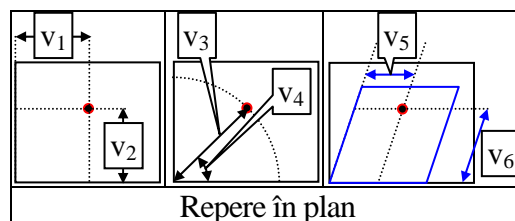
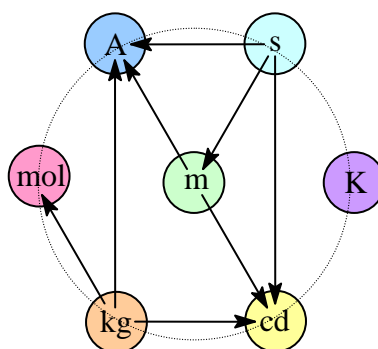
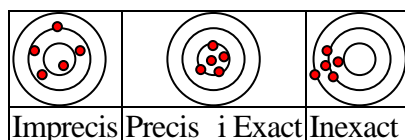


Lorentz JÄNTSCHI

Prezentarea și prelucrarea datelor experimentale



U.T.Press
2013

ISBN 978-973-662-912-9

Prezentarea și prelucrarea datelor experimentale

Mrimi și măsurarea lor	2
Mrimi	
Procesul de măsurare a unei mrimi	
Valoarea a unei mrimi. Unitate de măsură	
Sisteme de mrimi și sisteme de unități de măsură	9
Cantitatea de substanță	
Analiza dimensională	
Sisteme de unități de măsură ; sistemul internațional	
Evaluarea numerică a expresiilor matematice	15
Ecuații algebrice și transcendentale	
Rezolvarea ecuațiilor polinomiale	
Rezolvarea numerică a ecuațiilor transcendentale	
Rezolvarea sistemelor de ecuații liniare	
Metode iterative	
Sisteme de ecuații neliniare	
Integrarea numerică	
Regula 1/3 a lui Simpson de integrare numerică	
Elemente de teoria probabilităților	24
Măsurători statistice pentru populații finite	
Distribuții statistice din măsurători repetate	
Legi de distribuție statistice ale acestora	
Acordul între observație și model: statistici	
Minimizarea erorii de acord	
Utilizarea momentelor	
Utilizarea momentelor centrale	
Folosind statisticile referitoare la populație	
Estimare pe baza anseii maxime	
Statistica Benford	
Statistica Jarque-Bera	
Statistica Kolmogorov-Smirnov	
Distribuția Kolmogorov	
Testul Kolmogorov-Smirnov	
Statistica Anderson-Darling	
Statistica Pearson-Fisher Chi Square	
Testul χ^2 independență, omogenitate și asociere	
Probleme frecvente în aplicarea testului χ^2	
Măsurători ale tendinței centrale	
Binary and multinomial variables analysis: binomial confidence intervals	38
(binomial design and model)	
(binomial confidence intervals)	
(contingency of binomials)	
(multinomial distribution)	
Ordinal variables analysis: ranks statistic	43
(order statistics)	
(Fisher-Yates correlation coefficient)	
(generalized correlation coefficient)	
(Jackknife method)	
(Bootstrap method)	
(tied values and fractional ranking)	
(Spearman ρ)	
(Kendall τ)	
(Goodman and Kruskal's γ)	
(Hoeffding D)	
(other statistics)	
(ordered contingencies)	
(polychoric correlation)	
Linear regression analysis: linear models	51
Aplicații	55
Referințe	66

Mărimi și măsurarea lor

Mărimi

O **mărimă** este rezultatul unei măsurători efectuate asupra unei observabile cu scopul de a colecta valoarea unei proprietăți.

Se poate imagina **spațiul de observare** ca având o structură de arbore (a se vedea *Structura spațiului de observare*) care exprimă relațiile de apartenență dintre observabile în care la bază se află Universul (ca întreg spațiul de observare) iar la suprafață (aproape de noi în calitate de observatori) se află compușii chimici - ca formă de reprezentare a materiei cu compoziție (de atomi) și relații (între acestea) bine definite.

Structur	Proprietate
- Univers	Întreg spațiul de observare
- Energie Radiant	Viteza comparabilă cu cea a luminii
+ Radiații ca	Diferențiate prin intermediul proprietăților
- Materie	Întreg spațiul de observare nerelativistic
- Corp	Viteza mult mai mică decât a luminii
- Ansamblu de materiale	Compoziție (chimică) variabilă și discontinuă
- Materiale	Compoziție (chimică) variabilă și continuă
- Amestec de substanțe	Compoziție (chimică) bine definită
+ Substanță eterogenă	Compoziție (chimică) variabilă
- Soluție	Stare de agregare bine definită
+ Aliaj	Amestec de metale în stare lichidă sau solidă
- Substanță omogenă	Compoziție (chimică) constantă
+ Compus chimic	Structură chimică bine definită și unică

Fig. 1. Structura spațiului de observare

Procesul de măsurare a unei mărimi

Procesul de observare este o activitate de colectare a cunoștințelor cu ajutorul simțurilor sau instrumentelor. Se presupune existența unui observator și a unei observabile. Procesul de observare transferă o formă abstractă a cunoașterii de la observabil la observator (ca de exemplu, sub formă de numere sau imagini).

Măsurarea cuprinde două operațiuni serializate: observarea și înregistrarea rezultatelor observației. Măsurarea depinde de natura obiectului observat (material) sau fenomen (imaterial), de metoda de măsurare și de modul de înregistrare a rezultatelor observației.

Măsurarea presupune identificarea anterioară a elementului sau a elementelor care fac obiectul investigației și rezultatul măsurării este o proprietate a elementului observat. O serie de măsurători presupune existența unei colecții de elemente distincte - **mulțime** - în care ordinea poate să nu fie relevantă. Mulțimea vid (\emptyset) este mulțimea cu nici un element în ea.

Proprietatea (ca urmare a unei serii de observații) înregistrată cu exact una din exact două valori numite nefavorabil (și scrisă ca F sau 0) și favorabil (și scrise ca T sau 1), respectiv, dă o **valoare de adevăr**. Mulțimea de valori de adevăr ($\{0,1\}$ sau $\{T, F\}$) este o mulțime în care elementele sunt ordonate în mod convențional ($0 < 1$, $F < T$). Negația logică (!) este operațiunea (informațională) ce schimbă valoarea de adevăr, în timp ce identitatea logică () lasă valoarea de adevăr neschimbată și se exprimă faptul că rezultatul unei operații de măsurare pe două elemente este același. Folosind proprietatea 'valoare de adevăr' pe elementele unei mulțimi conceptul de **submulțime** este raționalizat. Apartenența este o proprietate a unui element de a fi (\in) sau a nu fi (\notin) într-o mulțime.

Asocierea totală scrisă ca $S1 \times S2$ și definită prin $S1 \times S2 = \{(e1, e2) \mid E1 \in S1, S2 \in e2\}$ este produsul cartezian al mulțimilor $S1$ și $S2$, iar o submulțime a $S1 \times S2$ este numită relație binară. Dacă $S1=S2$ relațiile sunt numite endo-relații. Următorul tabel redă (endo)relații (binare) cu proprietăți speciale (a se vedea *Relații binare: proprietăți și reprezentanți*).

