 21.4, 21.4, 21.4, 21.5\}\), so \(n = 12\) and \(s = n - 1\).

The empirical mean is:
\[
\bar{\xi} = \frac{20.0 + 20.2 + 20.4 + 20.7 + 20.7 + 21.0 + 21.1 + 21.3 + 21.4 + 21.4 + 21.4 + 21.5}{12} = 20.925 ,
\]
the empirical quadratic mean:
\[
\bar{\xi}^2 = \frac{20.0^2 + 20.2^2 + 20.4^2 + 20.7^2 + 20.7^2 + 21.0^2}{12} + \frac{21.1^2 + 21.3^2 + 21.4^2 + 21.4^2 + 21.4^2 + 21.5^2}{12} \approx 438.100 \, 833 ,
\]
the empirical variance and dispersion:
\[
\sigma^2 = \left(\frac{\bar{\xi}^2}{n}\right) - \left(\frac{\bar{\xi}}{n}\right)^2 \approx 438.101 - 20.925^2 \approx 0.2454 ,
\]
\[
\sigma = \sqrt{\left(\frac{\bar{\xi}^2}{n}\right) - \left(\frac{\bar{\xi}}{n}\right)^2} \approx \sqrt{0.2454} \approx 0.4954 ,
\]
the corrected empirical variance and dispersion:
\[ (\sigma^*)^2 = \frac{n}{n-1} \cdot \left( \left( \frac{\xi^2}{n} \right) - \left( \frac{\bar{\xi}}{n} \right)^2 \right) \approx \frac{12}{11} \cdot (438.101 - 20.925^2) \approx 0.2677 \]

\[ \sigma^* = \sqrt{\frac{n}{n-1} \cdot \left( \left( \frac{\xi^2}{n} \right) - \left( \frac{\bar{\xi}}{n} \right)^2 \right)} \approx \sqrt{0.2677} = 0.5174 \]

**Definition II.6** Any function \( g(\xi_1, \ldots, \xi_n) \) of the sample \((\xi_1, \ldots, \xi_n)\) is called **statistical function**, or shortly **statistic**. \( \square \)

**Remark II.7** Many formulas use the advantage of datasets which are "symmetric to the origin", more precisely having mean \( \bar{\xi} = 0 \). This can be achieved by a little trick, which is worth learning. Let the original dataset (real numbers) be \( \Xi = \{\xi_i : i = 1, \ldots, n\} \) and denote \( \bar{\xi} \) its mean (a fixed real number). Now, prepare the modified dataset \( \Xi' := \{\xi_i - \bar{\xi} : i = 1, \ldots, n\} \), i.e. subtract \( \bar{\xi} \) from each data. Then clearly \( \bar{\xi}' = 0 \). Most of the further calculations allow this transformation.

Recall the similar transformation standardizing a r.v. \( \xi \) as \( \xi_{st} = \frac{\xi - M(\xi)}{D(\xi)} \) resulting \( M(\xi_{st}) = 0 \) and \( D(\xi_{st}) = 1 \). Similarly, a dataset \( \Xi \) can also be standardized as

\[ \Xi_{st} := \left\{ \frac{\xi_i - \bar{\xi}}{\sigma_\xi} : i = 1, \ldots, n \right\} \quad (3.8) \]

resulting similarly \( \bar{\xi}_{st} = 0 \) and \( \sigma_{\xi_{st}} = 1 \).

However, not each further calculations allow this transformation.
Chapter 4

Confidence intervals

Shortly: interval estimations (reliability intervals, "konfidencia =megbízhatósági intervallumok").

The general problem is:

Problem II.8 Give an interval \([a, b]\) of real numbers such that

\[ P(a < \gamma < b) \geq 1 - \varepsilon \]  \hspace{1cm} (4.1)\]

where \(\gamma\) is the parameter we are interested in and \(0 < \varepsilon < 1\) is given.

Definition II.9 The interval \([a, b]\) is the confidence (secure, "konfidencia, megbízhatósági") interval and \(1 - \varepsilon\) is the confidence level.

Remark II.10 Increasing \(n\) (the size of the sample) decreases \([a, b]\), but if decreasing \(\varepsilon\) then \([a, b]\) increases.

4.1 Interval for the probability

Problem II.11 Find the interval for \(p = P(A)\) for the event \(A\):

\[ P(a < p < b) \geq 1 - \varepsilon \]  \hspace{1cm} (4.2)\]
**Theorem II.12** If \( n \) independent experiments resulted \( k \) outcomes of \( A \) and \( n \) is large enough\(^1\), then

\[
[a, b] = \left[ \frac{k}{n} - \eta, \frac{k}{n} + \eta \right]\tag{4.3}
\]

where

\[
\eta = \frac{u_\varepsilon}{\sqrt{n}} \cdot \sqrt{\frac{k}{n} \cdot \left(1 - \frac{k}{n}\right)}\tag{4.4}
\]

and

\[
\Phi(u_\varepsilon) = 1 - \frac{\varepsilon}{2}\tag{4.5}
\]

(use table \( \Phi \)). \( \square \)

**Example II.13** Out of 30 pieces 10 is broken. Give an interval for \( p = P(\text{broken}) \) with confidence level 95%.

**Solution II.14** \( \varepsilon = 0.05 \) and \( \Phi(u_\varepsilon) = 1 - \frac{\varepsilon}{2} = 0.975 \) imply \( u_\varepsilon = 1.96 \).

Further:

\[
\eta = \frac{1.96}{\sqrt{30}} \cdot \sqrt{\frac{10}{30} \cdot \left(1 - \frac{10}{30}\right)} \approx 0.168690 ,
\]

\[
a \approx \frac{10}{30} - 0.168690 \approx 0.164643 ,
\]

\[
b \approx \frac{10}{30} + 0.168690 = 0.502023 ,
\]

so, by 95% we have

\[
P(0.164 < p < 0.502) \geq 0.95 . \quad \square \tag{4.6}
\]

**Remark II.15** \( i \) The interval \([a, b] = [0.164, 0.502]\) is fairly large since \( n \) is small and \( \varepsilon \) is small, too.

\( ii \) Theorem **II.12** is based on Moivre-Laplace’s theorem (see [SzII]). \( \square \)

---

\(^1\) \( n \) must be above 30, but \( n > 200 \) is preferable.
4.2 Interval for the mean when $\sigma$ is known

Problem II.16  Give an interval for $m = M(\xi)$ if $\xi$ is normal (Gaussian) and $\sigma = D(\xi)$ and $\varepsilon$ both are given:

$$P(a \leq m \leq b) \geq 1 - \varepsilon \quad . \quad (4.7)$$

Theorem II.17

$$[a, b] = \left[ \xi - u_\varepsilon \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{\xi} + u_\varepsilon \cdot \frac{\sigma}{\sqrt{n}} \right] \quad (4.8)$$

where $u_\varepsilon$ satisfies (4.5).  \[ \square \]

Example II.18  $\xi$ is normal with $\sigma = 3$ and the sample is:  \{\xi_1, ..., \xi_n\} =


Give an interval for 95% confidence.

Solution II.19  So $n = 12$, $D(\xi) = \sigma = 3$, $m = M(\xi) = ?$, $\varepsilon = 5\% = 0.05$, $\Phi(u_{0.05}) = 0.975$ and $u_{0.05} = 1.96$.  Using (3.1) and (4.8) we have

$$\bar{\xi} = \frac{20.0 + 20.2 + 20.4 + 20.7 + 20.7 + 21.0 + 21.1 + 21.3 + 21.4 + 21.4 + 21.4 + 21.5}{12} = 20.925,$$

$$\frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{12}} \approx 0.866025,$$

$$a \approx 20.925 - 1.96 \cdot 0.866025 \approx 19.227591,$$

$$b \approx 20.925 + 1.96 \cdot 0.866025 = 22.622409 .$$

So

$$P(19.228 < m < 22.622) > 1 - \varepsilon = 0.95 \quad . \quad (4.9)$$
4.3 Interval for the mean when $\sigma$ is unknown

**Problem II.20** Give an interval for $m = M(\xi)$ if $\xi$ is normal (Gaussian) and $\varepsilon$ is given but $\sigma = D(\xi)$ is unknown.

**Theorem II.21** After finding $t_\varepsilon$ in the table of the Student- (or t-) distribution with degree of freedom $s = n - 1$ we have

$$[a, b] = \left[\bar{\xi} - t_\varepsilon \cdot \frac{\sigma^*}{\sqrt{n}}, \quad \bar{\xi} + t_\varepsilon \cdot \frac{\sigma^*}{\sqrt{n}}\right] \quad (4.10)$$

i.e.

$$P(a < M(\xi) < b) > 1 - \varepsilon. \quad \square \quad (4.11)$$

**Example II.22** Let the sample be:


and let $1 - \varepsilon = 95\%$.

**Solution II.23** $n = 12$, $s = n - 1 = 11$, $m = M(\xi) = ?$, $\varepsilon = 5\% = 0.05$, so $t_{0.05} = 2.201$ (from the table). We calculated $\bar{\xi}, (\bar{\xi}^2)$ and $\sigma^*$ in example II.5, so:

$$\frac{\sigma^*}{\sqrt{n}} \approx \frac{0.5174}{\sqrt{12}} \approx 0.1494,$$

$a \approx 20.925 - 2.201 \cdot 0.1494 \approx 20.5962,$

$b \approx 20.925 + 2.201 \cdot 0.1494 \approx 21.2538,$

and finally

$$P(20.596 < M(\xi) < 21.254) > 1 - \varepsilon = 0.95. \quad (4.12)$$
4.4 Interval for the dispersion

Problem II.24 Give an interval for $\sigma = D(\xi)$ if $\xi$ is normal (Gaussian) and $\varepsilon$ is given.

Theorem II.25 For the variance we have

$$[a^2, b^2] = \left[\frac{n \cdot (\sigma^*)^2}{\chi^2_{\varepsilon/2}}, \frac{n \cdot (\sigma^*)^2}{\chi^2_{1-\varepsilon/2}}\right]$$

(4.13)

i.e.

$$P(a^2 < D^2(\xi) < b^2) > 1 - \varepsilon$$

(4.14)

and for the dispersion

$$[a, b] = \left[\frac{\sqrt{n} \cdot \sigma^*}{\chi^2_{\varepsilon/2}}, \frac{\sqrt{n} \cdot \sigma^*}{\chi^2_{1-\varepsilon/2}}\right]$$

(4.15)

i.e.

$$P(a < D(\xi) < b) > 1 - \varepsilon$$

(4.16)

where $\chi^2_{\varepsilon/2}$ and $\chi^2_{1-\varepsilon/2}$ are from the table of the $\chi^2$ or chi-square distribution with degree of freedom $s = n - 1$.

Example II.26 The confidence level is 95% and the sample is: $X_1, \ldots, X_n = 20.0, 20.2, 20.4, 20.7, 21.0, 21.1, 21.3, 21.4, 21.4, 21.4, 21.5$.

Solution II.27 $n = 12$, the degree of freedom is $s = n - 1 = 11$, $\varepsilon = 5\% = 0.05$.

Using table $\chi^2$ we find ($\varepsilon/2 = 0.025$, $1 - \varepsilon/2 = 0.975$, $s = 11$):

$$\chi^2_{\varepsilon/2} = \chi^2_{0.025} \approx 21.920 \quad \text{and} \quad \chi^2_{1-\varepsilon/2} = \chi^2_{0.975} \approx 3.816,$$

(4.17)

so

$$\chi^2_{0.025} \approx \sqrt{21.920} \approx 4.6819 \quad \text{and} \quad \chi^2_{0.975} \approx \sqrt{3.816} \approx 1.9535.$$ 

(4.18)

We calculated $\xi$, $(\xi^2)$ and $\sigma^*$ in Example II.5, so

$$a^2 = \frac{n \cdot (\sigma^*)^2}{\chi^2_{\varepsilon/2}} \approx \frac{12 \cdot 0.2677}{21.920} \approx 0.1466 \quad \Rightarrow \quad a \approx \sqrt{0.1466} \approx 0.3829,$$

$$b^2 = \frac{n \cdot (\sigma^*)^2}{\chi^2_{1-\varepsilon/2}} \approx \frac{12 \cdot 0.2677}{3.816} \approx 0.8418 \quad \Rightarrow \quad b \approx \sqrt{0.8418} \approx 0.9175,$$

so

$$P(0.1466 < D^2(\xi) < 0.8418) > 1 - \varepsilon = 0.95$$

(4.19)

and

$$P(0.3829 < D(\xi) < 0.9175) > 1 - \varepsilon = 0.95.$$ 

(4.20)
Chapter 5

Point estimations and hypothesis testing

5.1 General notions

Definition II.28  

i) Any statistical function \( g(x_1, \ldots, x_n) \) is an estimation ("becs-lés") of the parameter \( a \) (of a r.v. \( \xi \)), and it is often denoted by \( \hat{a}(x_1, \ldots, x_n) \), or shortly by \( \hat{a} \).

ii) The estimation \( \hat{a} = g(x_1, \ldots, x_n) \) is unbiased (un-distorted, not-deformed, "torzítatlan") if its mean equals to \( a = a(\xi) \), i.e.

\[
M(\hat{a}) = a .
\]  

(5.1)

iii) The estimation \( \hat{a} \) is consistent ("konzisztens", "következetes") if

\[
\forall \varepsilon, \delta > 0 \ (\exists n_0) \ (\forall n > n_0) \ P( |\hat{a}(x_1, \ldots, x_n) - a| \geq \varepsilon ) < \delta .
\]  

(5.2)

iv) The estimation \( \hat{a}_1 \) is more efficient/ effective ("hatásos") than \( \hat{a}_2 \) for the same parameter \( a \) if

\[
D(\hat{a}_1) < D(\hat{a}_2) .
\]

Remark II.29  

The exact value of \( a \) is unknown in general.

Example II.30  

By the Laws (Theorems) of Large Numbers we know, that

i) \( \hat{p} := \frac{k}{n} \) (relative frequency) is an unbiased estimation of the probability \( p \),

37
ii) \( \hat{\xi} = \bar{\xi} := \frac{\xi_1 + \ldots + \xi_n}{n} \) (average) is an unbiased estimation of the mean \( M(\xi) \).

iii) \( (\sigma_n^*)^2 := \frac{\sum_{i=1}^{n} (\bar{\xi} - \xi_i)^2}{n-1} \) (corrected empirical variance) is an unbiased estimation of the variance \( D^2(\xi) \).

\[ \Box \]

**Remark II.31** Be careful: the denominator of \((\sigma_n^*)^2\) is \(n - 1\), instead of \(n\).

**Definition II.32**

i) Any statement or assumption on \( \xi \) (and \( \eta \)), a **hypothesis** ("hipotézis, feltételezés"). The hypothesis we investigate is denoted by \( H_0 \) and called **base- or null-hypothesis** ("nullhipotézis"), its negation is denoted by \( H \) and called **alternative hypothesis** ("ellenhipotézis").

ii) The algorithm for deciding the hypothesis is called a **test** ("próba").

iii) After our calculations either \( H_0 \) is **accepted** ("elfogadjuk") or \( H_0 \) is **rejected** ("elvetjük"), i.e. \( H \) is accepted. \[ \Box \]

We may have two types of errors after our calculations:

**Definition II.33**

**Type I error** ("elsőfajú hiba") occurs when \( H_0 \) is true but we reject it.

**Type II error** ("másodfajú hiba") occurs when \( H_0 \) is not true but we accept it:

<table>
<thead>
<tr>
<th>( H_0 ) is accepted</th>
<th>( H_0 ) is rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0 ) is true</td>
<td>OK</td>
</tr>
<tr>
<td>( H_0 ) is false</td>
<td>Type II error</td>
</tr>
</tbody>
</table>

\[ \Box \]

**Remark II.34** The probability of type I error is usually denoted by \( \varepsilon \). The probability of type II error is hard to determine, but it usually tends to 0 if \( n \to \infty \).

**Remark II.35** Our main goal is to decrease type I errors: we want to avoid rejecting \( H_0 \) when \( H_0 \) is true (e.g. not kicking out any student who had prepared for the exam)! Of course, this could be fulfilled by accepting \( H_0 \) in all cases, i.e. setting \( \varepsilon := 0 \), but it would be a nonsense! So we have to balance \( \varepsilon \) in somehow - read further.

**Definition II.36** The **significance level of a test** ("megbízhatósági szint") is \( 1 - \varepsilon \) (where \( \varepsilon \) is the probability of type I error). \[ \Box \]
5.2. PARAMETRIC TESTS

Remark II.37 i) The word "significance level" means "important, essential, reliable,..." (in Hungarian: "szignifikancia- vagy megbízhatósági szint, szignifikáns, jelentős").
   ii) Most of the tests (see below) start with giving the significance level or \( \varepsilon \) (probability of type I error).
   iii) Decreasing \( \varepsilon \) makes type I error smaller and the test more reliable, however type II error increases at the same time when the sample size \( (n) \) is fixed. Increasing \( n \) type II error tends to 0.
   iv) In general, choosing the significance level to be 95\% is a suitable choice.

Definition II.38 i) If the hypothesis is quantitative (usually on some characteristics of \( \xi \), e.g. "\( M(\xi) = m_0 \)"), then the estimation and the test are called parametric ("paraméteres"), otherwise they are nonparametric ("nemparaméteres").
   ii) If the hypothesis is an equality, its test must be a two-sided test ("kétoldali próba").
      If the hypothesis is an inequality, its test must be a one-sided test ("egyoldali próba"). □

Example II.39 Some hypotheses (for details see the subsections below):
   i) \( H_0 : M(\xi) = m_0 \) (\( m_0 \in \mathbb{R} \) is a given number), so \( H : M(\xi) \neq m_0 \). This hypothesis needs a parametric and two-sided test.
   ii) \( H_0 : M(\xi) \leq m_0 \) (\( m_0 \in \mathbb{R} \) is a given number), so \( H : M(\xi) > m_0 \). This hypothesis needs a parametric and one-sided test.
   iii) \( H_0 : "\xi \) is a normal distribution". This hypothesis needs a nonparametric test. □

Remark II.40 In practice \( H_0 \) must contain the equality sign (\( = \) or \( \leq \) or \( \geq \)) and \( H \) (the negation of \( H_0 \)) may contain only the signs \( \neq \), \(< \) and \( > \).

5.2 Parametric tests

5.2.1 u- test for the mean of one sample when \( \sigma \) is known
("Egymintás u-próba")
   \( \xi \) is normal, \( \sigma \) is known, \( m_0 \) and \( \varepsilon \) are given (\( m_0 \in \mathbb{R} \), \( (\xi_1, \ldots, \xi_n) \) is the sample.
Algorithm II.41 For the two-sided test $H_0 : M(\xi) = m_0$

\[ i) \text{calculate } u_{sz} := \sqrt{n} \cdot \frac{\xi - m_0}{\sigma}, \]

\[ ii) \text{find } u_\varepsilon \in \mathbb{R}^+ \text{ such that } \Phi(u_\varepsilon) = 1 - \frac{\varepsilon}{2}, \]

\[ iii) \text{accept } H_0 \text{ in the case } |u_{sz}| \leq u_\varepsilon \text{ with significance } 1 - \varepsilon \]
\[ \text{or reject } H_0 \text{ in the case } |u_{sz}| > u_\varepsilon \text{ with significance } 1 - \varepsilon. \]

Algorithm II.42 For one-sided tests: $H_0 : M(\xi) \geq / \leq m_0$

\[ i) \text{calculate } u_{sz} := \sqrt{n} \cdot \frac{\xi - m_0}{\sigma}, \]

\[ ii) \text{find } u_\varepsilon \in \mathbb{R}^+ \text{ such that } \Phi(u_\varepsilon) = 1 - \varepsilon, \]

\[ iii) \text{accept } H_0 : M(\xi) \leq m_0 \text{ in the case } u_{sz} \leq u_\varepsilon \text{ with significance } 1 - \varepsilon \]
\[ \text{or reject } H_0 \text{ in the case } u_{sz} > u_\varepsilon \text{ with significance } 1 - \varepsilon. \]

\[ iv) \text{accept } H_0 : M(\xi) \geq m_0 \text{ in the case } -u_\varepsilon \leq u_{sz} \text{ with significance } 1 - \varepsilon \]
\[ \text{or reject } H_0 \text{ in the case } -u_\varepsilon > u_{sz} \text{ with significance } 1 - \varepsilon. \]

Remark II.43 If the dispersion $\sigma$ is unknown, theoretically the $t$-test (see below) is applicable, but for large samples ($n > 30$) the $u$-test can also be used, but use $\sigma^*$ instead of $\sigma$.

Example II.44 Let $m_0 = 1200$, $\sigma = 3$ and $\bar{\xi} = \{1193, 1198, 1203, 1191, 1195, 1196, 1199, 1191, 1201, 1196, 1193, 1198, 1204, 1196, 1198, 1200\}$.

Decide the hypothesis $H_0 : M(\xi) = m_0$ with significance level 99.9%.

Solution II.45 Two sided test. So $\varepsilon = 0.001$, $\Phi(u_\varepsilon) = 1 - \frac{\varepsilon}{2} = 0.9995$ and $u_\varepsilon = 3.29$. Further $n = 16$, $\bar{\xi} = (1193 + 1198 + 1203 + 1191 + 1195 + 1196 + 1199 + 1191 + 1201 + 1196 + 1193 + 1198 + 1204 + 1196 + 1198 + 1200)/16 = 1197$, so
\[ u_{sz} = \sqrt{16 \cdot \frac{1197 - 1200}{3}} = -4. \]
Since $|u_{sz}| = 4 > u_\varepsilon = 3.29$ we must reject $H_0$ with significance 99.9%.

Example II.46 Let $m_0 = 70$, $\sigma$ is unknown, $n = 36$, $\bar{\xi} = 68.5$ and $\sigma^* = 6$.

Decide the hypothesis $H_0 : M(\xi) \geq m_0$ with significance level 95%.

Solution II.47 One sided test. Though the dispersion ($\sigma$) is unknown, but the sample is large enough ($n > 30$), so the $u$-test can also be used. So $\varepsilon = 0.05$, $\Phi(u_\varepsilon) = 1 - 0.05 = 0.95$ and $u_\varepsilon = 1.65$.
\[ u_{sz} = \sqrt{n} \cdot \frac{\bar{\xi} - m_0}{\sigma^*} = \sqrt{36} \cdot \frac{68.5 - 70}{6} = -1.5. \]
Since $-u_\varepsilon = -1.65 \leq u_{sz} = -1.5$ we have to accept the hypothesis $H_0 : M(\xi) \geq m_0$ with significance level 95%.
5.2.2  \textit{t-} test for the mean of one sample when \(\sigma\) is unknown

("Egymintás t-próba")

\(\xi\) is normal, \(\sigma\) is unknown, \(m_0\) and \(\varepsilon\) are given (\(m_0 \in \mathbb{R}\)), \((\xi_1, \ldots, \xi_n)\) is the sample.

\textbf{Algorithm II.48}  \textit{For the two-sided test} \(H_0 : M(\xi) = m_0 :\)
\begin{enumerate}[i)]  
  \item calculate \(t_{sz} := \sqrt{n} \cdot \frac{\bar{\xi} - m_0}{\sigma^*}\)
  \item find \(t_\varepsilon \in \mathbb{R}^+\) in the table of the Student-distribution for \(\beta = p = 1 - \frac{\varepsilon}{2}\) and degree of freedom \(s = n - 1\),
  \item accept \(H_0\) in the case \(|t_{sz}| \leq t_\varepsilon\) with significance \(1 - \varepsilon\), or reject \(H_0\) in the case \(|t_{sz}| > t_\varepsilon\) with significance \(1 - \varepsilon\).
\end{enumerate}

\textbf{Algorithm II.49}  \textit{For one-sided tests} \(H_0 : M(\xi) \geq / \leq m_0 :\)
\begin{enumerate}[i)]  
  \item calculate \(t_{sz} := \sqrt{n} \cdot \frac{\bar{\xi} - m_0}{\sigma^*}\),
  \item find \(t_\varepsilon \in \mathbb{R}^+\) in the table of the Student-distribution for \(\beta = p = 1 - \varepsilon\) and degree of freedom \(s = n - 1\),
  \item accept \(H_0\) : \(M(\xi) \leq m_0\) in the case \(t_{sz} \leq t_\varepsilon\) with significance \(1 - \varepsilon\) or reject \(H_0\) in the case \(t_{sz} > t_\varepsilon\) with significance \(1 - \varepsilon\).
  \item accept \(H_0\) : \(M(\xi) \geq m_0\) in the case \(-t_\varepsilon \leq t_{sz}\) with significance \(1 - \varepsilon\) or reject \(H_0\) in the case \(-t_\varepsilon > t_{sz}\) with significance \(1 - \varepsilon\).
\end{enumerate}

\textbf{Remark II.50}  \textit{For large samples }\(n > 30\) the \(u\)-test can also be applied but we use \(\sigma^*\) instead of \(\sigma\).

\textbf{Example II.51}  Let the sample be \(\bar{\xi} = \{1.51, 1.49, 1.54, 1.52, 1.54\}\). Decide the hypothesis \(H_0 : M(\xi) = 1.50\) with significance level 95%.

