Mathematical Statistics and Stochastic Processes

István Szalkai
szalkai@almos.uni-pannon.hu
Pannon University, Veszprém, Hungary

11.11.2019.

Contents

Introduction

Mathematical Statisztics and *Stochastic Processes* became extremaly 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 neccessary. 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 ("...").

$\square$ = 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,

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

$\mathbf{a}$, $\overrightarrow{a}$, $\underline{a}$ = vectors,

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

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

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

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

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

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

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

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

F_ξ , F , $G, H : \mathbb{R} \to \mathbb{R}$ = distribution functions ("eloszlásfüggvények"),

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

f' , $\dfrac{df}{dx}$, $\dfrac{d}{dx}f$ = derivatives of f ,

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

$D(\xi)$, $\sigma(\xi)$, σ_ξ = *dispersion* of ξ ("ξ szórása"),

$D^2(\xi)$, $\sigma^2(\xi)$, σ_ξ^2 , $var(\xi)$ = *variance* of ξ ("ξ szórásnégyzete").

$\xi^* := \dfrac{\xi - M(\xi)}{D(\xi)}$ is the standardized version of ξ .

Remark .1 $\ln(x)$, $\log(x)$ *usually denote the natural logarithm (base e) and* $\lg(x)$ *the* $\log_{10}(x)$, *but different books, programs and users can use other choiches, please check it in each situation. However, in most applications there is no substant 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 measurings at an experiment, so it is better to consider the r.v. *vector* of data $\overrightarrow{\zeta} = (\zeta_1, ..., \zeta_n)$ instead of a set or separate r.v. $\{\zeta_1, ..., \zeta_n\}$.

Chapter 1

Two - dimensional random variables and independence

Definition I.1 $\overrightarrow{\zeta} : \Omega \to \mathbb{R}^2$ *is a **2 dimensional r.v.** or a vector-r.v.* □

Explanations: $\overrightarrow{\zeta} = (\xi, \eta) = \begin{pmatrix} \xi \\ \eta \end{pmatrix}$, i.e. $\overrightarrow{\zeta}(\omega) = (\ \xi(\omega), \eta(\omega)\ )$ for $\omega \in \Omega$, so ξ and η are the *coordinate (function)s* of $\overrightarrow{\zeta}$.

In fact, ξ and η are *any two* r.v. as you like: $\xi, \eta : \Omega \to \mathbb{R}$.

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

1.1 General definitions

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

$$F_{\overrightarrow{\zeta}} : \mathbb{R}^2 \to \mathbb{R} \ , \qquad F_{\overrightarrow{\zeta}}(x, y) := P(\xi < x, \eta < y) \ . \tag{1.1}$$

□

In what follows, we simply write ζ and $F_{\boldsymbol{\zeta}}$ instead of $\overrightarrow{\zeta}$ and $F_{\overrightarrow{\zeta}}$.

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

Definition I.4 *By the theorem above ξ and η are called the **marginal** (or border) **distributions of** $\overrightarrow{\zeta}$, ("*határeloszlás*" or "*peremeloszlás*").*

Definition I.5 *ξ and η are* ***independent*** *(of each-other) if*

$$\forall x, y \in \mathbb{R} \quad F_\zeta(x, y) = F_\xi(x) \cdot F_\eta(y) \ . \tag{1.2}$$

□

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

For the following notions ξ and η do *not* need to have a common distribution function.

Definition I.6 *The* ***covariance*** *(in Hungarian: "*kovariencia*") of ξ and η is:*

$$\mathbf{cov}(\xi, \eta) := M(\ (\xi - m_\xi) \cdot (\eta - m_\eta)\ ) \tag{1.3}$$

where $m_\xi = M(\xi)$ and $m_\eta = M(\eta)$, or, without abbreviations

$$\mathbf{cov}(\xi, \eta) := M(\ (\xi - M(\xi)) \cdot (\eta - M(\eta))\ ) \ .$$

$cov(\xi, \eta)$ is also denoted by $\sigma_{\xi,\eta}$. □

Remark I.7 *"co-variance" literaly means* varying together *("együtt változás"). $cov(\xi, \eta)$ really detects the changing measure of ξ and η . Look: $\xi - M(\xi)$ and $\eta - M(\eta)$ are the differences of ξ and η 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 "ξ and η move in the* same *direction" (concerning to their means), i.e. ξ and η help and strenghten 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 "ξ and η move in* other *directions", i.e. ξ and η 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. ξ, η and $a, b, c, d \in \mathbb{R}$ real numbers (constant r.v.) we have*

(o) $\quad cov(\xi, \eta) = M(\xi \cdot \eta) - M(\xi) \cdot M(\eta)$,

(i) *if $M(\xi) = 0$ then $cov(\xi, \eta) = M(\xi \cdot \eta)$* ,

(ii) *if ξ and η are independent, then* $\quad cov(\xi, \eta) = 0$,

(iii) but the reverse implication is not true in general,
however *it is true for normal distributions,*

(iv) $D^2(\xi+\eta) = D^2(\xi) + D^2(\eta) + 2\cdot cov(\xi,\eta)$ *for any two r.v.* ξ *and* η ,

(v) $cov(\xi,\xi) = D^2(\xi)$ ***(auto/self covarience**, "saját/ön- kovariencia"),*

(vi) $cov(\xi,\eta) = cov(\eta,\xi)$ ***(symmetry**, "szimmetrikusság"),*

(vii) $cov(a\xi+b, c\eta+d) = ac\cdot cov(\xi,\eta)$,

(viii) $cov(\xi,\eta) = cov(\xi - M(\xi)\ ,\ \eta - M(\eta))$,

(ix) $cov(a\xi+b, a\xi+b) = a^2D^2(\xi)$,

(x) $cov(a,\eta) = 0$,

(xi) $cov(a_1\xi_1 + a_2\xi_2\ ,\ b_1\eta_1 + b_2\eta_2) =$
$= a_1b_1cov(\xi_1,\eta_1) + a_1b_2cov(\xi_1,\eta_2) + a_2b_1cov(\xi_2,\eta_1) + a_2b_2cov(\xi_2,\eta_2)$.

Proof. (o) by definition $cov(\xi,\eta) =$

$$= M((\xi - m_\xi)\cdot(\eta - m_\eta)) = M(\xi\eta) - M(\xi m_\eta) - M(\eta m_\xi) + M(m_\xi m_\eta)$$
$$= M(\xi\eta) - m_\eta\cdot M(\xi) - m_\xi\cdot M(\eta) + m_\xi m_\eta$$
$$= M(\xi\eta) - m_\eta\cdot m_\xi \quad - m_\xi\cdot m_\eta \quad + m_\xi m_\eta$$
$$= M(\xi\eta) - m_\xi m_\eta = M(\xi\cdot\eta) - M(\xi)\cdot M(\eta) .$$

(i) follows from (o).

(ii) if ξ and η are independent then $M(\xi\cdot\eta) = M(\xi)\cdot M(\eta)$ (see [SzI1]).

(iii) we do not prove it here.

(iv) $D^2(\xi+\eta) = M\left([\xi+\eta-m_\xi-m_\eta]^2\right) =$

$$= M\left([\xi-m_\xi]^2\right) + M\left([\eta-m_\eta]^2\right) + 2\cdot M((\xi-m_\xi)\cdot(\eta-m_\eta))$$
$$= D^2(\xi) + D^2(\eta) + 2\cdot cov(\xi,\eta) .$$

(v) by definition $cov(\xi,\xi) := M\left((\xi-m_\xi)^2\right) = D^2(\xi)$.

(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

$$cov\,(a\xi + b, c\eta + d) = M\,(ac\,(\xi - m_\xi)\,(\eta - m_\eta))$$
$$= ac \cdot M\,((\xi - m_\xi)\,(\eta - m_\eta)) = ac \cdot cov\,(\xi, \eta)\ .$$

(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) $cov\,(a, \eta) = M\,(a \cdot \eta) - M\,(a) \cdot M\,(\eta) = a \cdot M\,(\eta) - a \cdot M\,(\eta) = 0$. ■

Remark I.9 *(o) Clearly* $(\xi \cdot \eta)\,(\omega) = \xi\,(\omega) \cdot \eta\,(\omega)$ *for* $\omega \in \Omega$.

(ii) and (iii) say that calculating $cov\,(\xi, \eta)$ *can not decide the independence of* ξ *and* η *, in the case* $cov\,(\xi, \eta) = 0$ *we can only say that* ξ *and* η *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. ξ, η *, since (iv) is valid for* any *r.v.* ξ *and* η *(see also [SzI1]).*

(vii) Clearly $cov\,(\xi, \eta)$ *changes when we change* measure units *(cm or km), since such a change zooms (in or out) the fluctuations of* ξ *and* η*. For this reason* $cov\,(\xi, \eta)$ differs *from* $cov\,(\xi^*, \eta^*)$ *where* $\xi^* = \frac{\xi - M(\xi)}{D(\xi)}$ *and* $\eta^* = \frac{\eta - M(\eta)}{D(\eta)}$ *are the* standard versions *of* ξ *and* η *. This phenomenom is called "*$cov\,(\xi, \eta)$ *is* not *normed" or "depends upon the scales" ("skálafüggő"). The normed version of* $cov\,(\xi, \eta)$ *is the correlation coefficient (see below).*

(viii) must be clear by everyday thinking: the covarience ("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* ξ *, nor* ξ *with* a *.*

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

First example: Let ξ be a uniform (continuous) r.v. on the interval [-1,1] and let $\eta = \xi^2$, clearly ξ and η are *not* independent (please check). However, by (o)

$$cov\,(\xi, \eta) = M\left(\xi \cdot \xi^2\right) - M\,(\xi) \cdot M\left(\xi^2\right) = M\left(\xi^3\right) - M\,(\xi) \cdot M\left(\xi^2\right) = 0 - 0 = 0$$

since $M\left(\xi^3\right) = M\,(\xi) = 0$. Similarly $cov\left(\xi, \xi^2\right) = 0$ 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 $\mathrm{Im}\,(X) = \{0, 2\}$, $\mathrm{Im}\,(Y) = \{0, 1, 2\}$, $P(X = 0, Y = 1) = \frac{1}{2}$, $P(X = 2, Y = 0) = P(X = 2, Y = 2) = \frac{1}{4}$ and the other possibilities are zero:

X\Y	0	1	2	Σ
0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
2	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
Σ	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

So $P(X=0)=P(X=2)=\frac{1}{2}$, $P(Y=0)=P(Y=2)=\frac{1}{4}$ and $P(Y=1)=\frac{1}{2}$. Further $M(X)=M(Y)=1$ and $M(X\cdot Y)=0+0+2\cdot 2\cdot\frac{1}{4}=1$ so $cov(X,Y)=0$, i.e. X and Y are uncorrelated. On the other hand X and Y are not independent, since

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

(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 .
ξ and η are uncorrelated since

$$cov(\xi,\eta)=cov(X+Y,X-Y)=cov(X,X)-cov(X,Y)+cov(Y,X)-cov(Y,Y)$$

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

However ξ and η 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

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

See also: https://en.wikipedia.org/wiki/Covariance Subsection 3.4 = ,
https://en.wikipedia.org/wiki/Covariance#Uncorrelatedness_and_independence ,
https://en.wikipedia.org/wiki/Correlation_and_dependence ,
https://hu.wikipedia.org/wiki/Kovariancia (in Hungarian),
https://de.wikipedia.org/wiki/Kovarianz_(Stochastik) (in German).

Remark I.11 *The main* disadvantage *of cov is property (vii): depends on the* scales *(measure units) a and c of ξ and η . 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*

$$R(\xi,\eta):=\frac{cov(\xi,\eta)}{D(\xi)\cdot D(\eta)}\ . \tag{1.4}$$

Other notations are $r(\xi,\eta)$ and $\rho\{\xi,\eta\}$. $\square$

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* ξ *and* η *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:*

$$\xi \text{ and } \eta \text{ are independent} \iff R(\xi, \eta) = 0 \ ,$$

(v) $|R(\xi, \eta)| = 1$ *if and only if* ξ *and* η *are "the same":*

$$\eta = a \cdot \xi + b \qquad \text{for some } a, b \in \mathbb{R} \ , \ a \neq 0 \ . \tag{1.5}$$

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

$$m_\eta = M(\eta) = M(a\xi + b) = aM(\xi) + b = am_\xi + b$$

and the definition the *enumerator is*

$$cov(\xi, \eta) = M((\xi - m_\xi)(\eta - m_\eta)) = M((\xi - m_\xi)(a\xi + b - (am_\xi + b)))$$

$$= M((\xi - m_\xi)(a(\xi - m_\xi))) = M\left(a(\xi - m_\xi)^2\right) = a \cdot D^2(\xi) \ ,$$

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) $\left(\sum\limits_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum\limits_{i=1}^{n} x_i^2\right) \cdot \left(\sum\limits_{i=1}^{n} y_i^2\right)$ for any $x_1, y_1, ..., x_n, y_n \in \mathbb{R}$ real numbers and $n \in \mathbb{N}$,

(C) $\left(\sum\limits_{i=1}^{\infty} x_i y_i\right)^2 \leq \left(\sum\limits_{i=1}^{\infty} x_i^2\right) \cdot \left(\sum\limits_{i=1}^{\infty} y_i^2\right)$ for any $x_1, y_1, ..., x_n, y_n, ... \in \mathbb{R}$ sequences, if the sums are finite,

(BS) $\left(\int\limits_a^b f(x) g(x)\, dx\right)^2 \leq \left(\int\limits_a^b f^2(x)\, dx\right) \cdot \left(\int\limits_a^b g^2(x)\, dx\right)$ for any functions $f, g : \mathbb{R} \to \mathbb{R}$, if the integrals are finite.

In general: $\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \cdot \langle \mathbf{y}, \mathbf{y} \rangle$ for any scalar product $\langle ., . \rangle$. □

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

The other direction is more difficult. ■

Remark I.15 *The main significancy 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 ξ and η " has no mathematical background or meaning, we feel and say similar sentences.*

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 η from ξ (and back, of ξ from η) 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: $Y[^oF] = 1.8 \cdot X[^oC] + 32$ and $X[^oC] = \frac{1}{1.8} Y[^oF] - \frac{32}{1.8} \approx 0.5556 \cdot Y[^oF] - 17.7778$.*

The quantities $\mathrm{cov}(\xi,\eta)$ 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

Definition I.17 *If $\mathrm{Im}(\xi) = \{x_1, x_2, ..., x_n, ...\}$ and $\mathrm{Im}(\eta) = \{y_1, y_2, ..., y_m, ...\}$ then the* ***distribution of*** $\overrightarrow{\zeta} = (\xi,\eta)$ *(or: the common/joint distribution of ξ and η) is the set of probabilities: $\{p_{i,j} : 1 \le i,j \le \infty\}$ where*

$$p_{i,j} := P(\xi = x_i \ , \ \eta = y_j) \ . \quad \square \tag{1.6}$$

Clearly

$$0 \le p_{i,j} \le 1 \quad \text{and} \quad \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{i,j} = 1 \ . \tag{1.7}$$

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

Definition I.18

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

are the ***marginal*** *(or border)* ***distributions*** *("peremeloszlások")* ***of*** $\overrightarrow{\zeta}$. $\square$

Theorem I.19 *In fact, the sets of probabilities*

$$\left\{q_i^{(\xi)} : 1 \leq i \leq \infty\right\} \quad \text{and} \quad \left\{q_j^{(\eta)} : 1 \leq j \leq \infty\right\} \tag{1.9}$$

are the distributions of ξ *and* η *.* □

Theorem I.20 *The discrete r.v.* ξ *and* η *are* ***independent*** *if and only if for every* $i, j \in \mathbb{N}$ *we have*

$$P(\xi = x_i \ , \ \eta = y_j) = P(\xi = x_i) \cdot P(\eta = y_j) \tag{1.10}$$

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 $F_\zeta(x, y) = \sum_{x_i < x} \sum_{y_j < y} p_{i,j}$ *for any* $x, y \in \mathbb{R}$,

$$F_\xi(x) = \sum_{x_i < x} q_i^{(\xi)} \quad \text{and} \quad F_\eta(y) = \sum_{y_j < y} q_j^{(\eta)} \ . \qquad \square$$

Theorem I.23 $M(\xi \cdot \eta) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{i,j} \cdot x_i \cdot y_j$,

$$M(\xi) = \sum_{i=1}^{\infty} q_i^{(\xi)} \cdot x_i \quad \text{and} \quad M(\eta) = \sum_{j=1}^{\infty} q_j^{(\eta)} \cdot y_j \ . \qquad \square$$

1.3 Summary and an example

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

$\xi \backslash \eta$	y_1	y_2	...	y_j	...	y_m	ξ_{marg}
x_1	$p_{1,1}$	$p_{1,2}$	...	$p_{1,j}$	...	$p_{1,m}$	$q_1^{(\xi)}$
x_2	$p_{2,1}$	$p_{2,2}$	...	$p_{2,j}$	...	$p_{2,m}$	$q_2^{(\xi)}$
...	...	...	...	...	...	...	...
x_i	$p_{i,1}$	$p_{i,2}$	...	$p_{i,j}$	...	$p_{i,m}$	$q_i^{(\xi)}$
...	...	...	...	...	...	...	...
x_n	$p_{n,1}$	$p_{n,2}$	...	$p_{n,j}$	...	$p_{n,m}$	$q_n^{(\xi)}$
η_{marg}	$q_1^{(\eta)}$	$q_2^{(\eta)}$	...	$q_j^{(\eta)}$	...	$q_n^{(\eta)}$	1

Table 1: *Two-dimensional finite discrete distribution*

As in the previous section, $\{x_1, x_2, ..., x_n\}$ and $\{y_1, y_2, ..., y_m\}$ are the values of ξ and η. The *joint distribution* of ξ and η 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, ξ and η are *not* independent. Independence requires (1.10) for *each* i *and* j (each row and each column).

Considering only the first and last column/row, we can find the *distributions* of the (one variable) r.v. ξ/η 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 $cov(\xi, \eta) = M(\xi \cdot \eta) - M(\xi) \cdot M(\eta)$ and $R(\xi, \eta) = \frac{cov(\xi,\eta)}{D(\xi) \cdot D(\eta)}$.

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*[3]*. Calculate $cov(X, Y)$, $R(X, Y)$ and estimate the measure of dependence of X and Y.*

$X \setminus Y$	1	2	3	4
10	*2*	*6*	*6*	*4*
20	*41*	*53*	*72*	*33*
30	*12*	*10*	*11*	*18*

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+...+12+10+11+18=268, so the common- and the marginal distributions are the following:*

[3] ξ and η were replaced to X and Y for technical reasons only.

$X \setminus Y$	1	2	3	4	$q_j^{(X)}$
10	*2/268*	*6/268*	*6/268*	*4/268*	*18/268*
20	*41/268*	*53/268*	*72/268*	*33/268*	*199/268*
30	*12/268*	*10/268*	*11/268*	*18/268*	*51/268*
$q_i^{(Y)}$	*55/268*	*69/268*	*89/268*	*55/268*	*268/268*

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

Means (expexted values):

*M(X*Y) = 10*1*(2/268) +10*2*(6/268) +10*3*(6/268) +10*4*(4/268) +*
*+20*1*(41/268) +20*2*(53/268) +20*3*(72/268) +20*4*(33/268) +*
*+30*1*(12/268) +30*2*(10/268) +30*3*(11/268) +30*4*(18/268) =*
*= 14490/268 ≈**54.0672,***

M(Y) = 1(55/268)+2*(69/268)+3*(89/268)+4*(55/268) = 680/268 ≈**2.5373,***

M(X) = 10(18/268)+20*(199/268)+30*(51/268) = 5690/268 ≈**21.2313,***

cov(X, Y)** = M(XY)-M(X)*M(Y) =* $14120/268^2$ *≈**0.1966.