Id	Nume	Defini ie	Reprezentan i
RE	Reflexive	$\forall a: (a,a) \in RE$	$=, \subseteq, ,$
CR	Co-reflexive	$\forall a,b: (a,b) \in CR \text{ then } a \sim b$	$=$
QR	Cvasi-reflexive	$(a,b) \in QR \text{ atunci } (a,a), (b,b) \in QR$	lim
IR	Ireflexive	$\forall a: (a,a) \notin IR$	$, \perp, <$
SY	Simetrice	$(a,b) \in SY \text{ atunci } (b,a) \in SY$	$=, CD, CM$
NS	Anti-simetrice	$(a,b), (b,a) \in NS \text{ atunci } a \sim b$	
AS	Asimetrice	$(a,b) \in AS \text{ atunci } (b,a) \notin AS$	IH, $<$
TS	Tranzitive	$(a,b), (b,c) \in TS \text{ atunci } (a,c) \in TS$	$=, <, \subseteq, , \Rightarrow, IH$
TL	Totale	$\forall a,b: (a,b) \in TL \text{ sau } (b,a) \in TL$	
TC	Tri-hotome	exact una dintre $(a,b) \in TL, (b,a) \in TL, a \sim b$	$<$
ED	Euclidiene	$(a,b), (a,c) \in ED \text{ atunci } (b,c) \in ED$	$=$
SE	Seriale	$\exists b : (a,b) \in SE$	
UQ	Unicitate	$(a,b), (a,c) \in UQ \text{ atunci } b \sim c$	$f(\cdot)$
EQ	Echivalen	atunci RE, SY, TS	$=, \sim, \subseteq, CM, CD, $
PO	Ordine par ial	atunci RE, NS, TS	$ $
TO	Ordine total	atunci PO, TL	Alfabet,
WO	Bine ordonate	atunci TO, SE	
\perp	Co-prime	cel mai mare divizor comun este 1	
VT	Adev rul vacuos	`daca A atunci B` cÂnd A = Fals	
=	Egal	atunci RE, CR, SY, NS, TS, ED, EQ	
	Mai mic sau egal	atunci RE, NS, TS, TL, SE, PO, TO	
<	Mai mic	atunci IR, NS, AS, TS, TC, SE	
\subseteq	Submul ime	RE, NS, TS, SE, PO	
	Diferit	IR, SI	
DI	Distan euclidian	RE, SI, TS, ED, SE, EQ	
IH	Mo tenere	AS, TS	
CM	Congruen modulo n	EQ	
CD	Congruen div n	EQ	
lim	Limita unei serii	RE, QR	
$f(\cdot)$	Func ie matematic	SE, UQ	
inj	Func ie injectiv	$a \sim b \text{ atunci } f(a) \sim f(b)$	
srj	Func ie surjectiv	$\exists x : b=f(a)$	
bij	Func ie bijectiv	INJ, SRJ	

Tab. 1. Rela ii binare: propriet i i reprezentan i

Similaritatea Între conceptul de *func ie matematică* i *func ia de măsurare* este evident cÂnd analiz m propriet ile rela iilor care se stabilesc Între mul imea observabilelor i mul imea valorilor asociate din spa iul informa ional. Ca i În cazul func iilor matematice, atunci cÂnd sunt efectuate m sur tori experimentale sunt asigurate dou propriet i Între elemente observate i propriet ile lor Înregistrate. i anume, pentru toate elementele observate avem Înregistr ri ale propriet ilor lor atunci cÂnd facem m sur tori - fiind asigurat *serializarea* (SE, v. *Rela ii binare: proprietă i și reprezentan i*). O m sur ne ofer (Într-un anumit moment de timp i spa iu), o pies informa ional (o Înregistrare) i unicitatea (UQ, v. *Rela ii binare: proprietă i și reprezentan i*) fiind asigurat de asemenea. Nici o alt proprietate cunoscut (matematic) a rela iilor nu este, În general, valabil pentru func ii matematice i nici pentru func ia de m surare, a a Încât putem spune c ceea func ia de m surare face prin intermediul informa iilor este expresia unei func ii matematice (vezi *Colectarea datelor experimentale este o func ie matematică*).

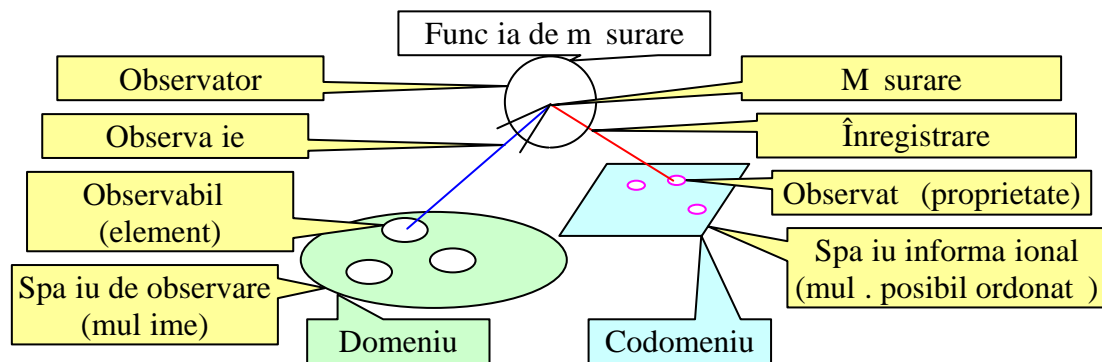


Fig. 2. Colectarea datelor experimentale este o funcție matematică

Pentru o mulțime finită S o **funcție de numerotare** poate fi definită iterativ după cum urmează : $S_0=S$; $S_1=S \setminus \{s_1\}$; ... ; $S_i=S \setminus \{s_i\}$; ... ; (etc.). Funcția $f(i)=s_i$ este o funcție de numerotare pe mulțimea S , și arată că orice mulțime finită este numărabilă. Alegerea elementelor s_1, \dots, s_i, \dots din mulțimea S este instrumentul specific de măsurare. Este implicit observația, înregistrarea, și are ca efect construirea unei submulțimi reunind elementele rămase.

S-a arătat mai sus că conceptul de funcție matematică este legat de conceptul de măsurare. Mai departe, funcția de numerotare este instrumentul specific, cu care se face ordonarea în spațiul informațional. Mai mult, în cazul în care mulțimea S conține n elemente (desigur, ar trebui numărate mai întâi), atunci există exact $n!$ posibilități de a enumera elementele sale prin intermediul funcției de numerotare. În afară de numerotarea implicită, funcția de măsurare aduce în spațiul informațional valoarea unei proprietăți observate.

Pentru două (presupus) finite mulțimi A (spațiul nostru de observare) și B (spațiul nostru informațional) sunt exact $|B|^{|A|}$ posibilități de a defini (construi) funcții matematice $f:A \rightarrow B$ (și ne amintim, posibilități de măsurare) care asociază elementele din A cu elemente din B .

Pentru o observație cu 0 și 1 ($|B|=2$) asupra unei mulțimi cu n elemente ($|A|=n$) avem un rezultat al numărării ($|A|=n$), un rezultat al posibilităților de enumerare ($|A|=n!$) și un rezultat al posibilităților de observare ($|B|^{|A|}=2^n$). Se poate verifica imediat că $n < n!$. Pentru $n > 3$ și mai mult, $n < 2^n < n!$ pentru $n > 4$. Chiar mai mult decât atât, pentru n $n \ll 2^n \ll n!$, adică $\lim_n (n/2^n) = \lim_n (2^n/n!) = 0$.