\textbf{Solution II.52}  Two sided test. \(n = 5, \ s = 4\),
\[
\bar{\xi} = \frac{1.51 + 1.49 + 1.54 + 1.52 + 1.54}{5} = 1.52,
\]
\[
\bar{\xi}^2 = \frac{1.51^2 + 1.49^2 + 1.54^2 + 1.52^2 + 1.54^2}{5} = 2.31076,
\]
\[
\sigma^* = \sqrt{\frac{5}{5-1} \cdot (2.31076 - 1.52^2)} = 0.02121,
\]
\[
t_{sz} = \sqrt{n} \cdot \frac{\bar{\xi} - m_0}{\sigma^*} = \sqrt{5} \cdot \frac{1.52 - 1.50}{0.02121} = 2.1085,
\]
\[
\varepsilon = 0.05, \ \beta = p = 1 - \frac{\varepsilon}{2} = 0.975, \ t_\varepsilon = 2.78.
\]

Since \(|t_{sz}| = 2.1085 < t_\varepsilon = 2.78\) we must accept \(H_0\) with significance 95%.
Example II.53 Let the sample be $\bar{\xi} = \{3.1, 2.8, 1.5, 1.7, 2.4, 2.0, 3.3, 1.6\}$. Decide the hypothesis $H_0: M(\xi) \geq 3.1$ with significance level 98%.

Solution II.54 One sided test. $n = 8$, $s = 7$,

$\bar{\xi} = \frac{3.1 + 2.8 + 1.5 + 1.7 + 2.4 + 2.0 + 3.3 + 1.6}{8} = 2.3$,

$\bar{\sigma}^2 = \frac{3.1^2 + 2.8^2 + 1.5^2 + 1.7^2 + 2.4^2 + 2.0^2 + 3.3^2 + 1.6^2}{8} = 5.725$,

$\sigma^* = \sqrt{\frac{8}{8-1}} \cdot (5.725 - 2.3^2) \approx 0.7051$,

$t_{sz} = \sqrt{n} \cdot \frac{\bar{\xi} - m_0}{\sigma^*} = \sqrt{8} \cdot \frac{2.3 - 3.1}{0.7051} \approx -3.2091$,

$\varepsilon = 0.02$, $p = 1 - \varepsilon = 0.98$, $t_\varepsilon = 2.52$.

Since $t_{sz} \approx -3.2091 < -t_\varepsilon = -2.52$ we must reject $H_0$ with significance 98%.

5.2.3 k-test for the dispersion of one sample

("Egymintás szórás-próba"")

$\xi$ is normal, $\sigma$ is unknown, $\varepsilon$ and $\sigma_0$ are given ($\sigma_0 \in \mathbb{R}^+$), $(\xi_1, \ldots, \xi_n)$ is the sample.

i) For all the cases below the calculated test function is:

$$k_{sz} := \frac{(n - 1) \cdot (\sigma^*)^2}{\sigma_0^2},$$

the degree of freedom is $s = n - 1$. Then

Algorithm II.55 For the two-sided test $H_0: D(\xi) = \sigma_0$

ii) find $k_{\varepsilon/2} = \chi^2_{n-1, \varepsilon/2} \in \mathbb{R}^+$ and $k_{1-\varepsilon/2} = \chi^2_{n-1, 1-\varepsilon/2} \in \mathbb{R}^+$ in the table of the $\chi^2$-distribution for $\beta = \frac{\varepsilon}{2}$ and $\beta = 1 - \frac{\varepsilon}{2}$.

iii) accept $H_0$ in the case $k_{1-\varepsilon/2} \leq k_{sz} \leq k_{\varepsilon/2}$ with significance $1 - \varepsilon$, or reject $H_0$ in the case either $k_{sz} < k_{1-\varepsilon/2}$ or $k_{\varepsilon/2} < k_{sz}$ with significance $1 - \varepsilon$. $\square$
5.2. PARAMETRIC TESTS

Algorithm II.56 For the one-sided test $H_0 : D(\xi) \geq \sigma_0$

ii) find $k_{1-\varepsilon} = \chi^2_{n-1, 1-\varepsilon} \in \mathbb{R}^+$ in the table of the $\chi^2$-distribution for $\beta = 1 - \varepsilon$,

iii) accept $H_0$ in the case $k_{1-\varepsilon} \leq k_{sz}$ with significance $1 - \varepsilon$, or reject $H_0$ in the case $k_{sz} < k_{1-\varepsilon}$ with significance $1 - \varepsilon$.

Algorithm II.57 For the one-sided test $H_0 : D(\xi) \leq \sigma_0$

ii) find $k_\varepsilon = \chi^2_{n-1, \varepsilon} \in \mathbb{R}^+$ in the table of the $\chi^2$-distribution for $\beta = \varepsilon$,

iii) accept $H_0$ in the case $k_{sz} \leq k_\varepsilon$ with significance $1 - \varepsilon$, or reject $H_0$ in the case $k_\varepsilon < k_{sz}$ with significance $1 - \varepsilon$.

Example II.58 Decide $H_0 : D(\xi) = 1.1$ when, $\sigma^* = 1.3$, $n = 10$ and $\varepsilon = 0.1$.

Solution II.59 Two sided test: $\sigma_0 = 1.1$, $\beta = \frac{\varepsilon}{2} = 0.05$, $k_\varepsilon = 16.919$, $1 - \frac{\varepsilon}{2} = 0.975$, $k_{1-\varepsilon} = 2.7$, $k_{sz} = \frac{9 \cdot 1.3^2}{1.1^2} \approx 12.57$, $k_{1-\varepsilon} < k_{sz} < k_\varepsilon$, so $H_0$ is accepted.

Example II.60 Decide $H_0 : D(\xi) \leq 1.1$ when, $\sigma^* = 1.3$, $n = 10$ and $\varepsilon = 0.1$.

Solution II.61 One sided test: $\sigma_0 = 1.1$, $\beta = \varepsilon = 0.1$, $k_\varepsilon = 14.684$, $k_{sz} = \frac{9 \cdot 1.3^2}{1.1^2} \approx 12.57 < k_\varepsilon$ so $H_0$ is accepted.

5.2.4 u- test for the means of two samples

("Kétmintás u-próba")

$\xi$ and $\eta$ are normal, $\varepsilon$ and $m_0$ are given ($m_0 \in \mathbb{R}$), $(\xi_1, \ldots, \xi_n)$ and $(\eta_1, \ldots, \eta_m)$ are large and independent samples, further let denote $\sigma_\xi := D(\xi)$ and $\sigma_\eta := D(\eta)$.

Here we will deal with hypothesis $M(\xi) - M(\eta) \nabla m_0$ where $\nabla$ can be any of $\geq$, $\leq$ or $=$.

Algorithm II.62 i) When $\sigma_\xi$ and $\sigma_\eta$ are known (for any-sided test) calculate

$$u_{sz} := \frac{\bar{\xi} - \bar{\eta} - m_0}{\sqrt{\frac{\sigma_\xi^2}{n} + \frac{\sigma_\eta^2}{m}}} \quad (5.4)$$

when \( \sigma_\xi \) and \( \sigma_\eta \) are not known (for any-sided test), calculate
\[
 u_{sz} := \frac{\bar{\xi} - \bar{\eta} - m_0}{\sqrt{\frac{\sigma_\xi^2}{n} + \frac{\sigma_\eta^2}{m}} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} \quad (5.5)
\]

\( i_2 \) For the two-sided test \( H_0 : M (\xi) - M (\eta) = m_0 \) find \( u_\varepsilon \in \mathbb{R}^+ \) such that \( \Phi (u_\varepsilon) = 1 - \frac{\varepsilon}{2} \).

\( ii_2 \) for one-sided tests \( H_0 : M (\xi) - M (\eta) \geq \) find \( u_\varepsilon \in \mathbb{R}^+ \) such that \( \Phi (u_\varepsilon) = 1 - \varepsilon \).

\( iii_1 \) For the two-sided test \( H_0 : M (\xi) - M (\eta) = m_0 \)
accept \( H_0 \) in the case \( |u_{sz}| \leq u_\varepsilon \) or reject \( H_0 \) in the case \( |u_{sz}| > u_\varepsilon \) with significance 1 - \( \varepsilon \).

\( iii_2 \) For the one-sided test \( H_0 : M (\xi) - M (\eta) \geq m_0 \)
accept \( H_0 \) in the case \( -u_\varepsilon \leq u_{sz} \) or reject \( H_0 \) in the case \( -u_\varepsilon > u_{sz} \) with significance 1 - \( \varepsilon \).

\( iii_3 \) For the one-sided test \( H_0 : M (\xi) - M (\eta) \leq m_0 \)
accept \( H_0 \) in the case \( u_{sz} \leq u_\varepsilon \) or reject \( H_0 \) in the case \( u_{sz} > u_\varepsilon \) with significance 1 - \( \varepsilon \).

**Example II.63** Let \( n = 10 \), \( \bar{\xi} = 40.1 \), \( \sigma_\xi = 5.48 \), \( m = 8 \), \( \bar{\eta} = 38.3 \), \( \sigma_\eta = 6.32 \).
Decide \( M (\xi) = M (\eta) \) with significance level 95%.

**Solution II.64** Two-sided test and \( \sigma_\xi \), \( \sigma_\eta \) are known. \( H_0 : M (\xi) - M (\eta) = 0 \),
\( m_0 = 0 \), \( \varepsilon = 0.05 \), \( \Phi (u_\varepsilon) = 1 - \frac{\varepsilon}{2} = 0.975 \), so \( u_\varepsilon = 1.96 \).

Now \( u_{sz} = \frac{40.1 - 38.3 - 0}{\sqrt{\frac{5.48^2}{10} + \frac{6.32^2}{8}}} \approx 0.6366 < u_\varepsilon \),
and \( H_0 \) is accepted with significance level 95%.

**Example II.65** Let \( n = 225 \), \( \bar{\xi} = 57 \), \( \sigma_\xi = 12 \), \( m = 250 \), \( \bar{\eta} = 60 \), \( \sigma_\eta = 15 \).
Decide \( M (\xi) \geq M (\eta) \) with significance level 98%.

**Solution II.66** One-sided test and \( \sigma_\xi \), \( \sigma_\eta \) are known. \( H_0 : M (\xi) - M (\eta) \geq 0 \),
\( m_0 = 0 \), \( \varepsilon = 0.02 \), \( \Phi (u_\varepsilon) = 1 - \varepsilon = 0.98 \), so \( u_\varepsilon = 2.05 \).

Now \( u_{sz} = \frac{57 - 60 - 0}{\sqrt{\frac{12^2}{225} + \frac{15^2}{250}}} \approx -2.417 < -u_\varepsilon \),
so we reject \( H_0 \) with significance level 98%.
Example II.67 Let \( n = 40 \), \( \bar{\xi} = 102 \), \( \sigma_\xi = 5.648 \), \( m = 35 \), \( \bar{\eta} = 95 \), \( \sigma_\eta = \sigma_\eta = 5.648 \). Decide \( M(\xi) \leq M(\eta) + 4 \) with significance level 99%.

Solution II.68 One-sided test and \( \sigma_\xi = \sigma_\eta \) are known. \( H_0 : M(\xi) - M(\eta) \leq 4 \), \( m_0 = 4 \), \( \varepsilon = 0.01 \), \( \Phi(u_\varepsilon) = 1 - \varepsilon = 0.99 \), so \( u_\varepsilon = 2.33 \).

Now \( u_{sz} = \frac{102 - 95 - 4}{\sqrt{\frac{5.648^2}{40} + \frac{5.648^2}{35}}} \approx 2.2949 < u_\varepsilon \),

so we accept \( H_0 \) with significance level 98%.

5.2.5 \( \tau \)-test for the means of two samples when \( \sigma_1 = \sigma_2 \)

("Kétmintás \( \tau \)-próba")

\( \xi \) and \( \eta \) are normal, only the equality \( \sigma_1 = \sigma_2 \) is known (but we do not know either \( \sigma_1 \) or \( \sigma_2 \)), \( \varepsilon \) is given, \( (\xi_1, ..., \xi_n) \) and \( (\eta_1, ..., \eta_m) \) are not large samples. (For large samples the \( u \)-test can also be used.)

Algorithm II.69 For the two-sided test \( H_0 : M(\xi) = M(\eta) \)

i) calculate

\[
t_{sz} := \frac{\bar{\xi} - \bar{\eta}}{\sqrt{(n-1)\sigma^2_\xi + (m-1)\sigma^2_\eta}} \cdot \sqrt{\frac{nm(n+m-2)}{n+m}} 
\]  

(5.6)

ii) find \( t_\varepsilon \in \mathbb{R}^+ \) in the table of the Student-distribution for \( p = 1 - \frac{\varepsilon}{2} \) and degree of freedom \( s = n + m - 2 \).

iii) accept \( H_0 \) in the case \( |t_{sz}| \leq t_\varepsilon \) with significance \( 1 - \varepsilon \).

or reject \( H_0 \) in the case \( |t_{sz}| > t_\varepsilon \) with significance \( 1 - \varepsilon \). \( \square \)

Algorithm II.70 For the two-sided test \( H_0 : M(\xi) - M(\eta) = m_0 \) (where \( m_0 \in \mathbb{R} \) any number)

i) calculate

\[
t_{sz} := \frac{\bar{\xi} - \bar{\eta} - m_0}{\sqrt{\sigma^2_\xi \cdot (n-1) + \sigma^2_\eta \cdot (m-1)}} \cdot \sqrt{\frac{1 \cdot \frac{1}{n} + \frac{1}{m}}{n + m - 2}} 
\]  

(5.7)
ii) find $t_\varepsilon \in \mathbb{R}^+$ in the table of the Student-distribution for $p = 1 - \frac{\varepsilon}{2}$ and degree of freedom $s = n + m - 2$, 

iii) accept $H_0$ in the case $|t_{sz}| \leq t_\varepsilon$ with significance $1 - \varepsilon$.

or reject $H_0$ in the case $|t_{sz}| > t_\varepsilon$ with significance $1 - \varepsilon$. □

Example II.71 Let $\vec{\xi} = \{300, 301, 303, 288, 294, 296\}$ and $\vec{\eta} = \{305, 317, 308, 300, 314, 316\}$. Decide the hypothesis $H_0 : M(\xi) = M(\eta)$ with significance level 99%.

Solution II.72

$$\bar{\xi} = \frac{300 + 301 + 303 + 288 + 294 + 296}{6} = 297,$$

$$\sigma^2_\xi = \frac{6}{6-1} \cdot (88234.3 - 297^2) \approx 30.39,$$

$$\bar{\eta} = \frac{305 + 317 + 308 + 300 + 314 + 316}{6} = 310,$$

$$\sigma^2_\eta = \frac{6}{6-1} \cdot (96138.3 - 310^2) \approx 45.9,$$

$$t_{sz} = \frac{297 - 310}{\sqrt{\frac{36 \cdot 10}{6 + 6}}} \cdot \sqrt{\frac{36 \cdot 10}{6 + 6}} \approx -3.643,$$

$n = m = 6$, $s = 6 + 6 - 2 = 10$, $\varepsilon = 0.01$, $\beta = p = 1 - \frac{\varepsilon}{2} = 0.995$, $t_\varepsilon = 3.17$. Since $|t_{sz}| \approx 3.643 > t_\varepsilon = 3.17$ we must reject $H_0$ with significance 99%.

5.2.6 F-test for the dispersions of two samples

whether $\sigma_1 = \sigma_2$

("Kétmintás F-próba")

$\xi$ and $\eta$ are normal, $H_0 : D(\xi) = D(\eta)$, $\varepsilon$ is given, $(\xi_1, ..., \xi_n)$ and $(\eta_1, ..., \eta_m)$ are the samples.

Algorithm II.73 i) if $\sigma^2_\xi > \sigma^2_\eta$ then let $F_{sz} := \frac{\sigma^2_\xi}{\sigma^2_\eta}$, otherwise let $F_{sz} := \frac{\sigma^2_\eta}{\sigma^2_\xi}$ (i.e. $F_{sz} > 1$ always holds),
5.3 Nonparametric tests

Remark II.76 The most widely used nonparametric test is Pearson’s chi-squared test, i.e. shortly the $\chi^2$ (“khi-négyzet”) test. It is important to know, that while the previous tests can be used for small and medium size samples as well, the $\chi^2$ test works only for large samples.

As in hypothesis tests, the significance level $1 - \epsilon$ is always given.

5.3.1 Goodness of fit

("illeszkedésvizsgálat"), GFI = goodness of fit index ("az illeszkedés jósága mutató"). See also the section "Normality test".

$H_0$: The sample $\xi$ fits the discrete distribution $(p_1, ..., p_k)$.

In detail: Does the sample $(\xi_1, ..., \xi_n)$ fits into $k$ mutually exclusive classes with probabilities $p_i$ ($i = 1, ..., k$), i.e. is $\{A_1, ..., A_k\}$ a complete system of events with $P(A_i) = p_i$?

Algorithm II.77 i) count the occurrences in $A_i$ (i.e. how many $\xi_j$ is in $A_i$) and denote these numbers by $a_i$,

ii) calculate

$$\chi^2_{sz} := \sum_{i=1}^{k} \frac{(a_i - np_i)^2}{np_i} = n \cdot \sum_{i=1}^{k} \left( \frac{a_i}{n} - p_i \right)^2,$$  \hspace{1cm} (5.8)
iii) find $\chi^2$ in the "Chi-squared" table (the degree of freedom is $k-1$),

iv) accept $H_0$ in the case $|\chi^2_{sz}| \leq \chi^2_\alpha$ with significance $1-\alpha$,
or reject $H_0$ in the case $|\chi^2_{sz}| > \chi^2_\alpha$ with significance $1-\alpha$.

Example II.78 We tossed 4 coins (together) 160 times and get the distribution of the heads as:

<table>
<thead>
<tr>
<th>nu. of heads ($i$)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency ($a_i$)</td>
<td>5</td>
<td>35</td>
<td>67</td>
<td>41</td>
<td>12</td>
<td>160</td>
</tr>
</tbody>
</table>

Are the coins fair with significance 95% ?

Solution II.79 The coins are fair $\iff \xi := \"the number of heads\"$ is a binomial distribution with $p = \frac{1}{2}$.

\[
p_0 = \binom{4}{0} \cdot \left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^4 = \left(\frac{1}{2}\right)^4 = \frac{1}{16} = 0.0625 ,
\]

\[
p_1 = \binom{4}{1} \cdot \left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^3 = \left(\frac{1}{2}\right)^4 = \frac{4}{16} = 0.25 ,
\]

\[
p_2 = \binom{4}{2} \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^2 = \left(\frac{1}{2}\right)^4 = \frac{6}{16} = 0.375 ,
\]

\[
p_3 = \binom{4}{3} \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^1 = \left(\frac{1}{2}\right)^4 = \frac{4}{16} = 0.25 ,
\]

\[
p_4 = \binom{4}{4} \cdot \left(\frac{1}{2}\right)^4 = \left(\frac{1}{2}\right)^4 = \frac{1}{16} = 0.0625 .
\]

\[
\chi^2_{sz} = \sum_{i=1}^{k} \frac{(a_i - np_i)^2}{np_i} = \frac{5^2}{10} + \frac{5^2}{40} + \frac{7^2}{60} + \frac{1^2}{40} + \frac{2^2}{10} \approx 4.3667 ,
\]

\[
s = 5-1 = 4, \ \ \ \varepsilon = 0.05, \ \ \ \chi^2_\varepsilon = 9.488, \ \ \ H_0 \text{ is accepted since } \chi^2_{sz} < \chi^2_\varepsilon . \ \ \ \Box
\]
5.3. NONPARAMETRIC TESTS

5.3.2 Homogenity

("homogenitás, azonosság")

\( H_0 \): The complete systems of events \( \{A_1, ..., A_k\} \) and \( \{B_1, ..., B_k\} \) determined by \( \xi \) and \( \eta \) are the same.

In detail: The sample is the union of \((\xi_1, ..., \xi_n)\) and \((\eta_1, ..., \eta_m)\), i.e. and the equality of \( \xi \) and \( \eta \) is the question.

Algorithm II.80 One sided test.

i) count the occurrences of \( \xi \) in \( A_i \) and of \( \eta \) in \( B_i \) and denote these numbers by \( a_i \) and \( b_i \) \((i = 1, ..., k)\),

ii) calculate

\[
\chi^2_{sz} := \frac{1}{mn} \sum_{i=1}^{k} \frac{(ma_i - nb_i)^2}{a_i + b_i},
\]

(5.9)

iii) find \( \chi^2 \) in the "Chi-squared" table (the degree of freedom now is \( k - 1 \), \( \beta = \epsilon \)),

iv) accept \( H_0 \) in the case \( |\chi^2_{sz}| \leq \chi^2_\epsilon \) with significance \( 1 - \epsilon \),

or reject \( H_0 \) in the case \( |\chi^2_{sz}| > \chi^2_\epsilon \) with significance \( 1 - \epsilon \). \( \square \)

Example II.81 Decide homogenity with significance 95% for the below samples:

<table>
<thead>
<tr>
<th></th>
<th>A_1</th>
<th>A_2</th>
<th>A_3</th>
<th>A_4</th>
<th>A_5</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \xi )</td>
<td>51</td>
<td>64</td>
<td>26</td>
<td>18</td>
<td>21</td>
<td>180</td>
</tr>
<tr>
<td>( \bar{\eta} )</td>
<td>72</td>
<td>51</td>
<td>33</td>
<td>23</td>
<td>21</td>
<td>200</td>
</tr>
</tbody>
</table>

Solution II.82 \( n = 51 + 64 + 26 + 18 + 21 = 180, m = 72 + 51 + 33 + 23 + 21 = 200 \),

\( s = 5 - 1 = 4 \), \( \epsilon = 0.05 \), \( \chi^2_{s, \epsilon} = \chi^2_{4, 0.05} = 9.488 \),

further

\[
\chi^2_{sz} = \frac{1}{180 \cdot 200} \left( \frac{(200 \cdot 51 - 180 \cdot 72)^2}{51 + 72} + \frac{(200 \cdot 64 - 180 \cdot 51)^2}{64 + 51} + \frac{(200 \cdot 26 - 180 \cdot 33)^2}{26 + 33} + \frac{(200 \cdot 18 - 180 \cdot 23)^2}{18 + 23} + \frac{(200 \cdot 21 - 180 \cdot 21)^2}{21 + 21} \right)
\]

\( \approx 5.458 < \chi^2_{s, \epsilon} \),

so \( H_0 \) is accepted.
5.3.3 Independence

("függetlenség")

\(H_0:\) The complete systems of events \(\{A_1, \ldots, A_k\}\) and \(\{B_1, \ldots, B_\ell\}\) determined by \(\xi\) and \(\eta\) are independent.

**In detail:** The sample is \(\zeta = ((\xi_1, \eta_1), \ldots, (\xi_n, \eta_n))\), i.e. \(n\) many double measurements are, and the dependence between \(\xi\) and \(\eta\) is the question.

**Algorithm II.83**

1. make the table of the occurrences in \(A_i\) vs. \(B_j\) and denote these by \(c_{ij}\),
2. calculate the marginal distributions \((a_1, \ldots, a_k)\) and \((b_1, \ldots, b_\ell)\),
3. calculate
   \[
   \chi^2_{sz} := \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^\ell \frac{(nc_{ij} - a_i b_j)^2}{a_i b_j},
   \tag{5.10}
   \]
   (use the vertices of the rectangles in the table for the enumerator),
4. find \(\chi^2\) in the "Chi-squared" table (the degree of freedom now is \((k-1)(\ell-1)\)),
5. accept \(H_0\) in the case \(|\chi^2_{sz}| \leq \chi^2\) with significance \(1 - \varepsilon\),
or reject \(H_0\) in the case \(|\chi^2_{sz}| > \chi^2\) with significance \(1 - \varepsilon\).

**Example II.84** Is there a connection with significance 95% between gender and success on the basis of the table?

<table>
<thead>
<tr>
<th>(\xi) (\eta)</th>
<th>success</th>
<th>unsuccess</th>
</tr>
</thead>
<tbody>
<tr>
<td>man</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>woman</td>
<td>34</td>
<td>26</td>
</tr>
</tbody>
</table>

**Solution II.85** So \(n = 28+12+34+26 = 100\), \(k = \ell = 2\), \(s = (k-1)(\ell-1) = 1\), \(\varepsilon = 0.05\),

<table>
<thead>
<tr>
<th>(\xi) (\eta)</th>
<th>success</th>
<th>unsuccess</th>
<th>(b_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>man</td>
<td>28</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>woman</td>
<td>34</td>
<td>26</td>
<td>60</td>
</tr>
<tr>
<td>(a_i)</td>
<td>62</td>
<td>38</td>
<td>100</td>
</tr>
</tbody>
</table>

\[
\chi^2_{sz} = \frac{1}{100} \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n \cdot c_{ij} - a_i \cdot b_j)^2}{a_i \cdot b_j}
= \frac{1}{100} \left( \frac{(100-28-40-62)^2}{40-62} + \frac{(100-12-40-38)^2}{40-38} + \frac{(100-34-60-62)^2}{60-62} + \frac{(100-26-60-38)^2}{60-38} \right) \approx 1.8110 ,
\]

\(\chi^2_{sz} < \chi^2 = 3.84\), so \(H_0\) is accepted: no connection between gender and success with significance 95%.
5.3.4 Test for correlation

A frequent and important question is: "is there any connection between the normal r.v. $\xi$ and $\eta$?"