*Since **cov(X, Y)>0** , X and Y strenghten each other, the move "in the same" direction.*

Dispersions and R(X, Y):

$M(Y^2) = (1^2)*(55/268)+(2^2)*(69/268)+(3^2)*(89/268)+(4^2)*(55/268) = 2012/268 \approx \mathbf{7.5075},$

$M(X^2) = (10^2)*(18/268)+(20^2)*(199/268)+(30^2)*(51/268) \approx \mathbf{475.0000},$

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

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

$$\mathbf{R}(\mathbf{X}, \mathbf{Y}) = \frac{cov\,(X, Y)}{D(Y)D(X)} = \frac{0.1966}{1.0342 * 4.9224} \approx 0.0386 \ .$$

Since R(X, Y) is small (≈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*** $\overrightarrow{\zeta}$ *is the common/joint density function of* (ξ, η) *, i.e. the function* $h : \mathbb{R}^2 \to \mathbb{R}^{+,0}$ *such that for any* $a, b, c, d \in \mathbb{R} \cup \{-\infty, +\infty\}$ *,* $a \le b$ *and* $c \le d$ *we have*

$$\boxed{P(a \le \xi \le b\ ,\ c \le \eta \le d) = \int_a^b \int_c^d h(x, y)\ dy\ dx\ .} \tag{1.11}$$

□

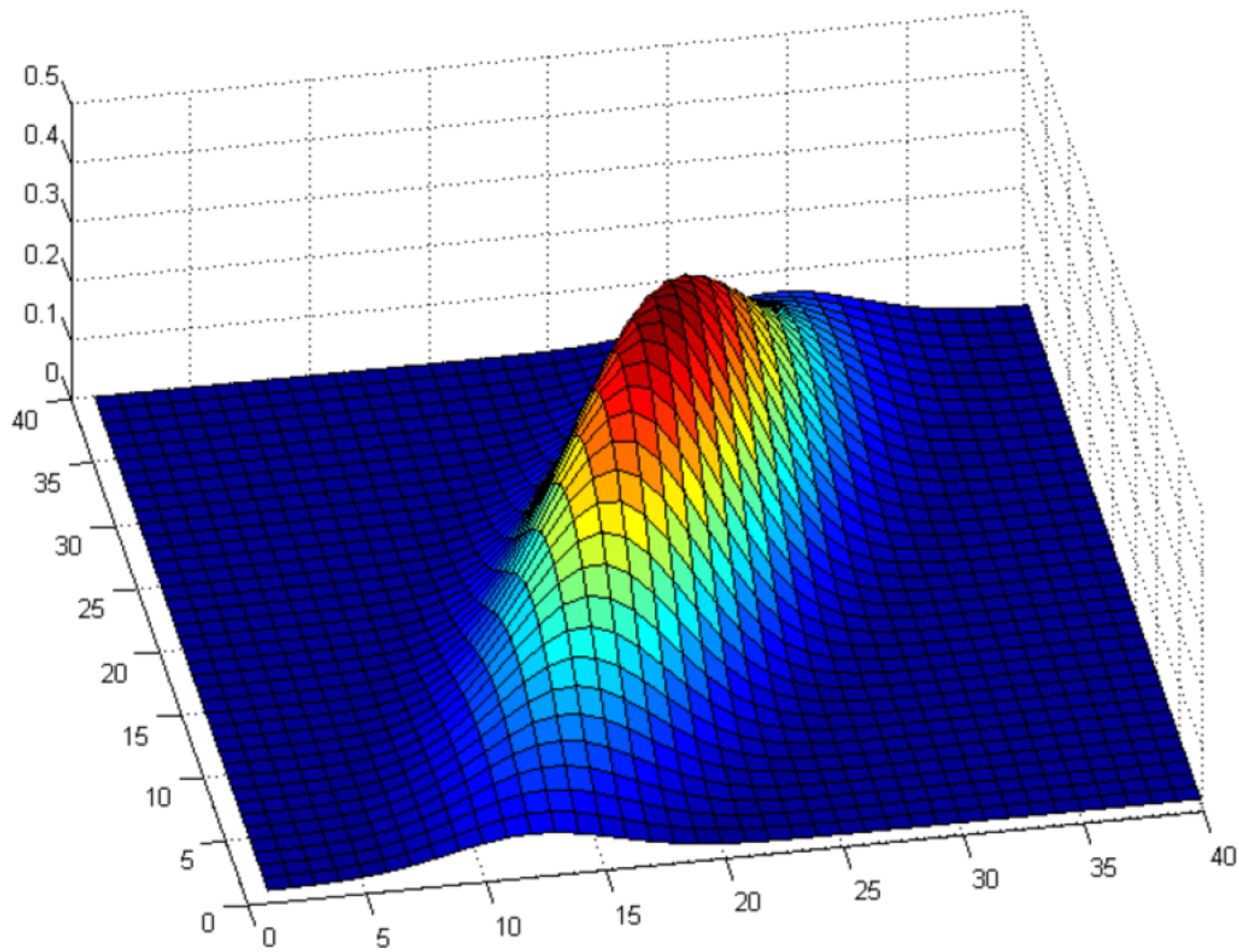

Figure 1: *A typical 2-dimensional continuous density function*

Remark I.27 Any *function* $h : \mathbb{R}^2 \to \mathbb{R}$ *is suitable if* $0 \le h(x,y)$ *and*

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)\ dydx = 1\ . \tag{1.12}$$

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 ξ and η . Further (by (1.11))

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

Theorem I.28 *The continuous r.v.* ξ *and* η *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 $a, b, c, d \in \mathbb{R} \cup \{-\infty, +\infty\}$, $a \le b$ *and* $c \le d$ *we have*

$$P(a \le \xi \le b\ ,\ c \le \eta \le d) = P(a \le \xi \le b) \cdot P(c \le \eta \le d) \tag{1.16}$$

i.e.

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

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

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

$M(\xi) = \int_{-\infty}^{\infty} x \cdot f_\xi(x)\ dx$ *and* $M(\eta) = \int_{-\infty}^{\infty} y \cdot f_\eta(y)\ dy$. $\square$

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

$$P(\xi = x \mid \eta = y) = \frac{P(\xi = x \;\&\; \eta = y)}{P(\eta = y)}, \tag{1.18}$$

$$P(\eta = y \mid \xi = x) = \frac{P(\xi = x \;\&\; \eta = y)}{P(\xi = x)} \tag{1.19}$$

and

$$P(\xi < x \mid \eta < y) = \frac{P(\xi < x \cap \eta < y)}{P(\eta < y)}. \tag{1.20}$$

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

$$P(\xi = x_i \mid \eta = y_j) = \frac{p_{i,j}}{q_j^{(\eta)}}, \quad P(\eta = y_j \mid \xi = x_i) = \frac{p_{i,j}}{q_i^{(\xi)}} \tag{1.21}$$

$$P(\xi = x_i \mid \eta \le y_j) = \frac{\sum\limits_{\ell=1}^{j} p_{i,\ell}}{\sum\limits_{\ell=1}^{j} q_\ell^{(\eta)}} \quad \textit{and} \quad P(\xi \le x_i \mid \eta \le y_j) = \frac{\sum\limits_{s=1}^{i}\sum\limits_{\ell=1}^{j} p_{s,\ell}}{\sum\limits_{\ell=1}^{j} q_\ell^{(\eta)}}, \tag{1.22}$$

for continuous r.v.

$$P(\xi < b \mid \eta < d) = \frac{\int\limits_{-\infty}^{b}\int\limits_{-\infty}^{d} h(x,y)\; dy\; dx}{\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{d} h(x,y)\; dy\; dx}. \tag{1.23}$$

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

$$F_\xi(x|y) = P(\xi < x \mid \eta = y) \quad \textit{and} \quad F_\eta(y|x) = P(\eta < y \mid \xi = x). \tag{1.24}$$

For continuous r.v. the ***conditional density functions*** *are*

$$f_\xi(x|y) = \frac{h(x,y)}{f_\eta(y)} \quad \textit{and} \quad f_\eta(y|x) = \frac{h(x,y)}{f_\xi(x)} \tag{1.25}$$

for the conditions "$\eta = y$" and "$\xi = x$", respectively. □

Theorem I.31 *For continuous r.v.*

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

further

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

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

$$M(\xi \mid \eta = y_j) = \sum_{i=1}^{\infty} x_i \cdot P(\xi = x_i \mid \eta = y_j) = \frac{1}{q_j^{(\eta)}} \sum_{i=1}^{\infty} x_i \cdot p_{i,j} \tag{1.28}$$

and

$$M(\eta \mid \xi = x_i) = \sum_{j=1}^{\infty} y_j \cdot P(\eta = y_j \mid \xi = x_i) = \frac{1}{q_i^{(\xi)}} \sum_{j=1}^{\infty} y_j \cdot p_{i,j} \ , \tag{1.29}$$

for continuous r.v.:

$$\begin{aligned} M(\xi \mid \eta = y) &= \int_{-\infty}^{+\infty} x \cdot f(x|y)\, dx \ , \\ M(\eta \mid \xi = x) &= \int_{-\infty}^{+\infty} y \cdot g(y|x)\, dy \ , \end{aligned} \tag{1.30}$$

which can also be written as

$$M(\xi \mid \eta = y) = \frac{1}{f_\eta(y)} \cdot \int_{-\infty}^{+\infty} x \cdot h(x,y)\, dx \tag{1.31}$$

and

$$M(\eta \mid \xi = x) = \frac{1}{f_\xi(x)} \cdot \int_{-\infty}^{+\infty} y \cdot h(x,y)\, dy \ . \tag{1.32}$$

$\square$

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 Covarience and independence

Definition I.33 $\overrightarrow{\xi} : \Omega \to \mathbb{R}^n$ *is an* ***n- dimensional r.v.*** *or a vector-r.v.* $\square$

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

In fact, $\xi_1, ..., \xi_n$ are *any* n r.v. as you like.

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

The dimension n can also be denoted by μ and by any other letter.

Definition I.34 $M\left(\overrightarrow{\xi}\right) := (\ M(\xi_1), ..., M(\xi_n)\ ) \in \mathbb{R}^n$ *is an* n *-dimensional vector.* $\square$

Definition I.35 *For* $\overrightarrow{\xi} : \Omega \to \mathbb{R}^n$ *and* $\overrightarrow{\eta} : \Omega \to \mathbb{R}^m$ *the* ***covariance matrix*** *("kovariencia mátrix") is*

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

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

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

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

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

(iii) $\mathbf{C}$ *is positive semidefinite*[1])*,*

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

then $cov\left(\overrightarrow{\eta}, \overrightarrow{\eta}\right) = \mathbf{A} \cdot cov\left(\overrightarrow{\xi}, \overrightarrow{\xi}\right) \cdot \mathbf{A}^T$. □

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*

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

or, in modern notation

---

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

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.*□

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

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* ξ *and* η *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-dimemsional

Definition I.39 *For any* k *-dimensional r.v.* $\overrightarrow{\xi} = (\xi_1, ..., \xi_k)$ *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* $\mathbf{A} \in \mathbb{R}^{n\times k}$ *and* $\mathbf{m} \in \mathbb{R}^n$ *the following* n *-dimensional r.v.* $\overrightarrow{\eta} := \mathbf{A}\cdot\overrightarrow{\xi} + \mathbf{m}$ *is called* 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*[2]*, 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* $\overrightarrow{\xi} = (\xi_1, ..., \xi_n)$ *is an* n *-dimensional **normal (Gaussian)** r.v. if the joint density function is*

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

□

[2] **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 $\underline{x}^T A\underline{x} > 0$ for each $\underline{x} \in \mathbb{R}^n$ where $\underline{x}^T A\underline{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}x_ix_j$. □

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

$$\xi := \text{the number of occurences of } A \ .$$

Then we have, taking $q = 1 - p$

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

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

$$\xi_2 := \text{the number of occurences of } \overline{A}$$

and have

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

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 $(2 \le n)$

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

$\xi_i := X_i :=$ *number of* A_i *occuring for* $i = 1, ..., n$.

Then $\overrightarrow{\xi} = (\xi_1, ..., \xi_n)$ *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 $k_1, ..., k_n \in \mathbb{N}$*

$$P(\xi_1 = k_1, ..., \xi_n = k_n) = \begin{cases} \dfrac{m!}{k_1! \cdot ... \cdot k_n!} \cdot p_1^{k_1} \cdot ... \cdot p_n^{k_n} & \text{if } k_1 + ... + k_n = m \\ \\ 0 & \text{otherwise} \end{cases}$$

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

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

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

$$\binom{m}{k_1, ..., k_n} = \frac{m!}{k_1! \cdot ... \cdot k_n!} \ . \tag{2.6}$$

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

$$\xi := \text{the number of occurences of } A = A_1 \ .$$

Then we have

$$P(\xi = k) = \frac{\binom{M_1}{k}\binom{N-M_1}{m-k}}{\binom{N}{m}} \quad \text{for } 0 \le k \le m \ . \tag{2.7}$$

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 $(2 \le n)$

Definition I.45 *Let* $A_1 \cup^* A_2 \cup^* ... \cup^* A_n = H$, $|A_i| = M_i$, $\sum_{i=1}^{n} M_i = N = |H|$ *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*

$\xi_i := X_i :=$ *number of* A_i *occuring,* without repetitions/putting back *for* $i = 1, ..., n$. *Then* $\overrightarrow{\xi} = (\xi_1, ..., \xi_n)$ *is called* n ***-dimensional binomial / multinomial / Bernoulli r.v.*** □

Theorem I.46 *The distribution is: for any nonnegative integers* $k_1, ..., k_n \in \mathbb{N}$

$$P(\xi_1 = k_1, ..., \xi_n = k_n) = \begin{cases} \dfrac{\binom{M_1}{k_1} \cdot ... \cdot \binom{M_n}{k_n}}{\binom{N}{m}} & \text{if } k_1 + ... + k_n = m \\ 0 & \text{otherwise} \end{cases} \qquad \square$$

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* $x_1, ..., x_n$.
*ii) A statistical **sample** ("minta") is n many r.v.* $(\xi_1, ..., \xi_n)$ *OR* $(X_1, ..., X_n)$.
*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)*

$$\hat{\xi} = \bar{\xi} := \frac{\xi_1 + ... + \xi_n}{n} \tag{3.1}$$

*is the **empirical** (greek)/ **practical** ("tapasztalati") **average/ mean/ expected value.***

ii) $\widehat{(\xi^2)} = \overline{(\xi^2)} := \dfrac{\xi_1^2 + ... + \xi_n^2}{n}$ *is the **empirical squared mean.***

*iii) The **empirical variance** and **dispersion** are*

$$\sigma^2 := \frac{1}{n}\sum_{i=1}^{n}\left(\xi_i - \overline{\xi}\right)^2 = \frac{\left(\xi_1 - \overline{\xi}\right)^2 + ... + \left(\xi_n - \overline{\xi}\right)^2}{n} \tag{3.2}$$

and

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i - \overline{\xi}\right)^2} \quad , \tag{3.3}$$

*iv) The **corrected** ("korrigált, javított") **empirical variance** and **dispersion** are*

$$(\sigma^*)^2 := \frac{n}{n-1}\cdot\sigma^2 = \frac{\left(\xi_1 - \overline{\xi}\right)^2 + ... + \left(\xi_n - \overline{\xi}\right)^2}{n-1} \tag{3.4}$$

and

$$\sigma^* = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\xi_i - \overline{\xi}\right)^2} = \sqrt{\frac{n}{n-1}} \cdot \sigma \quad . \tag{3.5}$$

□

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

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

Theorem II.4

$$\sigma^2 = \overline{\left(\xi^2\right)} - \left(\overline{\xi}\right)^2 = \frac{\xi_1^2 + ... + \xi_n^2}{n} - \left(\overline{\xi}\right)^2 \ , \tag{3.6}$$

$$\sigma = \sqrt{\overline{\left(\xi^2\right)} - \left(\overline{\xi}\right)^2} \quad \textit{and so} \quad \sigma^* = \sqrt{\frac{n}{n-1}\left(\overline{\left(\xi^2\right)} - \left(\overline{\xi}\right)^2\right)} \ . \tag{3.7}$$

□

Example II.5 *Let* $\{\xi_1, ..., \xi_n\} =$
$= \{20.0,\ 20.2,\ 20.4,\ 20.7,\ 20.7,\ 21.0,\ 21.1,\ 21.3,\ 21.4,\ 21.4,\ 21.4,\ 21.5\}$,
so $n = 12$ *and* $s = n-1$.
The empirical mean is:

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

the empirical quadratic mean:

$$\overline{\left(\xi^2\right)} = \frac{20.0^2 + 20.2^2 + 20.4^2 + 20.7^2 + 20.7^2 + 21.0^2}{12} + \frac{21.1^2 + 21.3^2 + 21.4^2 + 21.4^2 + 21.4^2 + 21.5^2}{12} \approx 438.100\,833 \ ,$$

the empirical variance and dispersion:

$$\sigma^2 = \overline{\left(\xi^2\right)} - \left(\overline{\xi}\right)^2 \approx 438.101 - 20.925^2 \approx 0.2454 \, ,$$

$$\sigma = \sqrt{\overline{\left(\xi^2\right)} - \left(\overline{\xi}\right)^2} \approx \sqrt{0.245\,4} \approx 0.4954 \ ,$$

the corrected empirical variance and dispersion:

$$(\sigma^*)^2 = \frac{n}{n-1}\cdot\left(\overline{(\xi^2)} - \left(\overline{\xi}\right)^2\right) \approx \frac{12}{11}\cdot(438.101 - 20.925^2) \approx 0.2677\ ,$$

$$\sigma^* = \sqrt{\frac{n}{n-1}\cdot\left(\overline{(\xi^2)} - \left(\overline{\xi}\right)^2\right)} \approx \sqrt{0.2677} = 0.5174\ .$$

Definition II.6 *Any function $g(\xi_1, ..., \xi_n)$ of the sample $(\xi_1, ..., \xi_n)$ 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 $\Xi = \{\xi_i : i = 1, ..., n\}$ and denote $\bar{\xi}$ its mean (a fixed real number). Now, prepare the modified dataset $\Xi' := \left\{\xi_i - \bar{\xi} : i = 1, ..., n\right\}$, i.e. substract $\bar{\xi}$ from each data. Then clearly $\overline{\xi'} = 0$. Most of the further calculations allow this transformation.*

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

$$\Xi_{st} := \left\{\frac{\xi_i - \bar{\xi}}{\sigma_\xi} : i = 1, ..., n\right\} \tag{3.8}$$

resulting similarly $\overline{\xi_{st}} = 0$ 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*

$$P(a < \gamma < b) \quad \geq \quad 1 - \varepsilon \tag{4.1}$$

where γ *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* ε *then* $[a, b]$ *increases.*

4.1 Interval for the probability

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

$$P(a < p < b) \quad \geq \quad 1 - \varepsilon \tag{4.2}$$

Theorem II.12 *If n independent experiments resulted k outcomes of A and n is large enough*[1]*, then*

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

where

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

and

$$\Phi\left(u_\varepsilon\right) = 1 - \frac{\varepsilon}{2} \tag{4.5}$$

(use table Φ). □

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

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

Further:

$$\eta = \frac{1.96}{\sqrt{30}} \cdot \sqrt{\frac{10}{30} \cdot \left(1 - \frac{10}{30}\right)} \approx 0.168\,690 \ ,$$

$$a \approx \frac{10}{30} - 0.168\,690 \approx 0.164\,643 \ ,$$

$$b \approx \frac{10}{30} + 0.168\,690 = 0.502\,023 \ ,$$

so, by 95% *we have*

$$P\left(0.164 < p < 0.502\right) \quad \geq \quad 0.95 \ . \qquad \square \tag{4.6}$$

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

ii) Theorem II.12 is based on Moivre-Laplace's theorem (see [SzI1]). □

---

[1] n must be above 30, but $n > 200$ is preferable.

4.2 Interval for the mean when σ is known

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

$$P(a \le m \le b) \quad \ge \quad 1 - \varepsilon \quad . \tag{4.7}$$

Theorem II.17

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

where u_ε *satisfies (4.5).* □

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

$= \{20.0,\ 20.2,\ 20.4,\ 20.7,\ 20.7,\ 21.0,\ 21.1,\ 21.3,\ 21.4,\ 21.4,\ 21.4,\ 21.5\}$.

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*

$\overline{\xi} =$

$$\frac{20.0 + 20.2 + 20.4 + 20.7 + 20.7 + 21.0 + 21.1 + 21.3 + 21.4 + 21.4 + 21.4 + 21.5}{12}$$

$= 20.925$,

$\frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{12}} \approx 0.866\,025$,

$a \approx 20.925 - 1.96 \cdot 0.866\,025 \approx 19.227\,591$,

$b \approx 20.925 + 1.96 \cdot 0.866\,025 = 22.622\,409$.

So

$$P(19.228 < m < 22.622) \quad > \quad 1 - \varepsilon = 0.95 \ . \tag{4.9}$$

4.3 Interval for the mean when σ is unknown

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

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

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

i.e.

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

Example II.22 *Let the sample be:*
$X_1, ..., X_n =$ 20.0, 20.2, 20.4, 20.7, 20.7, 21.0, 21.1, 21.3, 21.4, 21.4, 21.4, 21.5
and let $1 - \varepsilon = 95\%$.

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

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

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

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

and finally

$$P(20.596 < M(\xi) < 21.254) \quad > \quad 1 - \varepsilon = 0.95 \ . \tag{4.12}$$

4.4 Interval for the dispersion

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

Theorem II.25 *For the variance we have*

$$\left[a^2, b^2\right] = \left[\frac{n \cdot (\sigma^*)^2}{\chi^2_{\varepsilon/2}} \quad , \quad \frac{n \cdot (\sigma^*)^2}{\chi^2_{1-\varepsilon/2}}\right] \tag{4.13}$$

i.e.

$$P\left(a^2 < D^2(\xi) < b^2\right) \quad > \quad 1 - \varepsilon \tag{4.14}$$

and for the dispersion

$$[a, b] = \left[\frac{\sqrt{n} \cdot \sigma^*}{\chi_{\varepsilon/2}} \; , \; \frac{\sqrt{n} \cdot \sigma^*}{\chi_{1-\varepsilon/2}}\right] \tag{4.15}$$

i.e.

$$P\left(a < D(\xi) < b\right) \quad > \quad 1 - \varepsilon \tag{4.16}$$

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

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

Solution II.27 $n = 12$ *, the degree of freedom is* $s = n - 1 = 11$ *,* $\varepsilon = 5\% = 0.05$ *. Using table χ^2 we find ($\varepsilon/2 = 0.025$, $1 - \varepsilon/2 = 0.975$, $s = 11$):*

$$\chi^2_{\varepsilon/2} = \chi^2_{0.025} \approx 21.920 \quad \text{and} \quad \chi^2_{1-\varepsilon/2} = \chi^2_{0.975} \approx 3.816 \ , \tag{4.17}$$

so

$$\chi_{0.025} \approx \sqrt{21.920} \approx 4.6819 \quad \text{és} \quad \chi_{0.975} \approx \sqrt{3.816} \approx 1.953\,5 \ . \tag{4.18}$$

We calculated $\overline{\xi}$, $\overline{(\xi^2)}$ and σ^* in Example II.5, so

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

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

so

$$P\left(0.1466 < D^2(\xi) < 0.8418\right) \quad > \quad 1 - \varepsilon = 0.95 \tag{4.19}$$

and

$$P\left(0.3829 < D(\xi) < 0.9175\right) \quad > \quad 1 - \varepsilon = 0.95 \ . \tag{4.20}$$

Chapter 5

Point estimations and hypothesis testing

5.1 General notions

Definition II.28 *i) Any statistical function $g(\xi_1, ..., \xi_n)$ is an* ***estimation*** *("becslés") of the* ***parameter*** *a (of a r.v. ξ), and it is often denoted by $\hat{a}(\xi_1, ..., \xi_n)$, or shortly by $\hat{a}$.*

ii) The estimation $\hat{a} = g(\xi_1, ..., \xi_n)$ is ***unbiased*** *(un-distorted, not-deformed, "torzítatlan") if its mean equals to $a = a(\xi)$, i.e.*

$$M(\hat{a}) = a \ . \tag{5.1}$$

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

$(\forall \varepsilon, \delta > 0)\ (\exists n_0)\ (\forall n > n_0)$

$$P\left(\ |\hat{a}(\xi_1, ..., \xi_n) - a| \ \geq \varepsilon\right) < \delta \ . \tag{5.2}$$

iv) The estimation $\hat{a}_1$ is ***more efficient/ effective*** *("hatásos") than $\hat{a}_2$ for the* same *parameter a if* $\quad D(\hat{a}_1) < D(\hat{a}_2)$. □

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) $\hat{p} := \dfrac{k}{n}$ (relative frequency) *is an unbiased estimation of the probability p ,* *ii)* $\hat{\xi} = \bar{\xi} := \dfrac{\xi_1 + ... + \xi_n}{n}$ *(*average*) is an unbiased estimation of the mean* $M(\xi)$ *.*

iii) $(\sigma_n^*)^2 := \dfrac{\sum_{i=1}^{n} (\bar{\xi} - \xi_i)^2}{n-1}$ *(*corrected empirical variance*)*

is an unbiased estimation of the variance $D^2(\xi)$ *.* □

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* ξ *(and* η*), a* ***hypothesis*** *("hipotézis, feltételezés"). The hypothesis we inverstigate 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.* □

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:*

	H_0 is **true**	H_0 is **false**
H_0 is **accepted**	*OK*	Type **II** error
H_0 is **rejected**	Type **I** error	*OK*

□

Remark II.34 *The probability of type* ***I*** *error is usually denoted by* ε *.*
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 couse, 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* ε *in somehow - read further.*

Definition II.36 *The* ***significance level of a test*** *("megbízhatósági szint") is* $1 - \varepsilon$ *(where* ε *is the probability of type* ***I*** *error).* □

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 ε (probability of type I error).