Dacă o observație cu 0 și 1 este 'cel mai simplu tip de observație' atunci o observație ce înregistrează în spațiul informațional numere reale este cel mai complex tip de observație.

Presupunând că rezultatul observației este un număr real, putem folosi o pereche formată dintr-un bit (0 sau 1) consemnând semnul și un număr real pozitiv pentru a echivala conținutul din spațiul informațional (numărul real cu semn). Mai mult, se poate construi o funcție matematică bijectivă (care aduce o corespondență 1:1) între orice număr real pozitiv $[0, \infty)$ și un număr real din intervalul $[0,1)$: $f:[0,1) \rightarrow [0, \infty)$, $f(x)=1+1/(1-x)$. Verificarea că este o funcție bijectivă pentru domeniul de definiție se poate face verificând că $f(x)=1/(x-1)^2 > 0$. Rezultă deci că o codificare formată dintr-un semn (un bit) și o succesiune de 0 și 1 (reprezentarea în baza 2 a oricărui număr real subunitar) reflectă în totalitate orice număr real. Trecând însă din nou la limită dimensiunea spațiului observațional ($n \rightarrow \infty$) "puterea" reprezentării prin numere reale ($f:A \rightarrow \mathbb{R}$) este de aceeași cardinalitate cu cea a reprezentării cu numere întregi ($f:A \rightarrow \mathbb{Z}$) sau binar ($f:A \rightarrow \{0,1\}$), 2^0 , unde 0 este cardinalitatea mulțimii numerelor naturale. Acest simplu fapt ne arată că chiar dacă se măsoară calitatea reprezentării prin numere reale, rezoluția reprezentării este în continuare insuficientă pentru a egala calitatea enumerării ($0!$).

Valoarea unei mrimi. Valoarea numerică. Unitate de măsur

O primă consecință imediată a calității reprezentării din spațiul informațional este existența degenerării. Degenerarea este reprezentarea prin intermediul aceleiași valori a rezultatului observației asupra a două elemente distincte (diferite).

Această degenerare este uneori un avantaj (când se pun în evidență similitudinile între proprietățile a două elemente) alteori un dezavantaj (când măsurarea care a avut ca scop evidențierea diferențelor între cele două elemente a eșuat în a-și atinge scopul).

O a doua consecință imediată a calității reprezentării din spațiul informațional este că dacă degenerarea nu poate fi evitată prin funcția de măsurare, încă poate fi diminuată prin scala de măsurare. Ar trebui remarcat faptul că nu toate scalele de măsurare induc relații de ordine în spațiul informațional. Exemple naturale sunt grupa de sânge și aminoacizii care constituie codul genetic, și anume sunt situații când codificarea din spațiul informațional nu exprimă o relație de ordine (naturale) între valorile măsurate.

Fie o mulțime cu două elemente ($C=\{a,b\}$) și formulăm ipoteza că ordinea nu este relevantă între ele. Mulțimea submulțimilor lui C este $S_C=\{\{\},\{a\},\{b\},\{a,b\}\}$. Un ordin natural în mulțimea S_C este definit prin cardinalitatea submulțimii. Cardinalitatea ca relație de ordine nu este strictă, pentru că există două submulțimi cu același număr de elemente: $0=|\{\}\|<|\{a\}|=1=|\{b\}\|<|\{a,b\}|=2$. S-ar putea întreba: "Ce tip de scală de măsurare definește cardinalitatea?" - Pentru a oferi un răspuns util trebuie să ne întoarcem la măsurare și noi ar trebui să întrebăm mai întâi: "Ce caracteristici se doresc a fi evaluate?". În cazul în care răspunsul la a doua întrebare este numărul de elemente în subgrupul observat, atunci cardinalitatea este bine definită a fi **cantitativă** - fiind dotată cu o relație de ordine. În cazul în care diferențierea între submulțimile lui C este scopul dorit, atunci cardinalitatea submulțimii nu este suficientă. S-ar putea construi în continuare un alt experiment menit să diferențieze submulțimile pentru care apare degenerarea (în cazul de mai sus pentru submulțimile cu un element) și o nouă funcție de măsurare ar da răspunsul la întrebarea "Submulțimea conține elementul 'a'?" (complementar cu răspunsul la întrebarea "Submulțimea conține elementul b?"). Aceasta este o măsurătoare tipic **calitativă** - cum s-a potrivit.

O altă consecință derivată din căutarea după submulțimile unei mulțimi este că **scala de măsură** care se intenționează să se aplicare trebuie să fie de cel puțin verificată din punct de vedere al consistenței cu scopul propus.

Mai mult, chiar și atunci când nu există relații de ordine, pot exista alte relații (cum ar fi complementul logic $\{a\}=\{a,b\}\setminus\{b\}$ în mulțimea submulțimilor mulțimii $\{a,b\}$), care aduce în spațiul informațional faptul că nu întotdeauna rezultatele măsurătorilor sunt independente unul față de altul.

Mergând mai departe, tabelul de mai jos clasifică după complexitate (definit de către operațiile permise între valorile înregistrate) scalele de măsurare (a se vedea *Scale de măsură*).

Scal	Tip	Operații	Structur	Statistici	Exemple
Binomial	Logic	"=", "!"	Algebră Booleană [¹]	Moda, Fisher Exact [²]	Dead/Alive Fe ele unei monezi
(multi) Nomi(n)al	Discret	"="	Mulțime standard	Moda, Chi squared	Sistemul de grupe de sânge ABO Clasificarea organismelor vii
Ordinal	Discret	"=", "<"	Algebră comutativă	Mediana, Rangul	Numărul de atomi în molecule
Interval	Continuu	" ", "-"	Spațiu afin (uni-dimensional)	Media, StDev, Corelația, Regresia, ANOVA	Scala de temperatură
Raport	Continuu	" ", "-", "*"	Spațiu vectorial (uni-dimensional)	GeoMean, HarMean, CV, Logaritmi	Dulcea relativă la sucroză pH Scala distanțelor Time scale Energy scale

Tab.2. Scale de măsurare

O scală de măsurare este **nominală** dacă între valorile sale o relație de ordine nu poate fi definită. De obicei scala nominală de măsurare este destinată să fie utilizată pentru măsurători calitative. Scala **binară** (sau binomial) este cu doar două valori posibile (între care există o relație de ordine),

cum ar fi: {Da,Nu}, {Viu, Mort},{Vivo,Vitro}, {prezent, absent}, {alcan saturat, alt tip de compus}, {num r întreg, num r neîntreg}. Scala nominal cu mai mult de două valori posibile este numită multinomial. Scala multinomial de m surare are un număr finit de valori posibile și independent de numărul lor, operează relația de complementaritate. Astfel, pentru {0,A,B,AB} grupe sanguine o valoare diferită de oricare dintre cele trei, sigur este cea de a patra. O serie finită de valori poate fi considerată o scală **ordinală** dacă între valorile lor posibile se poate defini o relație de ordine (natural). Dacă presupunem că "Absent"<"Prezent", "Fals"<"Adevărat", "0"<"1", "Negativ"<"Nenegativ", "Nepozitiv"<"Pozitiv", atunci toate aceste scale de măsurare sunt ordinale. Mai mult, un exemplu de scală ordinală cu trei valori este: "Negativ"<"Zero"<"Pozitiv". Un alt lucru important cu privire la scalele ordinale este că nu sunt necesare, cu o cardinalitate finită. Dar este necesară existența unei relații de ordine definite prin "Succesorul unui element" (al unei valori) și complementul acesteia "Predecesorul unui element" (al unei valori). În scala **interval** distanța (sau diferența) între valorile posibile are un sens. De exemplu diferența între 30° și 40° pe scala de temperatură are aceeași semnificație cu diferența între 70° și 80°. Intervalul între două valori este interpretabil (are un sens fizic). Acesta este motivul pentru care are sens calcularea valorii medii a unei variabile de tip interval, ceea ce însă nu are sens pentru valorile unei scale ordinale. În același timp (vezi *Termometrul cu mercur și scale de temperatură*) cum ar fi 80° nu este de două ori mai fierbinte decât 40° (așa cum 2m sunt de două ori mai mulți decât 1m), pentru scalele interval raportul dintre două valori nu are nici un sens.