The base hypothesis usually is: "$H_0 : \text{no correlation between } \xi \text{ and } \eta."$

In other words, $H_0$ says that $r_{\xi,\eta} = 0$.

Algorithm II.86 Calculate $r$ from the dataset as described in section 6.3 "Estimating the correlation coefficient", using (6.20) or (6.21), and calculate

$$t_{sz} = r \cdot \sqrt{\frac{s}{1 - r^2}}$$

(5.11)

where $s = n - 2$ is the degree of freedom.

Pick the critical value $t_\varepsilon$ from the Student t-table, using $s$ and $\varepsilon$.

If $|t_{sz}| \leq t_\varepsilon$ then accept $H_0$, otherwise reject it.

Example II.87 Suppose that $n = 14$ and $r = 0.818505$. Then $s = 12$ and

$t_{sz} = 0.818505 \cdot \sqrt{\frac{12}{1 - 0.818505^2}} \approx 4.9354$. For $\varepsilon_1 = 5\%$ and $\varepsilon_2 = 1\%$ we have $t_{0.05} = 2.179$ and $t_{0.01} = 3.055$. Since $t_{sz} > t_{0.05}$ and $t_{sz} > t_{0.01}$ we have to reject $H_0$ for both $\varepsilon$.

Remark II.88 See also the formulae (6.20) and (6.21) and their role in Section 6.3 "Estimating the correlation coefficient".

5.3.5 Normality testing

Now the base hypothesis is: "$H_0 : \xi \text{ is normal.}\"

Let us mention first the old but illustrative method, called the "Ruler Method" ("vonalzós módszer", see section 6.5.1), which will be explained in more detail in section 6.5 "Nonlinear regressions - linearizing methods" and in [SzI2].

If we are given the dataset $\Xi = \{(x_i, y_i) : i = 1, ..., n\}$ where $x_i$ are arbitrary real numbers and $y_i$ are the measured (or: approximated) value of the probability $P(\xi < x_i)$ then the points must (almost) fit the graph of the distribution function $F_{m,\sigma}(x)$. Have in mind that not only $m$ and $\sigma$ are unknown but even the normality
of $\xi$ is in question! Though we can plot the dataset $\Xi$ to the (usual) coordinate system, how to decide whether they are on (or close to) such a curve?

Since $F_{m,\sigma}$ is a strictly monotone increasing function, we can suitably transform the coordinate system (rarely speaking: we "expand" the $y$ axis in a suitable manner) such that the graphs of all the normal density functions $F_{m,\sigma}$ became (straight) lines, as you can see on the next Figure! This coordinate system is called Gaussian or normal. Placing your ruler on the figure you can justify whether the dataset $\Xi$ fits a line or not, and equivalently, whether the r.v. $\xi$ (measured by $\Xi$) is normal. Moreover, from the "usual" formula $\bar{y} = a\bar{x} + b$ of this line the parameters $m$ and $\sigma$ can be calculated.

Sorry, Excel and many other applications can not handle normal coordinate systems but the webpage [HM] can, please try it! You can find a normal coordinate grid on my webpage as well:

https://math.uni-pannon.hu/~szalkai/koordinata/Gauss-papir-L140-szines.gif

![Figure 2: Gaussian coordinate system](https://math.uni-pannon.hu/~szalkai/koordinata/Gauss-papir-L140-szines.gif)
5.3. NONPARAMETRIC TESTS

Idea of the "modern" algorithm: For any continuous density function $f_0$ (or cumulative distribution function $F_0$) we may ask "is $\xi$ having the distribution function $f_\xi = f_0$ , i.e. $F_\xi = F_0$ ".

For deciding this, divide $\text{Im} (\xi)$ into intervals $[x_{i-1}, x_i)$ with the points $x_0, x_1, ..., x_r$ for $i = 1, ..., r$. Now use the method of Section "Goodness of fit" for the virtual events $A_i$ as: $P (A_i) = F (x_i) - F (x_{i-1}) = p_i$.

Example II.89 We tossed 5 dices many times. The number of occurrences of different sums of the dots is shown in the table. Decide with significance level 95% whether this distribution is normal.

<table>
<thead>
<tr>
<th>sum</th>
<th>freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt; 10$</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>13</td>
<td>55</td>
</tr>
<tr>
<td>14</td>
<td>70</td>
</tr>
<tr>
<td>15</td>
<td>90</td>
</tr>
<tr>
<td>16</td>
<td>95</td>
</tr>
<tr>
<td>17</td>
<td>99</td>
</tr>
<tr>
<td>18</td>
<td>96</td>
</tr>
<tr>
<td>19</td>
<td>85</td>
</tr>
<tr>
<td>20</td>
<td>75</td>
</tr>
<tr>
<td>21</td>
<td>58</td>
</tr>
<tr>
<td>22</td>
<td>99</td>
</tr>
<tr>
<td>23</td>
<td>12</td>
</tr>
<tr>
<td>24</td>
<td>14</td>
</tr>
<tr>
<td>25</td>
<td>22</td>
</tr>
</tbody>
</table>

Solution II.90 $n = 15 + 20 + 30 + 40 + 55 + 70 + 90 + 95 + 99 + 98 + 96 + 85 + 75 + 58 + 35 + 33 + 19 + 22 = 1035$. By symmetry the sum of the dots on 5 dices has mean $M (\xi) = 5 \cdot 3.5 = 17.5 = m$ , the range is $[a, b] = [5, 30]$ , so we assume $\sigma = 2.5$ since by the "3$\sigma$ -rule" we have $P (|\xi - M (\xi)| < 3\sigma) > 0.997$.

For simplifying our calculations use the intervals $[x_0, x_1) = [5, 10)$ , $[x_1, x_2) = [10, 15)$ , $[x_2, x_3) = [15, 20)$ , $[x_3, x_4) = [20, 25)$ , $[x_4, x_5) = [25, 31)$, so we have the following empirical frequency table:

<table>
<thead>
<tr>
<th>nu. of interval (i)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>total (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency (n)</td>
<td>0.0145</td>
<td>0.2077</td>
<td>0.4618</td>
<td>0.2763</td>
<td>0.0396</td>
<td>1035</td>
</tr>
<tr>
<td>relative freq. $(\frac{n}{N})$</td>
<td>0.0145</td>
<td>0.2077</td>
<td>0.4618</td>
<td>0.2763</td>
<td>0.0396</td>
<td>1</td>
</tr>
</tbody>
</table>

The theoretical probabilities are $p_i = F_{m,\sigma} (x_i) - F_{m,\sigma} (x_{i-1})$ , so

$p_1 = F_{m,\sigma} (10) - F_{m,\sigma} (5) = \Phi \left( \frac{10 - 17.5}{2.5} \right) - \Phi \left( \frac{5 - 17.5}{2.5} \right) = \Phi (-3) - \Phi (-5)$

$= (1 - 0.9987) - (1 - 0.9999) = 0.0012$

$p_2 = F_{m,\sigma} (15) - F_{m,\sigma} (10) = \Phi \left( \frac{15 - 17.5}{2.5} \right) - \Phi \left( \frac{10 - 17.5}{2.5} \right) = \Phi (-1) - \Phi (-3)$

$= (1 - 0.8413) - (1 - 0.9987) = 0.1574$

$p_3 = F_{m,\sigma} (20) - F_{m,\sigma} (15) = \Phi \left( \frac{20 - 17.5}{2.5} \right) - \Phi \left( \frac{15 - 17.5}{2.5} \right) = \Phi (1) - \Phi (-1)$

$= 0.8413 - (1 - 0.8413) = 0.6826$

$1) m - 3\sigma = 17.5 - 3 \cdot 2.5 = 10, m + 3\sigma = 17.5 + 3 \cdot 2.5 = 25$. On the other hand: $\sigma_{1\text{ dice}} = \sqrt{\frac{91}{6} - \left( \frac{3}{2} \right)^2} \approx 1.708$ and $\sigma_{5\text{ dice}} = \sqrt{5} \cdot \sigma_{1\text{ dice}} \approx 3.819$. 
\[ p_4 = F_{m,\sigma}(25) - F_{m,\sigma}(20) = \Phi\left(\frac{25-17.5}{2.5}\right) - \Phi\left(\frac{20-17.5}{2.5}\right) = \Phi(3) - \Phi(1) = 0.9987 - 0.8413 = 0.1574 \]

\[ p_5 = F_{m,\sigma}(31) - F_{m,\sigma}(25) = \Phi\left(\frac{31-17.5}{2.5}\right) - \Phi\left(\frac{25-17.5}{2.5}\right) = \Phi(5.4) - \Phi(3) = 0.9999 - 0.9987 = 0.0012 \]

The following table compares empirical and theoretical probabilities:

<table>
<thead>
<tr>
<th>( i )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_i/n )</td>
<td>0.0145</td>
<td>0.2077</td>
<td>0.4618</td>
<td>0.2763</td>
<td>0.0396</td>
<td>1</td>
</tr>
<tr>
<td>( p_i )</td>
<td>0.0012</td>
<td>0.1574</td>
<td>0.6826</td>
<td>0.1574</td>
<td>0.0012</td>
<td>0.9998</td>
</tr>
</tbody>
</table>

\[
\chi^2_{sz} = \sum_{i=1}^{k} \frac{(a_i - np_i)^2}{np_i} = n \sum_{i=1}^{k} \frac{(a_i - np_i)^2}{p_i} = 1035 \cdot \left( \frac{(0.0145 - 0.0012)^2}{0.0012} + \frac{(0.2077 - 0.1574)^2}{0.1574} + \frac{(0.4618 - 0.6826)^2}{0.6826} + \frac{(0.2763 - 0.1574)^2}{0.1574} + \frac{(0.0396 - 0.0012)^2}{0.0012} \right) \approx 1.5535.
\]

Further: \( \varepsilon = 0.05 \), \( s = 5 - 1 = 4 \), \( \chi^2 = 9.488 \), so \( H_0 \) is accepted since \( \chi^2_{sz} < \chi^2_{\varepsilon} \).

End of the solution.  

The "real" probabilities of sums of 5 dices are shown in the Figure below.

![Figure 3: Probabilities of sums of 5 dices](image-url)
Chapter 6

Regression and the least square method

Literary the word "regression" ("regresszió"), or "regression toward the mean" means "turning back", "back looking, -hitting" ("visszatérés, -útés, -tekintés"). The term was first used by Galton \(^1\) when investigating human and biological data. He observed, for example, that the height of children tend to back to the average of the population: if the parents are higher/shorter than the average, then their children are (in average) shorter/higher than their parents, i.e. closer to the average. Of course, this phenomenon is true only in statistical meaning: it is true only for most of the parents and children, i.e. with probability close (but not equal) to 1.

In mathematical statistics we are interested in the type of the connection of two random variables \(\xi\) and \(\eta\) ("new" and "old", "input" and "output", etc.). The covariency \(\text{cov}(\xi, \eta)\) and correlation \(R(\xi, \eta)\) measure only the magnitude of the dependency, now we are interested in the type of the dependency (see the forthcoming sections).


**Remark II.91** If the common/joint distribution function \(F(x, y)\) for \(\xi\) and \(\eta\) is known, the theoretical answer to the above question is easy: the best answer is to approximate \(\eta\) with \(\xi\) is

\[
\eta = m_2(\xi)
\]

(6.1)

---

\(^1\) Sir **Francis Eugene Galton** (1822-1911) English mathematician.
where the function \( m_2 : \mathbb{R} \to \mathbb{R} \) is the conditional mean
\[
m_2(x) = M(\eta \mid \xi = x)
\] (6.2)
which was defined in Section 1.5 "Conditional probability".

The function \( m_2 \) is called \textit{regression function of first kind} (elsőfajú regressziós függvény).

In the case \( \xi \) and \( \eta \) have a normal joint distribution, \( m_2 \) is a linear function:
\[
m_2(x) = ax + b, \text{ i.e. } \eta = a\xi + b \text{ for some real numbers } a, b \in \mathbb{R} \text{ (which can be computed from the mean and variance of } \xi \text{ and } \eta).}
\]

However, in practice we have to find much easier methods for calculating the connection between \( \xi \) and \( \eta \). In what follows, \( \xi \) and \( \eta \) are any r.v. on a (common) sample space \( \Omega \).

Theoretically we deal with random variables \( \xi \) and \( \eta \), but in practice we have only a set of (measured) corresponding data \( \xi_i \) and \( \eta_i \) as \( \{(\xi_i, \eta_i) : i = 1, \ldots, n\} \). As in the Introduction of Statistics we learned, \( \xi_i \) and \( \eta_i \) are, in fact, real numbers (in our notepad), we could write \( x_i \) and \( y_i \) instead. Since after repeated measurements they often vary, they are called r.v. in theory. This is the reason that most of the theorems have \textit{two versions} (see e.g. Theorem II.95): one for r.v. and the other for the dataset \( \{(\xi_i, \eta_i) : i = 1, \ldots, n\} \). If you like, you can (advised to) think of \( \xi_i \) and \( \eta_i \) as real numbers, or even \( x_i \) and \( y_i \).

In mathematics we use(d) variables \( x \) and \( y \) as \( y = f(x) \), but in the context of \( \xi \) and \( \eta \) we have to write them like \( g(\xi), \eta \approx g(\xi), (a\xi + b) - \eta_i \), etc. In this chapter we mix these two notations, you can also turn \( \xi \) and \( \eta \) to \( x \) and \( y \) if you like.

### 6.1 The general case

First we define the general problem we want to solve in this chapter. The general problem and solution methods will be explained in the special cases.

**Definition II.92** We are given the r.v. \( \eta \) and \( \xi \), or the dataset
\[
\{(\xi_i, \eta_i) : i = 1, \ldots, n\}.
\] (6.3)

We are looking for the function \( g : \mathbb{R} \to \mathbb{R} \) such that the r.v. \( g(\xi) \) is the closest one to \( \eta \). The difference is measured by
\[
M((g(\xi) - \eta)^2)
\] (6.4)
and by
\[ \sum_i [g(\xi_i) - \eta_i]^2 \] (6.5)
respectively, i.e. we want to minimize the quantities in (6.4) and in (6.5).
More precisely, we have to choose \( g \) from a given type of functions with parameters, i.e. in fact
\[ g(\xi) = g(\xi, a_1, ..., a_m) \] (6.6)
and we have to find the parameter values which minimize (6.4) and (6.5).

Remark II.93 (i) The quantities (6.4) and (6.5) are similar to the definition of the variance. Again, the square eliminates + and – values, and corrects the magnitude of small and large numbers.

(ii) The problem and the solution are called Least Squares Method (“legkisebb négyzetek módszere”), since we want to minimize the mean (sum) of squares of the differences of \( g(\xi_i) \) and \( \eta_i \). There is a slight similarity between (6.4) and the definition of the variance.

6.2 Linear regression

("Lineáris regresszió")

The easiest formula is \( g(x) = ax + b \ (a, b \in \mathbb{R}) \). The approximation question "\( \eta \approx a\xi + b \)" can be raised for any r.v. \( \xi \) and \( \eta \), the error is investigated in the next section "Regression and covariance" in Theorem II.102, graphical illustration is detailed in the section "The ruler method".

Other approximations, like
\[ \eta \approx a_0 + a_1\xi + a_2\xi^2 + ... + a_n\xi^n \] (6.7)
(polinomial regression) can also be applied in various applications. Let us emphasize, that enlarging the number of the unknown parameters \( a_0, ...a_n \) (not only in polinomial but also in other types of regression), in general does not increase the accuracy of the approximation of \( \eta \), since \( a_0, ...a_n \) are all not real numbers but random variables.

Problem II.94 Determine \( a, b \in \mathbb{R} \) such that
\begin{itemize}
  \item[i)] \( M ([a\xi + b - \eta]^2) \) or
  \item[ii)] \( \sum_{i=1}^{n} (a\xi_i + b - \eta_i)^2 \)
\end{itemize}
is minimal:
Theorem II.95  

i) For \( M \left( [a \xi + b - \eta]^2 \right) \) minimal we have

\[
a = \frac{M(\xi \eta) - M(\xi) M(\eta)}{M(\xi^2) - M^2(\xi)}
\]  

(6.8)

and

\[
b = M(\eta) - a \cdot M(\xi)
\]

(6.9)

or, in another forms:

\[
a = \frac{\text{cov}(\xi, \eta)}{D^2(\xi)} = R(\xi, \eta) \cdot \frac{D(\eta)}{D(\xi)}
\]  

(6.10)

and

\[
b = M(\eta) - M(\xi) \cdot \frac{M(\xi \eta) - M(\xi) M(\eta)}{M(\xi^2) - M^2(\xi)}.
\]  

(6.11)

ii) For \( \sum_{i=1}^{n} (a \xi_i + b - \eta_i)^2 \) minimal we have

\[
a = \frac{n \cdot \sum_{i=1}^{n} \xi_i \eta_i - \left( \sum_{i=1}^{n} \xi_i \right) \left( \sum_{i=1}^{n} \eta_i \right)}{n \cdot \sum_{i=1}^{n} \xi_i^2 - \left( \sum_{i=1}^{n} \xi_i \right)^2}
\]  

(6.12)
6.2. LINEAR REGRESSION

and

\[ b = \frac{1}{n} \left( \sum_{i=1}^{n} \eta_i - a \cdot \sum_{i=1}^{n} \xi_i \right), \tag{6.13} \]

or, in another forms:

\[ a = \frac{\bar{\xi} \bar{\eta} - \bar{\xi} \cdot \bar{\eta}}{\sigma_{\xi}^2} = \frac{\sum_{i=1}^{n} (\xi_i - \bar{\xi}) (\eta_i - \bar{\eta})}{\sum_{i=1}^{n} (\xi_i - \bar{\xi})^2}, \tag{6.14} \]

and

\[ b = \bar{\eta} - a \cdot \bar{\xi} \tag{6.15} \]

Remark II.96 (i) We listed all possible formulae for \( a \) and \( b \), please choose your favourite one! Or, you might use any computer program, like Excel, to calculate \( a \) and \( b \).

(ii) In the case \( M(\xi) = 0 \) or \( \sum_{i=1}^{n} \xi_i = 0 \), i.e. when the dataset \( \{\xi_i : i = 1, ..., n\} \) is symmetric to the origin, the above formulas have much simpler form

\[ a = \frac{\sum_{i=1}^{n} \xi_i \cdot \eta_i}{\sum_{i=1}^{n} \xi_i^2} \quad \text{and} \quad b = \frac{1}{n} \sum_{i=1}^{n} \eta_i. \tag{6.16} \]

The symmetric property can be easily achieved by using \( \xi' := \xi - m_\xi \) and \( \xi'_i := \xi_i - \bar{\xi} \) instead of \( \xi \) and \( \xi_i \) where \( m_\xi = M(\xi) \) and \( \bar{\xi} = \frac{1}{n} \sum_{i=1}^{n} \xi_i \).

(iii) The function

\[ \eta = R \cdot \frac{\sigma_\eta}{\sigma_\xi} \cdot (x - m_\xi) + m_\eta \tag{6.17} \]

is called regression function of second kind (másodfajú regressziós függvény), which corresponds to \eqref{6.10} and \eqref{6.9}, of course \( R = R(\xi, \eta) \).

In the special case, when the regression function of first kind is a linear function, then these two kinds of regression functions \eqref{6.1} and \eqref{6.17} coincide.

Proof. of Theorem II.95

i) We have to find the minimum value of the two-variable function

\[ F(a, b) := M ([a\xi + b - \eta] \cdot \xi^2) \] . It is wellknown, that in this case the partial derivatives must be zero: \( \frac{\partial F}{\partial a} = 0 \) and \( \frac{\partial F}{\partial b} = 0 \), this system of equalities (see \eqref{6.18}...
below) has the solution shown in (6.8) and (6.9). In detail:

\[ F(a, b) = M(a^2 \xi^2 + b^2 + \eta^2 + 2ab \xi - 2a \xi \eta - 2b \eta) = a^2 M(\xi^2) + b^2 + M(\eta^2) + 2ab M(\xi) - 2a M(\xi \eta) - 2b M(\eta), \]

\[ \frac{\partial F}{\partial a} = 2a M(\xi^2) + 2b M(\xi) - 2M(\xi \eta), \]

\[ \frac{\partial F}{\partial b} = 2b + 2a M(\xi) - 2M(\eta), \]

so the system of equalities we have to solve is:

\[ \begin{align*}
& a M(\xi^2) + b M(\xi) = M(\xi \eta) \\
& a M(\xi) + b = M(\eta)
\end{align*} \quad (6.18) \]

The solution is

\[ a = \frac{\det \begin{bmatrix} M(\xi \eta) & M(\xi) \\ M(\eta) & 1 \end{bmatrix}}{\det \begin{bmatrix} M(\xi^2) & M(\xi) \\ M(\xi) & 1 \end{bmatrix}} = \frac{M(\xi \eta) - M(\xi) M(\eta)}{M(\xi^2) - M^2(\xi)} = \frac{\text{cov}(\xi, \eta)}{D^2(\xi)}, \]

\[ b = M(\eta) - a \cdot M(\xi), \]

justifying (6.8) and (6.9).

One can easily check that the (unique) solution of (6.18) is (6.8) and (6.9). However do not forget, that the equalities \( \frac{\partial F}{\partial a} = \frac{\partial F}{\partial b} = 0 \) are only necessary conditions for the extreme value(s) of \( F \), one should check that the solution ((6.8),(6.9)) really gives a minimum. However:

\[ \frac{\partial^2 F}{\partial a^2} = 2M(\xi^2), \quad \frac{\partial^2 F}{\partial b^2} = 2, \quad \frac{\partial^2 F}{\partial ab} = 2M(\xi), \]

\[ \Delta(a, b) = 4M(\xi^2) - 4M^2(\xi) = 4D^2(\xi) > 0 \quad \text{and} \quad \frac{\partial^2 F}{\partial a^2} = 2M(\xi^2) > 0. \]

**ii)** Since the real numbers \( \xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_n \) are given (fixed), we have to find the minimum value of the two-variable function \( H(a, b) := \sum_{i=1}^{n} (a \xi_i + b - \eta_i)^2 \), similarly to case i):

\[ H(a, b) = \sum_{i=1}^{n} \left(a^2 \xi_i^2 + 2ab \xi_i - 2a \xi_i \eta_i + b^2 - 2b \eta_i + \eta_i^2\right) = \]

\[ = a^2 \sum_{i=1}^{n} \xi_i^2 + 2ab \sum_{i=1}^{n} \xi_i - 2a \sum_{i=1}^{n} \xi_i \eta_i + nb^2 - 2b \sum_{i=1}^{n} \eta_i + \sum_{i=1}^{n} \eta_i^2, \]
6.2. LINEAR REGRESSION

\[ \frac{\partial H}{\partial a} = 2a \sum_{i=1}^{n} \xi_i^2 + 2b \sum_{i=1}^{n} \xi_i - 2 \sum_{i=1}^{n} \xi_i \eta_i, \]

\[ \frac{\partial H}{\partial b} = 2a \sum_{i=1}^{n} \xi_i + 2nb - 2 \sum_{i=1}^{n} \eta_i, \]

so the system of equalities:

\[ \begin{align*}
    a \sum_{i=1}^{n} \xi_i^2 + b \sum_{i=1}^{n} \xi_i &= \sum_{i=1}^{n} \xi_i \eta_i \\
    a \sum_{i=1}^{n} \xi_i + bn &= \sum_{i=1}^{n} \eta_i
\end{align*} \]

has the solution

\[ a = \frac{\det \begin{bmatrix} \sum_{i=1}^{n} \xi_i \eta_i & \sum_{i=1}^{n} \xi_i \\ \sum_{i=1}^{n} \eta_i & n \end{bmatrix}}{\det \begin{bmatrix} \sum_{i=1}^{n} \xi_i^2 & \sum_{i=1}^{n} \xi_i \\ \sum_{i=1}^{n} \xi_i & n \end{bmatrix}} = \frac{n \cdot \left( \sum_{i=1}^{n} \xi_i \eta_i \right) - \left( \sum_{i=1}^{n} \xi_i \right) \left( \sum_{i=1}^{n} \eta_i \right)}{n \cdot \sum_{i=1}^{n} \xi_i^2 - \left( \sum_{i=1}^{n} \xi_i \right)^2}, \]

\[ b = \frac{1}{n} \left( \sum_{i=1}^{n} \eta_i - a \cdot \sum_{i=1}^{n} \xi_i \right) = \bar{\eta} - a \cdot \bar{\xi}, \]

which coincide with (6.12) and (6.13). Checking whether ((6.12), (6.13)) solve (6.19) and really give (absolute) minimum of \( H \) is left to the Reader.