*iii) Decreasing ε 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 ξ , 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 hypoteses (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 : " ξ is a normal distibution". This hypothesis needs a nonparamteric 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 σ is known

("Egymintás u-próba")

ξ is normal, σ is known, m_0 and ε are given ($m_0 \in \mathbb{R}$), $(\xi_1, ..., \xi_n)$ is the sample.

Algorithm II.41 *For the* two-sided test $H_0 : \; M(\xi) = m_0$
i) calculate $u_{sz} := \sqrt{n} \cdot \dfrac{\bar{\xi} - m_0}{\sigma}$,
ii) find $u_\varepsilon \in \mathbb{R}^+$ *such that* $\Phi(u_\varepsilon) = 1 - \dfrac{\varepsilon}{2}$,
iii) accept H_0 *in the case* $|u_{sz}| \le u_\varepsilon$ *with significance* $1 - \varepsilon$
or reject H_0 *in the case* $|u_{sz}| > u_\varepsilon$ *with significance* $1 - \varepsilon$. □

Algorithm II.42 *For* one-sided tests: $H_0 : \; M(\xi) \ge / \le m_0$
i) calculate $u_{sz} := \sqrt{n} \cdot \dfrac{\bar{\xi} - m_0}{\sigma}$,
ii) find $u_\varepsilon \in \mathbb{R}^+$ *such that* $\Phi(u_\varepsilon) = 1 - \varepsilon$,
iii) accept $H_0 : M(\xi) \le m_0$ *in the case* $u_{sz} \le u_\varepsilon$ *with significance* $1 - \varepsilon$
or reject H_0 *in the case* $u_{sz} > u_\varepsilon$ *with significance* $1 - \varepsilon$.
iv) accept $H_0 : M(\xi) \ge m_0$ *in the case* $-u_\varepsilon \le u_{sz}$ *with significance* $1 - \varepsilon$
or reject H_0 *in the case* $-u_\varepsilon > u_{sz}$ *with significance* $1 - \varepsilon$. □

Remark II.43 *If the dispersion* σ *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* σ^* *instead of* σ .

Example II.44 *Let* $m_0 = 1200$, $\sigma = 3$ *and* $\overrightarrow{\xi} = \{1193, 1198, 1203, 1191, 1195, 1196, 1199, 1191, 1201, 1196, 1193, 1198, 1204, 1196, 1198, 1200\}$.
Decide the hypothesis $H_0 : M(\xi) = m_0$ *with siginificance level* 99.9% .

Solution II.45 Two sided test. *So* $\varepsilon = 0.001$, $\Phi(u_\varepsilon) = 1 - \dfrac{\varepsilon}{2} = 0.9995$ *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 $u_{sz} = \sqrt{16} \cdot \dfrac{1197 - 1200}{3} = -4$.
Since $|u_{sz}| = 4 > u_\varepsilon = 3.29$ *we must reject* H_0 *with significance* 99.9% .

Example II.46 *Let* $m_0 = 70$, σ *is unknown,* $n = 36$, $\bar{\xi} = 68.5$ *and* $\sigma^* = 6$. *Decide the hypothesis* $H_0 : M(\xi) \ge m_0$ *with siginificance level* 95% .

Solution II.47 One sided test. *Though the dispersion* (σ) *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 $-u_\varepsilon = -1.65 \le u_{sz} = -1.5$ *we have to accept the hypothesis* $H_0 : M(\xi) \ge m_0$ *with siginificance level* 95% .

5.2.2 t- test for the mean of one sample when σ is unknown

("Egymintás t-próba")

ξ is normal, σ is *unknown*, m_0 and ε are given ($m_0 \in \mathbb{R}$), $(\xi_1, ..., \xi_n)$ is the sample.

Algorithm II.48 *For the* two-sided test $H_0 : \; M(\xi) = m_0$ *:*
i) calculate $t_{sz} := \sqrt{n} \cdot \dfrac{\bar{\xi} - m_0}{\sigma^*}$
ii) find $t_\varepsilon \in \mathbb{R}^+$ *in the table of the Student-distribution for* $\beta = p = 1 - \dfrac{\varepsilon}{2}$ *and degree of freedom* $s = n - 1$ *,*
iii) accept H_0 *in the case* $|t_{sz}| \le t_\varepsilon$ *with significance* $1 - \varepsilon$ *,*
or reject H_0 *in the case* $|t_{sz}| > t_\varepsilon$ *with significance* $1 - \varepsilon$ *.* □

Algorithm II.49 *For* one-sided tests $H_0 : \; M(\xi) \ge / \le m_0$
i) calculate $t_{sz} := \sqrt{n} \cdot \dfrac{\bar{\xi} - m_0}{\sigma^*}$ *,*
ii) find $t_\varepsilon \in \mathbb{R}^+$ *in the table of the Student-distribution for* $\beta = p = 1 - \varepsilon$ *and degree of freedom* $s = n - 1$ *,*

iii) accept $H_0 : M(\xi) \le m_0$ *in the case* $t_{sz} \le t_\varepsilon$ *with significance* $1 - \varepsilon$
or reject H_0 *in the case* $t_{sz} > t_\varepsilon$ *with significance* $1 - \varepsilon$ *.*

iv) accept $H_0 : M(\xi) \ge m_0$ *in the case* $-t_\varepsilon \le t_{sz}$ *with significance* $1 - \varepsilon$
or reject H_0 *in the case* $-t_\varepsilon > t_{sz}$ *with significance* $1 - \varepsilon$ *.* □

Remark II.50 *For large samples (*$n > 30$*) the* u *-test can also be applied but we use* σ^* *instead of* σ *.*

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

Solution II.52 Two sided test. $n = 5,\ s = 4$,

$$\bar{\xi} = \frac{1.51 + 1.49 + 1.54 + 1.52 + 1.54}{5} = 1.52,$$

$$\overline{\xi^2} = \frac{1.51^2 + 1.49^2 + 1.54^2 + 1.52^2 + 1.54^2}{5} = 2.31076 \ ,$$

$$\sigma^* = \sqrt{\frac{5}{5-1} \cdot (2.31076 - 1.52^2)} = 0.02121\,,$$

$$t_{sz} = \sqrt{n} \cdot \frac{\bar{\xi} - m_0}{\sigma^*} = \sqrt{5} \cdot \frac{1.52 - 1.50}{0.02121} = 2.1085 \ ,$$

$$\varepsilon = 0.05 \ ,\ \beta = p = 1 - \frac{\varepsilon}{2} = 0.975 \ ,\ t_\varepsilon = 2.78 \ .$$

Since $|t_{sz}| = 2.1085 < t_\varepsilon = 2.78$ *we must accept* H_0 *with significance* 95% *.*

Example II.53 *Let the sample be* $\overrightarrow{\xi} = \{3.1, 2.8, 1.5, 1.7, 2.4, 2.0, 3.3, 1.6\}$. *Decide the hypothesis* $H_0 : M(\xi) \geq 3.1$ *with siginificance level* 98% .

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

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

$\overline{\xi^2} = \dfrac{3.1^2 + 2.8^2 + 1.5^2 + 1.7^2 + 2.4^2 + 2.0^2 + 3.3^2 + 1.6^2}{8} = 5.725$,

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

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

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

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")

ξ is normal, σ is unknown, ε and σ_0 are given ($\sigma_0 \in \mathbb{R}^+$), $(\xi_1, ..., \xi_n)$ is the sample.

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

$$k_{sz} := \frac{(n-1) \cdot (\sigma^*)^2}{\sigma_0^2} , \tag{5.3}$$

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* χ^2 *-distribution for* $\beta = \dfrac{\varepsilon}{2}$ *and* $\beta = 1 - \dfrac{\varepsilon}{2}$,

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* $1 - \varepsilon$. □

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* χ^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* χ^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:* $\sigma_0 = 1.1$ *,* $\beta = \dfrac{\varepsilon}{2} = 0.05$ *,* $k_\varepsilon = 16.919$ *,* $1 - \dfrac{\varepsilon}{2} = 0.975$ *,* $k_{1-\varepsilon} = 2.7$ *,* $k_{sz} = \dfrac{9 \cdot 1.3^2}{1.1^2} \approx 12.57$ *,* $k_{1-\varepsilon} < k_{sz} < k_\varepsilon$ *, 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:* $\sigma_0 = 1.1$ *,* $\beta = \varepsilon = 0.1$ *,* $k_\varepsilon = 14.684$ *,*
$k_{sz} = \dfrac{9 \cdot 1.3^2}{1.1^2} \approx 12.57 < k_\varepsilon$ *so* H_0 *is accepted.*

5.2.4 u- test for the means of two samples

("Kétmintás u-próba")

ξ and η are normal, ε and m_0 are given ($m_0 \in \mathbb{R}$), $(\xi_1, ..., \xi_n)$ and $(\eta_1, ..., \eta_m)$ 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 ∇ can be any of $\geq$, $\leq$ or $=$.

Algorithm II.62 i_1*) When* σ_ξ *and* σ_η *are* known *(for* any-sided test*) calculate*

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

i_2) *when* σ_ξ *and* σ_η *are* not *known (for* any-sided test*), calculate*

$$u_{sz} := \frac{\bar{\xi} - \bar{\eta} - m_0}{\sqrt{\frac{\sigma_\xi^{*2} \cdot (n-1) + \sigma_\eta^{*2} \cdot (m-1)}{n+m-2}} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} \tag{5.5}$$

ii_1) *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 $H_0 : M(\xi) - M(\eta) \geq / \leq m_0$ *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* σ_ξ , σ_η *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 $u_{sz} = \dfrac{40.1 - 38.3 - 0}{\sqrt{\frac{5.48^2}{10} + \frac{6.32^2}{8}}} \approx 0.6366 < u_\varepsilon$,

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* σ_ξ , σ_η *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 $u_{sz} = \dfrac{57 - 60 - 0}{\sqrt{\frac{12^2}{225} + \frac{15^2}{250}}} \approx -2.417 < -u_\varepsilon$,

so we reject H_0 *with significance level* 98% .

Example II.67 *Let* $n = 40$, $\bar{\xi} = 102$, $\sigma_\xi = 5.648$, $m = 35$, $\bar{\eta} = 95$, $\sigma_\xi = \sigma_\eta = 5.648$. *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 $u_{sz} = \dfrac{102 - 95 - 4}{\sqrt{\frac{5.648^2}{40} + \frac{5.648^2}{35}}} \approx 2.2949 < u_\varepsilon$,

so we accept H_0 *with significance level* 98% .

5.2.5 t- test for the means of two samples when $\sigma_1 = \sigma_2$

("Kétmintás t-próba")

ξ and η are normal, only the *equality* $\sigma_1 = \sigma_2$ is known (but we do not know either σ_1 or σ_2), ε 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

$$t_{sz} := \frac{\bar{\xi} - \bar{\eta}}{\sqrt{(n-1)\sigma_\xi^{*2} + (m-1)\sigma_\eta^{*2}}} \cdot \sqrt{\frac{nm(n+m-2)}{n+m}} \tag{5.6}$$

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* $1 - \varepsilon$. □

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

$$t_{sz} := \frac{\bar{\xi} - \bar{\eta} - m_0}{\sqrt{\frac{\sigma_\xi^{*2} \cdot (n-1) + \sigma_\eta^{*2} \cdot (m-1)}{n+m-2}} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} \tag{5.7}$$

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

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

Solution II.72 $\bar{\xi} = \dfrac{300 + 301 + 303 + 288 + 294 + 296}{6} = 297$,

$$\overline{\xi^2} = \frac{300^2 + 301^2 + 303^2 + 288^2 + 294^2 + 296^2}{6} \approx 88234.\dot{3} ,$$

$$\sigma_\xi^{*2} = \frac{6}{6-1} \cdot \left(88234.\dot{3} - 297^2\right) \approx 30.3\dot{9} ,$$

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

$$\overline{\eta^2} = \frac{305^2 + 317^2 + 308^2 + 300^2 + 314^2 + 316^2}{6} = 96138.\dot{3} ,$$

$$\sigma_\eta^{*2} = \frac{6}{6-1} \cdot \left(96138.\dot{3} - 310^2\right) \approx 45.\dot{9} ,$$

$$t_{sz} = \frac{297 - 310}{\sqrt{5 \cdot 30.40 + 5 \cdot 46}} \cdot \sqrt{\frac{36 \cdot 10}{6+6}} \approx -3.643 ,$$

$n = m = 6$, $s = 6 + 6 - 2 = 10$, $\varepsilon = 0.01$, $\beta = p = 1 - \frac{\varepsilon}{2} = 0.995$, $t_\varepsilon = 3.17$.

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")

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

Algorithm II.73 *i) if* $\sigma_\xi^{*2} > \sigma_\eta^{*2}$ *then let* $F_{sz} := \dfrac{\sigma_\xi^{*2}}{\sigma_\eta^{*2}}$, *otherwise let* $F_{sz} := \dfrac{\sigma_\eta^{*2}}{\sigma_\xi^{*2}}$
(i.e. $F_{sz} > 1$ *always holds),*

ii) find $F_\varepsilon \in \mathbb{R}^+$ *in the table of the F-distribution for the given* ε *in the row* $m-1$ *in the column* $n-1$ *,*
iii) accept H_0 *in the case* $|F_{sz}| \leq F_\varepsilon$ *with significance* $1-\varepsilon$ *.*
or reject H_0 *in the case* $|F_{sz}| > F_\varepsilon$ *with significance* $1-\varepsilon$ *.* □

Example II.74 *Let* $\overrightarrow{\xi} = \{11.9, 12.1, 12.8, 12.2, 12.5, 11.9, 12.5, 11.8, 12.4, 12.9\}$ *,*
$\overrightarrow{\eta} = \{12.1, 12.0, 12.9, 12.2, 12.7, 12.6, 12.6, 12.8, 12.0, 13.1\}$ *.*
Decide the hypothesis $H_0 : D(\xi) = D(\eta)$ *with siginificance level* 95% *.*

Solution II.75 $n = m = 10$ *,* $\overline{\xi} = 12.3$ *,* $\sigma_\xi^{*2} \approx 0.1467$ *,* $\overline{\eta} = 12.5$ *,* $\sigma_\xi^{*2} \approx 0.1578$ *,* $F_{sz} = \dfrac{0.1578}{0.1467} \approx 1.0756$ *. The* 9 *'th row and* 9 *'th column of the* F *table shows* $F_\varepsilon = 3.18$ *. Since* $|F_{sz}| \approx 1.0756 \leq F_\varepsilon = 3.18$ *we accept the hypothesis* H_0 *.*

5.3 Nonparametric tests

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

As in hypothesis tests, the significance level $1-\varepsilon$ 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 $\overrightarrow{\xi}$ 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 occurences in* A_i *(i.e. how many* ξ_j *is in* A_i*) and denote these numbers by* a_i *,*
ii) calculate

$$\chi_{sz}^2 := \sum_{i=1}^{k} \frac{(a_i - np_i)^2}{np_i} = n \cdot \sum_{i=1}^{k} \frac{\left(\frac{a_i}{n} - p_i\right)^2}{p_i} , \tag{5.8}$$

iii) find χ^2_ε in the "Chi-squared" table (the degree of freedom is $k-1$),

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

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

nu. of heads (i)	0	1	2	3	4	*total*
frequency (a_i)	5	35	67	41	12	160

.

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 $p=\frac{1}{2}$.*

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

$$p_1 = \binom{4}{1}\cdot\left(\frac{1}{2}\right)^1\cdot\left(\frac{1}{2}\right)^{4-1} = \binom{4}{1}\cdot\left(\frac{1}{2}\right)^4 = \frac{4}{16} = 0.25 \ ,$$

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

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

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

i	0	1	2	3	4
a_i	5	35	67	41	12
$n\cdot p_i$	10	40	60	40	10
$(a_i - n\cdot p_i)^2$	5^2	5^2	7^2	1^2	2^2

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

$s=5-1=4$, $\varepsilon=0.05$, $\chi^2_\varepsilon = 9.488$, *$H_0$ is accepted since $\chi^2_{sz} < \chi^2_\varepsilon$.* □

5.3.2 Homogenity

("homogenitás, azonosság")

H_0 : The complete systems of events $\{A_1, ..., A_k\}$ and $\{B_1, ..., B_k\}$ determined by ξ and η 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 ξ and η is the question.

Algorithm II.80 *One sided test.*
i) count the occurences of $\overrightarrow{\xi}$ *in* A_i *and of* $\overrightarrow{\eta}$ *in* B_i *and denote these numbers by* a_i *and* b_i $(i = 1, ..., k)$*,*

ii) calculate

$$\chi^2_{sz} := \frac{1}{mn} \sum_{i=1}^{k} \frac{(ma_i - nb_i)^2}{a_i + b_i} \quad , \tag{5.9}$$

iii) find χ^2_ε *in the "Chi-squared" table (the degree of freedom now is* $(k-1)$*,* $\beta = \varepsilon$*),*

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

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

	A_1	A_2	A_3	A_4	A_5	n
$\overrightarrow{\xi}$	*51*	*64*	*26*	*18*	*21*	*180*

	B_1	B_2	B_3	B_4	B_5	m
$\overrightarrow{\eta}$	*72*	*51*	*33*	*23*	*21*	*200*

Solution II.82 $n = 51+64+26+18+21 = 180$, $m = 72+51+33+23+21 = 200$, $s = 5-1 = 4$, $\varepsilon = 0.05$, $\chi^2_{s,\varepsilon} = \chi^2_{4,0.05} = 9.488$,
further

$$\chi^2_{sz} = \frac{1}{180 \cdot 200} \left(\frac{(200 \cdot 51 - 180 \cdot 72)^2}{51 + 72} + \frac{(200 \cdot 64 - 180 \cdot 51)^2}{64 + 51} + \right.$$

$$\left. + \frac{(200 \cdot 26 - 180 \cdot 33)^2}{26 + 33} + \frac{(200 \cdot 18 - 180 \cdot 23)^2}{18 + 23} + \frac{(200 \cdot 21 - 180 \cdot 21)^2}{21 + 21} \right)$$

$\approx 5.458 < \chi^2_{s,\varepsilon}$,

so H_0 *is accepted.*

5.3.3 Independence

("függetlenség")

H_0 : The complete systems of events $\{A_1, ..., A_k\}$ and $\{B_1, ..., B_\ell\}$ determined by ξ and η are *independent.*

In detail: The sample is $\overrightarrow{\zeta} = ((\xi_1, \eta_1), ..., (\xi_n, \eta_n))$, i.e. n many double measurements are, and the *dependence* between ξ and η is the question.

Algorithm II.83 *i) make the* table *of the occurences in* A_i *vs.* B_j *and denote these by* c_{ij} *,*
ii) calculate the marginal distributions $(a_1, ..., a_k)$ *and* $(b_1, ..., b_\ell)$*,*
iii) calculate

$$\chi^2_{sz} := \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{(nc_{ij} - a_i b_j)^2}{a_i b_j} , \tag{5.10}$$

(use the vertices of the rectangles in the table for the enumerator),
iv) find χ^2_ε *in the "Chi-squared" table (the degree of freedom now is* $(k-1)(\ell-1)$*),*
v) accept H_0 *in the case* $|\chi^2_{sz}| \le \chi^2_\varepsilon$ *with significance* $1-\varepsilon$ *,*
or reject H_0 *in the case* $|\chi^2_{sz}| > \chi^2_\varepsilon$ *with significance* $1-\varepsilon$ *.* □

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

$\xi \setminus \eta$	*success*	*unsuccess*
man	28	12
woman	34	26

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

$\xi \setminus \eta$	*success*	*unsuccess*	b_j
man	28	12	40
woman	34	26	60
a_i	62	38	100

$$\chi^2_{sz} = \frac{1}{100} \cdot \sum_{i=1}^{2} \sum_{i=1}^{2} \frac{(n \cdot c_{i,j} - a_i \cdot b_j)^2}{a_i \cdot b_j}$$

$$= \frac{1}{100} \left(\frac{(100 \cdot 28 - 40 \cdot 62)^2}{40 \cdot 62} + \frac{(100 \cdot 12 - 40 \cdot 38)^2}{40 \cdot 38} + \frac{(100 \cdot 34 - 60 \cdot 62)^2}{60 \cdot 62} + \frac{(100 \cdot 26 - 60 \cdot 38)^2}{60 \cdot 38} \right) \approx 1.8110 ,$$

$\chi^2_{sz} < \chi^2_\varepsilon = 3.84$, *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. ξ *and* η ?"