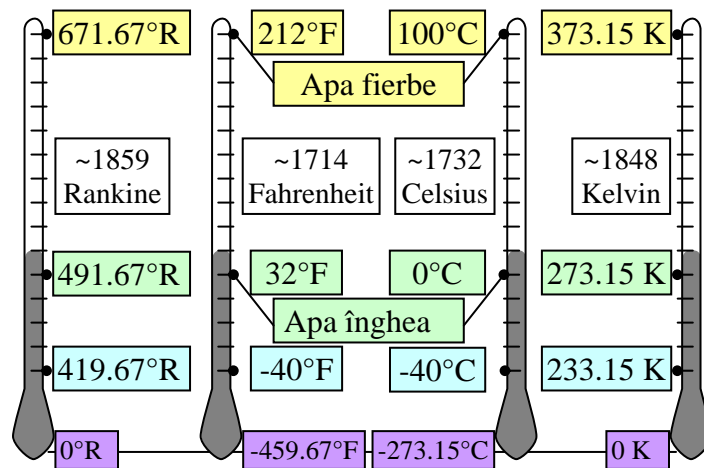
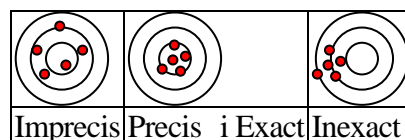


Fig. 3. Termometrul cu mercur și scale de temperatură

În cele din urmă, pe scalele de tip **raport**, valorile 0 și/sau 1 au întotdeauna o semnificație. Ipoteza este că cea mai mică valoare observabilă este 0. Rezultă prin urmare faptul că dacă două valori sunt luate pe o scală raport, putem calcula raportul lor, și, de asemenea, această măsurare posedă o scală de măsurare de tip raport.

În cele din urmă, trebuie să remarcăm că măsurimile măsurate pe o scală raport sunt **aditive**, în timp ce măsurimile măsurate pe celelalte scale sunt **neaditive**.

Trebuie menționat faptul că scala de măsurare nu dă precizia de măsurare, și din acest punct de vedere este ilustrativ exemplul de mai jos (v. *Precizie și exactitate*).



Tab. 3. Precizie și exactitate

De scala pe care a fost măsurată proprietatea depinde modul în care datele se pot prelucra și interpreta. Așa cum s-a ilustrat (v. *Precizie și exactitate*) **precizia** și **exactitatea** unei măsurători sunt la fel de importante ca și valoarea măsurată în sine. Acesta este motivul pentru care se obișnuiește să se exprime valoarea unei măsurători împreună cu precizia sa de măsurare.

Desigur că există mai multe modalități de a exprima precizia unei măsurători. De exemplu, o măsurătoare cu un aparat de măsurare nu poate depăși precizia pentru care aparatul a fost construit.

Din punctul de vedere al scalei de măsurare, o variabilă (din spațiul informațional) care numără moleculele dintr-un spațiu (fizic) dat este "la fel" de tip raport ca și o variabilă ce măsoară temperatura mediului (fizic) în care aceste molecule sunt localizate, chiar dacă rezultatul acestor două operații de măsurare nu are aceeași precizie sau precizie comparabilă (sau nu pare a avea).

Este de dorit în mod evident ca scala de măsurare să încorporeze cât mai multe caracteristici ale variabilei măsurate, precum și să posede numai caracteristicile găsite în variabilă măsurată, pentru că, în caz contrar, scala de măsurare devine o sursă de eroare.

Pe de altă parte, există limite. Pentru a arăta aceasta, este suficient să facem apel la o serie de probleme nerezolvate în fizică [3]:

- ÷ Este 'spațiu-timp-ul' fundamental continuu sau discret?
- ÷ Există în natură mai mult de 4 dimensiuni 'spațiu-timp'?
- ÷ Sunt motive (fizice) să ne așteptăm la existența altor universuri care să fie fundamental neobservabile?
- ÷ Prin ce diferă spațiul de timp?
- ÷ Există particule purtătoare de sarcină magnetică cum sunt electronii pentru sarcina electrică?
- ÷ Care este cel mai greu posibil nucleu atomic stabil sau instabil?

Simpla enumerare de mai sus ne arată că noi trebuie să operăm cu incertitudini.

În acest sens, este extrem de util să ne amintim de existența **principiului incertitudinii** [4]: Principiul lui Heisenberg al incertitudinii stabilește prin inegalități (precise) că anumite perechi de proprietăți (fizice) cum sunt poziția și momentul nu pot fi simultan cunoscute (măsurate) cu o precizie mare arbitrară. Cu cât mai precis o proprietate este măsurată, cu atât mai puțin precis cea de-a doua proprietate poate fi furnizată de un experiment de măsurare. Principiul incertitudinii stabilește că minimul pentru produsul incertitudinilor celor două proprietăți este egal cu jumătate din constanta lui Planck redusă ($\hbar = h/2$).

În sensul celor de mai sus, este perfect justificat să se definească starea unei observabile prin intermediul unei **funcții de undă** având ca domeniu un spațiu-timp real iar ca codomeniu o coordonată complexă a cărei amplitudine semnifică probabilitatea unei configurații a sistemului.

Într-adevăr, în 1926 Schrödinger [5] formulează ecuația ondulatorie a mecanicii cuantice a cărei soluție este o funcție de probabilitate (Ecuația lui Schrödinger): $i\hbar\partial\Psi/\partial t = \hat{H}\Psi$, unde Ψ este funcia de undă și d amplitudinea probabilității pentru diferite configurații ale sistemului la diferite momente de timp ($|\Psi(x,y,z,t)|^2$ este densitatea de probabilitate de a găsi particula la coordonata (x,y,z) și momentul de timp t); i/\hbar este operatorul energiei; i este unitatea imaginară ($i^2 = -1$); \hbar este constanta lui Planck redusă ($\hbar = h/2$); $h = 6.62606 \cdot 10^{-34}$ J·s; \hat{H} este operatorul Hamilton ($\hat{H} = -\hbar^2\nabla^2/2m$); ∇^2 : operatorul Laplace ($\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$).

Este dificil de înțeles acest lucru pentru o stare fizică, ceea ce a făcut ca autorul să explice plastic acest fapt într-o corespondență cu un coleg. Exemplul a devenit faimos și a rămas sub numele de "pisica lui Schrödinger" (v. *Pisica lui Schrödinger*).