Now we show that \( a \) is equivalent to (6.14), (6.15) is obvious.

\[ \sum_{i=1}^{n} \left( \xi_i - \bar{\xi} \right) \left( \eta_i - \bar{\eta} \right) = \sum_{i=1}^{n} \left( \xi_i \eta_i - \xi_i \bar{\eta} - \bar{\xi} \eta_i + \bar{\xi} \bar{\eta} \right) \]

\[ = \sum_{i=1}^{n} \left( \xi_i - \bar{\xi} \right)^2 = \sum_{i=1}^{n} \left( \xi_i^2 - 2\xi_i \bar{\xi} + (\bar{\xi})^2 \right) \]

\[ = \sum_{i=1}^{n} \xi_i \eta_i - \bar{\eta} \cdot \sum_{i=1}^{n} \xi_i - \bar{\xi} \cdot \sum_{i=1}^{n} \eta_i + n \cdot \bar{\xi} \bar{\eta} \]

\[ = \frac{\sum_{i=1}^{n} \xi_i^2 - 2\xi \cdot \sum_{i=1}^{n} \xi_i + n \cdot (\bar{\xi})^2}{\frac{n \cdot \bar{\xi} \eta - \bar{\eta} \cdot \bar{\xi} - \bar{\xi} \cdot \bar{\eta} + n \cdot \bar{\xi} \cdot \bar{\eta}}{n \cdot \bar{\xi}^2 - 2\xi \cdot n \cdot \bar{\xi} + n \cdot (\bar{\xi})^2}} \]

\[ = \frac{n \cdot \bar{\xi} \eta - \bar{\eta} \cdot n \cdot \bar{\xi} - \bar{\xi} \cdot n \cdot \bar{\eta} + n \cdot \bar{\xi} \cdot \bar{\eta}}{n \cdot \bar{\xi}^2 - 2\xi \cdot n \cdot \bar{\xi} + n \cdot (\bar{\xi})^2}, \]  (*)
CHAPTER 6. REGRESSION AND THE LEAST SQUARE METHOD

one hand \((*) = \frac{\xi \eta - \bar{\xi} \cdot \bar{\eta}}{\xi^2 - (\bar{\xi})^2} = \frac{\xi \eta - \bar{\xi} \cdot \bar{\eta}}{\sigma^2_{\xi}}\),

other hand \((*) = \frac{n \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i \eta_i - n \cdot \frac{1}{n^2} \cdot \sum_{i=1}^{n} \xi_i \cdot \sum_{i=1}^{n} \eta_i}{n \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i^2 - n \cdot \left(\frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i\right)^2} = \frac{\sum_{i=1}^{n} \xi_i \eta_i - \frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i \cdot \sum_{i=1}^{n} \eta_i}{\sum_{i=1}^{n} \xi_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} \xi_i\right)^2} = \frac{n \cdot \sum_{i=1}^{n} \xi_i \eta_i - \left(\sum_{i=1}^{n} \xi_i\right) \cdot \left(\sum_{i=1}^{n} \eta_i\right)}{n \cdot \sum_{i=1}^{n} \xi_i^2 - \left(\sum_{i=1}^{n} \xi_i\right)^2} = \frac{n \cdot \sum_{i=1}^{n} \xi_i \eta_i - \left(\sum_{i=1}^{n} \xi_i\right) \cdot \left(\sum_{i=1}^{n} \eta_i\right)}{n \cdot \sum_{i=1}^{n} \xi_i^2 - \left(\sum_{i=1}^{n} \xi_i\right)^2} = \frac{\sum_{i=1}^{n} \xi_i \eta_i - \frac{1}{n} \cdot \sum_{i=1}^{n} \xi_i \cdot \sum_{i=1}^{n} \eta_i}{\sum_{i=1}^{n} \xi_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} \xi_i\right)^2}.

End of Proof. ■

Remark II.97 Though using the formulae from (6.12) to (6.15) of Theorem II.95 one can compute \(a\) and \(b\) for the line \(ax + b\). However these computations are difficult for large or many datasets. For approximate values of \(a\) and \(b\) the "Ruler Method" was applied a couple of years ago (before the computers). Roughly speaking, plot the data \((x_i, y_i)\) to a grid on a suitable coordinate system, and fit a (straight) ruler to your drawing. This method is detailed in subsections 6.5.1 "The Ruler Method" and after, for various coordinate systems.

6.3 Estimating the correlation coefficient

Before investigating the connecttion between regression and covariance, first we have to learn how to approximate \(R(\xi, \eta)\) from the dataset (6.3). If you are interested in r.v. \(\xi\) and \(\eta\), you may skip this section.

\(R(\xi, \eta)\) was introduced and discussed (theoretically) in Definition I.12 in Section 1.1. Now we have to give an empirical estimation for \(R(\xi, \eta)\).

By \(R(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{D(\xi)D(\eta)} = \frac{\text{M}[\xi \cdot (\xi - M(\xi)) \cdot (\eta - M(\eta))]}{D(\xi)D(\eta)}\) our choice is

\[r_{\xi, \eta} = \frac{1}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2}} \cdot \frac{1}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (\eta_i - \bar{\eta})^2}} = \frac{1}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2}} \cdot \frac{1}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (\eta_i - \bar{\eta})^2}} = \frac{\sum_{i=1}^{n} [(\xi_i - \bar{\xi}) \cdot (\eta_i - \bar{\eta})]}{n \cdot \sigma_{\xi} \cdot \sigma_{\eta}}, \quad (6.20)\]
which is equivalent to the easier (for hand-calculations) formula
\[
r_{\xi,\eta} = \frac{n \cdot \sum_{i=1}^{n} \xi_i \eta_i - \left( \sum_{i=1}^{n} \xi_i \right) \left( \sum_{i=1}^{n} \eta_i \right)}{\sqrt{n \cdot \sum_{i=1}^{n} \xi_i^2 - \left( \sum_{i=1}^{n} \xi_i \right)^2} \cdot \sqrt{n \cdot \sum_{i=1}^{n} \eta_i^2 - \left( \sum_{i=1}^{n} \eta_i \right)^2}}. 
\] (6.21)

The above (6.20) and (6.21) formulae are in strict connection with (5.11) in Subsection 5.3.4 "Test for correlation" in Section 5.3.

Example II.98 Consider the morning and afternoon values of our activity for 10 days. Does any connection exist between them?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning (ξ)</td>
<td>8.2</td>
<td>9.6</td>
<td>7.0</td>
<td>9.4</td>
<td>10.9</td>
<td>7.1</td>
<td>9.0</td>
<td>6.6</td>
<td>8.4</td>
<td>10.5</td>
</tr>
<tr>
<td>Afternoon (η)</td>
<td>8.7</td>
<td>9.6</td>
<td>6.9</td>
<td>8.5</td>
<td>11.3</td>
<td>7.6</td>
<td>9.2</td>
<td>6.3</td>
<td>8.4</td>
<td>12.3</td>
</tr>
</tbody>
</table>

Solution II.99 \( n = 10 \), \( \sum \xi_i = 86.7 \), \( \sum \xi_i^2 = 771.35 \), \( \sum \eta_i = 88.8 \), \( \sum \eta_i^2 = 819.34 \), \( \sum \xi_i \eta_i = 792.92 \), so
\[
r = \frac{10 \cdot 792.92 - 86.7 \cdot 88.8}{\sqrt{10 \cdot 771.35 - (86.7)^2} \cdot \sqrt{10 \cdot 819.34 - (88.8)^2}} \approx 0.9357.
\]
This means, that the connection between \( \xi \) and \( \eta \) is strong.

6.4 Regression and covariance

In Section 1.1 "Two dimensional ... General definitions" and in subsection 5.3.3 "Independence" we discussed how the value of the correlation coefficient \( R(\xi, \eta) \) depends on the strength of the connection between \( \xi \) and \( \eta \). In this section we investigate this dependency in more detail.

Definition II.100 Let
\[
\omega := a\xi + b - \eta \quad \text{and} \quad \omega_i := a\xi_i + b - \eta_i \tag{6.22}
\]
the error - random variable and the error - data, i.e. the difference between \( a\xi + b \) and \( \eta \), and between \( a\xi_i + b_i \) and \( \eta \). □
Recall, that in Theorem II.95 we achieved \( M(\omega^2) \) to be minimal, finding the suitable \( a \) and \( b \). Now we determine this minimal value of error.

**Proposition II.101** If \( a \) and \( b \) are determined as in Theorem II.95, then \( M(\omega) = 0 \) and \( \bar{\omega} = 0 \), so \( D^2(\omega) = M(\omega^2) \) and \( \sigma_a^2 = \bar{\omega}^2 \).

**Proof.** We use only \( b = M(\eta) - a \cdot M(\xi) \). Then
\[
M(\omega) = M(a\xi + M(\eta) - aM(\xi) - \eta) = aM(\xi) + M(\eta) - aM(\xi) - M(\eta) = 0,
\]
so \( D^2(\omega) = M(\omega^2) \) follows.

Similarly, using \( b = \bar{\eta} - a \cdot \bar{\xi} \) we have
\[
\bar{\omega} = a \cdot \bar{\xi} + b - \bar{\eta} = a \cdot \bar{\xi} + (\bar{\eta} - a \cdot \bar{\xi}) - \bar{\eta} = 0. \quad \blacksquare
\]

**Theorem II.102** If \( a \) and \( b \) are determined as in Theorem II.95, then

\[
\begin{align*}
\text{i)} & \quad D^2(\omega) = D^2(\eta) \cdot (1 - R^2(\xi, \eta)) \\
\text{ii)} & \quad \sigma_a^2 = \sigma_\eta^2 \cdot (1 - r_{\xi,\eta}^2) .
\end{align*}
\]

**Proof.** i) Using \( b = M(\eta) - a \cdot M(\xi) \) we have
\[
D^2(\omega) = M(\omega^2) = M([a\xi + b - \eta]^2)
= M([a\xi + M(\eta) - a \cdot M(\xi) - \eta]^2) = M([a \cdot (\xi - M(\xi)) - (\eta - M(\eta))]^2)
= M(a^2 \cdot (\xi - M(\xi))^2 + (M(\eta) - \eta)^2 - 2a \cdot (\xi - M(\xi)) \cdot (\eta - M(\eta)))
= a^2 \cdot D^2(\xi) + D^2(\eta) - 2a \cdot \text{cov}(\xi, \nu) = (*).
\]

Now use \( a = \frac{\text{cov}(\xi, \eta)}{D^2(\xi)} \) and continue as
\[
(*) = \frac{\text{cov}^2(\xi, \eta)}{D^2(\xi)} + D^2(\eta) - 2 \cdot \frac{\text{cov}^2(\xi, \eta)}{D^2(\xi)} = D^2(\eta) - \frac{\text{cov}^2(\xi, \eta)}{D^2(\xi)} = D^2(\eta) \cdot (1 - R^2(\xi, \eta)).
\]

ii) Using \( b = \bar{\eta} - a \cdot \bar{\xi} \) we have
\[
\begin{align*}
\sigma_a^2 &= \bar{\omega}^2 = \frac{1}{n} \sum_{i=1}^{n} \left[a\xi_i + (\bar{\eta} - a \cdot \bar{\xi}) - \eta_i\right]^2
= \frac{1}{n} \sum_{i=1}^{n} \left[a \cdot (\xi_i - \bar{\xi}) - (\eta_i - \bar{\eta})\right]^2
= \frac{1}{n} a^2 \cdot \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2 - 2a \cdot \frac{1}{n} \sum_{i=1}^{n} (\eta_i - \bar{\eta}) \cdot (\xi_i - \bar{\xi}) + \frac{1}{n} \sum_{i=1}^{n} (\eta_i - \bar{\eta})^2
= a^2 \cdot \sigma_\xi^2 - 2a \cdot \frac{1}{n} \sum_{i=1}^{n} (\eta_i - \bar{\eta}) (\xi_i - \bar{\xi}) + \sigma_\eta^2 .
\end{align*}
\]
Now use (6.14) twice and then (6.20) to continue
\[ a^2 \cdot \sigma^2_{\xi} - 2a^2 \cdot \frac{1}{n} \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2 + \sigma^2_{\eta} = -a^2 \cdot \sigma^2_{\xi} + \sigma^2_{\eta} \]
\[ = \sigma^2_{\eta} - \sigma^2_{\xi} \cdot \left( \frac{1}{n} \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2 \right)^2 \]
\[ = \sigma^2_{\eta} - \sigma^2_{\xi} \cdot \left( \frac{1}{n} \sum_{i=1}^{n} (\eta_i - \bar{\eta})^2 \right)^2 \]
\[ = \sigma^2_{\eta} - \sigma^2_{\xi} \cdot \left( \frac{r_{\xi \eta}}{\sigma_{\xi}} \right)^2 = \sigma^2_{\eta} \cdot (1 - r^2_{\xi \eta}). \]

**Remark II.103**

**i** First we can justify Theorem 1.14 from Section 1.1 "Two dimensional ... General definitions" stating \( R(\xi, \eta) = 1 \) if and only if \( \eta = a\xi + b \) for some numbers \( a, b \in \mathbb{R} \). By (6.23) we can conclude that \( R(\xi, \eta) = 1 \) exactly when \( D^2(\omega) = 0 \). We know from elementary probability theory, that \( D^2(\omega) = 0 \) corresponds to \( \omega = c \) (\( c \in \mathbb{R} \) constant), i.e. \( \omega = a\xi + b - \eta = c \) which is minimal exactly when \( c = 0 \) i.e. \( \eta = a\xi + b \). So, \( R(\xi, \eta) \) is "close to 1" just in case when the datapoints are almost on a (straight) line.

**ii** On the other hand, the case \( R(\xi, \eta) = 0 \) (i.e. \( \xi \) and \( \eta \) are uncorrelated) together with (6.10) implies \( a = 0 \), i.e. the (approximating) regression line must be horizontal, see Figure 4. In this case, e.g. by (6.25) \( D^2(\omega) = D^2(\eta) \) which must not be surprising, since, by the horizontal line the differences of \( \eta \) and \( b \) (\( = \omega \)) are equal to the differences of \( \eta \) and \( M(\eta) \) (see (6.9)).

**iii** Figure 6 shows different datasets with the same \( r = R(\xi, \eta) \), illustrating, that \( R(\xi, \eta) \) measures only (approximately) the magnitude of the correlation, not the exact correspondence between \( \xi \) and \( \eta \), \( r = 0.816 \) (this example by Anscombe\(^2\)).

**iv** The formula (6.23) is equivalent to
\[ |R(\xi, \eta)| = \sqrt{1 - \frac{D^2(\omega)}{D^2(\eta)}} \]
(6.25)
where, of course \( \omega \) is the error in (6.22) for the optimal parameters \( a \) and \( b \).

The formula \( \sqrt{1 - \frac{D^2(\omega)}{D^2(\eta)}} \) for any \( \xi \) and \( \eta \) (i.e. for any \( a \) and \( b \)) is often called **correlation index** ("korrelációs index") and denoted by \( I(\xi, \eta) \). Let us highlight that \( I(\xi, \eta) \) and \( |R(\xi, \eta)| \) correspond only when \( a \) and \( b \) are optimal.

---

\(^2\) Francis John Anscombe (1918-2001) was an English statistician.
Figure 5: Different regression values

Figure 6: Same regression values after Anscombe ($r = 0.816$)
6.5 Nonlinear regressions - linearizing methods

Some function equalities \( y = g(x) \) can be transformed to a linear connection

\[
\tilde{y} = \tilde{a} \cdot \tilde{x} + \tilde{b}
\]

(6.26)

for some transformed values \( \tilde{x} \) and \( \tilde{y} \) of \( x \) and \( y \), with real numbers \( \tilde{a} \) and \( \tilde{b} \). (In the language of statistics we write \( \eta = g(\xi) \) and \( \tilde{y} = \tilde{a} \xi + \tilde{b} \).) If, moreover the transformation of \( x \) and \( y \) to \( \tilde{x} \) and \( \tilde{y} \) can be done graphically (see below), then the simple but illustrative "ruler method" (see below) can be applied. Graphical transformation means that we do not draw the datapoints \((\xi_i, \eta_i)\) and/or the function \( \eta = g(\xi) \) in the usual Cartesian coordinate system but in another, modified one. (Examples with figures are given in the subsequent subsections.) In modified coordinate systems the values "\( x \)" and "\( y \)" are written not in the geometric (real) distance but in \( \tilde{x} \) and \( \tilde{y} \), i.e. we have logarithmic or other scales on the axes, instead of the usual equidistant ones. This results that the graph of the function \( y = g(x) \) is transformed to be linear. The theory of such "linearizing methods" is explained in [SzI2], a computer program (application) for drawings is in [HM]. Please, try it! Other computer programs, like Excel is familiar with some, but not all of these transformations. Illustrative applications can be learned in Section 5.3.5 "Normality testing" and in the subsequent ones.

After the transformation (6.26) we can apply the formulas of Theorem II.95 directly to the dataset \( \{(\xi_i, \eta_i) : i = 1, \ldots, n\} \) to get the values of \( \tilde{a} \) and \( \tilde{b} \) in (6.26).

Be careful: the error \( M \left( \tilde{b} + \tilde{a} \xi - \tilde{\eta} \right)^2 \) in (6.26) is not the same as in the original (6.4), even it might not be minimal at the same values at \( a, b \) and at \( \tilde{a}, \tilde{b} \) ! We make only simpler and approximate computations.

We give some more accurate investigations and computations of (6.4) in Section 6.6 Nonlinear regressions - direct methods.

---

3) We use here the accent \( \tilde{x} \) instead of \( \hat{x} \) since \( \hat{x} \) is used for another notion in Statistics.
6.5.1 The Ruler Method

Looking at Figure 4 in Section "Linear regression" we can imagine the following illustrative method for (straight) line fitting. After dotting the dataset to the coordinate grid, take a common ruler and fit it manually to the dataset, so that the ruler can fit the set of dots in the best ("closest") way. From the position of this ruler you can determine the slope ($a$) and the intersection value ($b$) of the wanted line $y = ax + b$. You might fit your ruler to the monitor of your computer when using [HM] or Excel. This method (modifying the coordinate scales) is widely used not only in statistics but in all natural sciences (physics, chemics, biology, astronomy, economy, etc.)

In the following subsections we learn several methods to transform various function graphs into (straight) lines, in order to apply either the formulas of Theorem 11.95 or to use "The Ruler Method" for those function graphs, too. On the webpage [HM] you can display (almost) any function in all coordinate systems. Please try it! Figure 2 in Section 5.3.5 Normality testing also used a coordinate transformation (called normal) to straighten cumulative distribution functions, the program (application) on [HM] can handle normal coordinate transformations, too.

6.5.2 Exponential regression

The function equality turns to

$$\eta = b \cdot a^c \xi$$

$$\lg(\eta) = \lg(b) + \xi \cdot c \cdot \lg(a),$$

or in short form to

$$\tilde{\eta} = \tilde{b} + \xi \cdot \tilde{a}$$

when applying $\lg$ to \eqref{6.27}, i.e. $\ell(x) = \lg(x)$, $\tilde{\eta} = \lg(\eta)$, $\tilde{\xi} = \lg(\xi)$, $\tilde{a} = c \cdot \lg(a)$ and $\tilde{b} = \lg(b)$.

This means that we can use the linear regression method to the (similarly transformed) dataset

$$\left(\xi_i, \tilde{\eta}_i\right) := (\xi_i, \lg \eta_i) \quad (i = 1, \ldots, n),$$

---

4) This approximative method was widely used till the mid of XX century for easier problems. See also the section "Normality Testing".

5) The equality $\eqref{6.27} \eta = b \cdot a^c \xi$ can be written in the form $\eta = b \cdot d^c$ where $d = a^c$, so $c$ can be eliminated.
6.5. NONLINEAR REGRESSIONS - LINEARIZING METHODS

so \( \tilde{a} \) and \( \tilde{b} \) can be computed from the formulae of Theorem II.95. Finally we must not forget to use

\[
a = \exp \left( \frac{\tilde{a}}{c} \right) = e^{\tilde{a}/c} \quad \text{and} \quad b = \exp \left( \tilde{b} \right) = e^{\tilde{b}}
\]

(6.31)
to get \( a \) and \( b \) (for the expression (6.27)).

Using semilogarithmic coordinate system, i.e. logarithmic scale one axe (now \( \eta \)) and usual (equidistant) scale on the other axe (now \( \xi \)).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.png}
\caption{Exponential function in Cartesian (left), in semilogarithmic (medium) coordinate systems, and its transform by (6.28) (right)}
\end{figure}

On the webpage [HM] you can display any exponential (and any other) function in the semilogarithmic coordinate system as well.


\footnote{The word "semi" means "half".}
6.5.3 Logarithmic regression

Now we have the function equality

$$\eta = a \cdot \lg (\xi) + b,$$  \hspace{1cm} (6.32)

which is itself linear in $\tilde{\xi} = \lg (\xi)$ and $\tilde{\eta} = \eta$, i.e. $\ell (x) = \lg (x)$, $\bar{a} = a$ and $\bar{b} = b$. This means, that we can use the linear regression method to the (similarly transformed) dataset in (6.30) and we immediately get $a$ and $b$.

We have to use semilogarithmic coordinate system again, but now we need logarithmic scale on the axe $\xi$ and equidistant scale on the axe $\eta$.

![Figure 8: Logarithmic function in Cartesian and in semilogarithmic coordinate systems](http://math.uni-pannon.hu/~szalkai/koordinata/semilog-uj-f.jpg) and on ![http://math.uni-pannon.hu/~szalkai/koordinata/semilog-uj-hata.jpg](http://math.uni-pannon.hu/~szalkai/koordinata/semilog-uj-hata.jpg) we supply semi-log coordinate drawings in high resolution.
### 6.5.4 Power regression

The function

\[ \eta = b \cdot \xi^a \]  

(6.33)

turns to

\[ \lg(\eta) = a \cdot \lg(\xi) + \lg(b) \]  

(6.34)

or in short form to

\[ \tilde{\eta} = a \cdot \tilde{\xi} + \tilde{b} \]  

(6.35)

where \( \tilde{\eta} = \lg(\eta), \tilde{\xi} = \lg(\xi), \tilde{a} = a \) and \( \tilde{b} = b \). Now use the linear regression method to the dataset \( (\tilde{\xi}_i, \tilde{\eta}_i) \) := \( (\lg \xi_i, \lg \eta_i) \), compute \( \bar{a} \) and \( \bar{b} \) from Theorem II.95, and use

\[ a = \bar{a} \quad \text{and} \quad b = \exp(\bar{b}) = e^{\bar{b}}. \]  

(6.36)

In this case we have to use the (double) logarithmic coordinate system, i.e. logarithmic scale on both axes.

On the Figure below we see power functions for different exponents.

![Figure 9: Power functions in Cartesian and in (double) logarithmic coordinate systems](http://math.uni-pannon.hu/~szalkai/koordinata/loglog-uj-f.jpg)

On [http://math.uni-pannon.hu/~szalkai/koordinata/loglog-uj-f.jpg](http://math.uni-pannon.hu/~szalkai/koordinata/loglog-uj-f.jpg) we supply a loglog coordinate drawing *in high resolution.*
6.5.5 Hiperbolic regression

The general hiperbolic function ("inverted relations", "fordított arányosságok") has the form

\[ \eta = \frac{\alpha \xi + \beta}{\gamma \xi + \delta} \]  
(6.37)

which can not be linearized, in general, since it has four undefined constants \((\alpha, \beta, \gamma, \delta)\). Though we can simplify by one of them (which is nonzero), e.g. by \(\alpha \neq 0\) gives (6.37) to

\[ \eta = \frac{\xi + \beta/\alpha}{(\gamma/\alpha) \cdot \xi + \delta/\alpha} = \frac{\xi + \beta'}{\gamma' \xi + \delta'} , \]  
(6.38)

i.e. we actually still have three undefined constants, which are still more than two.

So, we must eliminate one of the constants \(\alpha, \beta, \gamma, \delta\).

**Theorem II.104** The function (6.37) has the following forms when one of the parameters is zero (using \(\xi = 1/\xi\) and \(\eta = 1/\eta\):

I) if \(\alpha = 0\) (and \(\beta \neq 0\)) then \(1 - \eta = \frac{\gamma \cdot \xi + \delta}{\beta} \), i.e. \(\bar{\eta} = \gamma' \xi + \delta'\),

II) if \(\beta = 0\) (and \(\alpha \neq 0\)) then \(1 - \eta = \frac{\delta}{\alpha} \cdot \frac{1}{\xi} + \frac{\gamma}{\alpha} \), i.e. \(\bar{\eta} = \delta' \xi + \gamma'\),

III) if \(\gamma = 0\) (and \(\delta \neq 0\)) then \(\eta = \frac{\alpha}{\delta} \cdot \xi + \frac{\beta}{\delta} = \alpha' \xi + \beta'\),

IV) if \(\delta = 0\) (and \(\gamma \neq 0\)) then \(\eta = \frac{\alpha}{\gamma} + \frac{\beta}{\gamma} \cdot \frac{1}{\xi} \), i.e. \(\eta = \beta' \xi + \alpha'\).

**Proof.** I) If \(\alpha = 0\) (and \(\beta \neq 0\)) then \(\eta = \frac{\beta}{\gamma \xi + \delta} \iff \frac{1}{\eta} = \frac{\gamma \xi + \delta}{\beta} = \frac{\gamma \xi + \gamma'}{\beta} \) i.e. \(\bar{\eta} = \gamma' \xi + \delta'\).