The base hypothesis usually is: "H_0 : no correlation between ξ and η."
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*

$$t_{sz} = r \cdot \sqrt{\frac{s}{1 - r^2}} \tag{5.11}$$

where $s = n - 2$ *is the degree of freedom.*
Pick the critical value t_ε *from the Student t-table, using* s *and* ε*.*
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* ε*.*

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 : ξ 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 σ are unknown but even the normality of ξ is in question! Though we can plot the dataset Ξ 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 Ξ fits a line or not, and equivalently, whether the r.v. ξ (measured by Ξ) is normal. Moreover, from the "usual" formula $\breve{y} = a\breve{x} + b$ of this line the parameters m and σ 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

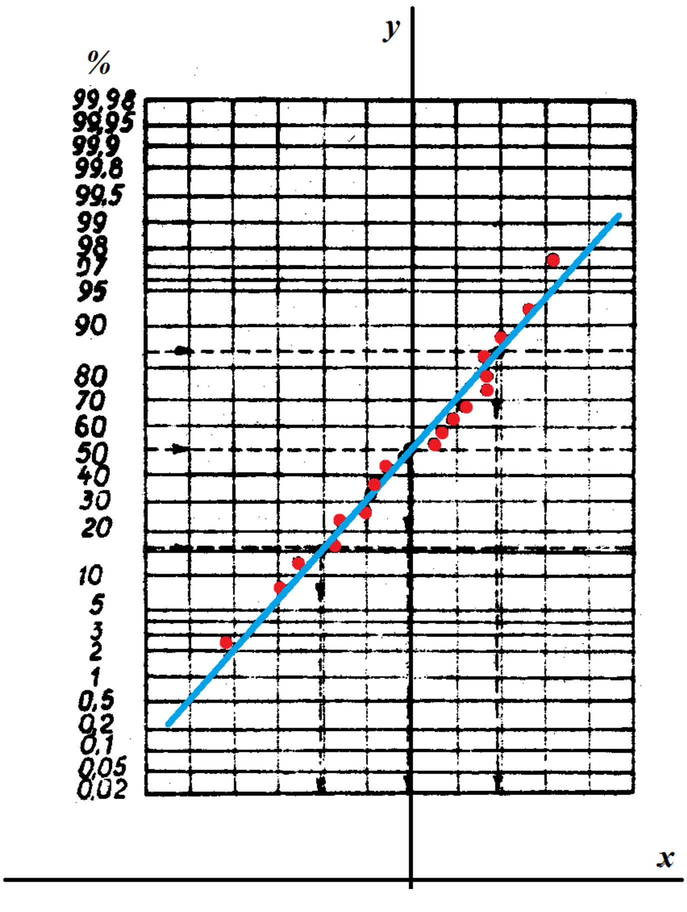

Figure 2: *Gaussian coordinate system*

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

For deciding this, divide Im (ξ) 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 occurences of different* sums of the dots *is shown in the table. Decide with significance level 95% whether this distribution is normal.*

sum	*<10*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*25<*
freq.	*15*	*20*	*30*	*40*	*55*	*70*	*90*	*95*	*99*	*98*	*96*	*85*	*75*	*58*	*35*	*33*	*19*	*22*

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 " 3σ -rule" we have*[1] $P(|\xi - M(\xi)| < 3\sigma) > 0.997$.
For simplifing 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:

nu. of interval (i)	*1*	*2*	*3*	*4*	*5*	*total (n)*
frequency (a_i)	*15*	*215*	*478*	*286*	*41*	*1035*
relative freq. $(\frac{a_i}{n})$	0.0145	0.2077	0.4618	0.2763	0.0396	*1*

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: $\sigma_{1\ dice} = \sqrt{\frac{91}{6} - \left(\frac{7}{2}\right)^2} \approx 1.708$ and $\sigma_{5\ dice} = \sqrt{5} \cdot \sigma_{1\ dice} \approx 3.819$.

$p_4 = F_{m,\sigma}(25) - F_{m,\sigma}(20) = \Phi\left(\frac{25-17.5}{2.5}\right) - \Phi\left(\frac{20-17.5}{2.5}\right) = \Phi(3) - \Phi(1)$
$= 0.9987 - 0.8413 = 0.1574$
$p_5 = F_{m,\sigma}(31) - F_{m,\sigma}(25) = \Phi\left(\frac{31-17.5}{2.5}\right) - \Phi\left(\frac{25-17.5}{2.5}\right) = \Phi(5.4) - \Phi(3)$
$= 0.9999 - 0.9987 = 0.0012$

The following table compares empirical and theoretical probabilities :

i	1	2	3	4	5	*total*
a_i/n	0.0145	0.2077	0.4618	0.2763	0.0396	1
p_i	0.0012	0.1574	0.6826	0.1574	0.0012	0.9998

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

Further: $\varepsilon = 0.05$, $s = 5 - 1 = 4$, $\chi^2_\varepsilon = 9.488$,
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.

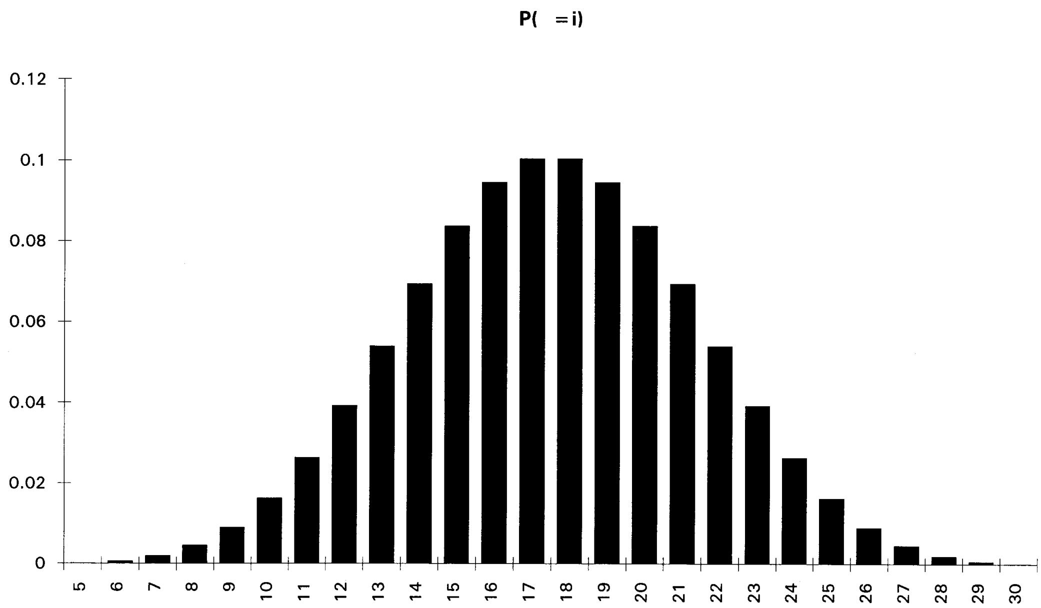

Figure 3: *Probabilities of sums of 5 dices*

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, -ütés, -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 ξ and η ("new" and "old", "input" and "output", etc.). The covariency $cov(\xi, \eta)$ 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).

See also: https://en.wikipedia.org/wiki/Francis_Galton ,
https://en.wikipedia.org/wiki/Regression_toward_the_mean ,
https://en.wikipedia.org/wiki/Bean_machine ,
https://hu.wikipedia.org/wiki/Galton-deszka ,
https://upload.wikimedia.org/wikipedia/commons/d/dc/Galton_box.webm .

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

$$\eta = m_2(\xi) \tag{6.1}$$

[1)] Sir **Francis Eugene Galton** (1822-1911) English mathematician.

where the function $m_2 : \mathbb{R} \to \mathbb{R}$ *is the* conditional mean

$$m_2(x) = M(\eta \mid \xi = x) \tag{6.2}$$

which was defined in Section 1.5 "Conditional probability".

The function m_2 *is called* ***regression function of first kind*** *(elsőfajú regressziós függvény).*

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

However, in practice we have to find much easier methods for calculating the connection between ξ *and* η *. In what follows,* ξ *and* η *are any r.v. on a (common) sample space* Ω *.* □

Theoretically we deal with random variables ξ and η , but in practice we have only a set of (measured) corresponding data ξ_i and η_i as $\{(\xi_i, \eta_i) : i = 1, ..., n\}$. As in the Introduction of Statistics we learned, ξ_i and η_i are, in fact, real numbers (in our notepad), we could write x_i and y_i instread. Since after repeated measurings they often vary, they are called r.v. in theory. This is the reason that most of the theorems have *two versions* (see e.g. Theorem II.95): one for r.v. and the other for the dataset $\{(\xi_i, \eta_i) : i = 1, ..., n\}$. If you like, you can (adviced to) think of ξ_i and η_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 ξ and η we have to write them like $g(\xi)$, $\eta \approx g(\xi)$, $(a\xi_i + b) - \eta_i$, etc. In this chapter we mix these two notations, you can also turn ξ and η 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.* η *and* ξ *, or the dataset*

$$\{(\xi_i, \eta_i) : i = 1, ..., n\} \ . \tag{6.3}$$

We are looking for the function $g : \mathbb{R} \to \mathbb{R}$ *such that the r.v.* $g(\xi)$ *is the* closest one *to* η *. The difference is measured by*

$$M\left([g(\xi) - \eta]^2\right) \tag{6.4}$$

and by

$$\sum_i [g(\xi_i) - \eta_i]^2 \tag{6.5}$$

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*

$$g(\xi) = g(\xi, a_1, ..., a_m) \tag{6.6}$$

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 η_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 $g(x) = ax + b \quad (a, b \in \mathbb{R})$. The approximation question "$\eta \approx a\xi + b$" can be raised *for any* r.v. ξ and η , 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

$$\eta \approx a_0 + a_1\xi + a_2\xi^2 + ... + a_n\xi^n \tag{6.7}$$

(**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 η , since $a_0, ...a_n$ are all not real numbers but *random variables.*

Problem II.94 *Determine $a, b \in \mathbb{R}$ such that*
i) $M\left([a\xi + b - \eta]^2\right)$ *or* ***ii)*** $\sum\limits_{i=1}^{n} (a\xi_i + b - \eta_i)^2$
is minimal:

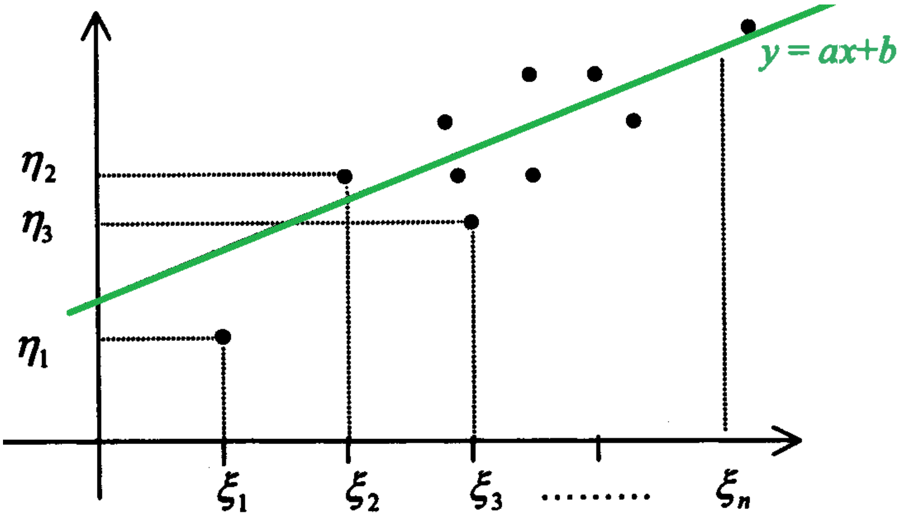

Figure 4: *Linear regression line*

Theorem II.95 ***i)*** *For* $M\left([a\xi + b - \eta]^2\right)$ *minimal we have*

$$a = \frac{M(\xi\eta) - M(\xi)\,M(\eta)}{M\left(\xi^2\right) - M^2(\xi)} \tag{6.8}$$

and

$$\boxed{b = M(\eta) - a \cdot M(\xi)}, \tag{6.9}$$

or, in another forms:

$$\boxed{a = \frac{cov(\xi,\eta)}{D^2(\xi)}} = R(\xi,\eta) \cdot \frac{D(\eta)}{D(\xi)} \tag{6.10}$$

and

$$b = M(\eta) - M(\xi) \cdot \frac{M(\xi\eta) - M(\xi)\,M(\eta)}{M\left(\xi^2\right) - M^2(\xi)}\ . \tag{6.11}$$

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

$$\boxed{a = \frac{n \cdot \sum\limits_{i=1}^{n} \xi_i\eta_i - \left(\sum\limits_{i=1}^{n} \xi_i\right)\left(\sum\limits_{i=1}^{n} \eta_i\right)}{n \cdot \sum\limits_{i=1}^{n} \xi_i^2 - \left(\sum\limits_{i=1}^{n} \xi_i\right)^2}} \tag{6.12}$$

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:

$$a = \frac{\overline{\xi\eta} - \bar{\xi}\cdot\bar{\eta}}{\sigma_\xi^2} = \frac{\sum\limits_{i=1}^{n}\left(\xi_i - \bar{\xi}\right)\left(\eta_i - \bar{\eta}\right)}{\sum\limits_{i=1}^{n}\left(\xi_i - \bar{\xi}\right)^2} \tag{6.14}$$

and

$$\boxed{b = \bar{\eta} - a\cdot\bar{\xi}} . \tag{6.15}$$

□

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* $\sum\limits_{i=1}^{n}\xi_i = 0$ *, i.e. when the dataset* $\{\xi_i : i = 1, ..., n\}$ *is* symmetric *to the origin, the above formulas have much simpler form*

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

The symmetric property can be easily achieved by using $\xi' := \xi - m_\xi$ *and* $\xi_i' := \xi_i - \bar{\xi}$ *instead of* ξ *and* ξ_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 (6.10) and (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 (6.1) and (6.17) coincide.*

Proof. of Theorem II.95:

i) We have to find the minimum value of the two-variable function $F(a, b) := M\left([a\xi + b - \eta]^2\right)$. 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 (6.18) below) has the solution shown in (6.8) and (6.9). In detail:

$$F(a,b) = M\left(a^2\xi^2 + b^2 + \eta^2 + 2ab\xi - 2a\xi\eta - 2b\eta\right) =$$
$$= a^2 M\left(\xi^2\right) + b^2 + M\left(\eta^2\right) + 2abM(\xi) - 2aM(\xi\eta) - 2bM(\eta) \ ,$$

$$\frac{\partial F}{\partial a} = 2aM\left(\xi^2\right) + 2bM(\xi) - 2M(\xi\eta) \ ,$$
$$\frac{\partial F}{\partial b} = 2b + 2aM(\xi) - 2M(\eta) \ ,$$

so the system of equalities we have to solve is:

$$\left.\begin{aligned} aM\left(\xi^2\right) + bM(\xi) &= M(\xi\eta) \\ aM(\xi) + b &= M(\eta) \end{aligned}\right\} \tag{6.18}$$

The solution is

$$a = \frac{\det\begin{bmatrix} M(\xi\eta) & M(\xi) \\ M(\eta) & 1 \end{bmatrix}}{\det\begin{bmatrix} M\left(\xi^2\right) & M(\xi) \\ M(\xi) & 1 \end{bmatrix}} = \frac{M(\xi\eta) - M(\xi)\,M(\eta)}{M\left(\xi^2\right) - M^2(\xi)} = \frac{cov(\xi,\eta)}{D^2(\xi)} \ ,$$

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

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 *neccessary* conditions for the extreme value(s) of F, one should check that the solution ((6.8),(6.9)) really gives a minimum. However:

$$\frac{\partial^2 F}{\partial a^2} = 2M\left(\xi^2\right) \ , \qquad \frac{\partial^2 F}{\partial b^2} = 2 \ , \qquad \frac{\partial^2 F}{\partial ab} = 2M(\xi) \ ,$$

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

ii) Since the *real numbers* $\xi_1, ..., \xi_n$, $\eta_1, ..., \eta_n$ 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) :

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

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

so the system of equalities:

$$\left.\begin{aligned} a\sum_{i=1}^{n}\xi_i^2 + b\sum_{i=1}^{n}\xi_i = \sum_{i=1}^{n}\xi_i\eta_i \\ a\sum_{i=1}^{n}\xi_i + bn = \sum_{i=1}^{n}\eta_i \end{aligned}\right\} \tag{6.19}$$

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.

$$\frac{\sum_{i=1}^{n}\left(\xi_i - \bar{\xi}\right)\left(\eta_i - \bar{\eta}\right)}{\sum_{i=1}^{n}\left(\xi_i - \bar{\xi}\right)^2} = \frac{\sum_{i=1}^{n}\left(\xi_i\eta_i - \xi_i\bar{\eta} - \bar{\xi}\eta_i + \bar{\xi}\bar{\eta}\right)}{\sum_{i=1}^{n}\left(\xi_i^2 - 2\xi_i\bar{\xi} + \left(\bar{\xi}\right)^2\right)} =$$

$$= \frac{\sum_{i=1}^{n}\xi_i\eta_i - \bar{\eta}\cdot\sum_{i=1}^{n}\xi_i - \bar{\xi}\cdot\sum_{i=1}^{n}\eta_i + n\cdot\bar{\xi}\bar{\eta}}{\sum_{i=1}^{n}\xi_i^2 - 2\bar{\xi}\cdot\sum_{i=1}^{n}\xi_i + n\cdot\left(\bar{\xi}\right)^2} =$$

$$= \frac{n\cdot\overline{\xi\eta} - \bar{\eta}\cdot n\cdot\bar{\xi} - \bar{\xi}\cdot n\cdot\bar{\eta} + n\cdot\bar{\xi}\cdot\bar{\eta}}{n\cdot\overline{\xi^2} - 2\bar{\xi}\cdot n\cdot\bar{\xi} + n\cdot\left(\bar{\xi}\right)^2} = \boxed{\frac{n\cdot\overline{\xi\eta} - n\cdot\bar{\xi}\cdot\bar{\eta}}{n\cdot\overline{\xi^2} - n\cdot\left(\bar{\xi}\right)^2}} \ , \qquad (*)$$

one hand $\quad (*) = \boxed{\dfrac{\overline{\xi\eta} - \bar{\xi}\cdot\bar{\eta}}{\overline{\xi^2} - \left(\bar{\xi}\right)^2} = \dfrac{\overline{\xi\eta} - \bar{\xi}\cdot\bar{\eta}}{\sigma_\xi^2}}$,

other hand $\quad (*) = \dfrac{n\cdot\dfrac{1}{n}\cdot\sum\limits_{i=1}^{n}\xi_i\eta_i - n\cdot\dfrac{1}{n^2}\cdot\sum\limits_{i=1}^{n}\xi_i\cdot\sum\limits_{i=1}^{n}\eta_i}{n\cdot\dfrac{1}{n}\cdot\sum\limits_{i=1}^{n}\xi_i^2 - n\cdot\left(\dfrac{1}{n}\cdot\sum\limits_{i=1}^{n}\xi_i\right)^2}$

$$= \frac{\sum\limits_{i=1}^{n}\xi_i\eta_i - \dfrac{1}{n}\cdot\sum\limits_{i=1}^{n}\xi_i\cdot\sum\limits_{i=1}^{n}\eta_i}{\sum\limits_{i=1}^{n}\xi_i^2 - \dfrac{1}{n}\left(\sum\limits_{i=1}^{n}\xi_i\right)^2} = \frac{n\cdot\sum\limits_{i=1}^{n}\xi_i\eta_i - \left(\sum\limits_{i=1}^{n}\xi_i\right)\cdot\left(\sum\limits_{i=1}^{n}\eta_i\right)}{n\cdot\sum\limits_{i=1}^{n}\xi_i^2 - \left(\sum\limits_{i=1}^{n}\xi_i\right)^2} .$$

End of Proof. $\blacksquare$

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 connetcion 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. ξ and η, 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 $\quad R(\xi,\eta) = \frac{cov(\xi,\eta)}{D(\xi)D(\eta)} = \frac{M[(\xi - M(\xi))\cdot(\eta - M(\eta))]}{D(\xi)D(\eta)} \quad$ our choice is

$$r_{\xi,\eta} = \frac{\dfrac{1}{n}\sum\limits_{i=1}^{n}\left[\left(\xi_i - \overline{\xi}\right)\left(\eta_i - \overline{\eta}\right)\right]}{\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}\left(\xi_i - \overline{\xi}\right)^2}\cdot\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}\left(\eta_i - \overline{\eta}\right)^2}} = \frac{\dfrac{1}{n}\sum\limits_{i=1}^{n}\left[\left(\xi_i - \overline{\xi}\right)\left(\eta_i - \overline{\eta}\right)\right]}{\sigma_\xi\cdot\sigma_\eta} , \qquad (6.20)$$

which is equivalent to the easier (for hand-calculations) formula

$$r_{\xi,\eta}=\frac{n\cdot\sum\limits_{i=1}^{n}\xi_i\eta_i-\left(\sum\limits_{i=1}^{n}\xi_i\right)\left(\sum\limits_{i=1}^{n}\eta_i\right)}{\sqrt{n\cdot\sum\limits_{i=1}^{n}\xi_i^2-\left(\sum\limits_{i=1}^{n}\xi_i\right)^2}\cdot\sqrt{n\cdot\sum\limits_{i=1}^{n}\eta_i^2-\left(\sum\limits_{i=1}^{n}\eta_i\right)^2}}\ . \tag{6.21}$$

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?*

	1	2	3	4	5	6	7	8	9	10
Morning (ξ)	8.2	9.6	7.0	9.4	10.9	7.1	9.0	6.6	8.4	10.5
Afternoon (η)	8.7	9.6	6.9	8.5	11.3	7.6	9.2	6.3	8.4	12.3

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 ξ and η 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 ξ and η . In this section we investigate this dependency in more detail.

Definition II.100 *Let*

$$\omega:=a\xi+b-\eta \quad \textit{and} \quad \omega_i:=a\xi_i+b-\eta_i \tag{6.22}$$

the ***error - random variable*** *and the* ***error - data****, i.e. the difference between $a\xi+b$ and η , and between $a\xi_i+b_i$ and η .* □

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* $\sigma_\omega^2 = \overline{\omega^2}$.

Proof. We use *only* $\quad b = M(\eta) - a \cdot M(\xi)$.
Then

$M(\omega) = M(a\xi + M(\eta) - aM(\xi) - \eta) = aM(\xi) + M(\eta) - aM(\xi) - M(\eta) = 0$,

so $\quad D^2(\omega) = M(\omega^2) \quad$ follows.