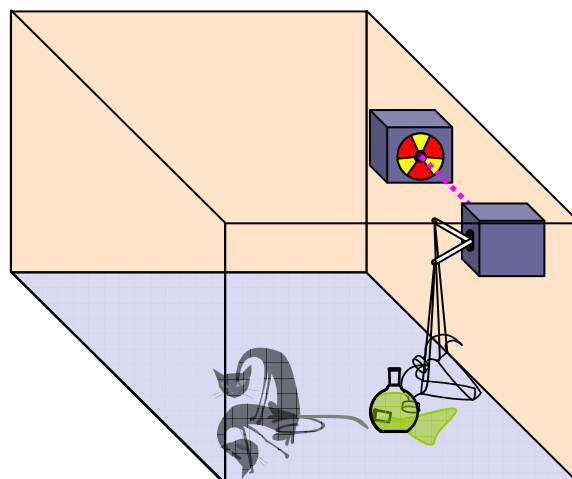


Fig. 4. Pisica lui Schrödinger

În esență, ceea ce experimentul denumit "pisica lui Schrödinger" exprimă este că odată supus un sistem unei protecții de decoerență cuantică este imposibil să se evalueze starea sistemului fără ca simultan cu evaluarea stării acestuia să se strice înseși starea de decoerență în care se află.

Structura spațiului observabilă a cum a fost ea prezentată (v. *Structura spațiului de observare*) descrie doar structura materialelor până la nivelul de compus chimic, întrucât acesta nu este ultimul nivel de structurare intrinsecă. La rândul său, compusul chimic posedă o structură care este alcătuită din atomi. Nici atomul nu este ultimul nivel de structură, fiind la rândul său alcătuit din nucleoni și electroni.

Privind problema din alt unghi, următorul nivel de rafinament (celui ilustrat de *Structura spațiului de observare*) sunt compușii chimici (v. *Nivele de rafinament ale conceptului de compus chimic*) definiți în sensul unei structuri chimice definite și unice. Raționalizarea structurii chimice se face prin intermediul formulelor chimice. În acest sens, formele de reprezentare ale structurii chimice se pot desfășura în continuare astfel:

Structură	Proprietate
Compus chimic	Structură moleculară definită și unică
Formulă brută	Numărul de atomi din fiecare element în raport cu unul dintre elemente
Formulă moleculară	Numărul de atomi ai fiecărui element cuprinși într-o moleculă
Formulă rațională	Exprime grupele structurale din moleculă
Formulă geometrică	Exprime geometria moleculei

Tab. 4. Nivele de rafinament ale conceptului de compus chimic

Nici măcar ultimul nivel de rafinament (formula geometrică) nu este întotdeauna suficient pentru a reda fidel structura moleculară. În acest sens, un exemplu simplu în care cunoscând distanțele între atomi și unghiurile pe care legăturile între acestea le formează nu este suficient pentru a accepta că referim o structură moleculară definită și unică este butanul și anume conformerii acestuia "Gauche g-" și "Gauche g+" care au proprietatea de a răsuci diferit lumina polarizată (v. *Conformerii butanului*).

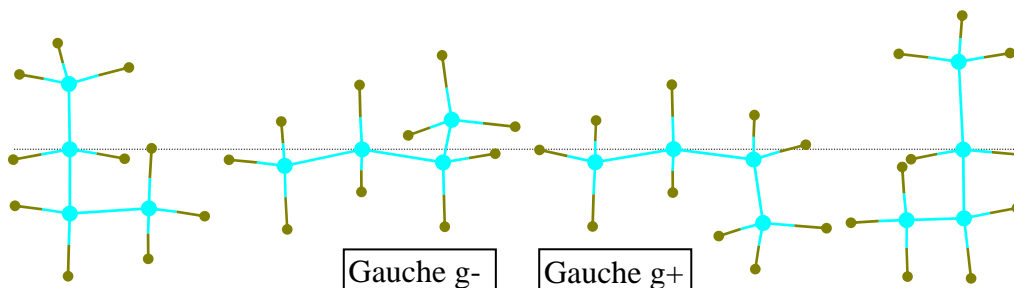


Fig. 5. Conformerii butanului

Sisteme de mrimi i sisteme de unit i de m sur

Cantitatea de substan

Indiferent de nivelul de structur (molecular , atomic , subatomic) la care ne referim num rul de particule (compu i chimici, atomi, electroni) la nivel microscopic (observabil cu ochiul liber sau cu instrumente de mrire) cuprins într-un spa iu de volum definit este imens. Din acest motiv pentru a face referire la macrocantit i este nevoie de o unitate de m sur corespunz toare. Aceasta este molul.

Molul este cantitatea de particule (molecule, ioni, atomi, electroni, altele asemenea sau grupuri ale acestora) al c ror tip trebuie specificat i al c ror num r este egal cu num rul de atomi de carbon existen i în 0.012 kg (12g) din izotopul acestuia ^{12}C .

Astfel, cantitatea de particule (impropriu spus "cantitate de substan ") se poate exprima prin intermediul num rului de particule (N) sau prin intermediul num rului de moli (n) iar între aceste dou modalit i de exprimare exist rela ia:

$$n = N/N_A$$

în care N_A este num rul lui Avogadro i exprim valoarea aproximativ a num rului de atomi de carbon existen i în 0.012 kg (12g) din izotopul acestuia ^{12}C : $N_A = 6.02214 \cdot 10^{23} \text{ mol}^{-1}$.

Prin intermediul cantit ii de substan o serie de propriet i observate au caracter intensiv i extensiv:

$$X_m = \frac{X}{n}$$

în care X nominalizeaz oricare *proprietate extensivă* (care depinde de cantitatea de substan) iar X_m nominalizeaz *proprietatea intensivă* corespondent (care nu mai depinde de cantitatea de substan).

Energia ca atribut al unei substan e i consecin a structurii sale atomice, moleculare sau agregate este o mrim extensiv în timp ce energia specific este corespondentul intensiv al energiei. Similar, energia liber - eliberat sau absorbit într-un proces este o mrim extensiv în timp ce poten ialul chimic este mrim extensiv asociat . Capacitatea caloric este cantitatea de c ldur ce produce schimbarea temperaturii cu 1K i este o mrim extensiv , i capacitatea caloric specific este mrim extensiv asociat .

Masa este proprietate extensiv (M), iar masa molar (M_m) este proprietate intensiv . Volumul (V) este o proprietate extensiv în timp ce volumul molar (V_m) este o proprietate intensiv . Concentra ia (molar , molal , procentual) este o mrim extensiv :

$$n = \frac{M}{M_m} = \frac{V}{V_m}, c_M = \frac{n}{V_s}, c_m = \frac{n}{m_s}, c_{\%m} = \frac{m_d}{m_s}, c_{\%v} = \frac{V_d}{V_s}$$

Se poate remarca c concentra ia molar variaz cu temperatura, deoarece volumul variaz cu temperatura, în timp ce molalitatea este o mrim independent de temperatur . Se nume te o solu ie diluat , o solu ie ce con ine cel mult $10^{-2} \text{ mol l}^{-1}$ de solut. În solu iile diluate ionii de solut sunt separa i de cel pu in 10 molecule de solvent. O alt mrim frecvent utilizat la amestecuri este frac ia molar x_j (a componentului j) din amestecul cu J ($j \in J$) componen i:

$$x_j = \frac{n_j}{\sum_j n_j}$$

Se poate demonstra c frac ia molar este o mrim intensiv . Astfel, fie un amestec P cu compozi ia exprimat prin raportul num rului de molecule din fiecare component j în amestec $1: 2: \dots : J$ (cum ar fi pentru $\text{C}_2\text{O}_4\text{H}_2$, $1: 2: 3 = 2:4:2 = 1:2:1$), i num rul de moli n.