II) If \(\beta = 0\) (and \(\alpha \neq 0\)) then \(\eta = \frac{\alpha \xi}{\gamma \xi + \delta} \iff \frac{1}{\eta} = \frac{\gamma \xi + \delta}{\alpha \xi} = \frac{\gamma \xi + \gamma'}{\alpha \xi} = \frac{\gamma}{\alpha} + \frac{\gamma'}{\alpha} \xi \) i.e. \(\bar{\eta} = \delta' \xi + \gamma'\).

III) If \(\gamma = 0\) (and \(\delta \neq 0\)) then \(\eta = \frac{\alpha \xi + \beta}{\delta} = \frac{\alpha}{\delta} \cdot \xi + \frac{\beta}{\delta} \) i.e. \(\eta = \alpha' \xi + \beta'\).

IV) If \(\delta = 0\) (and \(\gamma \neq 0\)) then \(\eta = \frac{\alpha \xi + \beta}{\gamma \xi} = \frac{\alpha \xi + \beta}{\gamma \xi} = \frac{\alpha}{\gamma} + \frac{\beta}{\gamma} \cdot \frac{1}{\xi} \) i.e. \(\eta = \beta' \xi + \alpha'\).  ■
The above Theorem helps us to transform the dataset \{ (\xi_i, \eta_i) : i = 1, \ldots, n \} to the appropriate one \{ (\xi'_i, \eta'_i) : i = 1, \ldots, n \}, how to solve the linearized regression problem \( \hat{\eta} = a\xi + \hat{b} \) by Theorem II.95 and after how to get the constants \( \alpha, \beta, \gamma, \delta \) in (6.37) from \( \hat{a} \) and \( \hat{b} \).

**Corollary II.105**

I) If \( \alpha = 0 \) (and \( \beta \neq 0 \)) then use the dataset \( (\xi'_i, \eta'_i) := (\xi_i, \frac{1}{\eta_i}) \), and after Theorem II.95 let \( \alpha = 0 \), \( \beta = 1 \), \( \gamma = \hat{a} \) and \( \delta = \hat{b} \).

II) If \( \beta = 0 \) (and \( \alpha \neq 0 \)) then use the dataset \( (\xi'_i, \eta'_i) := (\frac{1}{\xi_i}, \frac{1}{\eta_i}) \), and after Theorem II.95 let \( \alpha = 1 \), \( \beta = 0 \), \( \gamma = \hat{b} \) and \( \delta = \hat{a} \).

III) If \( \gamma = 0 \) (and \( \delta \neq 0 \)) then use the dataset \( (\xi'_i, \eta'_i) := (\xi_i, \frac{1}{\eta_i}) \), (unchanged) and after Theorem II.95 let \( \alpha = \hat{a} \), \( \beta = \hat{b} \), \( \gamma = 0 \) and \( \delta = 1 \).

IV) If \( \delta = 0 \) (and \( \gamma \neq 0 \)) then use the dataset \( (\xi'_i, \eta'_i) := (\frac{1}{\xi_i}, \eta_i) \), and after Theorem II.95 let \( \alpha = \hat{b} \), \( \beta = \hat{a} \), \( \gamma = 1 \) and \( \delta = 0 \).

**Proof. I)** The system of equations \( \frac{\gamma}{\beta} = \hat{a} \) and \( \frac{\delta}{\beta} = \hat{b} \) has the solution \( \beta = 1 \), \( \gamma = \hat{a} \) and \( \delta = \hat{b} \).

The other cases are similar.

We can use the transformations of Theorem II.104 also for drawing linear graphs of (6.37) on special coordinate systems: one or both (or none) of the axes are reciprocal.

**Corollary II.106**

I) if \( \alpha = 0 \) (and \( \beta \neq 0 \)) then use normal (equidistant) axe for \( \xi \) and reciprocal axe for \( \eta \),

II) if \( \beta = 0 \) (and \( \alpha \neq 0 \)) then use reciprocal scale on both axes,

III) if \( \gamma = 0 \) (and \( \delta \neq 0 \)) then (6.37) is already linear, so use the traditional Cartesian axes,

IV) if \( \delta = 0 \) (and \( \gamma \neq 0 \)) then use reciprocal axe for \( \xi \) and normal (equidistant) one for \( \eta \).

One example for Case II) is shown below:
We draw your attention to that Excel can not draw reciprocial coordinate system but [HM] can. Please try it! [HM] can handle all of the four cases above. On https://math.uni-pannon.hu/~szalkai/koordinata/reciprok-skala-160.gif we supply a reciprocial coordinate drawing in high resolution.

Remark II.107 We can observe on the Figure above, that the origin of the Cartesian coordinate system moved to the "infinity", along the (straight) line, in both directions, and further, the intersection points ("tengelymetszetek") of the linear graph with the axes (in the reciprocial coordinate system) correspond to the asymptotes of the ("original") hyperbola (in the Cartesian coordinate system).

6.5.6 Logit-probit regression

In pharmacy and in marketing statistics the following relation is investigated ($a, b$ can be any real parameters):

$$\eta = \frac{e^{a\xi+b}}{1 + e^{a\xi+b}} = 1 - \frac{1}{1 + e^{a\xi+b}},$$ (6.39)

which is closely related to the normal distribution. Here $\xi$ can be any real number but $0 < \eta < 1$.

Since the inverse of the function $y = 1 - \frac{1}{1+e^x}$ is $x = \ln \left( \frac{y}{1-y} \right)$, applying \ln \left( \frac{y}{1-y} \right) to (6.39) we get

$$\ln \left( \frac{\eta}{1-\eta} \right) = a\xi + b.$$ (6.40)
This means, that we can write \( \bar{\eta} = \ln \left( \frac{\eta}{1-\eta} \right) \), \( \bar{\xi} = \xi \) and apply the formule of Theorem II.95 to the dataset \((\bar{\xi}_i, \bar{\eta}_i) := (\xi_i, \ln(\frac{\eta_i}{1-\eta_i}))\) to compute \( \bar{a} = a \) and \( \bar{b} = b \).

![Figure 11: The function \( 1 - \frac{1}{1+e^{ax+b}} \) (blue) and its inverse \( \ln(\frac{x}{1-x}) \) (red).](image)

The functions \( \frac{e^{ax+b}}{1+e^{ax+b}} \) are symmetric to the point \((-\frac{b}{a}, \frac{1}{2})\), so \( \frac{e^x}{1+e^x} \) is symmetric to \((0, \frac{1}{2})\) (like \(\Phi\)).

We should use the transformation \( \ln(\frac{y}{1-y}) \) on the \(y\) axe so that the functions \( y = 1 - \frac{1}{1+e^{ax+b}} \) can have straight line graphs, details can be found in [SzI2]. Unfortunately neither Excel nor [HM] can make this transformation. The construction and the shape of the Figure 12 below is similar to the Gaussian coordinate system on Figure 2.
6.6 Nonlinear regressions - direct methods

When no linearizing method is applicable, we have to minimalize (6.4) directly. In very few fortunate cases we might get the solution directly.

When \( g \) is a polynomial, the regression is called **parabolic**. Here we introduce only the **quadratic** (second order) regression.
6.6. NONLINEAR REGRESSIONS - DIRECT METHODS

6.6.1 Quadratic regression

Now we consider the function
\[ \eta = a + b\xi + c\xi^2. \] (6.41)

0) In the case \( b = c = 0 \) we have a power function, which was dealt in a previous section.

I) In the case \( \sum x_i = 0 \) we have to solve the following system of linear equations for \( a, b, c \):

\[
\begin{align*}
  b &= \frac{\sum_{i=1}^{n} \xi_i \eta_i}{\sum_{i=1}^{n} \xi_i^2}, \quad \begin{cases}
    an + c \sum_{i=1}^{n} \xi_i^2 = \sum_{i=1}^{n} \eta_i \\
    a \sum_{i=1}^{n} \xi_i^2 + c \sum_{i=1}^{n} \xi_i = \sum_{i=1}^{n} \xi_i \cdot \eta_i
  \end{cases}
\end{align*}
\] (6.42)

II) For the general case we have to minimize the function

\[
F(a, b, c) = \sum_{i=1}^{n} \left[ a \xi_i^2 + b \xi_i + c - \eta_i \right]^2
\]

\[\begin{align*}
  &= \sum_{i=1}^{n} a^2 \xi_i^4 + 2ab \xi_i^3 + 2ac \xi_i^2 - 2a \xi_i^2 \eta_i + b^2 \xi_i^2 + 2bc \xi_i - 2b \xi_i \eta_i + c^2 - 2c \eta_i + \eta_i^2 \\
  &= a^2 A + b^2 B + c^2 C + ab D + ac E + bc F - aG - bH - cI + J
\end{align*}\]

where
\[
\begin{align*}
  A &= \sum_{i=1}^{n} \xi_i^4, \quad B = \sum_{i=1}^{n} \xi_i^3, \quad C = n, \quad D = 2 \xi_i^3, \quad E = 2 \sum_{i=1}^{n} \xi_i^2, \\
  F &= 2 \sum_{i=1}^{n} \xi_i, \quad G = 2 \sum_{i=1}^{n} \xi_i^2 \eta_i, \quad H = 2 \sum_{i=1}^{n} \xi_i \eta_i, \quad J = 2 \sum_{i=1}^{n} \eta_i, \quad J = \sum_{i=1}^{n} \eta_i^2.
\end{align*}\] (6.43)

Now
\[
\begin{align*}
  \frac{dF}{da} &= 2Aa + bD + cE - G = 0 \\
  \frac{dF}{db} &= 2Bb + aD + cF - H = 0 \\
  \frac{dF}{dc} &= 2Cc + aE + bF - I = 0
\end{align*}
\]

\[
\begin{align*}
  2aA + bD + cE &= G \\
  aD + 2bB + cF &= H \\
  aE + bF + 2cC &= I
\end{align*}
\]

which is a system of linear equations, and has the solution
\[ a = \frac{\det \begin{bmatrix} G & D & E \\ H & 2B & F \\ I & F & 2C \end{bmatrix}}{\begin{bmatrix} 2A & D & E \\ D & H & F \\ E & I & 2C \end{bmatrix}}, \quad b = \frac{\det \begin{bmatrix} 2A & D & E \\ D & 2B & F \\ E & F & 2C \end{bmatrix}}{\begin{bmatrix} D & 2B & H \\ E & F & I \end{bmatrix}}, \quad c = \frac{\det \begin{bmatrix} 2A & D & G \\ D & 2B & H \\ E & F & 2C \end{bmatrix}}{\begin{bmatrix} 2A & D & E \\ E & F & 2C \end{bmatrix}}, \]

i.e.

\[ a = \frac{F^2G + 2CHD - FHE + 2BEI - FDI - 4BCG}{\text{den}}, \]
\[ b = \frac{HE^2 + 2CGD - FGE + 2AFI - DEI - 4ACH}{\text{den}}, \]
\[ c = \frac{D^2I + 2BGE - FGD - 4ABI - HDE + 2AFH}{\text{den}}, \]

where the common denominator is

\[ \text{den} = 2AF^2 - 2FDE + 2CD^2 + 2BE^2 - 8ABC. \]
Chapter 7

Mathematical background

For more details see other textbooks and courses.

**The main idea is the following.** When we calculate a test number, we make a statistic, i.e. a composite function $\eta = g(\xi) = g(\xi_1, \ldots, \xi_n)$ of the sample $\xi = (\xi_1, \ldots, \xi_n)$ (see Definition II.6).

For example, in the t-test we have:

$$t_{sz} := \eta = \sqrt{n} \cdot \frac{\bar{\xi} - m_0}{\sigma^*} = \sqrt{n} \cdot \frac{\frac{n}{n} \frac{\xi_1 + \ldots + \xi_n}{n} - m_0}{\sqrt{\frac{\xi_1^2 + \ldots + \xi_n^2}{n} - \left(\frac{\xi_1 + \ldots + \xi_n}{n}\right)^2}}.$$ (7.1)

If we know the distribution of each data $\xi_i$, then the distribution of $\eta = g(\bar{\xi})$ can be determined by mathematical methods and the critical values, like $t_\varepsilon = \beta$ satisfying

$$P(\eta < \beta) = 1 - \varepsilon$$ (7.2)

i.e.

$$P(\beta \leq \eta) = \varepsilon$$ (7.3)

can be computed and collected in tables.
CHAPTER 7. MATHEMATICAL BACKGROUND

Figure 13: Critical value

We only have to find these critical values in the tables, eg. at the end of this book.

To "statistic-fans" we outline the Student- or t- and the χ² - distribution below.

7.1 The Student- or t- distribution

Definition II.108 Let ζ and ξ₁,...,ξₙ ∼ N(0,1) (i.e. standard normal) independent r.v.-s. Then

\[ \theta = \frac{\zeta}{\sqrt{\frac{\sum_{i=1}^{n} \xi_i^2}{n}}} \]

is called Student- or t- distribution of degree of freedom n. □

Theorem II.109 The density function is

\[ f_\theta(x) = \frac{\Gamma \left( \frac{n+1}{2} \right)}{\sqrt{n\pi} \cdot \Gamma \left( \frac{n}{2} \right) \cdot \left( 1 + \frac{x^2}{n} \right)^{\frac{n+1}{2}}} \]

where

\[ \Gamma (x) := \int_0^\infty t^{x-1}e^{-t}dt \]
is the so called $\Gamma$ - function (especially $\Gamma (n) = (n - 1)!$ for $n \in \mathbb{N}$).

Further, $M (\theta) = 0$ does exist only for $n \geq 2$ , and $D^2 (\theta) = \frac{n}{n - 2}$ does exist only for $n \geq 3$ .

Figure 14: Student distributions for $n = 5, 15, 100$
### 7.2 The $\chi^2$ distribution

**Definition II.110** Let $\xi_1, \ldots, \xi_n \sim N(0,1)$ (i.e. standard normal) independent r.v.-s, then

$$\eta := \sum_{i=1}^{n} \xi_i^2$$

is called **chi-square distribution with parameter** $n$. \qed

**Theorem II.111** The density function is

$$f_\eta(x) = \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \cdot \Gamma \left( \frac{n}{2} \right)}$$

for $0 < x$. Further, $M(\eta) = n$ and $D^2(\xi) = 2n$ for all $n$. \qed

![Figure 15: $\chi^2$ distributions for $n=1$ and $n=2$](image-url)
7.2. THE $\chi^2$ DISTRIBUTION

Figure 16: $\chi^2$ distributions for several $n$
Part III

Stochastic Processes
Chapter 8

Introduction

When considering different phenomena changing, or following each other in time, first: these phenomena are stochastic (random, "véletlen", or simply too difficult to us), and second: they might have some connections among them at certain level. These sequences of random phenomena are called Stochastic Processes ("sztochasztikus folyamatok").

In this Chapter we only highlight the main definitions and main properties of the most important stochastic processes, more detailed introductions can be found in the books [KT1], [KT2] and [KT3].

8.1 Elementary notions

Definition III.1 Any sequence of random variables (r.v., "v.v.") \( \mathbf{\xi} = (\xi_t : t \in T) \) for some index-set \( T \) is called a stochastic process ("sztochasztikus folyamat"), or s.p. ("szt.f.") for short.

In case \( T = \mathbb{N} \) we write \( \mathbf{\xi} = (\xi_1, \xi_2, ..., \xi_t, ...) \) and say discrete ("diszkrét"), in case \( T = \mathbb{R} \) (\( t \in \mathbb{R} \)) we say continuous ("folytonos") stochastic process.

Remark III.2 i) In practice we measure the same quantity (\( \Omega \)) several times: in time moments \( t \in T \). Both discrete (separated, "elkülönült") and continuous measurements are well known in practice.

ii) Each measurement (r.v.) \( \xi_i \) can also be a vector (higher dimensional) r.v.: \( \xi_i = [\xi_i^{(1)}, ..., \xi_i^{(n)}] \).
8.2 Examples

Example III.3 If we throw a dice for each \( t \in \mathbb{N} \) we have \( \xi_t : \Omega \to \{1, 2, 3, 4, 5, 6\} \) and each \( \xi_t \) has the same distribution and they are mutually ("páronként") independent from each other.

8.2.1 The Brownian motion

("Brown-mozgás"), also called Wiener process ("Wiener-folyamat").

A particle keeps moving in the space and let \( \xi_t \) denote its place in time \( t \in \mathbb{R} \). We assume that its movement in the future is independent of the previous movement, and the distance of its movement is described by a normal distribution. In a general mathematical form we can write:

**Definition III.4** A (one-dimensional) *Brownian motion* ("Brown-mozgás") is a stochastic process such that:

a) for any time moments \( t_0 < t_1 < \ldots < t_n < \ldots \) the increments i.e. relative movements ("növekmények, relatív elmozdulások")

\[
\xi_i := \xi_i - \xi_{i-1}
\]  

(8.1)

are mutually independent r.v.,

(a process with this property is said to be a process with independent increments.)

b) the probability distribution of the general increment r.v.

\[
\eta (x) = \xi_{t+x} - \xi_t \quad (x \in \mathbb{R})
\]

(8.2)

depends only on \( x = \Delta t \) and neither on \( t_1 = t \) or on \( t_2 = t + x \).

c)

\[
P (\xi_t - \xi_s < x) = \frac{1}{\sqrt{2\pi B (t-s)}} \int_{-\infty}^{x} e^{\frac{-u^2}{2B (t-s)}} \, du = \]

(8.3)

\[
= \frac{1}{\sqrt{2\pi B (t-s)}} \int_{-\infty}^{x} \exp \left( \frac{-u^2}{2B (t-s)} \right) \, du \]  

(8.4)

for some constant \( B \in \mathbb{R}^+ \) and for all \( s < t \).

---

1) Robert **Brown** (1773-1858) Scottish botanist and palaeobotanist.

2) Norbert **Wiener** (1894 -1964) American mathematician.
8.2. EXAMPLES

Theorem III.5 i) Assuming $\xi_0 = 0$ we have $M(t) = 0$ and $D^2(t) = \sqrt{B \Delta_{t,s}}$ for all $t \in \mathbb{R}^+$. 

ii) For any $t_0 < t_1 < \ldots < t_n < t$ the conditional probability is 

$$P(\xi_t < x \mid \xi_{t_1} = x_1, \ldots, \xi_{t_n} = x_n) = \frac{1}{\sqrt{2\pi B(t-t_n)}} \int_{-\infty}^{x-x_n} \exp\left(-\frac{u^2}{2B(t-t_n)}\right) du.$$ 

$\square$

Remark III.6 i) According to c) of Definition III.4 and Theorem i) above we know, that the distance made by the particle in (any but fixed) time $\Delta_{t,s} = t - s$ has a normal distribution with mean $m = 0$ and dispersion $\sigma = \sqrt{B \Delta_{t,s}}$. This assumption is encouraged by the Central Limit Theorem (see [SzI1]).

ii) It is also reasonable to have that the distribution of $\xi_t - \xi_s$ and that of $\xi_{t+h} - \xi_{s+h}$ are the same for any fixed $0 < h$ if we assume the medium to be in equilibrium.

iii) It is also clear that the displacement (relative motion) $\xi_t - \xi_s$ should depend only on the length $t-s$ and not on the time $t$ when we begin the observation.

iv) Theorem ii) says that the exact place of the particle depends only on the latest known position $x_n$ and all the previous information $x_{n-1}, \ldots, x_1$ are unimportant.

v) Higher dimensional Brownian motions can be defined similarly, but you must not consider them coordinatewise Brownian motions.

vi) See also the Section Markov processes.

8.2.2 The Poisson process

("Poisson folyamat")

Fix an event $A \subset \Omega$ and for $t \in \mathbb{R}^+\cup \{0\}$ let $\xi_t$ count the number of times $A$ occurred in the time period $[0, t]$. So each $\xi_t$ is represented as a nondecreasing step function. Obviously $\xi_0 = 0$ can be assumed.

Example III.7 Many practical phenomena can be considered as a Poisson process. (These are based on the concept of the law of rare events.) For example: the number of x-rays emitted by a substance undergoing radioactive decay, the number of telephone calls originating in a given locality, the occurrence of accidents at a certain intersection, the occurrence of errors in a page of typing, breakdowns of a machine, the arrival of customers for service, ...

3) Siméon Denis Poisson (1781 - 1840) French mathematician, physician and statistician.
The mathematical definition is as follows:

**Definition III.8** A stochastic process is called **Poisson process** if

**a)** the increments are mutually independent r.v. (see (8.1)),

**b)** the general increment r.v. depends only on $\Delta t$ (see (8.2)),

**c)** the probability of at least one event happening in a time period of duration $h$ is

$$P(A \text{ in } h) = p(h) := a \cdot h + o(h) \quad \text{for } h \to 0$$

and for some fixed $a > 0$ (and $\frac{o(h)}{h} \to 0$ as usual),

**d)** the probability of two or more events happening in time $h$ is $o(h)$.

**Remark III.9** Postulate d) is only to exclude the possibility of the simultaneous occurrence of two or more events.

Let $P_m(t)$ denote the probability that exactly $m$ events occur in time $t$, i.e. $P_m(t) = P(\xi_t = m), \quad m = 0, 1, 2, ...$. Now d) can be stated in the form:

$$\sum_{m=2}^{\infty} P_m(t) = o(h)$$

and clearly $p(h) = \sum_{m=1}^{\infty} P_m(t)$. Some further calculations show that

$$P_0(t) = e^{-at} \quad \text{for } t \in \mathbb{R}^{+,0}.$$ 

Clearly $P_0(h) = 1 - p(h)$ and $P_1(h) = p(h) + o(h)$.

Finally, using $P_m(0) = 0$ for $m \in \mathbb{N}$ we get the following:

**Theorem III.10** For each $t \in \mathbb{R}^{+,0}$ and $m \in \mathbb{N}$

$$P(\xi_t = m) = P_m(t) = \frac{(at)^m}{m!} e^{-at}$$

where $a$ is determined in (8.7). Therefore, $\xi_t$ follows a Poisson distribution with parameter $\lambda = at$ for each $t \in \mathbb{R}^{+,0}$.

The Poisson process often arises in a form where the time parameter is replaced by a suitable spatial ("térbeli") parameter (e.g. in 2- or 3- or in higher dimensions).
Example III.11 For example, consider a set $C \subseteq \mathbb{R}^d$ of points distributed in the space $\mathbb{R}^d$ ($1 \leq d$). For any (measurable, "mérhető") set $H \subseteq \mathbb{R}^d$ let

$$\zeta_H := N_H = |H \cap C| \tag{8.9}$$

denote the number of points (finite or infinite) from $C$ contained in $H$. We agree that $N_H$ is a random variable for each fixed set $H \subseteq \mathbb{R}^d$.

Definition III.12 The collection $\{N_H : H \subseteq \mathbb{R}^d \text{ is measurable} \}$ of random variables is said to be a homogeneous ("homogén") Poisson process if the following assumptions are fulfilled:

(i) the number of points in disjoint regions are independent r.v., that is $N_{H_1}$ and $N_{H_2}$ are independent if $H_1 \cap H_2 = \emptyset$,

(ii) for any subset $H \subseteq \mathbb{R}^d$ of finite volume ("térfogat") $N_H$ has a Poisson distribution with mean

$$\lambda = M(N_H) = a \cdot V(H) \tag{8.10}$$

where $V(H)$ is the (d-dimensional) volume of $H$ and $a \in \mathbb{R}^+$ is a fixed parameter.

Remark III.13 The parameter $a$ measures in a sense the intensity ("intenzitás, erősség") component of the distribution, which is independent of the size or shape of $H$.

Example III.14 Spatial ("térbeli") Poisson processes arise, for example

in considering the distribution of stars or galaxies in space,

in distribution of plants and animals on Earth,

in distribution of bacteria on a microscope slide,

etc.
Chapter 9

General stochastic processes

**Definition III.15** The stochastic processes s.p. ("sztochasztikus folyamat") \( \xi = (\xi_t : t \in T) \) are classified by:

- the state space ("állapottér") \( S \) where \( \xi_t : \Omega \to S \),
- the index or parameter set ("indexhalmaz, paraméterhalmaz") \( T \),
- the dependence relations ("függőségi viszonyok") among the r.v. \( \xi_t \).

\[ \square \]

9.1 The state space

("állapottér")

This is the "space" (set) \( S \) in which the possible values of each \( \xi_t \) "lie".

**Definition III.16** a) Finite ("véges") state spaces are of form \( S = \{s_0, s_1, ..., s_n\} \) for some \( n \in \mathbb{N} \).

i) In the case \( S = \{s_0, s_1, ..., s_n, \ldots\} \) or \( S = \mathbb{N} \) we refer to the process \( \xi \) as integer valued ("egészértékű") or alternatively as a discrete state ("diszkrét állapotú") process. These sets are also called enumerable or denumerable ("felsorolható, megszámlálható") sets.

ii) If \( S = \mathbb{R} \) the real line or a (real) interval \([a, b] \subset \mathbb{R}\) then we call \( \xi \) a real-valued ("valós értékű") stochastic process.

iii) If \( S \subseteq \mathbb{R}^k \) is a subset of \( \mathbb{R}^k \) (or possibly the whole \( \mathbb{R}^k \)) - the more dimensional space then \( \xi \) is said to be a \( k \)-vector ("\( k \)-vektor") process. \[ \square \]
As in case of a single r.v., the choice of the state space is not uniquely specified by the physical situation being described, although one particular choice usually stands out as most appropriate.