Similarly, using $\quad b = \bar{\eta} - a \cdot \bar{\xi} \quad$ we have

$\bar{\omega} = a \cdot \bar{\xi} + b - \bar{\eta} = a \cdot \bar{\xi} + \left(\bar{\eta} - a \cdot \bar{\xi}\right) - \bar{\eta} = 0$. ■

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

i)

$$D^2(\omega) = D^2(\eta) \cdot \left(1 - R^2(\xi, \eta)\right) \tag{6.23}$$

ii)

$$\sigma_\omega^2 = \sigma_\eta^2 \cdot \left(1 - r_{\xi,\eta}^2\right) \ . \tag{6.24}$$

Proof. i) Using $b = M(\eta) - a \cdot M(\xi)$ we have

$D^2(\omega) = M(\omega^2) = M\left([a\xi + b - \eta]^2\right)$

$= M\left([a\xi + M(\eta) - a \cdot M(\xi) - \eta]^2\right) = M\left([a \cdot (\xi - M(\xi)) - (\eta - M(\eta))]^2\right)$

$= M\left(a^2 \cdot (\xi - M(\xi))^2 + (M(\eta) - \eta)^2 - 2a \cdot (\xi - M(\xi)) \cdot (\eta - M(\eta))\right)$

$= a^2 \cdot D^2(\xi) + D^2(\eta) - 2a \cdot cov(\xi, \nu) = (*)$.

Now use $\qquad a = \dfrac{cov(\xi, \eta)}{D^2(\xi)} \qquad$ and continue as

$$(*) = \frac{cov^2(\xi, \eta)}{D^2(\xi)} + D^2(\eta) - 2 \cdot \frac{cov^2(\xi, \eta)}{D^2(\xi)} = D^2(\eta) - \frac{cov^2(\xi, \eta)}{D^2(\xi)}$$

$= D^2(\eta) \cdot (1 - R^2(\xi, \eta))$.

ii) Using $\quad b = \bar{\eta} - a \cdot \bar{\xi} \quad$ we have

$$\sigma_\omega^2 = \overline{\omega^2} = \frac{1}{n} \sum_{i=1}^{n} \left[a\xi_i + \left(\bar{\eta} - a \cdot \bar{\xi}\right) - \eta_i\right]^2 = \frac{1}{n} \sum_{i=1}^{n} \left[a \cdot \left(\xi_i - \bar{\xi}\right) - (\eta_i - \bar{\eta})\right]^2$$

$$= \frac{1}{n} a^2 \cdot \sum_{i=1}^{n} \left(\xi_i - \bar{\xi}\right)^2 - 2a \cdot \frac{1}{n} \sum_{i=1}^{n} (\eta_i - \bar{\eta}) \left(\xi_i - \bar{\xi}\right) + \frac{1}{n} \sum_{i=1}^{n} (\eta_i - \bar{\eta})^2$$

$$= a^2 \cdot \sigma_\xi^2 - 2a \cdot \frac{1}{n} \sum_{i=1}^{n} (\eta_i - \bar{\eta}) \left(\xi_i - \bar{\xi}\right) + \sigma_\eta^2 \ .$$

Now use (6.14) twice and then (6.20) to continue

$$= a^2 \cdot \sigma_\xi^2 - 2a^2 \cdot \frac{1}{n} \sum_{i=1}^{n} \left(\xi_i - \bar{\xi}\right)^2 + \sigma_\eta^2 = -a^2 \cdot \sigma_\xi^2 + \sigma_\eta^2$$

$$= \sigma_\eta^2 - \sigma_\xi^2 \cdot \left(\frac{\frac{1}{n} \sum_{i=1}^{n} \left(\xi_i - \bar{\xi}\right)\left(\eta_i - \bar{\eta}\right)}{\frac{1}{n} \sum_{i=1}^{n} \left(\xi_i - \bar{\xi}\right)^2} \right)^2 = \sigma_\eta^2 - \sigma_\xi^2 \cdot \left(\sigma_\eta \cdot \frac{r_{\xi,\eta}}{\sigma_\xi}\right)^2 = \sigma_\eta^2 \cdot \left(1 - r_{\xi,\eta}^2\right) .$$

■

Remark II.103 ***(o)*** *Figures 5 and 6 show some experimental datasets with* $r = R(\xi, \eta)$*. Using the fact* $|R(\xi, \eta)| \leq 1$ *we can conclude* $D^2(\omega) \leq D^2(\eta)$*.*

(i) *First we can justify Theorem I.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.* ξ *and* η *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 surprinsing, since, by the horizontal line the differences of* η *and* b *(*$= \omega$*) are equal to the differences of* η *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 correspondance between* ξ *and* η *,* $r = 0.816$ *(this example by* ***Anscombe***[2]*).*

(iv) *The formula (6.23) is equivalent to*

$$|R(\xi, \eta)| = \sqrt{1 - \frac{D^2(\omega)}{D^2(\eta)}} \tag{6.25}$$

where, of course ω *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 ξ *and* η *(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* $|R(\xi, \eta)|$ *corespond only when* a *and* b *are optimal.*

---

[2] Francis John **Anscombe** (1918-2001) was an English statistician.

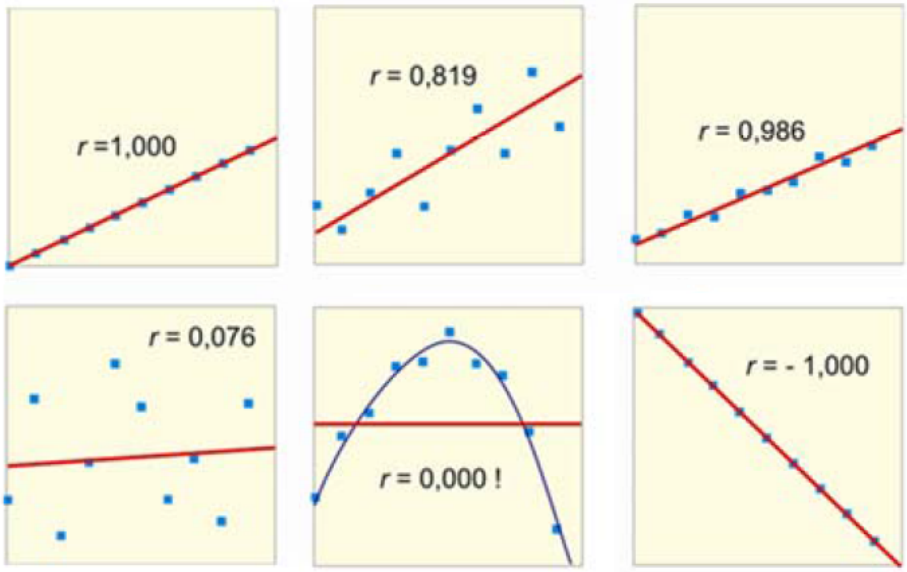

Figure 5: *Different regression values*

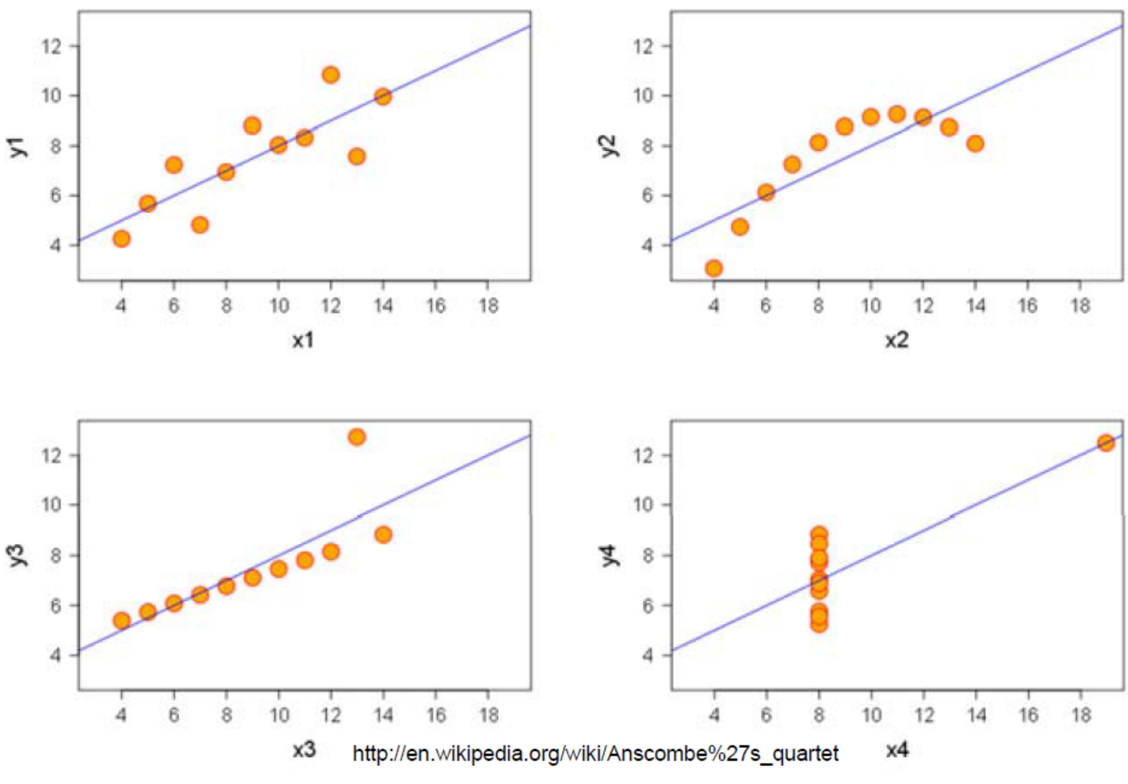

Figure 6: *Same regression values after* Anscombe ($r = 0.816$)
Source: https://en.wikipedia.org/wiki/Correlation_and_dependence

6.5 Nonlinear regressions - linearizing methods

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

$$\breve{y} = \breve{a} \cdot \breve{x} + \breve{b} \tag{6.26}$$

for some transformed values[3)] $\breve{x}$ and $\breve{y}$ of x and y , with real numbers $\breve{a}$ and $\breve{b}$. (In the language of statistics we write $\eta = g(\xi)$ and $\breve{\eta} = \breve{a}\breve{\xi} + \breve{b}$.) If, moreover the transformation of x and y to $\breve{x}$ and $\breve{y}$ 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 (ξ_i, η_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 $\breve{x}$ and $\breve{y}$, i.e. we have *logarihmic* 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 $\left\{(\breve{\xi}_i, \breve{\eta}_i) : i = 1, ..., n\right\}$ to get the values of $\breve{a}$ and $\breve{b}$ in (6.26).

Be careful: the error $M\left([\breve{b} + \breve{a}\breve{\xi} - \breve{\eta}]^2\right)$ 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 $\breve{a}, \breve{b}$! 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 $\breve{x}$ 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[4)]. 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 II.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 normal cumulative distribution functions, the program (application) on [HM] can handle normal coordinate transformations, too.

6.5.2 Exponential regression

The function equality[5)]

$$\eta = b \cdot a^{c \cdot \xi} \tag{6.27}$$

turns to

$$\lg(\eta) = \lg(b) + \xi \cdot c \cdot \lg(a) \ , \tag{6.28}$$

or in short form to

$$\breve{\eta} = \breve{b} + \xi \cdot \breve{a} \tag{6.29}$$

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

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

$$\left(\breve{\xi}_i, \breve{\eta}_i\right) := (\xi_i, \lg \eta_i) \qquad (i = 1, ..., n) \ , \tag{6.30}$$

---

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 (6.27) $\eta = b \cdot a^{c \cdot \xi}$ can be written in the form $\eta = b \cdot d^{\xi}$ where $d = a^c$, so c can be eliminated.

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

$$a = \exp\left(\frac{\breve{a}}{c}\right) = e^{\breve{a}/c} \quad \text{and} \quad b = \exp\left(\breve{b}\right) = e^{\breve{b}} \tag{6.31}$$

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

Using **semilogarithmic**[6)] coordinate system, i.e. logarithmic scale one axe (now η) and usual (equidistant) scale on the other axe (now ξ).

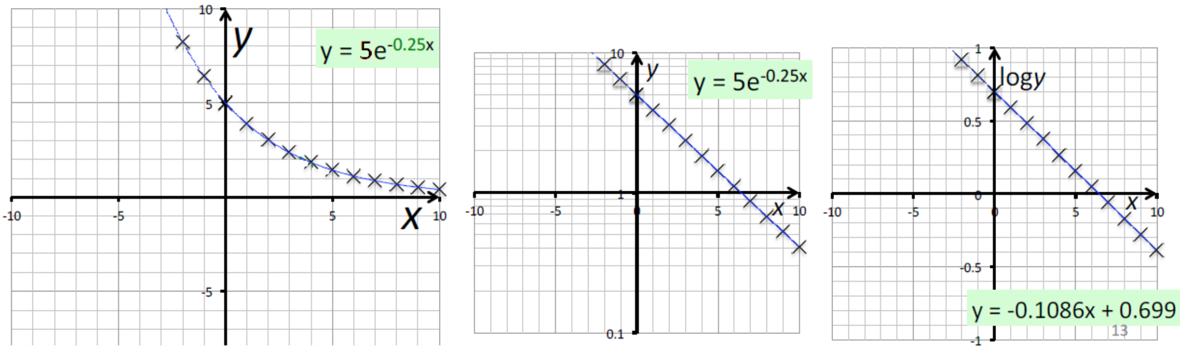

Figure 7: *Exponential function in* Cartesian (left), *in* semilogarithmic (medium) *coordinate systems, and its transform by* (6.28) (right)

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

On http://math.uni-pannon.hu/~szalkai/koordinata/semilog-uj-f.jpg and on http://math.uni-pannon.hu/~szalkai/koordinata/semilog-uj-hata.jpg we supply semi-log coordinate drawings *in high resolution.*

6) The word "*semi*" means "half".

6.5.3 Logarithmic regression

Now we have the function equality

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

which is itself linear in $\breve{\xi} = \lg(\xi)$ and $\breve{\eta} = \eta$, i.e. $\ell(x) = \lg(x)$, $\breve{a} = a$ and $\breve{b} = b$. 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 ξ and equidistant scale on the axe η .

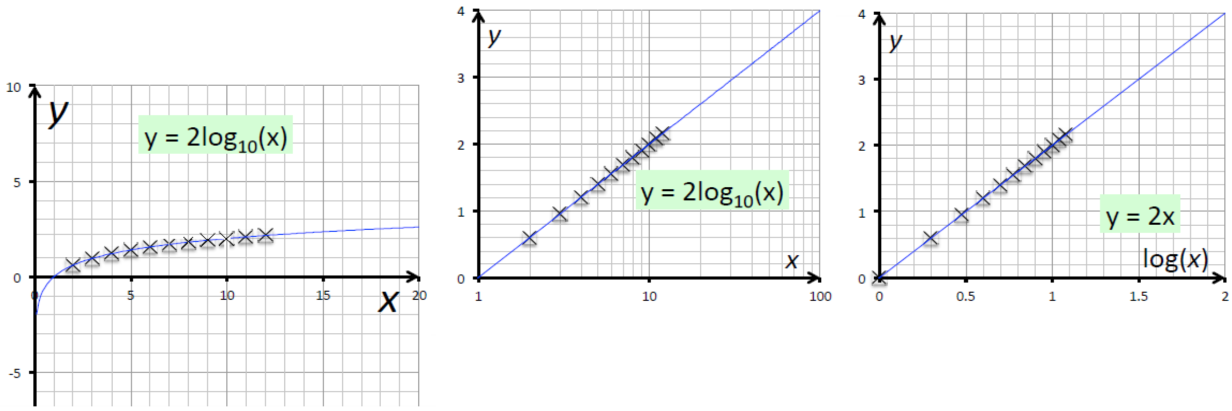

Figure 8: *Logarithmic function in* Cartesian *and in* semilogarithmic *coordinate systems*

On http://math.uni-pannon.hu/~szalkai/koordinata/semilog-uj-f.jpg and on 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

$$\eta = b \cdot \xi^a \tag{6.33}$$

turns to

$$\lg(\eta) = a \cdot \lg(\xi) + \lg(b) \tag{6.34}$$

or in short form to

$$\breve{\eta} = a \cdot \breve{\xi} + \breve{b} \tag{6.35}$$

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

$$a = \breve{a} \quad \textit{and} \quad b = \exp\left(\breve{b}\right) = e^{\breve{b}} \ . \tag{6.36}$$

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.

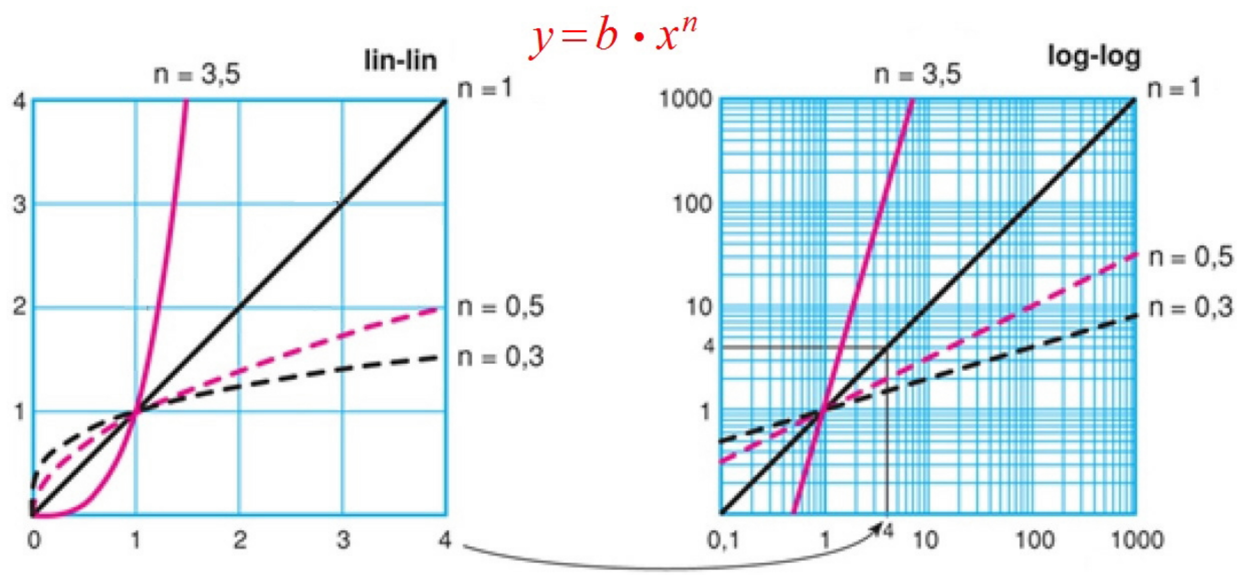

Figure 9: *Power functions in* Cartesian *and in* (double) logarithmic *coordinate systems*

On 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

$$\eta = \frac{\alpha\xi + \beta}{\gamma\xi + \delta} \tag{6.37}$$

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

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

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 $\breve{\xi} = 1/\xi$ and $\breve{\eta} = 1/\eta$):*

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

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

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. $\eta = \beta'\breve{\xi} + \alpha'$.*

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

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

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 $\eta = \frac{\alpha\xi + \beta}{\gamma\xi} = \frac{\alpha\xi}{\gamma\xi} + \frac{\beta}{\gamma\xi} = \frac{\alpha}{\gamma} + \frac{\beta}{\gamma} \cdot \frac{1}{\xi}$ i.e. $\eta = \beta'\breve{\xi} + \alpha'$. ∎

The above Theorem helps us to transform the dataset $\{(\xi_i, \eta_i) : i = 1, ..., n\}$ to the appropriate one $\left\{(\breve{\xi}_i, \breve{\eta}_i) : i = 1, ..., n\right\}$, how to solve the linearized regression problem $\breve{\eta} = \breve{a}\breve{\xi} + \breve{b}$ by Theorem II.95 and after how to get the constants $\alpha, \beta, \gamma, \delta$ in (6.37) from $\breve{a}$ and $\breve{b}$.

Corollary II.105

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

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

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

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

Proof. I) The system of equations $\dfrac{\gamma}{\beta} = \breve{a}$ and $\dfrac{\delta}{\beta} = \breve{b}$ has the solution $\beta = 1$, $\gamma = \breve{a}$ and $\delta = \breve{b}$.

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 **reciprocial**.