Din cele $N = n \cdot N_A$ molecule ale amestecului, pentru a respecta propor ia, num rul de molecule din componentul j este $N_j = N \cdot \frac{1}{j}$. Frac ia molar a amestecului este:

$$x_j = \frac{n_j}{\sum_j n_j} = \frac{N_j}{N_A} \bigg/ \sum_j \frac{N_j}{N_A} = \frac{N_j}{\sum_j N_j} = \frac{N \cdot \alpha_j / \sum_j \alpha_j}{\sum_j N \cdot \alpha_j / \sum_j \alpha_j} = \frac{N \cdot \alpha_j}{\sum_j N \cdot \alpha_j} = \frac{\alpha_j}{\sum_j \alpha_j}$$

Expresia rezultat nu depinde decât de compozi ie i nu depinde de num rul de moli sau molecule implicate a a c este o mrim intensiv .

Densitatea este o mrim intensiv . În cazul unui amestec cu J componen i:

$$\rho = \frac{\sum_j m_j}{\sum_j V_j} = \frac{\sum_j n_j M_j}{\sum_j V_j} = \frac{\sum_j n \cdot x_j M_j}{\sum_j V_j} = \frac{n \cdot \sum_j x_j M_j}{\sum_j V_j} = \frac{\sum_j x_j M_j}{\sum_j V_j / n} = \frac{\sum_j x_j M_j}{V_m}$$

În formula de mai sus intervin numai m rimi intensive (x_j , M_j și V_m) și astfel definește o m rime intensivă.

Presiunea, într-un sistem în echilibru, este o m rime intensivă atâta timp cât valoarea acesteia în sistem este egală cu valoarea acesteia în orice parte a acestuia.

Temperatura, într-un sistem în echilibru, este o m rime intensivă atâta timp cât valoarea acesteia în sistem este egală cu valoarea acesteia în orice parte a acestuia.

În final trebuie făcută remarca că conceptul de m rime intensivă referă macrocantități și își pierde sensul la nivel microscopic. Luând doar temperatura ca exemplu, în spațiu este de câteva grade Kelvin, în timp ce obiectele care se deplasează (cum ar fi o rachetă sau un meteorit) pot ajunge la temperaturi de câteva mii de grade Kelvin, așa cum rezultă din teoria cinetico-moleculară.

Analiza dimensională

În geometrie, pentru a identifica în mod unic poziția unui punct în plan avem nevoie de un reper. Față de acest reper, punctul de probă are două grade de libertate, ceea ce ne arată că poziția sa este în mod unic determinată de fixarea valorilor pentru două proprietăți geometrice ale acestuia. Alegerea reperului nu schimbă condiționarea anterioară. În figura de mai jos (*v. Repere în plan*), reperul poate fi de exemplu unul din colurile dreptunghiului. Cea mai evidentă modalitate este de a fixa valorile proiecțiilor din punctul de probă pe laturile dreptunghiului (fig. a) însă nu este necesar ca ambele valori să provină din același tip de măsurătoare (în fig. b o măsurătoare este de distanță, alta este de unghi) și nici nu este necesar să fie ortogonale (în fig. c oricare din laturile paralelogramului are o proiecție nenulă pe laturile învecinate).

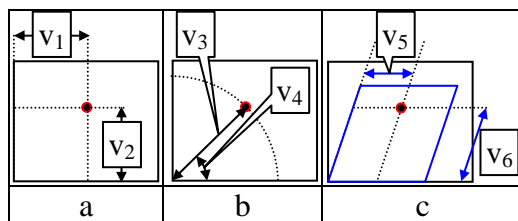


Fig.6. Repere în plan

Exemplul *Repere în plan* ne arată că modalitățile de a exprima prin valori propriile observate pot fi extrem de diferite. În același timp ne arată că ar putea exista, în fiecare caz în parte, o soluție de tipul celei ilustrate în fig. a, în care variabilele asociate celor două valori să fie ortogonale. Chiar însă și în acest caz, nimic nu ne oprește să considerăm că în cazul ilustrat de fig. b este de asemenea ortogonal, deci putem avea chiar mai multe soluții. În toate cazurile însă, am remarcat necesitatea impunerii a două valori fixe, ceea ce ne arată că dimensionalitatea sistemului este 2.

Problema poate fi extinsă la m rimi și relații altele decât geometrice. În general în timpul operăm cu m rimi și unități de măsură, și o **analiză dimensională** poate fi condusă dacă se alege un set redus de m rimi și unități pe baza cărora se poate exprima toate celelalte. Așa cum s-a arătat mai sus, problema nu are o singură soluție, însă unele soluții sunt preferate.

Analiza dimensională asupra m rimilor și relațiilor poate servi la verificarea corectitudinii definirii acestora.

În analiza dimensională a m rimilor și relațiilor fizice, se preferă ca m rimi de bază să fie lungimea, masa, timpul și sarcina electrică. Chiar și aici însă, o remarcă poate fi făcută, și anume că folosind constanta vitezei luminii în vid, se poate exprima o relație de dependență între lungime și timp, iar în ceea ce privește sarcina electrică, definiția sa utilizează lungimea, masa, și timpul.

Așa cum în *Repere în plan* se poate remarca, existența unor valori constante în proprietățile observate reduce dimensionalitatea sistemului. Dacă în oricare dintre cele trei cazuri se fixează o valoare din cele două, dimensionalitatea sistemului se reduce la 1, chiar dacă în fiecare caz în parte manifestarea reducerii dimensionalității produce efecte observabile diferite. Rezultă de aici că, indiferent de modalitatea de exprimare a valorii (deci a unității de măsură alese) de importanță sunt

valorile constantelor universale (v. *Constante universale*).

Constant	Explica ie	Valoare
Viteza luminii în vid	Distan a parcurs de lumin , în vid, în timp de 1 secund ; în fapt metrul este definit pe baza acestei valori	$c_0=299792458 \text{ m}\cdot\text{s}^{-1}$
Impedan a vidului	Este raportul între magnitudinea câmpului electric i magnetic al radia iilor electromagnetice ce traverseaz vidul	$Z_0 = 376.730313461\dots$
Constanta magnetic	Este permitivitatea magnetic a vidului	$\mu_0 = 4\pi \cdot 10^{-7} \text{ N}\cdot\text{A}^{-2}$
Constanta electric	Este permitivitatea electric a vidului	$\epsilon_0 = \frac{1}{\mu_0 c^2}$
Constanta gravita ional	Constanta care intervine în legea atrac iei universale	$G = 6.673\dots \cdot 10^{-11} \text{ m}^3 \cdot \text{kg}^{-1} \cdot \text{s}^{-2}$
Constanta lui Planck	Este cuanta de ac iune în mecanica cuantic exprimând raportul între energia i frecven a unei radia ii electromagnetice	$h = 6.626069\dots \cdot 10^{-34} \text{ J}\cdot\text{s}$
Lungimea Plank	Este aproximativ 10^{-20} mai mic decât diametrul unui proton i e considerat cea mai mic dimensiune posibil ; nu exist îns semnifica ie fizic dovedit a acesteia	$l_P = \sqrt{\frac{\hbar G}{c^3}}$
Masa Planck	Este masa unei g uri negre ipotetice a c rei raz este egal cu lungimea Plank; raza g urii negre ipotetice este astfel încât viteza necesar pentru a evada de pe suprafa este egal cu viteza luminii	$m_P = \sqrt{\frac{\hbar c}{G}}$
Timpul Planck	Este timpul necesar luminii s traverseze în vid o distan egal cu l_P	$t_P = \sqrt{\frac{\hbar G}{c^5}}$
Temperatura Planck	Este temperatura de la care teoria fizic actual înceteaz s mai opereze, deoarece nu avem cuno tin e despre gravita ia cuantic ; punctul în care se formeaz o 'gaur neagr ' de energie	$T_P = \frac{m_P c^2}{k_B}$