9.2 The index (parameter-) set

("indexhalmaz, paraméterhalmaz")

**Definition III.17**

i) If \( T = \mathbb{N} \cup \{0\} = \{0, 1, \ldots\} \) then we shall always say that \( \xi \) is a discrete time ("diszkrét idejű") stochastic process. When \( T \) is discrete we shall often write \( \xi_n \) instead of \( \xi_t \).

ii) If \( T = \mathbb{R}^+ = [0, \infty) \) then \( \xi \) is called a continuous time ("folytonos idejű") process.

iii) The case \( T = \{ \text{measurable sets} \} \subseteq \mathcal{P}(\mathbb{R}^d) \) and other cases are also possible. □

**Example III.18** We have already cited examples where the index set \( T \) is not one dimensional, e.g. spatial Poisson processes.

Another example is that of waves in oceans, where we may regard the latitude ("szélességi") and longitude ("hosszúsági") geographical ("földrajzi") coordinates as the value of \( t \) and \( \xi_t \) is then the height of the wave at the location \( t \in \mathbb{R}^2 \).

9.3 The mean-, dispersion- and autocovariance functions

("várható érték-, szórás- és kovariancia- függvények")

**Definition III.19** For any s.p. \( \xi \) the functions \( \{ M(\xi_t) : t \in T \} \), \( \{ D(\xi_t) : t \in T \} \) and \( \{ \text{cov}(\xi_t, \xi_s) : t, s \in T \} \) are called mean-, dispersion- and auto / self covariance functions ("várható érték / átlag, szórás- és auto / ön- kovariancia függvények"). □
Chapter 10

Classical types of stochastic processes

The in/dependencies ("függőségi viszonyok") among the r.v. $\xi_t$ are the most important properties of the stochastic processes.

10.1 Processes with stationary independent increments

("Független stacionárius [állandó] növekményű szt.f.")

Definition III.20 i) If the random variables

$$
\zeta_{t_1 t_2} := \xi_{t_2} - \xi_{t_1}, \quad \zeta_{t_2 t_3} := \xi_{t_3} - \xi_{t_2}, \quad \ldots, \quad \zeta_{t_n t_{n-1}} := \xi_{t_n} - \xi_{t_{n-1}}
$$

(10.1)

are independent for all choices of $t_1 < t_2 < \ldots < t_n$ (clearly $T = \mathbb{N}$ or $T = \mathbb{R}$), then we say that $\vec{\zeta}$ is a process with independent increments ("független növekményű").

ii) If the index set $T$ contains the smallest index $t_0$ (i.e. $T = \mathbb{N}$ or $T = [t_0, \infty)$), then it is also assumed that (expanding (10.1))

$$
\zeta_{t_0} := \xi_{t_0}, \quad \zeta_{t_0 t_1} := \xi_{t_1} - \xi_{t_0}, \quad \zeta_{t_1 t_2} := \xi_{t_2} - \xi_{t_1}, \quad \ldots, \quad \zeta_{t_n t_{n-1}} := \xi_{t_n} - \xi_{t_{n-1}}
$$

(10.2)

are (also) independent.

\[\square\]
Remark III.21 If the index set is discrete, that is $T = \mathbb{N}$, then a process with independent increments reduces to a sequence of independent r.v. $\zeta$ where

$$\zeta_0 = \xi_0 \quad \text{and} \quad \zeta_n = \xi_n - \xi_{n-1} \quad \text{for} \quad n = 1, 2, \ldots$$

(10.3)

in the sense that knowing the individual distributions of $\zeta_0$, $\zeta_1$, ... enables one to determine the joint distribution of any finite subset $\{\xi_{n_1}, \ldots, \xi_{n_m}\}$ of $\xi$. Especially

$$\xi_n = \zeta_0 + \ldots + \zeta_n \quad \text{for all} \quad n = 0, 1, 2, \ldots$$

(10.4)

Definition III.22 If the distribution of the increments or differences ("növekmények, különbségek")

$$\xi_{t+h} - \xi_t$$

(10.5)

depends only on the length $h$ of the interval and not on the time $t$ (for all $t \in T$ and $h \in \mathbb{R}^+$), then the process is said to have stationary increments ("stacionárus [állandó] növekményű").

For a process with stationary increments the distribution of $\xi_{t_1+h} - \xi_{t_1}$ is the same as the distribution of $\xi_{t_2+h} - \xi_{t_2}$ no matter what the values of $t_1$, $t_2$ and $h$. So, we can denote this distribution by

$$\vartheta_h := \xi_{t+h} - \xi_t$$

(10.6)

where $t \in T$ is arbitrary fixed index. □

Theorem III.23 If a process $\xi = \{\xi_t : t \in T\}$ where $T = [0, \infty)$ or $T = \mathbb{N}$ has stationary independent increments and has a finite mean (i.e. each all $M(\xi_t)$ does exists), then it is elementary to show that

$$M(\xi_t) = m_0 + m_1 \cdot t \quad (t \in T)$$

(10.7)

where $m_0 = M(\xi_0)$ and $m_1 = M(\xi_1) - m_0$.

Similarly

$$\sigma_t^2 = \sigma_0^2 + \sigma_1^2 \cdot t \quad (t \in T)$$

(10.8)

where

$$\sigma_0^2 = M[(\xi_0 - m_0)^2] = D^2(\xi_0)$$

(10.9)

and

$$\sigma_1^2 = M[(\xi_1 - m_1)^2] - \sigma_0^2 = D^2(\xi_1) - D^2(\xi_0) \quad .$$

(10.10)
10.2 Martingales

("Martingálok")

Definition III.25 Let \( \xi \) be a real-valued s.p. with discrete or continuous parameter set \( \mathbb{T} \). We say that \( \xi \) is a \textit{martingale} ("martingál") if

i) \( M(|\xi_t|) < \infty \) for all \( t \in \mathbb{T} \),

ii) for any \( n \in \mathbb{N} \), for any \( t_1 < t_2 < \ldots < t_n < t_{n+1} \) and for all values \( a_1, a_2, \ldots, a_n \in S \)

\[
M(\xi_{t_{n+1}} \mid \xi_{t_1} = a_1, \ldots, \xi_{t_n} = a_n) = a_n. \quad \Box \quad (10.11)
\]

Remark III.26 i) Observe the absolute value of \( \xi_t \) in i) and recall that i) is stronger than "\( \xi \) has a finite mean".

ii) Martingales may be considered as appropriate models for fair games in the sense that \( \xi_t \) denotes the amount of money that a player has at time \( t \). The martingale property ii) states then that the average amount a player will have at time \( t_{n+1} \), assuming that he has amount in the previous time \( t_n \), is equal to \( a_n \), regardless of what his past fortune (in the interval \([t_n, t_{n+1}]\) and before) was.

iii) The word "martingale" originally meant a gambling strategy in which one doubles the stake after each loss.

Claim III.27 i) One can easily verify that if \( \zeta_i \) are independent r.v. and \( M(\zeta_i) = 0 \), then the process

\[
\xi_n = \zeta_1 + \ldots + \zeta_n \quad (n \in \mathbb{N}) \quad (10.12)
\]

is a discrete martingale.

ii) Similarly, if \( \xi_t \) for \( 0 \leq t \) has independent increments whose means are 0 then \( \xi_t \) is a continuous time martingale.
10.3 Markov processes

(”Markofolyamat”)

Definition III.28 A process $\xi$ is said to be a **Markov s.p.** (”Markofolyamat”) if

$$P(a < \xi_t \leq b \mid \xi_{t_1} = a_1, \ldots, \xi_{t_n} = a_n) = P(a < \xi_t \leq b \mid \xi_{t_n} = a_n)$$

(10.13)

for all $t \in T$ whenever $t_1 < t_2 < \ldots < t_n < t$ and for all values $a_1, a_2, \ldots, a_n \in S$.

For discrete state ($S = \{s_0, s_1, \ldots, s_n, \ldots\}$) and discrete time ($T = \mathbb{N} \cup \{0\}$) the assumption (10.13) can be written easier:

Definition III.29 A process $\xi$ is said to be a **discrete Markov s.p.** (”diszkret Markofolyamat”) or a **Markov-chain** (”Markov-lánc”) if

$$P(\xi_{t_{n+1}} = a_{n+1} \mid \xi_{t_1} = a_1, \ldots, \xi_{t_n} = a_n) = P(\xi_{t_{n+1}} = a_{n+1} \mid \xi_{t_n} = a_n)$$

(10.14)

for all $t_1 < t_2 < \ldots < t_n < t_{n+1} \in T$ and for all $a_1, a_2, \ldots, a_n \in S$.

Remark III.30 i) Roughly speaking a Markov s.p. is one with the property that, if the value of $\xi_t$ is given, then the values of $\xi_s$ for $s > t$ do not depend on the values of $\xi_u$ for $u < t$. That is the probability of any particular future behaviours of the process, when its present state ($\xi_t$) is known exactly, is not altered by additional knowledge concerning its past behaviour.

We should make it clear, however, that if our knowledge of the present state ($\xi_t$) of the process is imprecise, then the probability of some future behaviour will be altered by additional information in general, relating to the past behaviour of the system.

ii) Note that a Markov s.p. having a finite or denumerable state space $S$ is called a **Markov chain** (”Markov-lánc”).

Example III.31 **Discrete Brownian motion as partial sums of independent r.v.’s** (”Diszkret Brown-mozgás, mint független v.v. részletösszege”)

Let a particle keep moving on the real line on the integer points $\mathbb{Z}$, starting from $0$, and suppose that it moves in the $n$’th moment 50% to the left and 50% to the right. If all the steps are independent, then $\xi$ is a discrete Markov s.p. where

$$\xi_n = \eta_1 + \ldots + \eta_n \quad 1 \leq n$$

(10.15)

and $\eta_i$ are independent and the values of $\eta_i$ are $\pm 1$ with probability $0.5$ (i.e. $\eta_i : \Omega \rightarrow \{-1, +1\}$, $P(\eta_i = -1) = P(\eta_i = +1) = 0.5$).

1) Andrey Andreyevich Markov (1856-1922) a Russian mathematician.
10.3. MARKOV PROCESSES

Claim III.32 In general, it is easy to prove, that partial sums \((10.15)\) of independent r.v. \(\eta_i\) are always a discrete Markov s.p. \(\square\)

Definition III.33 The Markov chain \(\xi\) in \((10.15)\) is called homogeneous ("homogén") if \(\eta_i\) all have the same distribution, otherwise \(\xi\) is inhomogeneous ("inhomogén"). \(\square\)

Example III.34 If we place reflecting mirrors ("visszaverő tükör") or back-kicking walls ("visszapattanó falak") to the points \(-K\) and \(K\), from where the particle ultimately (100\%) turns back, then we also get a Markov-chain.

Example III.35 Let \(N \in \mathbb{N}\) be fixed, let \(\eta_i\) be independent r.v. which have values \(\{0, 1, ..., N - 1\}\) with arbitrary probabilities. Now if we define \(\xi\) as \(\xi_0 = \eta_0\) and

\[
\xi_{n+1} = \begin{cases} 
\xi_n + \eta_n & \text{if } \xi_n + \eta_n < N \\
\xi_n + \eta_n - N & \text{if } \xi_n + \eta_n \geq N
\end{cases}
\] (10.16)

then \(\xi\) is also a Markov chain.

This example is called lower rounding ("lefelé kerekítés, csonkítás") against overfloating ("túlsordulás ellen").

Definition III.36 Let \(A \subset \mathbb{R}\) be an interval of the real line. Then the function

\[
P(x, s, t, A) := P(\xi_t \in A \mid \xi_s = x)
\] (10.17)

for \(t > s\) is called transition probability function ("átmenetvalószínűség-függvény") and is basic to the study of the structure of Markov s.p. \(\square\)

Claim III.37 We may express the condition \((10.13)\) also as follows:

\[
P(a < \xi_t \leq b \mid \xi_{t_1} = a_1, ..., \xi_{t_n} = a_n) = P(x_n, t_n, t, (a, b))
\] (10.18)

\(\square\)

It can be proved that the probability distribution of \((\xi_{t_1}, ..., \xi_{t_n})\) can be computed in terms of \((10.17)\) and the initial distribution function of \(\xi_{t_1}\).

Definition III.38 A Markov s.p. for which all realizations or sample functions \(\{\xi_t : t \in [0, \infty)\}\) are continuous functions, is called a diffusion process ("difúziós folyamat"). \(\square\)

Remark III.39 Poisson processes are continuous time Markov chains, and Brownian motions are diffusion processes.
For Markov chains the transition probability function, \((10.17)\) and \((10.18)\) can be written in easier form.

**Definition III.40** For a Markov chain \(\xi = \{\xi_n : n \in \mathbb{N}\}\)

i) the probabilities

\[
\begin{align*}
nP_{r, i, k} := P(\xi_{n+r} = k \mid \xi_n = i) 
\end{align*}
\]  

are called \textit{\(r\)-step transition probabilities} ("\(r\)-lépéses átmenetvalószínűségek"), shortly \textit{t.p.}, for \(r, n, i, k \in \mathbb{N}\).

ii) the (finite or infinite) matrix

\[
\begin{align*}
n\Pi_r := \begin{bmatrix}
nP_{1,1}^{(r)} & nP_{1,2}^{(r)} & \cdots \\
nP_{2,1}^{(r)} & nP_{2,2}^{(r)} & \cdots \\
\vdots & \vdots & \ddots
\end{bmatrix}
\end{align*}
\]  

is called \textit{transition probability matrix} ("átmenetvalószínűség mátrix").

For homogeneous Markov chains the index \(n\) is usually omitted. We also omit \(r\) in case \(r = 1\). \(\square\)

**Claim III.41** All the entries of \(n\Pi_r\) are probabilities \(\in [0,1]\) and each row has sum 1 since

\[
\begin{align*}
\sum_{k=1}^\infty nP_{i,k}^{(r)} = \sum_{k=1}^\infty P(\xi_{n+r} = k \mid \xi_n = i) = 1 . \(10.21)\)
\end{align*}
\]

**Definition III.42** Any quadratic ("négyzetes") matrix (either finite or infinite) with nonnegative entries is called \textit{stochastic matrix} ("sztochasztikus mátrix") if its each row has sum 1 (see \((10.20)\) and \((10.21)\)).

Moreover, if each column has sum 1, too, i.e. \(\sum_{i=1}^\infty nP_{i,k}^{(r)} = 1\), then the matrix is called a \textit{double stochastic matrix} ("kétszeresen sztochasztikus mátrix"). \(\square\)

**Claim III.43** Products of (double) stochastic matrixes is also a (double) stochastic one. \(\square\)

The following theorem is a fundamental one on Markov chains.
Theorem III.44 If the 1-step transition probabilities are independent of \( n \), then any \( r \)-step t.p. are also independent, and

\[ \Pi_r = (\Pi)^r \]  

(10.22)

i.e. the \( r \)-th power of the matrix \( \Pi = \Pi_1 \).

Remark III.45 i) The special case \( \Pi_r = \Pi^{r_1} \cdot \Pi^{r_2} \) of (10.23) for \( r_1 + r_2 = r \), i.e.

\[ p_{i,k}^{(r)} = \sum_{j=1}^{\infty} p_{i,j}^{(r_1)} \cdot p_{j,k}^{(r_2)} \]  

(10.23)

is often in use without mentioning and is called Markov equality ("Markov egyenlőség").

ii) The transition probabilities \( p_{i,k}^{(r)} \) are conditional probabilities ("feltételes valószínűségek"), so the unconditional ("feltétel nélküli") probabilities of \( \xi_n \)

\[ p_k(n) := P(\xi_n = k) \quad k \in \mathbb{N}, \quad n \in \mathbb{N} \cup \{0\} \]  

(10.24)

are called absolute probabilities ("abszolút valószínűségek") of \( \xi_n \).

Definition III.46 A Markov-chain \( \xi \) is called ergodic ("ergodikus") if all the limit probabilities

\[ P_k := \lim_{r \to \infty} p_{i,k}^{(r)} \]  

(10.25)

do exist, they are independent of \( i \), and

\[ \sum_{k=1}^{\infty} P_k = 1 \]  

(10.26)

Remark III.47 i) In general, the behaviour in which sample averages formed from a process converge to some underlying parameter of the process is termed ergodic. (See Remark III.49 too.)

ii) (10.26) says that the events

\[ A_k = \{ \lim_{r \to \infty} \xi_r = k \} \subset \Omega \]  

(10.27)

form a complete system of events ("teljes eseményrendszer").

The following result is a fundamental one in the theory of Markov chains.
Theorem III.48  **Ergodicity theorem of Markov** ("Markov ergodicitási tétele"): A homogeneous Markov chain $\xi$ having finitely many states ("véges állapotá") is ergodic if and only if

$$
\Pi = \begin{bmatrix}
p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\
p_{2,1} & p_{2,2} & \cdots & p_{2,N} \\
\vdots & \vdots & \ddots & \vdots \\
p_{N,1} & p_{N,2} & \cdots & p_{N,N}
\end{bmatrix}
$$

(see (10.20)) has a power $\Pi^v$ ($v \in \mathbb{N}$) in which at least one column contains only positive elements.

Further, the convergence in (10.25) is exponential:

$$
|p_{i,k}^n - P_k| \leq (1 - M\delta)^{n-1}
$$

where $M$ is the number of columns of $\Pi^v$ containing positive elements, $\delta$ is the least number in these columns. (Clearly $0 < M\delta \leq 1$.) \(\square\)

**Remark III.49**  

i) The assumption of ergodicity in (10.25) and in the previous theorem asserts the existence of a step number $v$ and of at least one state $s \in \mathbb{S}$ which state can be reached from any other state in at most $v$ steps with positive probability.

ii) Another meaning of ergodicity is that if starting from any state $s_i \in \mathbb{S}$, after a large number of steps the process reach the state $s_k$ with probability $P_k$ but independently of $s_i$. Moreover we have $\lim_{n \to \infty} p_k(n) = P_k$.

iii) By the Markov inequality (10.23) we get

$$
p_{i,k}^{n+1} = \sum_{j=1}^{N} p_{i,j}^{(n)} \cdot p_{j,k}^{(1)},
$$

and taking $n \to \infty$ we get

$$
P_k = \sum_{j=1}^{N} P_j P_{j,k} \quad \text{for } 1 \leq k \leq N.
$$

It is not hard to prove that the system of equalities above has a unique solution for the unknowns $P_k$ for $1 \leq k \leq N$. This system of equalities is often helpful in practice.
iv) If the matrix (10.20) for \( r = 1 \) -step is double stochastic and the process is ergodic then \( \Pi^n \) and \( \lim_{n \to \infty} \Pi^n \) are also double stochastic ones. Since all the elements of the \( k \)-th column \( P_k \), so \( P_k = 1/N \) (where \( N = |S| \)). This means, that the marginal distribution (after \( n \to \infty \)) is uniform ("egyenletes") on the numbers 1, ..., \( N \).

**Example III.50** Consider the practical problem of the volume of a water-puffer lake of a factory ("víztározó"), from \( [P] \).

Let \( K \) denote the volume of the lake, and let us try to use exactly (at most) \( M \) quantity water each year. Clearly we use less water if there is no \( M \) water in the lake, in this case we empty the lake. Suppose that \( K, M \) are integers and \( 0 < M < K \).

Denote \( \xi_t \) the water supply of the river in the \( t \) th year (\( t \in \mathbb{N} \)), i.e. \( \xi_1, ... \) are independent discrete r.v. with the same distribution, \( \text{Im}(\xi_t) = \mathbb{N} \) and let

\[
p_i := P(\xi_t = i) .
\]

(10.32)

Let further \( \zeta_t \) denote the water level of the lake at the end of the year (\( t \in \mathbb{N} \)), i.e. after we took out \( M \), and denote \( \zeta_0 \) the starting level.

Clearly the lake contains no more than \( K \) water in each moment, so we must have

\[
\zeta_{t+1} := \max \left\{ \min \left( \zeta_t + \xi_{t+1}, K \right) - M , \ 0 \right\}
\]

(10.33)

which implies

\[
\text{Im}(\zeta_t) = \{0, 1, ..., K - M\}
\]

(10.34)

and we let

\[
P_{i,j} := P(\zeta_{t+1} = j \mid \zeta_t = i)
\]

(10.35)

the possible water level in the next year.

For simplicity we assume

\[
M < K - M \quad \text{i.e.} \quad M < K/2 .
\]

(10.36)

**Solution III.51** Clearly

\[
P_{u,v} = 0 \quad \text{if} \quad u - M > v \quad \text{i.e.} \quad u - v > M ,
\]

(10.37)

or even, for suitable \( u, v, w \) (among others)

\[
0 \leq u, v \leq K - M \quad \text{and} \quad 0 \leq w = M + v - u
\]

(10.38)
\[ P_{u,v} = P(\xi_t = w) \iff u - M + w = v \]  
\[ P_{u,v} = p_{v+M-u} \]  
Further, for suitable \( j \)

\[ P_{j,K-M} = p_{K-j} + p_{K-j+1} + \ldots \]  
Finally, we have the following (large) system of equalities for \( P_{i,j} \):

\[
\begin{align*}
P_{0,0} &= p_0 + \ldots + p_M, \\
P_{0,1} &= p_{M+1}, \\
&\quad \ldots \\
P_{0,i} &= p_{M+i} \quad \text{(for } i = \zeta_{t+1} < K - M \text{ i.e. } M + i < K), \\
&\quad \ldots \\
P_{0,K-M-1} &= p_{K-1}, \\
P_{0,K-M} &= p_K + p_{K+1} + \ldots, \\
P_{1,0} &= p_0 + \ldots + p_{M-1}, \\
P_{1,1} &= p_M, \\
&\quad \ldots \\
P_{1,i} &= p_{M-1+i}, \\
&\quad \ldots \\
P_{1,K-M-1} &= p_{K-2}, \\
P_{1,K-M} &= p_{K-1} + p_K + \ldots, \\
&\quad \ldots \\
P_{j,0} &= p_0 + \ldots + p_{M-j}, \\
P_{j,1} &= p_{M+1-j}, \\
&\quad \ldots \\
P_{j,i} &= p_{M-j+i} \quad \text{(for } j \leq M), \\
&\quad \ldots \\
P_{j,K-M-1} &= p_{K-j-1}, \\
P_{j,K-M} &= p_{K-j} + p_{K-j+1} + \ldots, \\
&\quad \ldots \\
P_{M,0} &= p_0 \quad \text{(since } M < K - M \text{ i.e. } M < K/2), \\
P_{M,1} &= p_1, \\
&\quad \ldots \\
P_{M,i} &= p_i,
\end{align*}
\]
10.3. **MARKOV PROCESSES**

\[ P_{M,K-M-1} = p_{K-M-1} \quad \text{(since } M \leq K - M - 1), \]
\[ P_{M,K-M} = p_{K-M} + p_{K-M+1} + \cdots, \]
\[ P_{M+1,0} = 0, \]
\[ P_{M+1,1} = p_0, \]
\[ P_{M+1,i} = p_{i-1} \quad \text{(for } i \leq K - M - 1), \]
\[ P_{M+1,K-M-1} = p_{K-M-2}, \]
\[ P_{M+1,K-M} = p_{K-M-1} + p_{K-M} + \cdots, \]
\[ P_{M+\ell,0} = 0 \quad \text{(for } 1 \leq \ell \text{ and } M + \ell \leq K - M \text{ i.e. } \ell \leq K - 2M), \]
\[ P_{M+\ell,\ell-1} = 0 \quad \text{ (see (10.37))}, \]
\[ P_{M+\ell,\ell} = p_0, \]
\[ P_{M+\ell,\ell+i} = p_i \quad \text{(for } \ell + i \leq K - M - 1 \text{ i.e. } i \leq K - M - \ell - 1), \]
\[ P_{M+\ell,K-M-1} = p_{K-M-\ell-1}, \]
\[ P_{M+\ell,K-M} = p_{K-M-\ell} + p_{K-M-\ell+1} + \cdots, \]
\[ P_{K-M,0} = 0 \quad \text{(see (10.36))}, \]
\[ P_{K-M,1} = 0, \]
\[ P_{K-M,i} = 0 \quad \text{(for } i < (K - M) - M = K - 2M, \text{ see (10.37))}, \]
\[ P_{K-M,K-2M} = p_0 \quad \text{(by (10.40) } v + M - u = K - 2M + M - (K - M) = 0), \]
\[ P_{K-M,K-M-1} = p_{M-1}, \]
\[ P_{K-M,K-M} = p_M + p_{M+1} + \cdots. \]

**END of the Example.**
10.4 Stationary processes

("Stacionárius [állandó] folyamatok")

**Definition III.52** A s.p. \( \xi \) (where \( T \) can be any of the sets \((-\infty, \infty), [0, \infty), \mathbb{Z} \) or \( \mathbb{N} \setminus \{0\} \) is said to be **strictly stationary** ("erősen stacionárius") if the joint distribution functions of the families of random variables are

\[
\tilde{\xi}_{t+h} = (\xi_{t_1+h}, \xi_{t_2+h}, \ldots, \xi_{t_n+h}) \quad \text{and} \quad \tilde{\xi}_t = (\xi_{t_1}, \xi_{t_2}, \ldots, \xi_{t_n}), \quad (10.42)
\]

that is \( F_{\tilde{\xi}_{t+h}} \) and \( F_{\tilde{\xi}_t} : \mathbb{R}^n \to \mathbb{R} \) are the same for all \( h > 0 \) and arbitrary finite set of \( t_1, \ldots, t_n \in T \).