Corollary II.106

I) *if $\alpha = 0$ (and $\beta \neq 0$) then use normal (equidistant) axe for ξ and reciprocial axe for η ,*

II) *if $\beta = 0$ (and $\alpha \neq 0$) then use reciprocial 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 reciprocial axe for ξ and normal (equidistant) one for η .*

One example for Case II) is shown below:

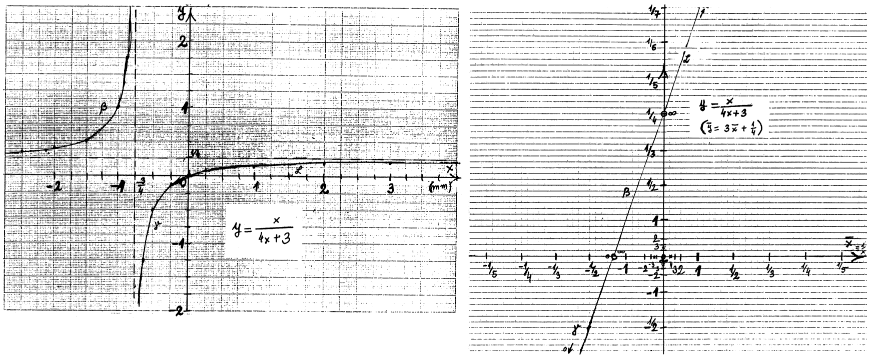

Figure 10: *Reciprocial function in* Cartesian *and in* reciprocial *coordinate systems*

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):

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

which is closely related to the normal distribution. Here ξ 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

$$\ln\left(\frac{\eta}{1-\eta}\right) = a\xi + b \ . \tag{6.40}$$

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

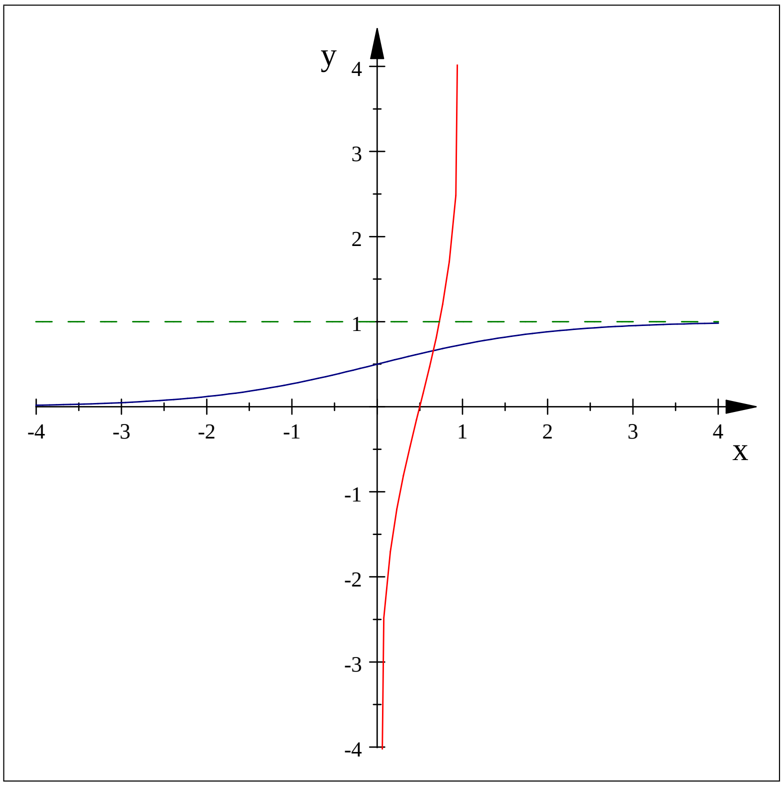

Figure 11: *The function* $1 - \dfrac{1}{1+e^x}$ (blue) and its inverse $\ln\left(\dfrac{x}{1-x}\right)$ *(red)*

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

We should use the transformation $\ln\left(\frac{y}{1-y}\right)$ on the y axe so that the functions $y = 1 - \dfrac{1}{1+e^{ax+b}}$ 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.

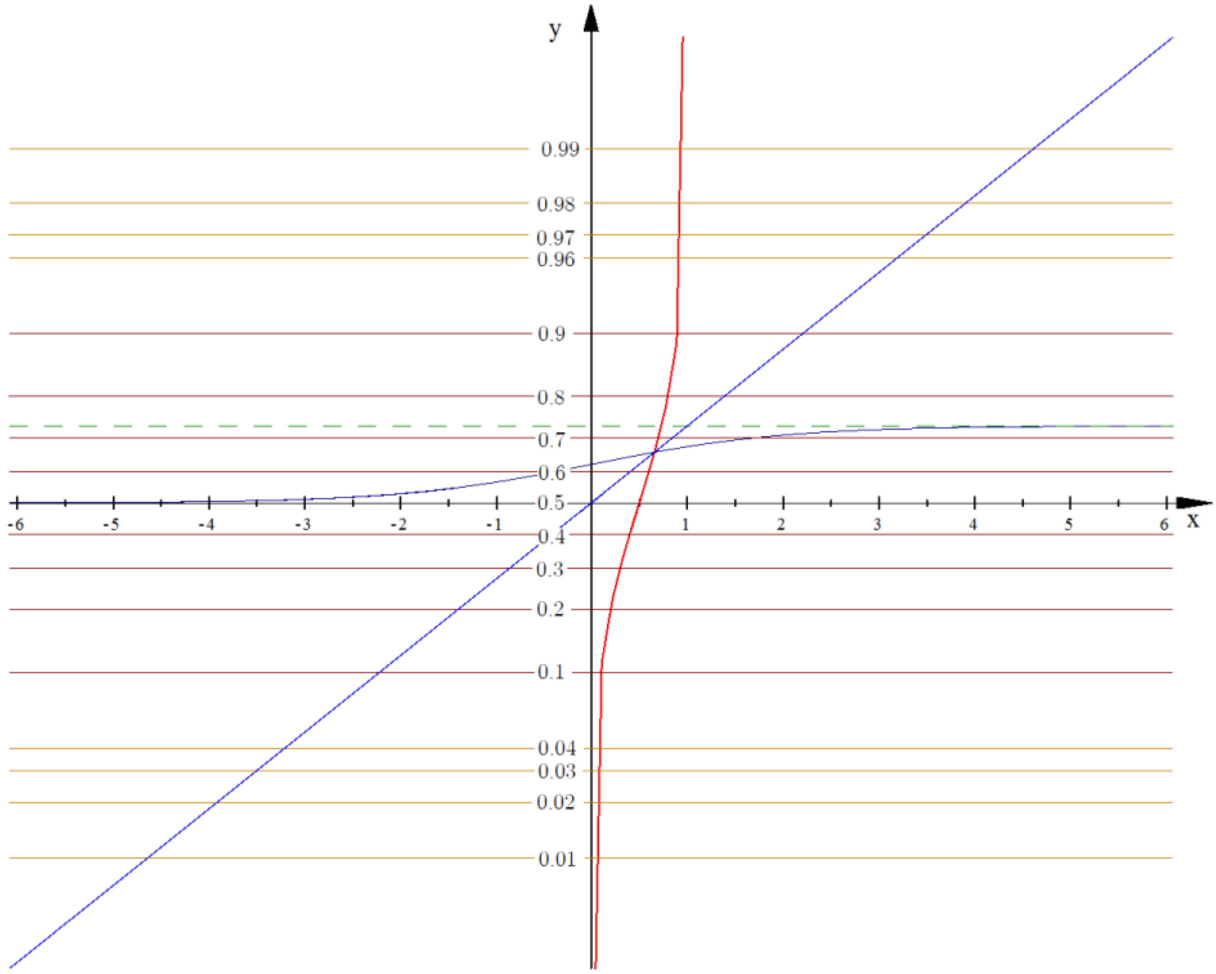

Figure 12: *The function* $1 - \dfrac{1}{1+e^x}$ (blue) *in the* logit-x *coordinate system*

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.

6.6.1 Quadratic regression

Now we consider the function

$$\eta = a + b\xi + c\xi^2 \ . \tag{6.41}$$

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 :

$$b = \frac{\sum\limits_{i=1}^{n} \xi_i \eta_i}{\sum\limits_{i=1}^{n} \xi_i^2} \quad , \qquad \left\{ \begin{array}{c} an + c\sum\limits_{i=1}^{n} \xi_i^2 = \sum\limits_{i=1}^{n} \eta_i \\ \\ a\sum\limits_{i=1}^{n} \xi_i^2 + c\sum\limits_{i=1}^{n} \xi_i^4 = \sum\limits_{i=1}^{n} \xi_i^2 \cdot \eta_i \end{array} \right. \ . \tag{6.42}$$

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

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

$$= \sum_{i=1}^{n} a^2\xi_i^4 + 2ab\xi_i^3 + 2ac\xi_i^2 - 2a\xi_i^2\eta_i + b^2\xi_i^2 + 2bc\xi_i - 2b\xi_i\eta_i + c^2 - 2c\eta_i + \eta_i^2 \ =$$

$$= a^2A + b^2B + c^2C + abD + acE + bcF - aG - bH - cI + J$$

where

$$\begin{aligned} A &= \sum_{i=1}^{n} \xi_i^4 \ , \ B = \sum_{i=1}^{n} \xi_i^2 \ , \ C = n \ , \ D = 2\xi_i^3 \ , \ E = 2\sum_{i=1}^{n} \xi_i^2 \ , \\ F &= 2\sum_{i=1}^{n} \xi_i \ , \ G = 2\sum_{i=1}^{n} \xi_i^2\eta_i \ , \ H = 2\sum_{i=1}^{n} \xi_i\eta_i \ , \ I = 2\sum_{i=1}^{n} \eta_i \ , \ J = \sum_{i=1}^{n} \eta_i^2 \ . \end{aligned} \tag{6.43}$$

Now

$$\left.\begin{array}{l} \frac{dF}{da} = 2Aa + bD + cE - G = 0 \\ \\ \frac{dF}{db} = 2Bb + aD + cF - H = 0 \\ \\ \frac{dF}{dc} = \ 2Cc + aE + bF - I = 0 \end{array}\right\} \Longleftrightarrow \left\{\begin{array}{c} 2aA + bD + cE = G \\ \\ aD + 2bB + cF = H \\ \\ aE + bF + 2cC = I \end{array}\right. \ ,$$

which is a system of linear equations, and has the solution

$$a = \frac{\det \begin{bmatrix} G & D & E \\ H & 2B & F \\ I & F & 2C \end{bmatrix}}{\det \begin{bmatrix} 2A & D & E \\ D & 2B & F \\ E & F & 2C \end{bmatrix}} , \; b = \frac{\det \begin{bmatrix} 2A & G & E \\ D & H & F \\ E & I & 2C \end{bmatrix}}{\det \begin{bmatrix} 2A & D & E \\ D & 2B & F \\ E & F & 2C \end{bmatrix}} , \; c = \frac{\det \begin{bmatrix} 2A & D & G \\ D & 2B & H \\ E & F & I \end{bmatrix}}{\det \begin{bmatrix} 2A & D & E \\ D & 2B & F \\ E & F & 2C \end{bmatrix}} ,$$

i.e.

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

where the common *denumerator* is

$$den = 2AF^2 - 2FDE + 2CD^2 + 2BE^2 - 8ABC \; .$$

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 $\eta = g(\boldsymbol{\xi}) = g(\xi_1, ..., \xi_n)$ of the sample $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)$ (see Definition II.6).

For example, in the t-test we have:

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

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

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

i.e.

$$P(\beta \leq \eta) = \varepsilon \tag{7.3}$$

can be computed and collected in tables.

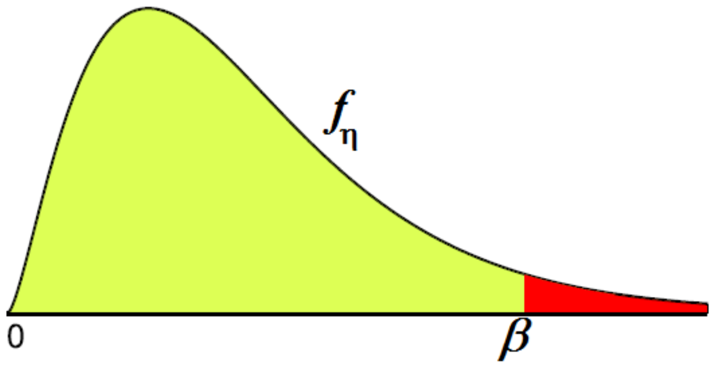

Figure 13: *Critical value*

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 χ^2 - distribution below.

7.1 The Student- or t- distribution

Definition II.108 *Let* ζ *and* $\xi_1, ..., \xi_n \sim N(0,1)$ *(i.e. standard normal) independent r.v.-s. Then*

$$\theta = \frac{\zeta}{\sqrt{\frac{\sum\limits_{i=1}^{n} \xi_i^2}{n}}} \tag{7.4}$$

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

Theorem II.109 *The density function is*

$$f_\theta(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \cdot \Gamma\left(\frac{n}{2}\right) \cdot \left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}} \tag{7.5}$$

where

$$\Gamma(x) := \int\limits_0^\infty t^{x-1} e^{-t} dt \tag{7.6}$$

is the so called Γ ***- function*** *(especially* $\Gamma(n) = (n-1)!$ *for* $n \in \mathbb{N}$*).*

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

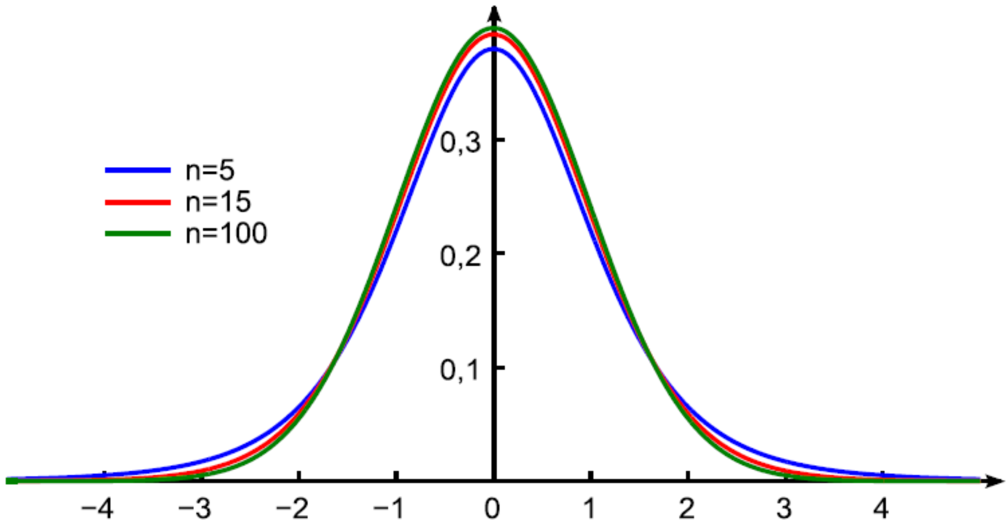

Figure 14: *Student distributions for* $n = 5,\ 15,\ 100$

7.2 The χ^2 distribution

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

$$\eta := \sum_{i=1}^{n} \xi_i^2 \tag{7.7}$$

is called ***chi-square distribution with parameter*** n **.** □

Theorem II.111 *The density function is*

$$f_\eta(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} \tag{7.8}$$

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

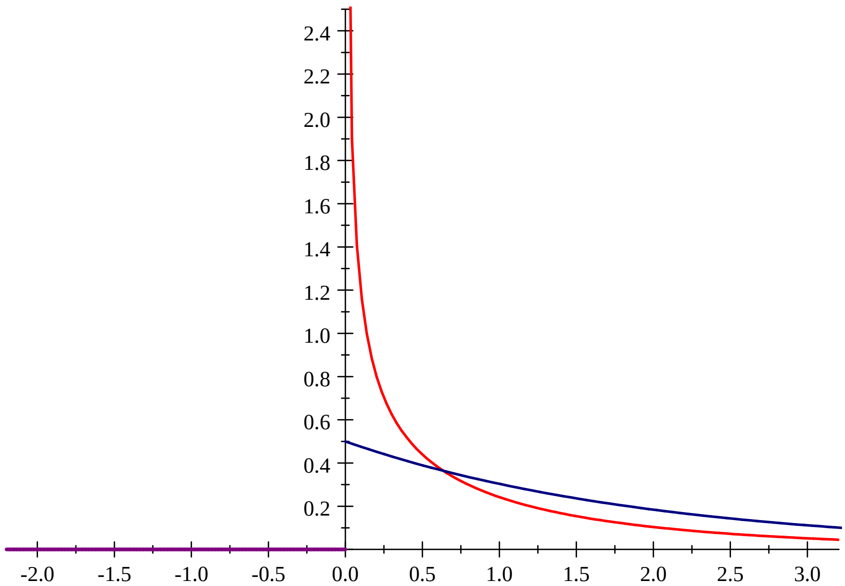

Figure 15: χ^2 *distributions for* **n=1** and **n=2**

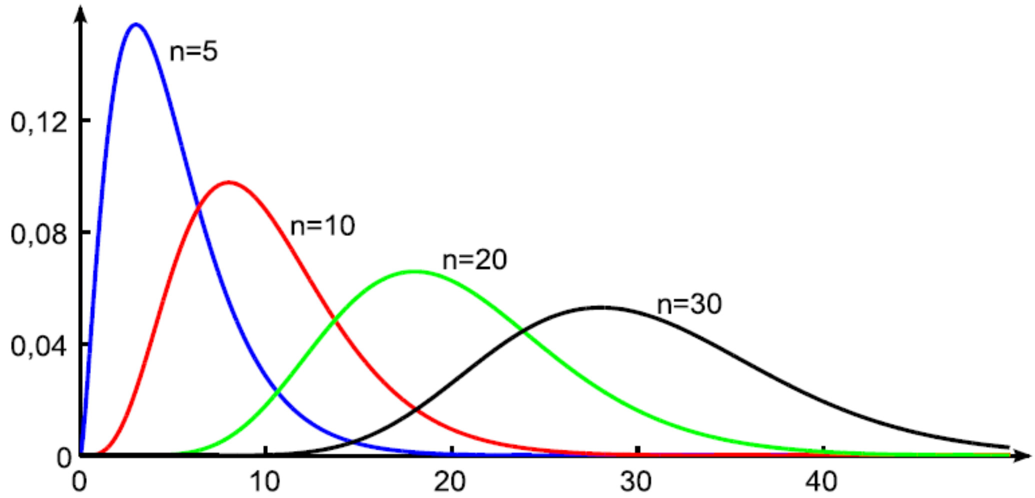

Figure 16: *χ^2 distributions for several n*

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.")* $\overrightarrow{\xi} = (\xi_t : t \in \mathbb{T})$ *for some* index-set $\mathbb{T}$ *is called a* ***stochastic process*** *("sztochasztikus folyamat"), or* ***s.p.*** *("szt.f.") for short.*

In case $\mathbb{T} = \mathbb{N}$ *we write* $\overrightarrow{\xi} = (\xi_1, \xi_2, ..., \xi_t, ...)$ *and say* ***discrete*** *("diszkrét"),*

in case $\mathbb{T} = \mathbb{R}$ *(*$t \in \mathbb{R}$*) we say* ***continuous*** *("folytonos") stochastic process.* □

Remark III.2 *i) In practice we measure the* same *quantity (*Ω*) several times: in time moments* $t \in \mathbb{T}$ *. Both discrete (separated, "elkülönült") and continuous measurements are well known in practice.*
ii) Each measurement (r.v.) ξ_i *can also be a vector (higher dimensional) r.v.:* $\xi_i = \left[\xi_i^{(1)}, ..., \xi_i^{(n)}\right]$.

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 ξ_t has the same distribution and they are mutually ("páronként") independent from each other.*

8.2.1 The Brownian motion

("Brown[1)]-mozgás"), also called Wiener[2)] process ("Wiener-folyamat").

A particle keeps moving in the space and let ξ_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* $t_0 < t_1 < ... < t_n < ...$ *the* ***increments*** *i.e. relative movements ("növekmények, relatív elmozdulások")*

$$\zeta_i := \xi_i - \xi_{i-1} \tag{8.1}$$

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.*

$$\eta(x) = \xi_{t+x} - \xi_t \quad (x \in \mathbb{R}) \tag{8.2}$$

depends only on $x = \Delta t$ and neither on $t_1 = t$ or on $t_2 = t + x$,

c)

$$P(\xi_t - \xi_s < x) = \frac{1}{\sqrt{2\pi B(t-s)}} \cdot \int_{-\infty}^{x} e^{\frac{-u^2}{2B(t-s)}} du = \tag{8.3}$$

$$= \frac{1}{\sqrt{2\pi B(t-s)}} \cdot \int_{-\infty}^{x} \exp\left(\frac{-u^2}{2B(t-s)}\right) du \tag{8.4}$$

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.

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

ii) For any $t_0 < t_1 < ... < t_n < t$ *the conditional probability is*

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

□

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} , ... , 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 proceeses.

8.2.2 The Poisson process

("Poisson[3] folyamat")

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

Example III.7 *Many pratical 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 occurence *of accidents at a certain intersection,*
the occurence *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* Δt *(see (8.2)),*

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

$$P(A \text{ in } h) = p(h) := a \cdot h + o(h) \quad \text{for } h \to 0 \tag{8.5}$$

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 occurence of two or more events.*

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

$$\sum_{m=2}^{\infty} P_m(t) = o(h) \qquad , \tag{8.6}$$

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

$$P_0(t) = e^{-at} \quad \text{for } t \in \mathbb{R}^{+,0} . \tag{8.7}$$

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* $m \in \mathbb{N}$

$$P(\xi_t = m) = P_m(t) = \frac{(at)^m}{m!} e^{-at} \tag{8.8}$$

where a *is determined in (8.7). Therefore,* ξ_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 $C \subset \mathbb{R}^d$ of points distributed in the space $\mathbb{R}^d$ $(1 \leq d)$. For any (measurable, "mérhető") set $H \subset \mathbb{R}^d$ 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 $H \subset \mathbb{R}^d$.

Definition III.12 *The collection $\{N_H : H \subset \mathbb{R}^d$ 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 $H \subset \mathbb{R}^d$ of finite volume ("térfogat") N_H has a Poisson distribution with mean

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

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")* $\overrightarrow{\xi} = (\xi_t : t \in \mathbb{T})$ *are classified by:*

- the ***state space*** *("állapottér")* $\mathbb{S}$ *where* $\xi_t : \Omega \to \mathbb{S}$ *,*

- the ***index*** *or* ***parameter set*** *("indexhalmaz, paraméterhalmaz")* $\mathbb{T}$ *,*

- the ***dependence relations*** *("függőségi viszonyok") among the r.v.* ξ_t *.* □

9.1 The state space

("állapottér")

This is the "space" (set) $\mathbb{S}$ in which the possible values of each ξ_t "lie".