Tab. 5. Constante universale

Constant	Explica ie	Valoare
Sarcina electric elementar	Sarcina electric transportat de un singur electron sau proton	$e = 1.6021765\dots \cdot 10^{-19} \text{ C}$
Energia Hartree	Energia electric poten ial a unui atom de hidrogen în starea sa fundamental i aproximativ dublul energiei sale de ionizare	$E_h = 4.359743\dots \cdot 10^{-18} \text{ J}$
Num rul lui Avogadro	Num rul de particule dintr-un mol de substan	$N_A = 6.02214 \cdot 10^{23} \text{ mol}^{-1}$
Constanta atomic de mas	A 12-a parte din masa unui atom de carbon 12	$m_u = 10^{-3} / N_A$
Constanta gazelor	Constanta de propor ionalitate din legea gazului ideal ($pV=nRT$)	$R=8.31446\dots \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$
Constanta Boltzmann	Stabile te propor ionalitatea între energie i temperatur în teoria cinetico-molecular ($\langle J \rangle = k_B \cdot T/2$, fiind energia unei particule, iar J num rul de componente ale energiei	$k_B = \frac{R}{N_A}$

Tab. 6. Alte constante frecvent utilizate

Pe lâng constantele universale o serie de alte constante sunt cunoscute i frecvent utilizate (v. *Alte constante frecvent utilizate*).

În tabelele 6 i 7 se observ c sunt relativ pu ine constante universale independente, celelalte, mult mai multe fiind ob inute din cele 'de baz '.

Un element important al analizei dimensionale îl reprezintă alegerea setului de mrimi de bază. De exemplu în cinematic un set de mrimi de bază este (masa, distanța, timpul) sau (M, L, T). Pe baza acestuia se poate exprima dimensionalitatea vitezei ($L \cdot T^{-1}$), accelerației ($L \cdot T^{-2}$), impulsului ($M \cdot L \cdot T^{-1}$) și a forței ($M \cdot L \cdot T^{-2}$). Setul de mrimi (distanța, viteza, timpul) nu este un set de bază pentru cinematic pentru că masa nu se poate exprima pe baza acestuia și de asemenea cele trei nu sunt independente ($V=L \cdot T^{-1}$).

Cu ajutorul analizei dimensionale se poate stabili corectitudinea unei ecuații din punctul de vedere al omogenității dimensionale în cazul în care sunt implicate o serie de mrimi diferite într-o expresie simplă se poate chiar stabili forma ecuației din analiza dimensională.

Sisteme de unități de măsură; sistemul internațional

Un sistem de unități de măsură este un set de unități care poate fi folosit pentru a specifica orice care poate fi măsurat.

Sistemele actuale de unități de măsură, MKS (metru, kilogram, secundă), CGS (centimetru, gram, secundă), FPS (picior, livră, secundă) diferă unul de celălalt doar prin alegerea unităților de măsură, mrimile măsurate fiind aceleași (distanță, masă, timp).

La acestea trei se adaugă și temperatura (măsurată în Kelvin, grade Celsius sau Fahrenheit), însă cum definiția actuală a distanței este făcută prin intermediul constantei vitezei luminii în vid ($d = c \cdot t$, c - constanta vitezei luminii în vid, v. *Constante universale*) practic tot prin intermediul a trei unități fundamentale (temperatura, timpul, masa) sunt construite și definițiile celorlalte de bază (v. *Relații între definițiile mărimilor sistemului fundamental de unități de măsură*).

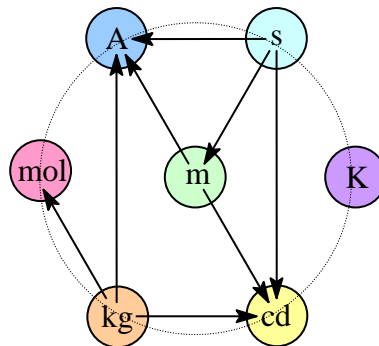


Fig. 7. Relații între definițiile mrimilor sistemului fundamental de unități de măsură

Revenind asupra mrimilor de bază, sistemul internațional de unități recunoaște un număr de 7 ca fiind mrimi de bază [6].

Mrimi			Unitate	
Nume	Dimensiune	Simbol	Nume	Simbol
masa	M	m	kilogram	kg
cantitatea de substanță	N	n	mol	mol
temperatura termodinamică		T	kelvin	K
timpul	T	t	secundă	s
lungimea	L	l, x, r	metru	m
curentul electric	I	I, i	amper	A
intensitatea luminoasă	J	I _v	candela	cd

Tab. 7. Mrimi și unități de bază ale sistemului internațional

Următoarele sunt definițiile mrimilor de bază din sistemul internațional de unități:

- ÷ **Kilogramul** este masa prototipului internațional pentru kilogram (materialul folosit fiind un aliaj de Platin cu 10±0.0001% Iridiu);
- ÷ **Molul** este cantitatea de substanță a unui sistem ce conține tot atâtea entități elementare câți atomi sunt în 0.0012 kg de Carbon cu masa atomică 12 (când molul este utilizat, entitățile elementare