**Remark III.53** This condition asserts that in essence the process is in probabilistic equilibrium ("egyensúly") and that the particular times at which we examine the s.p. are of no relevance. In particular the distribution of \( \xi_t \) is the same for each \( t \in T \).

The word stationary means "almost constant" ("majdnem állandó").

**Theorem III.54** The mean- and dispersion functions of stationary processes do not depend on \( t \in T : M(\xi_t) = M(\xi_0) \) and \( D(\xi_t) = D(\xi_0) \).

The autocovariance function depends on \((t-s)\):

\[
cov(\xi_t, \xi_s) = cov(\xi_{t-s}, \xi_0) \quad (10.43)
\]

for \( t, s \in T \).

**Definition III.55** A s.p. \( \xi \) is said to be

i) **wide sense stationary** ("gyengén stacionárius") if it possesses finite second moments (i.e. \( M(\xi_t^2) < \infty \)),

ii) **covariance stationary** ("stacionárius kovarianciájú") if \( cov(\xi_t, \xi_{t+h}) \) depends only on \( h \) for all \( t \in T \).

Recall, that \( cov(\zeta, \eta) = M(\zeta \cdot \eta) - M(\zeta) \cdot M(\eta) \).

**Claim III.56** A s.p. that has finite second moments is always covariance stationary, but there are covariance stationary processes that are not stationary.

**Remark III.57** Stationary processes are appropriate for describing many phenomena that occur in communication theory, astronomy, biology and sometimes in economics.
10.5. RENEWAL PROCESSES

Definition III.58 A Markov process is said to have stationary transition probabilities ("stacionáris átmenetvalószínűségű") if $\mathcal{P}(x,s,t,A)$, defined in (10.17), is a function only of $t - s$.

Remark III.59 Remember that $\mathcal{P}(x,s,t,A)$ of a Markov process is a conditional probability, which is given in the present state. Therefore there is no reason to expect that a Markov process with stationary transition probabilities is a stationary process, and this is indeed the case.

Neither the Poisson process nor the Brownian motion process is stationary. In fact, no nonconstant process with stationary independent increments is stationary. However, if $\{\xi_t : t \in [0, \infty)\}$ is a Brownian motion or a Poisson process, then $\zeta_t := \xi_{t+h} - \xi_t$ is a stationary process for any fixed $h \geq 0$.

10.5 Renewal processes

("Felújítási folyamatok")

Definition III.60 i) A renewal process ("felújítási folyamat") is a sequence $\tau = (\tau_n : n \in \mathbb{N})$ of independent and identically distributed positive r.v. representing the lifetimes of some "units" ("egységek"). The first unit is placed in operation at time 0, it fails at time $\tau_1$ and is immediately replaced by a new unit (with the same properties) which fails at time $\tau_1 + \tau_2$ and so on. The time of the $n$'th renewal ("felújítás") is

$$\sigma_n = \tau_1 + \ldots + \tau_n \quad (n \in \mathbb{N}) . \tag{10.44}$$

ii) A renewal counting ("számláló") process is $\nu = (\nu_t : t \in \mathbb{R}^+)$ where for $t \in \mathbb{R}^+$ and $n \in \mathbb{N}$

$$\nu_t = n \quad \overset{def}{\iff} \quad \sigma_n \leq t < \sigma_{n+1} . \tag{10.45}$$

Remark III.61 i) The renewal process $\sigma_n$ gives us the time moment of the $n$'th renewal, while a renewal counting process $\nu_t$ counts the number of renewals in the time interval $[0, t]$.

We often make no distinction between the renewal process and its counting process.

ii) Renewal processes occur in many applied areas such as management science, economics and biology. Renewal processes of equal importance often may be discovered embedded in other stochastic processes that, at first glance, seem unrelated.

iii) The Poisson process with parameter $\lambda$ is a renewal counting process for which the unit lifetimes have exponential distributions with common parameter $\lambda$. 

10.6 Point processes

("Pontfolyamatok")

Note: \( S \neq \emptyset \!\) 

Definition III.62 Let \( S \subseteq \mathbb{R}^n \) be a fixed set in the \( n \)-dimensional space and let \( \mathcal{A} \subseteq \mathcal{P}(S) \) be a family of subsets of \( S \). A \textbf{point process} ("pontfolyamat") is a s.p. indexed by the sets \( A \in \mathcal{A} \), that is \( \mathcal{T} = \mathcal{A} \), having the state space \( \mathbb{S} = \mathbb{N} \cup \{0\} \) (nonnegative integers). In other words: 
\[ \overrightarrow{\mathcal{S}} = \{ \xi_A : A \in \mathcal{A} \}. \]

Remark III.63 (i) Non-mathematicians please write \( \mathcal{A} = \mathcal{P}(S) \), i.e. let \( A \in \mathcal{A} \) mean "\( A \subseteq S \) is any (measurable) subset of \( S \) ".
(ii) We think a set of (enumerable) "points" \( C \subseteq S \) is being scattered over \( S \) in some random manner, and of
\[ \xi_A = N(A) := |A \cap C| \] (10.46)
as counting the number of points from \( C \) in the (measurable) set \( A \in \mathcal{A} \), i.e. \( A \subseteq S \).

Since \( N(A) \) is a counting function there are additional requirements on each realization.

Definition III.64 (continued):
i) if \( A_1 \cap A_2 = \emptyset \) and \( A_1 \cup A_2 \in \mathcal{A} \) then \( N(A_1 \cup A_2) = N(A_1) + N(A_2) \), i.e. \( \xi_{A_1 \cup A_2} = \xi_{A_1} + \xi_{A_2} \).
ii) if \( \emptyset \in \mathcal{A} \) then \( N(\emptyset) = 0 \), i.e. \( \xi_\emptyset = 0 \). \( \Box \)

Clearly ii) follows from i).

Definition III.65 Suppose \( S \) is a set in the real line (or plane or 3-dimensional space) and for every subset \( A \subseteq S \) let \( V(A) \) be the length (area, volume, resp.) of \( A \). Then
\[ \overrightarrow{\mathbb{N}} = \{ \nu_A : A \subseteq S \} \] (10.47)
is a \textbf{homogeneous Poisson point process of intensity} \( \lambda > 0 \) ("\( \lambda \) intenzitású (erősségű) Poisson-pontfolyamat") if \( \mathcal{A} = \mathcal{P}(S) \) (power set) and
10.7. MOVING AVERAGE PROCESSES

(i) for each $A \subset S$ we have $\nu_A := N(A)$ (more precisely: $\nu_A := |A \cap C|$) has a Poisson distribution with parameter $\lambda \cdot V(A)$ and $\lambda \in \mathbb{R}^+$ is (any) fixed positive real number.

(ii) for every finite collection $\{A_1, \ldots, A_n\} \subset \mathcal{A}$ of mutually ("páronként") disjoint subsets of $S$ the r.v.’s $\nu_{A_1}, \ldots, \nu_{A_n}$ are independent.

Remark III.66 (i) The above (i) says that the number of points from $C$ in $A$ do not depend on the shape of $A$ but the parameter in this Poisson distribution has linear dependency with the volume of $A$.

(ii) Every Poisson process $\{\xi_t : t \in [0, \infty]\}$ defines a Poisson point process on $S = [0, \infty)$. In fact, for any interval subset $A = (s, t)$ for $s < t$ we use

$$N(A) := \xi_t - \xi_s.$$  

(iii) Poisson point processes arise in considering the distribution of stars or galaxies in space, the planes distribution of plants and animals, or of bacteria on a slide, etc.

10.7 Moving average processes

("mozgóátlag folyamatok")

Definition III.67 Let $\vec{\zeta} = \{\zeta_n : n = 0, \pm 1, \pm 2, \ldots\}$ i.e. $\mathbb{T} = \mathbb{Z}$ (integers) be uncorrelated r.v. having a common mean $\mu$ and variance $\sigma^2$. Let $m \in \mathbb{N}$ and $a_1, a_2, \ldots, a_m \in \mathbb{R}$ be any fixed numbers and consider the process $\vec{\xi} = \{\xi_n : n \in \mathbb{Z}\}$ where

$$\xi_n = a_1\zeta_n + a_2\zeta_{n-1} + \ldots + a_m\zeta_{n-m+1} \quad \text{for } n \in \mathbb{Z}.$$  \hspace{1cm} (10.48)

Now the s.p. $\vec{\xi}$ is called a moving average processes ("mozgóátlag folyamat").

Remark III.68 The naming "moving average" refers to the application when the original s.p. $\vec{\zeta}$ has extreme low and high (expected) values and perhaps periodic or "seasonable" ("szezonális"), and these huge differences are decreased and the extreme alterations are smoothed by taking the (weighted) average of $m$ consecutive r.v. $\zeta_n, \ldots, \zeta_{n-m+1}$. So, the s.p. $\vec{\xi}$ contains the averages of these consecutive r.v. $\zeta_i$, and goes on, i.e. moves. The usual arithmetic mean

\hspace{1cm} \[2\] Consider e.g. the changes of the numbers of tourists in the four seasons of years, or your working attitude from Monday to Sunday and of the next weeks.
CHAPTER 10. CLASSICAL TYPES OF STOCHASTIC PROCESSES

("számtani/aritmetikai közép") uses \( a_1 = \ldots = a_m = \frac{1}{m} \) and weighted arithmetic means ("súlyozott számtani közép") need \( a_1 + \ldots + a_m = 1 \), however in (10.48) the numbers \( a_i \) can be any real numbers!

\( \xi_n \) are in blue and \( \xi_{n+v} \) are in red

Claim III.69 \[ M(\xi_n) = \mu \cdot (a_1 + \ldots + a_m) \]

and \[ D^2(\xi_n) = \sigma^2 \cdot (a_1^2 + \ldots + a_m^2). \]

For the covariance we have

Theorem III.70

\[ \text{cov} \left( \xi_n, \xi_{n+v} \right) = E \left[ \left( \xi_n - \mu \cdot \sum_{i=1}^{m} a_i \right) \left( \xi_{n+v} - \mu \cdot \sum_{i=1}^{m} a_i \right) \right] = \]

\[
\begin{cases}
\sigma^2 \cdot (a_m a_{m-v} + \ldots + a_{v+1} a_v) & \text{if } v \leq m - 1 \\
0 & \text{if } v \geq m
\end{cases}
\]

(10.49)

Since the covariance between \( \xi_n \) and \( \xi_{n+v} \) depends only on \( v \) and not on \( n \), the process \( \xi \) is covariance stationary.
Remark III.71 A common case is the "moving average" with a standardized variance in which $a_k = 1/\sqrt{m}$ for $k = 1, \ldots, m$. Now the covariance function becomes

$$R(v) = \begin{cases} \sigma^2 \cdot \left(1 - \frac{v}{m}\right) & \text{if } v \leq m - 1 \\ 0 & \text{if } v \geq m \end{cases}.$$  \tag{10.50}

10.8 Autoregressive processes

("autoregressziós folyamatok")

Definition III.72 Let $\{\zeta_n : n = 0, \pm 1, \pm 2, \ldots\}$ i.e. $\mathbb{T} = \mathbb{Z}$ (integers) be a covariance stationary process (see Def. III.55). Then, for any real number $\lambda \in \mathbb{R}$, $|\lambda| < 1$ the r.v. defined by

$$\xi_n = \zeta_n - \lambda \cdot \zeta_{n-1}$$  \tag{10.51}

are uncorrelated ("korrelálatlanok") with zero means and a common variance $\sigma^2$. The s.p. defined in (10.51) is called an autoregressive process of order one ("elsőrendű autoregressziós folyamat"). \hfill \square

Remark III.73 i) The word "regression" (latin) originally means "going back to the past, using the old things". As usual, "auto" (greek) means "self". In (10.51) $\lambda$ gives the "measure" of the autoregression.

ii) Recall, that "$\xi_i$ and $\xi_j$ are uncorrelated" only means that $\text{cov}(\xi_i, \xi_j) = 0$ which is weaker than "$\xi_i$ and $\xi_j$ are independent".

From (10.51) we may write

$$\zeta_n = \lambda \cdot \zeta_{n-1} + \xi_n = \cdots = \lambda^k \cdot \zeta_{n-k} + \sum_{j=0}^{k-1} \lambda^j \xi_{n-j} \quad \text{for } k \leq n.$$  \tag{10.52}

Further we have

Theorem III.74

$$M \left[ \left( \zeta_n - \sum_{j=0}^{k-1} \lambda^j \xi_{n-j} \right)^2 \right] = M \left[ \left( \lambda^k \zeta_{n-k} \right)^2 \right] = \lambda^{2k} \cdot M \left[ \zeta_{n-k}^2 \right].$$  \tag{10.53}
\[ M \left[ \zeta_{n-k}^2 \right] \] is constant, i.e. independent of \( n \) and \( k \), since the process \( \zeta \) is stationary.

Moreover, using \(|\lambda| < 1\), the right hand side of (10.53) decreases to 0 at a geometric rate. Thus

**Theorem III.75**

\[
\zeta_n = \lim_{k \to \infty} \sum_{j=0}^{k-1} \lambda^j \xi_{n-j} = \sum_{j=0}^{\infty} \lambda^j \xi_{n-j} \quad \text{in m.s.} \tag{10.54}
\]

where **m.s.** means **mean square distance** ("négyzetes közép távolságből") limit.

Equation (10.54) provides a representation of the original process \( \zeta \) as a moving average process.

Since mean square convergence implies convergence of the means and second moments, we have

**Theorem III.76**

\[
M(\zeta_n) = \lim_{k \to \infty} M \left( \sum_{j=0}^{k-1} \lambda^j \xi_{n-j} \right) = 0 \tag{10.55}
\]

and

\[
M(\zeta_n^2) = \frac{\sigma^2}{1 - \lambda^2}. \quad \square \tag{10.56}
\]

Let us compute the covariance between \( \zeta_n \) and \( \zeta_{n+k} \).

**Theorem III.77**

\[
M(\zeta_n \cdot \zeta_{n+k}) = \sigma^2 \cdot \lambda^k \tag{10.57}
\]

and so

\[
cov(\zeta_n, \zeta_{n+k}) = \sigma^2 \cdot \left( \lambda^k - \frac{1}{1 - \lambda^2} \right) \tag{10.58}
\]

for \( k \in \mathbb{N} \). \quad \square

The generalization of (10.51) is:

**Definition III.78** Let \( \{\zeta_n : n \in \mathbb{N}\} \) be a sequence of zero mean uncorrelated random variables having a common variance \( \sigma^2 \). Then the (stationary) process

\[
\zeta_n = \lambda_1 \xi_1 + \lambda_2 \xi_2 + \ldots + \lambda_p \xi_p + \xi_n \tag{10.59}
\]

for \(|\lambda_i| < 1\) is called a **\( p \) th order autoregressive process** ("\( p \)-edrendű autoregressziós folyamat"). \quad \square
10.9 White noise processes

("fehérzaj folyamatok")

**Definition III.79** The s.p. \( \mathbf{\xi} = \{\xi_t : t \in \mathbb{T}\} \) is a white noise process ("fehérzaj folyamat") if the following holds:

For every finite subset \( H \subset \mathbb{T} \) we have that \( \xi_H = \{\xi_t : t \in H\} \) are standard independent normal (Gaussian) distributions. \( \square \)

**Claim III.80** Clearly, by the independency, the common density function of \( \xi_H \) for \( H = \{h_1, h_2, \ldots, h_{|H|}\} \) is

\[
f_{\xi_H}(x_{h_1}, x_{h_2}, \ldots, x_{h_{|H|}}) = \frac{1}{\sqrt{(2\pi)^{|H|}}} \cdot \exp \left( -\frac{x_{h_1}^2 - x_{h_2}^2 - \ldots - x_{h_{|H|}}^2}{2} \right) \tag{10.60}\]

since the (one dimensional) standard normal density function is

\[
f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} . \quad \square
\]
References

[HHKSz] Hajba Tamás, Harmati István, Környei László, Szalay Krisztina: Valószínűségszámítás és matematikai statisztika, (in Hungarian), Sections 13-16 and 22-26 (pages 419-528),
http://www.sze.hu/~harmati/valszam/Valszam_mat_stat.pdf


KarlinTaylorIntrodStochModeling.pdf, 
http://matap.dmae.upm.es/WebpersonalBartolo/Probabilidad/


[SzI1] Szalkai I.: Illustrative Probability Theory, 2018,
https://math.uni-pannon.hu/~szalkai/Illustrative-PrTh-1.doc,
https://math.uni-pannon.hu/~szalkai/Illustrative-PrTh-1.pdf

Tables
### Standard normal distribution $\Phi(t)$

| $t$ | 0     | 0.1   | 0.2   | 0.3   | 0.4   | 0.5   | 0.6   | 0.7   | 0.8   | 0.9   | 1.0   | 1.1   | 1.2   | 1.3   | 1.4   | 1.5   | 1.6   | 1.7   | 1.8   | 1.9   | 2.0   | 2.1   | 2.2   | 2.3   | 2.4   | 2.5   | 2.6   | 2.7   | 2.8   | 2.9   | 3.0   | 3.1   | 3.2   | 3.3   | 3.4   | 3.5   | 3.6   | 3.7   | 3.8   | 3.9   |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|     | 0.5000| 0.5040| 0.5080| 0.5120| 0.5160| 0.5199| 0.5239| 0.5279| 0.5319| 0.5359| 0.5400| 0.5440| 0.5480| 0.5520| 0.5560| 0.5599| 0.5639| 0.5679| 0.5719| 0.5759| 0.5800| 0.5840| 0.5880| 0.5920| 0.5960| 0.6000| 0.6040| 0.6080| 0.6120| 0.6160| 0.6200| 0.6240| 0.6280| 0.6320| 0.6360| 0.6400| 0.6440| 0.6480| 0.6520| 0.6560| 0.6600| 0.6640| 0.6680| 0.6720| 0.6760| 0.6800| 0.6840| 0.6880| 0.6920| 0.6960| 0.7000| 0.7040| 0.7080| 0.7120| 0.7160| 0.7200| 0.7240| 0.7280| 0.7320| 0.7360| 0.7400| 0.7440| 0.7480| 0.7520| 0.7560| 0.7600| 0.7640| 0.7680| 0.7720| 0.7760| 0.7800| 0.7840| 0.7880| 0.7920| 0.7960| 0.8000| 0.8040| 0.8080| 0.8120| 0.8160| 0.8200| 0.8240| 0.8280| 0.8320| 0.8360| 0.8400| 0.8440| 0.8480| 0.8520| 0.8560| 0.8600| 0.8640| 0.8680| 0.8720| 0.8760| 0.8800| 0.8840| 0.8880| 0.8920| 0.8960| 0.9000| 0.9040| 0.9080| 0.9120| 0.9160| 0.9200| 0.9240| 0.9280| 0.9320| 0.9360| 0.9400| 0.9440| 0.9480| 0.9520| 0.9560| 0.9600| 0.9640| 0.9680| 0.9720| 0.9760| 0.9800| 0.9840| 0.9880| 0.9920| 0.9960| 0.9999|
### Critical values for $\chi^2$ distribution

<table>
<thead>
<tr>
<th>$s \backslash \beta$</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.001</th>
<th>0.95</th>
<th>0.975</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,706</td>
<td>3,841</td>
<td>5,024</td>
<td>6,635</td>
<td>10,827</td>
<td>0,004</td>
<td>0,001</td>
</tr>
<tr>
<td>2</td>
<td>4,605</td>
<td>5,991</td>
<td>7,378</td>
<td>9,210</td>
<td>13,815</td>
<td>0,103</td>
<td>0,051</td>
</tr>
<tr>
<td>3</td>
<td>6,251</td>
<td>7,815</td>
<td>9,348</td>
<td>11,345</td>
<td>16,268</td>
<td>0,352</td>
<td>0,022</td>
</tr>
<tr>
<td>4</td>
<td>7,779</td>
<td>9,488</td>
<td>11,143</td>
<td>13,277</td>
<td>18,465</td>
<td>0,711</td>
<td>0,484</td>
</tr>
<tr>
<td>5</td>
<td>9,236</td>
<td>11,070</td>
<td>12,833</td>
<td>15,086</td>
<td>20,517</td>
<td>1,145</td>
<td>0,831</td>
</tr>
<tr>
<td>6</td>
<td>10,645</td>
<td>12,592</td>
<td>14,449</td>
<td>16,812</td>
<td>22,457</td>
<td>1,635</td>
<td>1,237</td>
</tr>
<tr>
<td>7</td>
<td>12,017</td>
<td>14,067</td>
<td>16,013</td>
<td>18,475</td>
<td>24,322</td>
<td>2,167</td>
<td>1,690</td>
</tr>
<tr>
<td>8</td>
<td>13,362</td>
<td>15,507</td>
<td>17,535</td>
<td>20,090</td>
<td>26,125</td>
<td>2,733</td>
<td>2,180</td>
</tr>
<tr>
<td>9</td>
<td>14,684</td>
<td>16,919</td>
<td>19,023</td>
<td>21,666</td>
<td>27,877</td>
<td>3,325</td>
<td>2,700</td>
</tr>
<tr>
<td>10</td>
<td>15,987</td>
<td>18,307</td>
<td>20,483</td>
<td>23,209</td>
<td>29,588</td>
<td>3,940</td>
<td>3,247</td>
</tr>
<tr>
<td>11</td>
<td>17,275</td>
<td>19,675</td>
<td>21,920</td>
<td>24,725</td>
<td>31,264</td>
<td>4,575</td>
<td>3,816</td>
</tr>
<tr>
<td>12</td>
<td>18,549</td>
<td>21,026</td>
<td>23,337</td>
<td>26,217</td>
<td>32,909</td>
<td>5,226</td>
<td>4,404</td>
</tr>
<tr>
<td>13</td>
<td>19,812</td>
<td>22,362</td>
<td>24,736</td>
<td>27,688</td>
<td>34,528</td>
<td>5,892</td>
<td>5,009</td>
</tr>
<tr>
<td>14</td>
<td>21,064</td>
<td>23,685</td>
<td>26,119</td>
<td>29,141</td>
<td>36,123</td>
<td>6,571</td>
<td>5,629</td>
</tr>
<tr>
<td>15</td>
<td>22,307</td>
<td>24,996</td>
<td>27,488</td>
<td>30,578</td>
<td>37,697</td>
<td>7,261</td>
<td>6,262</td>
</tr>
<tr>
<td>16</td>
<td>23,542</td>
<td>26,296</td>
<td>28,845</td>
<td>32,000</td>
<td>39,252</td>
<td>7,962</td>
<td>6,908</td>
</tr>
<tr>
<td>17</td>
<td>24,769</td>
<td>27,590</td>
<td>30,191</td>
<td>33,409</td>
<td>40,790</td>
<td>8,672</td>
<td>7,564</td>
</tr>
<tr>
<td>18</td>
<td>25,989</td>
<td>28,869</td>
<td>31,526</td>
<td>34,805</td>
<td>42,312</td>
<td>9,390</td>
<td>8,231</td>
</tr>
<tr>
<td>19</td>
<td>27,204</td>
<td>30,144</td>
<td>32,852</td>
<td>36,191</td>
<td>43,820</td>
<td>10,117</td>
<td>8,901</td>
</tr>
<tr>
<td>20</td>
<td>28,412</td>
<td>31,410</td>
<td>34,170</td>
<td>37,566</td>
<td>45,315</td>
<td>10,851</td>
<td>9,591</td>
</tr>
<tr>
<td>21</td>
<td>29,615</td>
<td>32,671</td>
<td>35,479</td>
<td>38,932</td>
<td>46,797</td>
<td>11,591</td>
<td>10,283</td>
</tr>
<tr>
<td>22</td>
<td>30,813</td>
<td>33,924</td>
<td>36,781</td>
<td>40,289</td>
<td>48,268</td>
<td>12,338</td>
<td>10,982</td>
</tr>
<tr>
<td>23</td>
<td>32,007</td>
<td>35,172</td>
<td>38,076</td>
<td>41,638</td>
<td>49,728</td>
<td>13,091</td>
<td>11,689</td>
</tr>
<tr>
<td>24</td>
<td>33,196</td>
<td>36,415</td>
<td>39,364</td>
<td>42,980</td>
<td>51,179</td>
<td>13,484</td>
<td>12,401</td>
</tr>
<tr>
<td>25</td>
<td>34,382</td>
<td>37,652</td>
<td>40,646</td>
<td>44,314</td>
<td>52,620</td>
<td>14,611</td>
<td>13,120</td>
</tr>
<tr>
<td>26</td>
<td>35,563</td>
<td>38,885</td>
<td>41,923</td>
<td>45,642</td>
<td>54,052</td>
<td>15,379</td>
<td>13,844</td>
</tr>
<tr>
<td>27</td>
<td>36,741</td>
<td>40,113</td>
<td>43,194</td>
<td>46,963</td>
<td>55,476</td>
<td>16,151</td>
<td>14,573</td>
</tr>
<tr>
<td>28</td>
<td>37,916</td>
<td>41,337</td>
<td>44,461</td>
<td>48,278</td>
<td>56,893</td>
<td>16,928</td>
<td>15,308</td>
</tr>
<tr>
<td>29</td>
<td>39,087</td>
<td>42,557</td>
<td>45,772</td>
<td>49,558</td>
<td>58,302</td>
<td>17,708</td>
<td>16,047</td>
</tr>
<tr>
<td>30</td>
<td>40,256</td>
<td>43,773</td>
<td>46,979</td>
<td>50,892</td>
<td>59,703</td>
<td>18,493</td>
<td>16,791</td>
</tr>
</tbody>
</table>
### Student $t$-distribution