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

i) In the case $\mathbb{S} = \{s_0, s_1, ..., s_n, ...\}$ *or* $\mathbb{S} = \mathbb{N}$ *we refer to the process* $\overrightarrow{\xi}$ *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 $\mathbb{S} = \mathbb{R}$ *the real line or a (real) interval* $[a, b] \subset \mathbb{R}$ *then we call* $\overrightarrow{\xi}$ *a* ***real-valued*** *("valós értékű") stochastic process.*

iii) If $\mathbb{S} \subseteqq \mathbb{R}^k$ *is a subset of* $\mathbb{R}^k$ *(or possibly the whole* $\mathbb{R}^k$*) - the more dimensional space then* $\overrightarrow{\xi}$ *is said to be a* k ***-vector*** *("*k *-vektor") process.* □

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* $\mathbb{T} = \mathbb{N} \cup \{0\} = \{0, 1, ...\}$ *then we shall always say that* $\overrightarrow{\xi}$ *is a* ***discrete time*** *("diszkrét idejű") stochastic process. When* $\mathbb{T}$ *is discrete we shall often write* ξ_n *instead of* ξ_t *.*

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

iii) The case $\mathbb{T} = \{measurable\ sets\} \subsetneqq \mathcal{P}\left(\mathbb{R}^d\right)$ *and other cases are also possible.* □

Example III.18 *We have already cited examples where the index set* $\mathbb{T}$ *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* ξ_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.* $\overrightarrow{\xi}$ *the* functions $\{M(\xi_t) : t \in \mathbb{T}\}$, $\{D(\xi_t) : t \in \mathbb{T}\}$ *and* $\{cov(\xi_t, \xi_s) : t, s \in \mathbb{T}\}$ *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. ξ_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*

$$\zeta_{t_1 t_2} := \xi_{t_2} - \xi_{t_1}, \quad \zeta_{t_2 t_3} := \xi_{t_3} - \xi_{t_2}, \ ..., \quad \zeta_{t_n t_{n-1}} := \xi_{t_n} - \xi_{t_{n-1}} \tag{10.1}$$

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

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

$$\zeta_{t_0} := \xi_{t_0} \ , \ \zeta_{t_0 t_1} := \xi_{t_1} - \xi_{t_0} \ , \ \zeta_{t_1 t_2} := \xi_{t_2} - \xi_{t_1}, \ ..., \ \zeta_{t_n t_{n-1}} := \xi_{t_n} - \xi_{t_{n-1}} \tag{10.2}$$

are (also) independent. □

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

$$\zeta_0 = \xi_0 \text{ and } \zeta_n = \xi_n - \xi_{n-1} \quad \text{for } n = 1, 2, ... \tag{10.3}$$

in the sense that knowing the individual distributions of ζ_0 *,* ζ_1 *, ... enables one to determine the joint distribution of any* finite subset $\{\xi_{n_1}, ..., \xi_{n_m}\}$ *of* $\overrightarrow{\xi}$ *.*
Especially

$$\xi_n = \zeta_0 + ... + \zeta_n \quad \text{for all } n = 0, 1, 2, ... \tag{10.4}$$

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

$$\xi_{t+h} - \xi_t \tag{10.5}$$

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

$$\vartheta_h := \xi_{t+h} - \xi_t \tag{10.6}$$

where $t \in \mathbb{T}$ *is arbitrary fixed index.* □

Theorem III.23 *If a process* $\overrightarrow{\xi} = \{\xi_t : t \in \mathbb{T}\}$ *where* $\mathbb{T} = [0, \infty)$ *or* $\mathbb{T} = \mathbb{N}$ *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*

$$M(\xi_t) = m_0 + m_1 \cdot t \quad (t \in \mathbb{T}) \tag{10.7}$$

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

$$\sigma^2_{\xi_t} = \sigma^2_0 + \sigma^2_1 \cdot t \quad (t \in \mathbb{T}) \tag{10.8}$$

where

$$\sigma^2_0 = M[(\xi_0 - m_0)^2] = D^2(\xi_0) \tag{10.9}$$

and

$$\sigma^2_1 = M[(\xi_1 - m_1)^2] - \sigma^2_0 = D^2(\xi_1) - D^2(\xi_0) \quad . \tag{10.10}$$

□

Remark III.24 *Both the Brownian motion process and the Poisson process have stationary independent increments.*

10.2 Martingales

("Martingálok")

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

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

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

$$M\left(\xi_{t_{n+1}} \mid \xi_{t_1} = a_1, ..., \xi_{t_n} = a_n\right) = a_n \ . \qquad \square \tag{10.11}$$

Remark III.26 *i) Observe the absolute value of* ξ_t *in i) and recall that i) is stronger than "*ξ_t has a finite mean*".*

ii) Martingales may be considered as appropriate models for fair games *in the sense that* ξ_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* ζ_i *are independent r.v. and* $M(\zeta_i) = 0$ *, then the process*

$$\xi_n = \zeta_1 + ... + \zeta_n \qquad (n \in \mathbb{N}) \tag{10.12}$$

is a discrete martingale.

ii) Similarly, if ξ_t *for* $0 \le t$ *has* independent increments *whose means are* 0 *then* $\overrightarrow{\xi}$ *is a continuous time martingale.*

10.3 Markov processes

("Markov[1)] folyamatok")

Definition III.28 *A process* $\overrightarrow{\xi}$ *is said to be* ***Markov s.p.*** *("Markov folyamat") if*

$$P\left(a < \xi_t \le b \mid \xi_{t_1} = a_1, ..., \xi_{t_n} = a_n\right) = P\left(a < \xi_t \le b \mid \xi_{t_n} = a_n\right) \qquad (10.13)$$

for all $t \in \mathbb{T}$ *whenever* $t_1 < t_2 < ... < t_n < t$ *and for all values* $a_1, a_2, ..., a_n \in \mathbb{S}$ *.*
□

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

Definition III.29 *A process* $\overrightarrow{\xi}$ *is said to be a* ***discrete Markov s.p.*** *("diszkrét Markov folyamat") or a* ***Markov-chain*** *("Markov-lánc") if*

$$P\left(\xi_{t_{n+1}} = a_{n+1} \mid \xi_{t_1} = a_1, ..., \xi_{t_n} = a_n\right) = P\left(\xi_{t_{n+1}} = a_{n+1} \mid \xi_{t_n} = a_n\right) \qquad (10.14)$$

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

Remark III.30 *i) Roughly speaking a Markov s.p. is one with the property that, if the value of* ξ_t *is given, then the values of* ξ_s *for* $s > t$ *do* not *depend on the values of* ξ_u *for* $u < t$ *. That is the probability of any particular future behaviours of the process, when its present state* (ξ_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 (ξ_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 $\mathbb{S}$ *is called a* ***Markov chain*** *("Markov-lánc").*

Example III.31 ***Discrete Brownian motion as partial sums of independent r.v.'s*** *("Diszkrét 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* $\overrightarrow{\xi}$ *is a discrete Markov s.p. where*

$$\xi_n = \eta_1 + ... + \eta_n \quad 1 \le n \qquad (10.15)$$

and η_i *are independent and the values of* η_i *are* ± 1 *with probability* 0.5
(i.e. $\eta_i : \Omega \to \{-1, +1\}$*,* $P(\eta_i = -1) = P(\eta_i = +1) = 0.5$ *.*

[1)] Andrey Andreyevich **Markov** (1856-1922) a Russian mathematician.

Claim III.32 *In general, it is easy to prove, that partial sums (10.15) of independent r.v.* η_i *are always a discrete Markov s.p.* □

Definition III.33 *The Markov chain* $\overrightarrow{\xi}$ *in (10.15) is called* ***homogeneous*** *("homogén") if* η_i *all have the same distribution, otherwise* $\overrightarrow{\xi}$ *is* ***inhomogeneous*** *("inhomogén").* □

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* η_i *be independent r.v. which have values* $\{0, 1, ..., N-1\}$ *with arbitrary probabilities. Now if we define* $\overrightarrow{\xi}$ *as* $\xi_0 = \eta_0$ *and*

$$\xi_{n+1} = \begin{cases} \xi_n + \eta_n & \text{if } \xi_n + \eta_n < N \\ \xi_n + \eta_n - N & \text{if } \xi_n + \eta_n \geq N \end{cases} \tag{10.16}$$

then $\overrightarrow{\xi}$ *is also a Markov chain.*

This example is called ***lower rounding*** *("lefelé kerekítés, csonkítás")* ***against overfloating*** *("túlcsordulás ellen").*

Definition III.36 *Let* $A \subset \mathbb{R}$ *be an* interval *of the real line. Then the function*

$$\mathcal{P}(x, s, t, A) := P(\xi_t \in A \mid \xi_s = x) \tag{10.17}$$

for $t > s$ *is called* ***transition probability function*** *("átmenetvalószínűség-függvény") and is basic to the study of the structure of Markov s.p.* □

Claim III.37 *We may express the condition (10.13) also as follows:*

$$P\left(a < \xi_t \leq b \mid \xi_{t_1} = a_1, ..., \xi_{t_n} = a_n\right) = \mathcal{P}(x_n, t_n, t, (a, b]) \ . \tag{10.18}$$

□

It can be proved that the probability distribution of $\left(\xi_{t_1}, ..., \xi_{t_n}\right)$ can be computed in terms of (10.17) and the initial distribution function of ξ_{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*** *("diffúziós folyamat").* □

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* $\overrightarrow{\xi} = \{\xi_n : n \in \mathbb{N}\}$

i) the probabilities

$$_n p_{i,k}^{(r)} := P\left(\xi_{n+r} = k \mid \xi_n = i\right) \tag{10.19}$$

are called r ***-step transition probabilities*** *("*r *-lépéses átmenetvalószínűségek"), shortly* ***t.p.****, for* $r, n, i, k \in \mathbb{N}$ *.*

ii) the (finite or infinite) matrix

$$_n\mathbf{\Pi}_r := \left[{_n p_{i,k}^{(r)}}\right]_{i,k} = \begin{bmatrix} {_n p_{1,1}^{(r)}} & {_n p_{1,2}^{(r)}} & \cdots \\ {_n p_{2,1}^{(r)}} & {_n p_{2,2}^{(r)}} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \tag{10.20}$$

is called ***transition probability matrix*** *("átmenetvalószínűség mátrix").*

For homogeneous Markov chains the index n *is usually omitted. We also omit* r *in case* $r = 1$ *.* □

Claim III.41 *All the entries of* $_n\Pi_r$ *are probabilities* $\in [0, 1]$ *and each row has sum* 1 *since*

$$\sum_{k=1}^{\infty} {_n p_{i,k}^{(r)}} = \sum_{k=1}^{\infty} P\left(\xi_{n+r} = k \mid \xi_n = i\right) = 1 \ . \quad \square \tag{10.21}$$

Definition III.42 *Any quadratic ("négyzetes") matrix (either finite or infinite) with nonnegative entries is called* ***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.* $\sum_{i=1}^{\infty} {_n p_{i,k}^{(r)}} = 1$ *, then the matrix is called a* ***double stochastic matrix*** *("kétszeresen sztochasztikus mátrix").* □

Claim III.43 *Products of (double) stochastic matrixes is also a (double) stochastic one.* □

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*

$$\Pi_r = (\Pi)^r \tag{10.22}$$

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 (10.22) for* $r_1 + r_2 = r$ *, i.e.*

$$p_{i,k}^{(r)} = \sum_{j=1}^{\infty} p_{i,j}^{(r_1)} \cdot p_{j,k}^{(r_2)} \tag{10.23}$$

is often in use without mentioning and is called ***Markov equality*** *("Markov egyenlőség").*

ii) The transition probabilities ${}_n p_{i,k}^{(r)}$ *are conditional probabilities ("feltételes valószínűségek"), so the unconditional ("feltétel nélküli") probabilities of* ξ_n

$$p_k(n) := P(\xi_n = k) \qquad k \in \mathbb{N}, \quad n \in \mathbb{N} \cup (0) \tag{10.24}$$

are called ***absolute probabilities*** *("abszolút valószínűségek") of* ξ_n *.* □

Definition III.46 *A Markov-chain* $\overrightarrow{\xi}$ *is called* ***ergodic*** *("ergodikus") if all the limit probabilities*

$$P_k := \lim_{r \to \infty} p_{i,k}^{(r)} \tag{10.25}$$

do exist, they are independent of i *, and*

$$\sum_{k=1}^{\infty} P_k = 1 \ . \qquad \square \tag{10.26}$$

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

$$A_k = \{\lim_{r \to \infty} \xi_r = k\} \subset \Omega \tag{10.27}$$

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 $\overrightarrow{\xi}$ having finitely many states ("véges állapotú") is ergodic if and only if*

$$\Pi = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\ p_{2,1} & p_{2,2} & \cdots & p_{1,N} \\ \cdots & \cdots & \cdots & \cdots \\ p_{N,1} & p_{N,2} & \cdots & p_{N,N} \end{bmatrix} \tag{10.28}$$

(see (10.20)) has a power Π^v ($v \in \mathbb{N}$) in which at least one column contains only positive *elements.*

Further, the convergence in (10.25) is exponential:

$$\left|p_{i,k}^r - P_k\right| \leq (1 - M\delta)^{\frac{n}{v}-1} \tag{10.29}$$

where M is the number of columns of Π^v containing positive elements, δ is the least number in these columns. (Clearly $0 < M\delta \leq 1$.) □

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 many 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

$$p_{i,k}^{n+1} = \sum_{j=1}^{N} p_{i,j}^{(n)} \cdot p_{j,k}^{(1)} \quad , \tag{10.30}$$

and taking $n \to \infty$ we get

$$P_k = \sum_{j=1}^{N} P_j P_{j,k} \quad \text{for } 1 \leq k \leq N \ . \tag{10.31}$$

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 Π^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 $N = |\mathbb{S}|$). 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 ξ_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, $\mathrm{Im}(\xi_t) = \mathbb{N}$ *and let*

$$p_i := P(\xi_t = i) \ . \tag{10.32}$$

Let further ζ_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 ζ_0 the starting level.*

Clearly the lake contains no more than K water in each moment, so we must have

$$\zeta_{t+1} := \max\left\{ \ \min\left(\zeta_t + \xi_{t+1}, K\right) - M \ , \ 0\right\} \tag{10.33}$$

which implies

$$\mathrm{Im}(\zeta_t) = \{0, 1, ..., K - M\} \tag{10.34}$$

and we let

$$P_{i,j} := P\left(\zeta_{t+1} = j \mid \zeta_t = i\right) \tag{10.35}$$

the possible water level in the next year.

For simplicity we assume

$$M < K - M \quad i.e. \quad M < K/2 \ . \tag{10.36}$$

Solution III.51 *Clearly*

$$P_{u,v} = 0 \quad if \quad u - M > v \quad i.e. \quad u - v > M \ , \tag{10.37}$$

or even, for suitable *u, v, w (among others)*

$$0 \le u, v \le K - M \qquad and \qquad 0 \le w = M + v - u \tag{10.38}$$

imply

$$P_{u,v} = P\left(\xi_t = w\right) \iff u - M + w = v \tag{10.39}$$

i.e.

$$P_{u,v} = p_{v+M-u} \ . \tag{10.40}$$

Further, for suitable j

$$P_{j,K-M} = p_{K-j} + p_{K-j+1} + ... \quad . \tag{10.41}$$

Finally, we have the following (large) system of equalities for $P_{i,j}$.

$P_{0,0} = p_0 + ... + p_M$,
$P_{0,1} = p_{M+1}$,
...
$P_{0,i} = p_{M+i}$ (for $i = \zeta_{t+1} < K - M$ i.e. $M + i < K$),
...
$P_{0,K-M-1} = p_{K-1}$,
$P_{0,K-M} = p_K + p_{K+1} + ...$,

$P_{1,0} = p_0 + ... + p_{M-1}$,
$P_{1,1} = p_M$,
...
$P_{1,i} = p_{M-1+i}$,
...
$P_{1,K-M-1} = p_{K-2}$
$P_{1,K-M} = p_{K-1} + p_K + ...$,

...
...
$P_{j,0} = p_0 + ... + p_{M-j}$,
$P_{j,1} = p_{M+1-j}$,
...
$P_{j,i} = p_{M-j+i}$ (for $j \leq M$),
...
$P_{j,K-M-1} = p_{K-j-1}$,
$P_{j,K-M} = p_{K-j} + p_{K-j+1} + ...$,
...
...
$P_{M,0} = p_0$ (since $M < K - M$ i.e. $M < K/2$),
$P_{M,1} = p_1$,
...
$P_{M,i} = p_i$,

...
$P_{M,K-M-1} = p_{K-M-1}$ (since $M \leq K - M - 1$),
$P_{M,K-M} = p_{K-M} + p_{K-M+1} + \dots$,

$P_{M+1,0} = 0$,
$P_{M+1,1} = p_0$,
...
$P_{M+1,i} = p_{i-1}$ (for $i \leq K - M - 1$),
...
$P_{M+1,K-M-1} = p_{K-M-2}$,
$P_{M+1,K-M} = p_{K-M-1} + p_{K-M} + \dots$,
...
...
$P_{M+\ell,0} = 0$ (for $1 \leq \ell$ and $M + \ell \leq K - M$ i.e. $\ell \leq K - 2M$),
...
$P_{M+\ell,\ell-1} = 0$ (see (10.37)),
$P_{M+\ell,\ell} = p_0$,
$P_{M+\ell,\ell+i} = p_i$ (for $\ell + i \leq K - M - 1$ i.e. $i \leq K - M - \ell - 1$),
...
$P_{M+\ell,K-M-1} = p_{K-M-\ell-1}$,
$P_{M+\ell,K-M} = p_{K-M-\ell} + p_{K-M-\ell+1} + \dots$,
...
...
$P_{K-M,0} = 0$ (see (10.36)),
$P_{K-M,1} = 0$,
...
$P_{K-M,i} = 0$ (for $i < (K - M) - M = K - 2M$, see (10.37)),
...
$P_{K-M,K-2M} = p_0$ (by (10.40) $v + M - u = K - 2M + M - (K - M) = 0$),
...
$P_{K-M,K-M-1} = p_{M-1}$,
$P_{K-M,K-M} = p_M + p_{M+1} + \dots$.

END of the Example.

10.4 Stationary processes

("Stacionárius [állandó] folyamatok")

Definition III.52 *A s.p.* $\overrightarrow{\xi}$ *(where* $\mathbb{T}$ *can be any of the sets* $(-\infty, \infty)$ *,* $[0, \infty)$ *,* $\mathbb{Z}$ *or* $\mathbb{N}\backslash\{0\}$*) is said to be* ***strictly stationary*** *("erősen stacionárius") if the joint distribution functions of the families of random variables are*

$$\bar{\xi}_{t+h} = \left(\xi_{t_1+h} \ , \ \xi_{t_2+h} \ , \ \dots \ , \ \xi_{t_n+h}\right) \quad and \quad \bar{\xi}_t = \left(\xi_{t_1} \ , \ \xi_{t_2} \ , \ \dots \ , \ \xi_{t_n}\right) \ , \tag{10.42}$$

that is $F_{\bar{\xi}_{t+h}}$ *and* $F_{\bar{\xi}_t} : \mathbb{R}^n \to \mathbb{R}$ *are* the same *for all* $h > 0$ *and arbitrary finite set of* $t_1, ..., t_n \in \mathbb{T}$ *.* □

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* ξ_t *is the* same *for each* $t \in \mathbb{T}$ *.*
The word stationary *means "almost constant" ("majdnem állandó").*

Theorem III.54 *The mean- and dispersion functions of stationary processes do not depend on* $t \in \mathbb{T}$ *:* $M(\xi_t) = M(\xi_0)$ *and* $D(\xi_t) = D(\xi_0)$ *.*
The autocovariance function depends on $(t - s)$ *:*

$$cov(\xi_t, \xi_s) = cov\left(\xi_{t-s}, \xi_0\right) \tag{10.43}$$

for $t, s \in \mathbb{T}$ *.* □

Definition III.55 *A s.p.* $\overrightarrow{\xi}$ *is said to be*

i) ***wide sense stationary*** *("gyengén stacionárius") if it possesses finite second moments (i.e.* $M\left(\xi_t^2\right) < \infty$*),*

ii) ***covariance stationary*** *("stacionárius kovarianciájú") if* $cov(\xi_t, \xi_{t+h})$ *depends only on* h *for all* $t \in \mathbb{T}$ *.* □

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 eonomics.*

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* $\overrightarrow{\tau} = (\tau_n : n \in \mathbb{N})$ *of independent and identically distributed positive r.v. representing the* <u>*lifetimes*</u> *of some "units" ("egységek"). The first unit is placed in operation at time 0 , it fails at time* τ_1 *and is immediately replaced by a new unit (with the* <u>*same*</u> *properties) which fails at time* $\tau_1 + \tau_2$ *and so on. The* <u>*time*</u> *of the n'th* ***renewal*** *("felújítás") is*

$$\sigma_n = \tau_1 + ... + \tau_n \quad (n \in \mathbb{N}) \ . \tag{10.44}$$

ii) A renewal ***counting*** *("számláló") process is* $\overrightarrow{\nu} = (\nu_t : t \in \mathbb{R}^{+,0})$ *where for* $t \in \mathbb{R}^{+,0}$ *and* $n \in \mathbb{N}$

$$\nu_t = n \overset{def}{\iff} \sigma_n \leq t < \sigma_{n+1} \ . \tag{10.45}$$

□

Remark III.61 *i) The renewal process* σ_n *gives us the* <u>*time moment*</u> *of the* n *'th renewal, while a renewal* counting *process* ν_t *counts the* <u>*number of renewals*</u> *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 λ is *a renewal counting process for which the unit lifetimes have exponential distributions with common parameter* λ *.*

10.6 Point processes

("Pontfolyamatok")

Note: $S \neq \mathbb{S}$!

Definition III.62 *Let* $S \subseteqq \mathbb{R}^n$ *be a* fixed *set in the* n *-dimensional space and let* $\mathcal{A} \subseteqq \mathcal{P}(S)$ *be a family of* subsets *of* S *. A* ***point process*** *("pontfolyamat") is a s.p. indexed by the sets* $A \in \mathcal{A}$ *, that is* $\mathbb{T} = \mathcal{A}$ *, having the state space* $\mathbb{S} = \mathbb{N} \cup \{0\}$ *(nonnegative integers). In other words:* $\overrightarrow{\xi} = \{\xi_A : A \in \mathcal{A}\}$ *.* □

Remark III.63 *(i) Non-mathematicians please write* $\mathcal{A} = \mathcal{P}(S)$ *, i.e. let* $A \in \mathcal{A}$ *mean "*$A \subseteqq S$ *is any (measurable) subset of* S *".*
(ii) We think a set of (enumerable) "points" $C \subset S$ *is being scattered over* S *in some random manner, and of*

$$\xi_A = N(A) := |A \cap C| \tag{10.46}$$

as counting the number of points from C *in the (measurable) set* $A \in \mathcal{A}$ *, i.e.* $A \subseteqq S$ *.*

Since $N(A)$ is a counting function there are additional requirements on each realization.

Definition III.64 *(continued):*
i) if $A_1 \cap A_2 = \varnothing$ *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 $\varnothing \in \mathcal{A}$ *then* $N(\varnothing) = 0$ *, i.e.* $\xi_{\varnothing} = 0$ *,* □

Clearly ii) follows from i).