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>0.20</th>
<th>0.10</th>
<th>0.05</th>
<th>0.02</th>
<th>0.01</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
<td>31.821</td>
<td>63.657</td>
<td>636.619</td>
</tr>
<tr>
<td>2</td>
<td>1.886</td>
<td>2.92</td>
<td>4.303</td>
<td>6.965</td>
<td>9.925</td>
<td>31.598</td>
</tr>
<tr>
<td>3</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.51</td>
<td>5.841</td>
<td>12.941</td>
</tr>
<tr>
<td>4</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td>4.604</td>
<td>8.61</td>
</tr>
<tr>
<td>5</td>
<td>1.476</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
<td>4.032</td>
<td>6.859</td>
</tr>
<tr>
<td>6</td>
<td>1.44</td>
<td>1.943</td>
<td>2.447</td>
<td>3.143</td>
<td>3.707</td>
<td>5.959</td>
</tr>
<tr>
<td>7</td>
<td>1.415</td>
<td>1.895</td>
<td>2.365</td>
<td>2.998</td>
<td>3.499</td>
<td>5.405</td>
</tr>
<tr>
<td>8</td>
<td>1.397</td>
<td>1.86</td>
<td>2.306</td>
<td>2.896</td>
<td>3.355</td>
<td>5.041</td>
</tr>
<tr>
<td>9</td>
<td>1.383</td>
<td>1.833</td>
<td>2.262</td>
<td>2.821</td>
<td>3.25</td>
<td>4.781</td>
</tr>
<tr>
<td>10</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
<td>3.169</td>
<td>4.587</td>
</tr>
<tr>
<td>11</td>
<td>1.363</td>
<td>1.796</td>
<td>2.201</td>
<td>2.718</td>
<td>3.106</td>
<td>4.437</td>
</tr>
<tr>
<td>12</td>
<td>1.356</td>
<td>1.782</td>
<td>2.179</td>
<td>2.681</td>
<td>3.055</td>
<td>4.318</td>
</tr>
<tr>
<td>13</td>
<td>1.35</td>
<td>1.771</td>
<td>2.16</td>
<td>2.65</td>
<td>3.012</td>
<td>4.221</td>
</tr>
<tr>
<td>14</td>
<td>1.345</td>
<td>1.761</td>
<td>2.145</td>
<td>2.624</td>
<td>2.977</td>
<td>4.14</td>
</tr>
<tr>
<td>15</td>
<td>1.341</td>
<td>1.753</td>
<td>2.131</td>
<td>2.602</td>
<td>2.947</td>
<td>4.073</td>
</tr>
<tr>
<td>16</td>
<td>1.337</td>
<td>1.746</td>
<td>2.12</td>
<td>2.583</td>
<td>2.921</td>
<td>4.015</td>
</tr>
<tr>
<td>17</td>
<td>1.333</td>
<td>1.74</td>
<td>2.11</td>
<td>2.567</td>
<td>2.898</td>
<td>3.965</td>
</tr>
<tr>
<td>18</td>
<td>1.33</td>
<td>1.734</td>
<td>2.101</td>
<td>2.552</td>
<td>2.878</td>
<td>3.922</td>
</tr>
<tr>
<td>19</td>
<td>1.328</td>
<td>1.729</td>
<td>2.093</td>
<td>2.539</td>
<td>2.861</td>
<td>3.883</td>
</tr>
<tr>
<td>20</td>
<td>1.325</td>
<td>1.725</td>
<td>2.086</td>
<td>2.528</td>
<td>2.845</td>
<td>3.85</td>
</tr>
<tr>
<td>21</td>
<td>1.323</td>
<td>1.721</td>
<td>2.08</td>
<td>2.518</td>
<td>2.831</td>
<td>3.819</td>
</tr>
<tr>
<td>22</td>
<td>1.321</td>
<td>1.717</td>
<td>2.074</td>
<td>2.508</td>
<td>2.819</td>
<td>3.792</td>
</tr>
<tr>
<td>23</td>
<td>1.319</td>
<td>1.714</td>
<td>2.069</td>
<td>2.5</td>
<td>2.807</td>
<td>3.767</td>
</tr>
<tr>
<td>24</td>
<td>1.318</td>
<td>1.711</td>
<td>2.064</td>
<td>2.492</td>
<td>2.797</td>
<td>3.745</td>
</tr>
<tr>
<td>25</td>
<td>1.316</td>
<td>1.708</td>
<td>2.06</td>
<td>2.485</td>
<td>2.787</td>
<td>3.725</td>
</tr>
<tr>
<td>26</td>
<td>1.315</td>
<td>1.706</td>
<td>2.056</td>
<td>2.479</td>
<td>2.779</td>
<td>3.707</td>
</tr>
<tr>
<td>27</td>
<td>1.314</td>
<td>1.703</td>
<td>2.052</td>
<td>2.473</td>
<td>2.771</td>
<td>3.69</td>
</tr>
<tr>
<td>28</td>
<td>1.313</td>
<td>1.701</td>
<td>2.048</td>
<td>2.467</td>
<td>2.763</td>
<td>3.674</td>
</tr>
<tr>
<td>29</td>
<td>1.311</td>
<td>1.699</td>
<td>2.045</td>
<td>2.462</td>
<td>2.756</td>
<td>3.659</td>
</tr>
<tr>
<td>30</td>
<td>1.31</td>
<td>1.697</td>
<td>2.042</td>
<td>2.457</td>
<td>2.75</td>
<td>3.646</td>
</tr>
<tr>
<td>40</td>
<td>1.303</td>
<td>1.684</td>
<td>2.021</td>
<td>2.423</td>
<td>2.704</td>
<td>3.551</td>
</tr>
<tr>
<td>60</td>
<td>1.296</td>
<td>1.671</td>
<td>2</td>
<td>2.39</td>
<td>2.66</td>
<td>3.46</td>
</tr>
<tr>
<td>120</td>
<td>1.289</td>
<td>1.658</td>
<td>1.98</td>
<td>2.358</td>
<td>2.617</td>
<td>3.373</td>
</tr>
<tr>
<td>$\infty$</td>
<td>1.282</td>
<td>1.645</td>
<td>1.96</td>
<td>2.326</td>
<td>2.576</td>
<td>3.291</td>
</tr>
</tbody>
</table>
10.9. **WHITE NOISE PROCESSES**

<table>
<thead>
<tr>
<th>$P$</th>
<th>0.90</th>
<th>0.95</th>
<th>0.975</th>
<th>0.98</th>
<th>0.99</th>
<th>0.995</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.08</td>
<td>6.31</td>
<td>12.71</td>
<td>15.89</td>
<td>31.82</td>
<td>63.66</td>
</tr>
<tr>
<td>2</td>
<td>1.89</td>
<td>2.92</td>
<td>4.30</td>
<td>4.85</td>
<td>6.96</td>
<td>9.92</td>
</tr>
<tr>
<td>3</td>
<td>1.64</td>
<td>2.35</td>
<td>3.18</td>
<td>3.48</td>
<td>4.54</td>
<td>5.84</td>
</tr>
<tr>
<td>4</td>
<td>1.53</td>
<td>2.13</td>
<td>2.78</td>
<td>3.00</td>
<td>3.75</td>
<td>4.60</td>
</tr>
<tr>
<td>5</td>
<td>1.48</td>
<td>2.02</td>
<td>2.57</td>
<td>2.76</td>
<td>3.36</td>
<td>4.03</td>
</tr>
<tr>
<td>6</td>
<td>1.44</td>
<td>1.94</td>
<td>2.45</td>
<td>2.61</td>
<td>3.14</td>
<td>3.71</td>
</tr>
<tr>
<td>7</td>
<td>1.41</td>
<td>1.89</td>
<td>2.36</td>
<td>2.52</td>
<td>3.00</td>
<td>3.50</td>
</tr>
<tr>
<td>8</td>
<td>1.40</td>
<td>1.86</td>
<td>2.31</td>
<td>2.45</td>
<td>2.90</td>
<td>3.36</td>
</tr>
<tr>
<td>9</td>
<td>1.38</td>
<td>1.83</td>
<td>2.26</td>
<td>2.40</td>
<td>2.82</td>
<td>3.25</td>
</tr>
<tr>
<td>10</td>
<td>1.37</td>
<td>1.81</td>
<td>2.23</td>
<td>2.36</td>
<td>2.76</td>
<td>3.17</td>
</tr>
<tr>
<td>11</td>
<td>1.36</td>
<td>1.80</td>
<td>2.20</td>
<td>2.33</td>
<td>2.72</td>
<td>3.11</td>
</tr>
<tr>
<td>12</td>
<td>1.36</td>
<td>1.78</td>
<td>2.18</td>
<td>2.30</td>
<td>2.68</td>
<td>3.05</td>
</tr>
<tr>
<td>13</td>
<td>1.35</td>
<td>1.77</td>
<td>2.16</td>
<td>2.28</td>
<td>2.65</td>
<td>3.01</td>
</tr>
<tr>
<td>14</td>
<td>1.35</td>
<td>1.76</td>
<td>2.14</td>
<td>2.26</td>
<td>2.62</td>
<td>2.98</td>
</tr>
<tr>
<td>15</td>
<td>1.34</td>
<td>1.75</td>
<td>2.13</td>
<td>2.25</td>
<td>2.60</td>
<td>2.95</td>
</tr>
<tr>
<td>16</td>
<td>1.34</td>
<td>1.75</td>
<td>2.12</td>
<td>2.24</td>
<td>2.58</td>
<td>2.92</td>
</tr>
<tr>
<td>17</td>
<td>1.33</td>
<td>1.74</td>
<td>2.11</td>
<td>2.22</td>
<td>2.57</td>
<td>2.90</td>
</tr>
<tr>
<td>18</td>
<td>1.33</td>
<td>1.73</td>
<td>2.10</td>
<td>2.21</td>
<td>2.55</td>
<td>2.88</td>
</tr>
<tr>
<td>19</td>
<td>1.33</td>
<td>1.73</td>
<td>2.09</td>
<td>2.20</td>
<td>2.54</td>
<td>2.86</td>
</tr>
<tr>
<td>20</td>
<td>1.33</td>
<td>1.72</td>
<td>2.09</td>
<td>2.20</td>
<td>2.53</td>
<td>2.85</td>
</tr>
<tr>
<td>21</td>
<td>1.32</td>
<td>1.72</td>
<td>2.08</td>
<td>2.19</td>
<td>2.52</td>
<td>2.83</td>
</tr>
<tr>
<td>22</td>
<td>1.32</td>
<td>1.72</td>
<td>2.07</td>
<td>2.18</td>
<td>2.51</td>
<td>2.82</td>
</tr>
<tr>
<td>23</td>
<td>1.32</td>
<td>1.71</td>
<td>2.07</td>
<td>2.18</td>
<td>2.50</td>
<td>2.81</td>
</tr>
<tr>
<td>24</td>
<td>1.32</td>
<td>1.71</td>
<td>2.06</td>
<td>2.17</td>
<td>2.49</td>
<td>2.80</td>
</tr>
<tr>
<td>25</td>
<td>1.32</td>
<td>1.71</td>
<td>2.06</td>
<td>2.17</td>
<td>2.49</td>
<td>2.79</td>
</tr>
<tr>
<td>26</td>
<td>1.31</td>
<td>1.71</td>
<td>2.06</td>
<td>2.16</td>
<td>2.48</td>
<td>2.78</td>
</tr>
<tr>
<td>27</td>
<td>1.31</td>
<td>1.70</td>
<td>2.05</td>
<td>2.16</td>
<td>2.47</td>
<td>2.77</td>
</tr>
<tr>
<td>28</td>
<td>1.31</td>
<td>1.70</td>
<td>2.05</td>
<td>2.15</td>
<td>2.47</td>
<td>2.76</td>
</tr>
<tr>
<td>29</td>
<td>1.31</td>
<td>1.70</td>
<td>2.05</td>
<td>2.15</td>
<td>2.46</td>
<td>2.76</td>
</tr>
<tr>
<td>30</td>
<td>1.31</td>
<td>1.70</td>
<td>2.04</td>
<td>2.15</td>
<td>2.46</td>
<td>2.75</td>
</tr>
<tr>
<td>35</td>
<td>1.31</td>
<td>1.69</td>
<td>2.03</td>
<td>2.13</td>
<td>2.44</td>
<td>2.72</td>
</tr>
<tr>
<td>40</td>
<td>1.30</td>
<td>1.68</td>
<td>2.02</td>
<td>2.12</td>
<td>2.42</td>
<td>2.70</td>
</tr>
<tr>
<td>45</td>
<td>1.30</td>
<td>1.68</td>
<td>2.01</td>
<td>2.12</td>
<td>2.41</td>
<td>2.69</td>
</tr>
<tr>
<td>50</td>
<td>1.30</td>
<td>1.68</td>
<td>2.01</td>
<td>2.11</td>
<td>2.40</td>
<td>2.68</td>
</tr>
<tr>
<td>60</td>
<td>1.30</td>
<td>1.67</td>
<td>2.00</td>
<td>2.10</td>
<td>2.39</td>
<td>2.66</td>
</tr>
<tr>
<td>70</td>
<td>1.29</td>
<td>1.67</td>
<td>1.99</td>
<td>2.09</td>
<td>2.38</td>
<td>2.65</td>
</tr>
<tr>
<td>80</td>
<td>1.29</td>
<td>1.66</td>
<td>1.99</td>
<td>2.09</td>
<td>2.37</td>
<td>2.64</td>
</tr>
<tr>
<td>90</td>
<td>1.29</td>
<td>1.66</td>
<td>1.99</td>
<td>2.08</td>
<td>2.37</td>
<td>2.63</td>
</tr>
<tr>
<td>100</td>
<td>1.29</td>
<td>1.66</td>
<td>1.98</td>
<td>2.08</td>
<td>2.36</td>
<td>2.63</td>
</tr>
<tr>
<td>200</td>
<td>1.29</td>
<td>1.65</td>
<td>1.97</td>
<td>2.07</td>
<td>2.35</td>
<td>2.60</td>
</tr>
</tbody>
</table>
### Critical values of F-distribution for $\varepsilon = 0.05$ (95%)

(degrees of freedom for the numerator)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>161.4</td>
<td>199.5</td>
<td>215.7</td>
<td>224.6</td>
<td>230.2</td>
<td>234.0</td>
<td>236.8</td>
<td>238.9</td>
<td>240.5</td>
<td>243.9</td>
</tr>
<tr>
<td>3</td>
<td>10.13</td>
<td>6.94</td>
<td>6.59</td>
<td>6.39</td>
<td>6.26</td>
<td>6.16</td>
<td>6.09</td>
<td>6.04</td>
<td>6.00</td>
<td>5.91</td>
</tr>
<tr>
<td>4</td>
<td>6.61</td>
<td>5.79</td>
<td>5.41</td>
<td>5.19</td>
<td>5.05</td>
<td>4.95</td>
<td>4.88</td>
<td>4.82</td>
<td>4.77</td>
<td>4.68</td>
</tr>
<tr>
<td>5</td>
<td>5.99</td>
<td>5.14</td>
<td>4.76</td>
<td>4.53</td>
<td>4.39</td>
<td>4.28</td>
<td>4.21</td>
<td>4.15</td>
<td>4.10</td>
<td>4.00</td>
</tr>
<tr>
<td>6</td>
<td>5.59</td>
<td>4.74</td>
<td>4.35</td>
<td>4.12</td>
<td>3.97</td>
<td>3.87</td>
<td>3.79</td>
<td>3.73</td>
<td>3.68</td>
<td>3.57</td>
</tr>
<tr>
<td>7</td>
<td>5.32</td>
<td>4.46</td>
<td>4.07</td>
<td>3.84</td>
<td>3.69</td>
<td>3.58</td>
<td>3.50</td>
<td>3.44</td>
<td>3.39</td>
<td>3.28</td>
</tr>
<tr>
<td>8</td>
<td>5.12</td>
<td>4.26</td>
<td>3.86</td>
<td>3.63</td>
<td>3.48</td>
<td>3.37</td>
<td>3.29</td>
<td>3.23</td>
<td>3.18</td>
<td>3.07</td>
</tr>
<tr>
<td>9</td>
<td>4.96</td>
<td>4.10</td>
<td>3.71</td>
<td>3.48</td>
<td>3.33</td>
<td>3.22</td>
<td>3.14</td>
<td>3.07</td>
<td>3.02</td>
<td>2.91</td>
</tr>
<tr>
<td>10</td>
<td>4.84</td>
<td>3.98</td>
<td>3.59</td>
<td>3.36</td>
<td>3.20</td>
<td>3.09</td>
<td>3.01</td>
<td>2.95</td>
<td>2.90</td>
<td>2.79</td>
</tr>
<tr>
<td>11</td>
<td>4.75</td>
<td>3.88</td>
<td>3.49</td>
<td>3.26</td>
<td>3.11</td>
<td>3.00</td>
<td>2.91</td>
<td>2.85</td>
<td>2.80</td>
<td>2.69</td>
</tr>
<tr>
<td>12</td>
<td>4.67</td>
<td>3.80</td>
<td>3.41</td>
<td>3.18</td>
<td>3.02</td>
<td>2.92</td>
<td>2.83</td>
<td>2.77</td>
<td>2.71</td>
<td>2.66</td>
</tr>
<tr>
<td>13</td>
<td>4.60</td>
<td>3.74</td>
<td>3.34</td>
<td>3.11</td>
<td>2.96</td>
<td>2.85</td>
<td>2.76</td>
<td>2.70</td>
<td>2.65</td>
<td>2.55</td>
</tr>
<tr>
<td>14</td>
<td>4.54</td>
<td>3.68</td>
<td>3.29</td>
<td>3.06</td>
<td>2.90</td>
<td>2.79</td>
<td>2.71</td>
<td>2.64</td>
<td>2.59</td>
<td>2.48</td>
</tr>
<tr>
<td>15</td>
<td>4.49</td>
<td>3.63</td>
<td>3.24</td>
<td>3.01</td>
<td>2.85</td>
<td>2.74</td>
<td>2.66</td>
<td>2.50</td>
<td>2.45</td>
<td>2.31</td>
</tr>
<tr>
<td>16</td>
<td>4.45</td>
<td>3.59</td>
<td>3.20</td>
<td>2.96</td>
<td>2.81</td>
<td>2.70</td>
<td>2.61</td>
<td>2.55</td>
<td>2.49</td>
<td>2.38</td>
</tr>
<tr>
<td>17</td>
<td>4.41</td>
<td>3.55</td>
<td>3.16</td>
<td>2.93</td>
<td>2.77</td>
<td>2.66</td>
<td>2.58</td>
<td>2.51</td>
<td>2.46</td>
<td>2.33</td>
</tr>
<tr>
<td>18</td>
<td>4.38</td>
<td>3.52</td>
<td>3.13</td>
<td>2.90</td>
<td>2.74</td>
<td>2.63</td>
<td>2.54</td>
<td>2.48</td>
<td>2.42</td>
<td>2.28</td>
</tr>
<tr>
<td>19</td>
<td>4.35</td>
<td>3.49</td>
<td>3.10</td>
<td>2.87</td>
<td>2.71</td>
<td>2.60</td>
<td>2.51</td>
<td>2.45</td>
<td>2.39</td>
<td>2.24</td>
</tr>
<tr>
<td>20</td>
<td>4.32</td>
<td>3.47</td>
<td>3.07</td>
<td>2.84</td>
<td>2.68</td>
<td>2.57</td>
<td>2.49</td>
<td>2.43</td>
<td>2.37</td>
<td>2.23</td>
</tr>
<tr>
<td>21</td>
<td>4.30</td>
<td>3.44</td>
<td>3.05</td>
<td>2.82</td>
<td>2.66</td>
<td>2.55</td>
<td>2.46</td>
<td>2.40</td>
<td>2.34</td>
<td>2.22</td>
</tr>
<tr>
<td>22</td>
<td>4.28</td>
<td>3.42</td>
<td>3.03</td>
<td>2.80</td>
<td>2.64</td>
<td>2.53</td>
<td>2.44</td>
<td>2.38</td>
<td>2.32</td>
<td>2.20</td>
</tr>
<tr>
<td>23</td>
<td>4.26</td>
<td>3.40</td>
<td>3.01</td>
<td>2.78</td>
<td>2.62</td>
<td>2.51</td>
<td>2.42</td>
<td>2.36</td>
<td>2.30</td>
<td>2.18</td>
</tr>
<tr>
<td>24</td>
<td>4.24</td>
<td>3.38</td>
<td>2.99</td>
<td>2.76</td>
<td>2.60</td>
<td>2.49</td>
<td>2.40</td>
<td>2.34</td>
<td>2.28</td>
<td>2.16</td>
</tr>
<tr>
<td>25</td>
<td>4.22</td>
<td>3.37</td>
<td>2.98</td>
<td>2.74</td>
<td>2.59</td>
<td>2.47</td>
<td>2.39</td>
<td>2.32</td>
<td>2.27</td>
<td>2.15</td>
</tr>
<tr>
<td>26</td>
<td>4.21</td>
<td>3.35</td>
<td>2.96</td>
<td>2.73</td>
<td>2.57</td>
<td>2.46</td>
<td>2.37</td>
<td>2.30</td>
<td>2.25</td>
<td>2.13</td>
</tr>
<tr>
<td>27</td>
<td>4.20</td>
<td>3.34</td>
<td>2.95</td>
<td>2.71</td>
<td>2.56</td>
<td>2.44</td>
<td>2.36</td>
<td>2.29</td>
<td>2.24</td>
<td>2.12</td>
</tr>
<tr>
<td>28</td>
<td>4.18</td>
<td>3.33</td>
<td>2.93</td>
<td>2.70</td>
<td>2.54</td>
<td>2.43</td>
<td>2.35</td>
<td>2.28</td>
<td>2.22</td>
<td>2.10</td>
</tr>
<tr>
<td>29</td>
<td>4.17</td>
<td>3.32</td>
<td>2.92</td>
<td>2.69</td>
<td>2.53</td>
<td>2.42</td>
<td>2.33</td>
<td>2.27</td>
<td>2.21</td>
<td>2.09</td>
</tr>
<tr>
<td>30</td>
<td>4.08</td>
<td>3.23</td>
<td>2.84</td>
<td>2.61</td>
<td>2.45</td>
<td>2.34</td>
<td>2.25</td>
<td>2.19</td>
<td>2.12</td>
<td>2.00</td>
</tr>
<tr>
<td>60</td>
<td>3.92</td>
<td>3.07</td>
<td>2.68</td>
<td>2.45</td>
<td>2.29</td>
<td>2.17</td>
<td>2.09</td>
<td>2.02</td>
<td>1.96</td>
<td>1.83</td>
</tr>
<tr>
<td>∞</td>
<td>3.84</td>
<td>2.99</td>
<td>2.60</td>
<td>2.37</td>
<td>2.21</td>
<td>2.09</td>
<td>2.01</td>
<td>1.94</td>
<td>1.88</td>
<td>1.75</td>
</tr>
</tbody>
</table>
Index

Anscombe, Francis John, 65
autocovariance, 7
correlation index, 65
covariance matrix
auto/self, 20
covariance, 6
dataset
symmetric to the origin, 29
degree of freedom, 27
distribution
border / marginal, 5
common / joint, 5
distribution function
common / joint, 5
double logarithmic coordinate system, 71
empirical
average / expected value / mean, 27
dispersion / variance, 27
corrected, 27
squared mean, 27
error
type I, 38
type II, 38
estimation
consistent, 37
efficient, 37
parameter, 37
unbiased, 37
Gamma function, 81
Galton, Sir Francis Eugene, 55
GFI, 47
hypothesis, 38
accepted, 38
alternative, 38
null-, 38
rejected, 38
independent
r.v., 6
interval
confidence, 31
estimations, 31
INDEX

joint distribution function, 5

Least Squares Method, 57

level
  confidence-, 31
  significance-, 38

m.s., 112

matrix
  positive definite, 20
  symmetric, 21

mean square distance, 112

multinomial coefficient, 23

normed covariance, 9

o(h), 90

parabola regression, 76

Pearson, Karl, 10, 47

polynomial coefficient, 23

polynomial regression, 57

positive definite
  matrix, 20

r.v.
  independent, 6
  uncorrelated, 8

reciprocal coordinate system, 73

regression function
  of first kind, 56

regression function
  of second kind, 59

ruler method, 68

sample (statistical), 27

sampling with repetitions, 23

sampling without repetitions, 23

selfcovarience, 7

semi-, 69

semilogarithmic coordinate system, 69

standardized
  dataset, 29

random variable, 29

statistic, 29

symmetric
  matrix, 21

symmetric to the origin, 29

test, 38
  nonparametric, 39
  one sided, 39
  parametric, 39
  two sided, 39

uncorrelated r.v., 8