Definition III.65 *Suppose* S *is a set in the real line (or plane os 3-dimensional space) and for every subset* $A \subset S$ *let* $V(A)$ *be the length (area, volume, resp.) of* A *. Then*

$$\overrightarrow{\nu} = \{\nu_A : A \subset S\} \tag{10.47}$$

is a ***homogeneous Poisson point process of intensity*** $\lambda > 0$ *("*λ *intenzitású (erősségű) Poisson-pontfolyamat") if* $\mathcal{A} = \mathcal{P}(S)$ *(power set) and* *(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 $\{A_1, ..., A_n\} \subset \mathcal{A}$ *of mutually ("páronként") disjoint subsets of* S *the r.v.'s* $\nu_{A_1}, ..., \nu_{A_n}$ *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 depency 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* $\overrightarrow{\zeta} = \{\zeta_n : n = 0, \pm 1, \pm 2, ...\}$ *i.e.* $\mathbb{T} = \mathbb{Z}$ *(integers) be* uncorrelated *r.v. having a common mean* μ *and variance* σ^2*. Let* $m \in \mathbb{N}$ *and* a_1 *,* a_2 *,...,*$a_m \in \mathbb{R}$ *be any* fixed *numbers and consider the process* $\overrightarrow{\xi} = \{\xi_n : n \in \mathbb{Z}\}$ *where*

$$\xi_n = a_1\zeta_n + a_2\zeta_{n-1} + ... + a_m\zeta_{n-m+1} \quad \text{for } n \in \mathbb{Z} \ . \tag{10.48}$$

Now the s.p. $\overrightarrow{\xi}$ *is called a* ***moving average processes*** *("mozgóátlag folyamat").* □

Remark III.68 *The naming "moving average" refers to the application when the original s.p.* $\overrightarrow{\zeta}$ *has extreme low and high (expected) values and perhaps periodic or "****seasonable****"* [2] *("szezonális"), and these huge differences are decreased and the extreme alterations are smoothed by taking the (weighted) average of* m *consequtive r.v.* ζ_n *, ... ,* ζ_{n-m+1} *. So, the s.p.* $\overrightarrow{\xi}$ *contains the averages of these consequtive r.v.* ζ_i *, and goes on, i.e. moves. The usual arithmetic mean*

[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.

("számtani/aritmetikai közép") uses $a_1 = ... = a_m = \frac{1}{m}$ *and weighted arithmetic means ("súlyozott számtani közép") need* $a_1 + ... + a_m = 1$ *, however in (10.48) the numbers* a_i *can be* any *real numbers!*

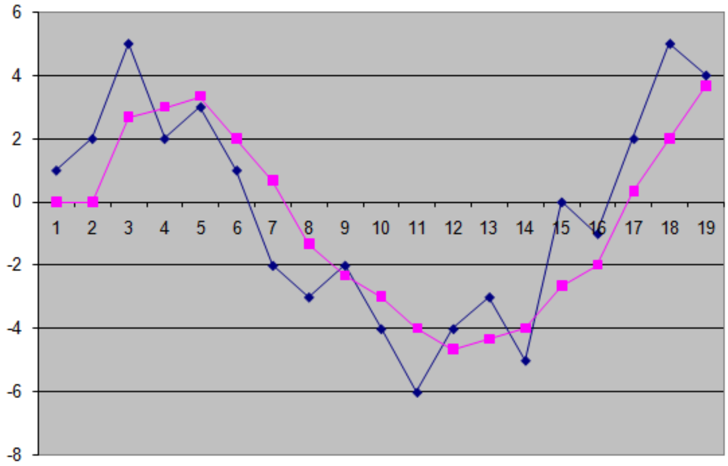

Figure 17: *Moving average*
ζ_n **are in blue** and ξ_n **are in red**

Claim III.69 $\quad M(\xi_n) = \mu \cdot (a_1 + ... + a_m)$

and

$D^2(\xi_n) = \sigma^2 \cdot (a_1^2 + ... + a_m^2)$. $\square$

For the covariance we have

Theorem III.70

$$cov\,(\xi_n, \xi_{n+v}) = E\left[\left(\xi_n - \mu \cdot \sum_{i=1}^{m} a_i\right)\left(\xi_{n+v} - \mu \cdot \sum_{i=1}^{m} a_i\right)\right] = \tag{10.49}$$

$$= \begin{cases} \sigma^2 \cdot (a_m a_{m-v} + ... + a_{v+1} a_v) & \text{if} \quad v \leq m-1 \\ 0 & \text{if} \quad v \geq m \end{cases} . \qquad \square$$

Since the covariance between ξ_n and ξ_{n+v} depends only on v and not on n , the process $\overrightarrow{\xi}$ 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* $k = 1, ..., m$. *Now the covariance function becomes*

$$R(v) = \begin{cases} \sigma^2 \cdot \left(1 - \frac{v}{m}\right) & \text{if} \quad v \le m-1 \\ 0 & \text{if} \quad v \ge m \end{cases} . \tag{10.50}$$

10.8 Autoregressive processes

("autoregressziós folyamatok")

Definition III.72 *Let* $\{\zeta_n : n = 0, \pm 1, \pm 2, ...\}$ *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 σ^2 . *The s.p. defined in (10.51) is called an* ***autoregressive process of order one*** *("elsőrendű autoregressziós folyamat").* □

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)* λ *gives the "measure" of the autoregression.*

ii) Recall, that "ξ_i *and* ξ_j *are uncorrelated" only means that* $\operatorname{cov}(\xi_i, \xi_j) = 0$ *which is weaker than "*ξ_i *and* ξ_j *are independent".*

From (10.51) we may write

$$\zeta_n = \lambda \cdot \zeta_{n-1} + \xi_n = ... = \lambda^k \cdot \zeta_{n-k} + \sum_{j=0}^{k-1} \lambda^j \xi_{n-j} \quad \text{for } k \le n . \tag{10.52}$$

Further we have

Theorem III.74

$$M\left[\left(\zeta_n - \sum_{j=0}^{k-1} \lambda^j \xi_{n-j}\right)^2\right] = M\left[\left(\lambda^k \zeta_{n-k}\right)^2\right] = \lambda^{2k} \cdot M\left[\zeta_{n-k}^2\right] . \qquad \square \tag{10.53}$$

$M\left[\zeta_{n-k}^2\right]$ is constant, i.e. independent of n and k, since the process $\overrightarrow{\zeta}$ is stationary.

Moreover, using $|\lambda| < 1$, the right hand side of (10.53) decreases to 0 at a geometric rate. Thus

Theorem III.75

$$\zeta_n = \lim_{k\to\infty} \sum_{j=0}^{k-1} \lambda^j \xi_{n-j} = \sum_{j=0}^{\infty} \lambda^j \xi_{n-j} \quad \textit{in } \boldsymbol{m.s.} \tag{10.54}$$

where ***m.s.*** *means* ***mean square distance*** *("négyzetes közép távolságban") limit.*

Equation (10.54) provides a representation of the original process $\overrightarrow{\zeta}$ as a moving average process.

Since mean square convergence implies convergence of the means and second moments, we have

Theorem III.76

$$M\left(\zeta_n\right) = \lim_{k\to\infty} M\left(\sum_{j=0}^{k-1} \lambda^j \xi_{n-j}\right) = 0 \tag{10.55}$$

and

$$M\left(\zeta_n^2\right) = \frac{\sigma^2}{1-\lambda^2} \ . \quad \square \tag{10.56}$$

Let us compute the covariance between ζ_n and ζ_{n+k} .

Theorem III.77

$$M(\zeta_n \cdot \zeta_{n+k}) = \sigma^2 \cdot \lambda^k \tag{10.57}$$

and so

$$cov\left(\zeta_n, \zeta_{n+k}\right) = \sigma^2 \cdot \left(\lambda^k - \frac{1}{1-\lambda^2}\right) \tag{10.58}$$

for $k \in \mathbb{N}$. $\square$

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* σ^2 *. Then the (stationary) process*

$$\zeta_n = \lambda_1\zeta_1 + \lambda_2\zeta_2 + ... + \lambda_p\zeta_p + \xi_n \tag{10.59}$$

for $|\lambda_i| < 1$ *is called a* ***p 'th order autoregressive process*** *("p -edrendű autoregressziós folyamat").* $\square$

10.9 White noise processes

("fehérzaj folyamatok")

Definition III.79 *The s.p.* $\overrightarrow{\xi} = \{\xi_t : t \in \mathbb{T}\}$ *is a* ***white noise process*** *("fehérzaj folyamat") if the following holds:*
for every finite *subset* $H \subset \mathbb{T}$ *we have that* $\boldsymbol{\xi}_H = \{\xi_t : t \in H\}$ *are standard independent normal (Gaussian) distributions.* □

Claim III.80 *Clearly, by the independency, the* common *density function of* $\boldsymbol{\xi}_H$ *for* $H = \left\{h_1, h_2, ..., h_{|H|}\right\}$ *is*

$$f_{\boldsymbol{\xi}_H}\left(x_{h_1}, x_{h_2}, ..., x_{h_{|H|}}\right) = \frac{1}{\sqrt{(2\pi)^{|H|}}} \cdot \exp\left(\frac{-x_{h_1}^2 - x_{h_2}^2 - ... - x_{h_{|H|}}^2}{2}\right) \qquad (10.60)$$

since the (one dimensional) standard normal density function is
$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}$ *.* □

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

[HM] **Horváth Márton:** *Computer program for Coordinatesystem Transformations*, 2015, https://math.uni-pannon.hu/~szalkai/HorvathM-program/

[KT1] **Karlin,S., Taylor,H.M.:** *A First Course in Stochastic Processes*, Academic Press New York, 1975.
http://labs.xjtudlc.com/labs/wldmt/reading%20list/books/.

[KT2] **Karlin,S., Taylor,H.M.:** *An Introduction to Stochastic Modeling,* Academic Press New York, 1988.
KarlinTaylorIntrodStochModeling.pdf ,
http://matap.dmae.upm.es/WebpersonalBartolo/Probabilidad/

[KT3] **Karlin,S., Taylor,H.M.:** *Sztochasztikus folyamatok* (*in Hungarian*), Gondolat Kiadó, Budapest, 1985.

[P] **Prékopa András:** *Valószínűségelmélet*, (*in Hungarian*) Műszaki Kiadó, 1980.

[SzG-EN] **Székely, G.J.:** *Paradoxes in Probability Theory and Mathematical Statistics*, Reidel, 1986.

[SzG-HU] **Székely J.Gábor:** Paradoxonok a véletlen matematikájában, Budapest: Műszaki Könyvkiadó, 1982, ISBN:963-10-4289-8, Typotex kiadó, 2004, 2010, ISBN: 978-963-2790-89-3

[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

[SzI2] **Szalkai I.:** *Unusual coordinate systems and linearizing methods* (*in Hungarian*), 2015, Haladvány kiadvány,
http://www.math.bme.hu/~hujter/150821.pdf

[W1] **Wikipedia:** *Wiener process*, https://en.wikipedia.org/wiki/Wiener_process

Tables

Standard normal distribution $(\Phi(t))$

t	0	1	2	3	4	5	6	7	8	9
0,0	,5000	,5040	,5080	,5120	,5160	,5199	,5239	,5279	,5319	,5359
0,1	,5398	,5438	,5478	,5517	,5557	,5596	,5636	,5675	,5714	,5754
0,2	,5793	,5832	,5871	,5910	,5948	,5987	,6026	,6064	,6103	,6141
0,3	,6179	,6217	,6255	,6293	,6331	,6368	,6406	,6443	,6480	,6517
0,4	,6554	,6591	,6628	,6664	,6700	,6736	,6772	,6808	,6844	,6879
0,5	,6915	,6950	,6985	,7019	,7054	,7088	,7123	,7157	,7190	,7224
0,6	,7258	,7291	,7324	,7357	,7389	,7422	,7454	,7486	,7518	,7549
0,7	,7580	,7612	,7642	,7673	,7704	,7734	,7764	,7794	,7823	,7852
0,8	,7881	,7910	,7939	,7967	,7996	,8023	,8051	,8078	,8106	,8133
0,9	,8159	,8186	,8212	,8238	,8264	,8289	,8315	,8340	,8365	,8389
1,0	,8413	,8438	,8461	,8485	,8508	,8531	,8554	,8577	,8599	,8621
1,1	,8643	,8665	,8686	,8708	,8729	,8749	,8770	,8790	,8810	,8830
1,2	,8849	,8869	,8888	,8907	,8925	,8944	,8962	,8980	,8997	,9015
1,3	,9032	,9049	,9066	,9082	,9099	,9115	,9131	,9147	,9162	,9177
1,4	,9192	,9207	,9222	,9236	,9251	,9265	,9279	,9292	,9306	,9319
1,5	,9332	,9345	,9357	,9370	,9382	,9394	,9406	,9418	,9429	,9441
1,6	,9452	,9463	,9474	,9484	,9495	,9505	,9515	,9525	,9535	,9545
1,7	,9554	,9564	,9573	,9582	,9591	,9599	,9608	,9616	,9625	,9633
1,8	,9641	,9649	,9656	,9664	,9671	,9678	,9686	,9693	,9699	,9706
1,9	,9713	,9719	,9726	,9732	,9738	,9744	,9750	,9756	,9761	,9767
2,0	,9772	,9778	,9783	,9788	,9793	,9798	,9803	,9808	,9812	,9817
2,1	,9821	,9826	,9830	,9834	,9838	,9872	,9846	,9850	,9854	,9857
2,2	,9861	,9864	,9868	,9871	,9875	,9878	,9881	,9884	,9887	,9890
2,3	,9893	,9896	,9898	,9901	,9904	,9906	,9909	,9911	,9913	,9916
2,4	,9918	,9920	,9922	,9925	,9927	,9929	,9931	,9932	,9934	,9936
2,5	,9938	,9940	,9941	,9943	,9945	,9946	,9948	,9949	,9951	,9952
2,6	,9953	,9955	,9956	,9957	,9959	,9960	,9961	,9962	,9963	,9964
2,7	,9965	,9966	,9967	,9968	,9969	,9970	,9971	,9972	,9973	,9974
2,8	,9974	,9975	,9976	,9977	,9977	,9978	,9979	,9979	,9980	,9981
2,9	,9981	,9982	,9982	,9983	,9984	,9984	,9985	,9985	,9986	,9986
3,0	,9987	,9987	,9987	,9988	,9988	,9989	,9989	,9989	,9990	,9990
3,1	,9990	,9991	,9991	,9991	,9992	,9992	,9992	,9992	,9993	,9993
3,2	,9993	,9993	,9994	,9994	,9994	,9994	,9994	,9995	,9995	,9995
3,3	,9995	,9995	,9995	,9996	,9996	,9996	,9996	,9996	,9996	,9997
3,4	,9997	,9997	,9997	,9997	,9997	,9997	,9997	,9997	,9997	,9998
3,5	,9998	,9998	,9998	,9998	,9998	,9998	,9998	,9998	,9998	,9998
3,6	,9998	,9998	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999
3,7	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999
3,8	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999	,9999
3,9	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0

Critical values for χ^2 distribution

s \ β	0,10	0,05	0,025	0,01	0,001	0,95	0,975
1	2,706	3,841	5,024	6,635	10,827	0,004	0,001
2	4,605	5,991	7,378	9,210	13,815	0,103	0,051
3	6,251	7,815	9,348	11,345	16,268	0,352	0,022
4	7,779	9,488	11,143	13,277	18,465	0,711	0,484
5	9,236	11,070	12,833	15,086	20,517	1,145	0,831
6	10,645	12,592	14,449	16,812	22,457	1,635	1,237
7	12,017	14,067	16,013	18,475	24,322	2,167	1,690
8	13,362	15,507	17,535	20,090	26,125	2,733	2,180
9	14,684	16,919	19,023	21,666	27,877	3,325	2,700
10	15,987	18,307	20,483	23,209	29,588	3,940	3,247
11	17,275	19,675	21,920	24,725	31,264	4,575	3,816
12	18,549	21,026	23,337	26,217	32,909	5,226	4,404
13	19,812	22,362	24,736	27,688	34,528	5,892	5,009
14	21,064	23,685	26,119	29,141	36,123	6,571	5,629
15	22,307	24,996	27,488	30,578	37,697	7,261	6,262
16	23,542	26,296	28,845	32,000	39,252	7,962	6,908
17	24,769	27,590	30,191	33,409	40,790	8,672	7,564
18	25,989	28,869	31,526	34,805	42,312	9,390	8,231
19	27,204	30,144	32,852	36,191	43,820	10,117	8,901
20	28,412	31,410	34,170	37,566	45,315	10,851	9,591
21	29,615	32,671	35,479	38,932	46,797	11,591	10,283
22	30,813	33,924	36,781	40,289	48,268	12,338	10,982
23	32,007	35,172	38,076	41,638	49,728	13,091	11,689
24	33,196	36,415	39,364	42,980	51,179	13,484	12,401
25	34,382	37,652	40,646	44,314	52,620	14,611	13,120
26	35,563	38,885	41,923	45,642	54,052	15,379	13,844
27	36,741	40,113	43,194	46,963	55,476	16,151	14,573
28	37,916	41,337	44,461	48,278	56,893	16,928	15,308
29	39,087	42,557	45,772	49,558	58,302	17,708	16,047
30	40,256	43,773	46,979	50,892	59,703	18,493	16,791

Student *t*- distribution

s \ ε	0,20	0,10	0,05	0,02	0,01	0,001
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,92	4,303	6,965	9,925	31,598
3	1,638	2,353	3,182	4,541	5,841	12,941
4	1,533	2,132	2,776	3,747	4,604	8,61
5	1,476	2,015	2,571	3,365	4,032	6,859
6	1,44	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,405
8	1,397	1,86	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,25	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
11	1,363	1,796	2,201	2,718	3,106	4,437
12	1,356	1,782	2,179	2,681	3,055	4,318
13	1,35	1,771	2,16	2,65	3,012	4,221
14	1,345	1,761	2,145	2,624	2,977	4,14
15	1,341	1,753	2,131	2,602	2,947	4,073
16	1,337	1,746	2,12	2,583	2,921	4,015
17	1,333	1,74	2,11	2,567	2,898	3,965
18	1,33	1,734	2,101	2,552	2,878	3,922
19	1,328	1,729	2,093	2,539	2,861	3,883
20	1,325	1,725	2,086	2,528	2,845	3,85
21	1,323	1,721	2,08	2,518	2,831	3,819
22	1,321	1,717	2,074	2,508	2,819	3,792
23	1,319	1,714	2,069	2,5	2,807	3,767
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,06	2,485	2,787	3,725
26	1,315	1,706	2,056	2,479	2,779	3,707
27	1,314	1,703	2,052	2,473	2,771	3,69
28	1,313	1,701	2,048	2,467	2,763	3,674
29	1,311	1,699	2,045	2,462	2,756	3,659
30	1,31	1,697	2,042	2,457	2,75	3,646
40	1,303	1,684	2,021	2,423	2,704	3,551
60	1,296	1,671	2	2,39	2,66	3,46
120	1,289	1,658	1,98	2,358	2,617	3,373
∞	1,282	1,645	1,96	2,326	2,576	3,291

Student t- test $P(X_f < a) = p$

$f \backslash p$	0,90	0,95	0,975	0,98	0,99	0,995
1	3,08	6,31	12,71	15,89	31,82	63,66
2	1,89	2,92	4,30	4,85	6,96	9,92
3	1,64	2,35	3,18	3,48	4,54	5,84
4	1,53	2,13	2,78	3,00	3,75	4,60
5	1,48	2,02	2,57	2,76	3,36	4,03
6	1,44	1,94	2,45	2,61	3,14	3,71
7	1,41	1,89	2,36	2,52	3,00	3,50
8	1,40	1,86	2,31	2,45	2,90	3,36
9	1,38	1,83	2,26	2,40	2,82	3,25
10	1,37	1,81	2,23	2,36	2,76	3,17
11	1,36	1,80	2,20	2,33	2,72	3,11
12	1,36	1,78	2,18	2,30	2,68	3,05
13	1,35	1,77	2,16	2,28	2,65	3,01
14	1,35	1,76	2,14	2,26	2,62	2,98
15	1,34	1,75	2,13	2,25	2,60	2,95
16	1,34	1,75	2,12	2,24	2,58	2,92
17	1,33	1,74	2,11	2,22	2,57	2,90
18	1,33	1,73	2,10	2,21	2,55	2,88
19	1,33	1,73	2,09	2,20	2,54	2,86
20	1,33	1,72	2,09	2,20	2,53	2,85
21	1,32	1,72	2,08	2,19	2,52	2,83
22	1,32	1,72	2,07	2,18	2,51	2,82
23	1,32	1,71	2,07	2,18	2,50	2,81
24	1,32	1,71	2,06	2,17	2,49	2,80
25	1,32	1,71	2,06	2,17	2,49	2,79
26	1,31	1,71	2,06	2,16	2,48	2,78
27	1,31	1,70	2,05	2,16	2,47	2,77
28	1,31	1,70	2,05	2,15	2,47	2,76
29	1,31	1,70	2,05	2,15	2,46	2,76
30	1,31	1,70	2,04	2,15	2,46	2,75
35	1,31	1,69	2,03	2,13	2,44	2,72
40	1,30	1,68	2,02	2,12	2,42	2,70
45	1,30	1,68	2,01	2,12	2,41	2,69
50	1,30	1,68	2,01	2,11	2,40	2,68
60	1,30	1,67	2,00	2,10	2,39	2,66
70	1,29	1,67	1,99	2,09	2,38	2,65
80	1,29	1,66	1,99	2,09	2,37	2,64
90	1,29	1,66	1,99	2,08	2,37	2,63
100	1,29	1,66	1,98	2,08	2,36	2,63
200	1,29	1,65	1,97	2,07	2,35	2,60

Critical values of F -distribution for $\varepsilon = 0{,}05$ (95%)
(degrees of freedom for the enumerator)

(degrees of freedom for the denumerator)

	1	2	3	4	5	6	7	8	9	12
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	243,9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,84	8,81	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,83	2,77	2,71	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,50	2,54	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,38
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,34
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,31
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,28
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,25
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,23
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,38	2,32	2,20
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,18
25	4,24	3,38	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,16
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,15
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,13
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,12
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,10
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,09
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,00
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,92
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,83
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,75

Index