- trebuie specificate și pot fi atomi, molecule, ioni, electroni, alte particule sau grupuri specificate de astfel de particule);
- ÷ **Kelvinul** este $1/273.16$ din temperatura termodinamică a punctului triplu al apei (definiția referă la apă cu următoarea compoziție în izotopi: $^2\text{H}:^1\text{H}=0.00015576$, $^{17}\text{O}:^{16}\text{O}=0.0003799$, $^{18}\text{O}:^{16}\text{O}=0.0020052$);
 - ÷ **Secunda** este durata corespunzătoare unui număr de 9192631770 perioade ale radiației corespunzătoare tranziției între cele două nivele hiperfine ale stării fundamentale a atomului de Cesium cu masa atomică 133 (definiția referă la atomul de cesiu la temperatura de repaus de 0K); **Hertzul** ([Hz], pentru **frecvență**) este definit implicit în definiția secunde, frecvența radiației corespunzătoare tranziției între cele două nivele hiperfine ale stării fundamentale a atomului de Cesium cu masa atomică 133 fiind de 9192631770 Hz ;
 - ÷ **Metru** este lungimea drumului parcurs de lumină în vid într-un interval de timp de $1/299792458$ secunde (rezultatul viteza luminii în vid este exact $c_0=299792458\text{ m}\cdot\text{s}^{-1}$);
 - ÷ **Amperul** este curentul electric constant care, dacă este menținut între doi conductori paraleli de lungime infinită, de grosime neglijabilă, plasați la 1m unul de celălalt în vid vor produce între acești conductori o forță egală cu $2\cdot 10^{-7}\text{ N}$ ($\text{kg}\cdot\text{m}\cdot\text{s}^{-2}$) pentru fiecare metru de lungime;
 - ÷ **Candela** este intensitatea luminoasă, într-o direcție dată, a unei surse care emite radiație monocromatică cu frecvența de $540\cdot 10^{12}\text{ Hz}$ (s^{-1}) și care are o intensitate radiantă în acea direcție de $1/683\text{ W/sr}$ (sr este simbolul pentru steradian, unitatea SI de măsură a unghiului solid; sfera are 4 sr ; 1sr este aria unei calote sferice cu aria 1 pe o sferă cu raza 1); **Lumenul** ([lm], pentru **flux luminos**) este definit implicit în definiția candelăi, fiind fluxul luminos produs de sursa care emite radiație monocromatică cu frecvența de $540\cdot 10^{12}\text{ Hz}$ (s^{-1}) în unghiul solid de 1sr ;
- Pe baza unităților de bază sunt construite unitățile derivate:
- ÷ **Radianul** ([rad], pentru **unghi**) este unghiul la centru dat de o deschidere de 1m pe un cerc cu o rază de 1m ;
 - ÷ **Steradianul** ([sr], pentru **unghi solid**) este unghiul solid la centru dat de o calotă sferică cu aria de 1m^2 pe o sferă cu o rază de 1m ;
 - ÷ **Newtonul** ([N], pentru **forță**) este forța necesară să accelereze 1kg de masă cu $1\text{m}\cdot\text{s}^{-2}$.
 - ÷ **Pascalul** ([Pa], pentru **presiune**) este presiunea exercitată de o forță de 1N pe o suprafață de 1m^2 ;
 - ÷ **Jouleul** ([J], pentru **energie**) este energia cheltuită în aplicarea unei forțe de 1N în deplasarea pe distanța de 1m , sau în traversarea curentului electric de 1A printr-o rezistență de 1Ω pentru 1s ;
 - ÷ **Wattul** ([W], pentru **putere**) este puterea necesară pentru a produce o energie de 1J pe durata de 1s ;
 - ÷ **Voltul** ([V], pentru **potențial electric**) este diferența în potențialul electric pe lungimea unui fir când un curent electric de 1A disipă 1W ;
 - ÷ **Ohmul** (Ω), pentru **rezistența electrică**) este rezistența între 2 puncte ale unui conductor când o diferență de potențial constant de 1V aplicată între aceste puncte produce în conductor un curent de 1A ; conductanța electrică este inversul rezistenței electrice; **Siemensul** ([S], pentru **conductanța electrică**) este conductanța între 2 puncte ale unui conductor când o diferență de potențial constant de 1V aplicată între aceste puncte produce în conductor un curent de 1A ;
 - ÷ **Coulombul** ([C], pentru **sarcina electrică**) este sarcina electrică transportată de un curent constant de 1A pe durata de 1s ;
 - ÷ **Faradul** ([F], pentru **capacitatea electrică**) este valoarea care produce o diferență de 1V când este încărcat cu 1C ;
 - ÷ **Luxul** ([lx], pentru **iluminare**) este iluminarea corespunzătoare unui flux luminos de 1lm care traversează o suprafață de 1m^2 ;
 - ÷ **Becquerelul** ([Bq], pentru **radioactivitate**) este activitatea unui material radioactiv în care are loc dezintegrarea a 1 nucleu atomic pe durata de 1s ;
 - ÷ **Grayul** ([Gy], pentru **doza absorbită de radiație ionizantă**) este doza absorbită când se transferă o energie de 1J unei cantități de materie de 1kg ;
 - ÷ **Sievertul** ([Sv], pentru **doza echivalentă de radiație ionizantă**) este doza absorbită când se observă apariția cancerului în 5.5% din cazuri;

÷ **Katalul** ([kat], pentru *activitatea catalitică*) este valoarea catalitică ce convertește 1 mol de reactant pe durata a 1s;

Multiplii și submultiplii unităților de măsură sunt standardizate (v. *Multiplii și submultiplii ai unităților de măsură*).

Divizor	10^0	10^1	10^2	10^3	10^6	10^9	10^{12}	10^{15}	10^{18}	10^{21}	10^{24}
Simbol	-	d	c	m	μ	n	p	f	a	z	y
Prefix	-	deci	centi	mili	micro	nano	pico	femto	atto	zepto	yocto
Multipliator	10^0	10^1	10^2	10^3	10^6	10^9	10^{12}	10^{15}	10^{18}	10^{21}	10^{24}
Simbol	-	da	h	k	M	G	T	P	E	Z	Y
Prefix	-	deca	hecto	kilo	mega	giga	tera	peta	exa	zetta	yotta

Tab. 8. Multiplii și submultiplii ai unităților de măsură

Evaluarea numeric a expresiilor matematice

Ecuații algebrice și transcendentale

O **funcție transcendentă** este o funcție care nu satisface o ecuație polinomială ai cărei coeficienți sunt ei înșiși polinoame, în contrast cu o funcție algebrică, care satisface o asemenea ecuație. Cu alte cuvinte, o funcție transcendentă este o funcție care "transcende" algebra, în sensul că aceasta nu poate fi exprimată în termenii unei secvențe finite de operații algebrice de adunare, înmulțire și extragerea rădăcinii. Exemple de funcții transcendente sunt funcția exponențială, logaritmul și funcțiile trigonometrice.

O **ecuație transcendentă** este o ecuație care conține o funcție transcendentă.

Un **număr algebric** este un număr rădăcină a unui polinom (nenul) într-o variabilă cu coeficienți raționali. Numerele care nu sunt algebrice sunt **numere transcendente**. Exemplele includ e și π .

Inversa funcției $f(w)=w \cdot e^w$ se numește **funcția W a lui Lambert**. Derivata funcției f , $f'(w) = e^w(1+w)$ ne arată că $w=-1$ este un punct de extrem pentru funcția f (de minim, $f(-1)=-e^{-1}$) și aici este de fapt de acest punct de extrem de o parte și de alta a acestuia există exact două soluții (w_1 și w_2) pentru ecuația $y=w \cdot e^w$. Din acest motiv inversa funcției f , funcția W este fie multivalorică (redând pentru un y două valori, w_1 și w_2) fie se restrânge domeniul de definiție al funcției astfel încât să devină (atât f cât și W) funcție bijectivă. Funcția Lambert este o sursă de numere transcendente, ceea ce înseamnă că soluțiile w asociate ecuației $y=w \cdot e^w$ unde y este un număr algebric exceptând pe $y=0$ sunt toate transcendente.

Funcția W a lui Lambert este de importanță teoretică, întrucât multe ecuații implicând exponențiale se rezolvă oferind soluțiile (indirecte) ca expresii implicând soluțiile funcției W .

Exemplul 1. Pentru $a, b > 0$ să se rezolve ecuația: $a^t = b \cdot t$. *Rezolvare.* $a^t = b \cdot t \iff 1 = b \cdot t / a^t \iff 1 = b \cdot t \cdot e^{-t \cdot \ln(a)} \iff 1/b = t \cdot e^{-t \cdot \ln(a)} \iff -\ln(a)/b = -t \cdot \ln(a) \cdot e^{-t \cdot \ln(a)}$. Dacă se obține w ca o soluție a ecuației $-\ln(a)/b = w \cdot e^w$ atunci $t = -w/\ln(a)$, care se exprimă formal astfel: $t = -W(-\ln(a)/b)/\ln(a)$.

Exemplul 2. Mai general, pentru $a > 0$ și $q > 0$, să se rezolve ecuația: $q^{a \cdot x + b} = c \cdot x + d$. *Indicație.* Se substituie $-t = a \cdot x + a \cdot d/c$ și $R = t \cdot q^t = -(a/c) \cdot q^{b - a \cdot d/c}$.

Exemplul 3. Să se rezolve ecuația $x^x = a$. *Rezolvare.* $x^x = a \iff x \cdot \ln(x) = \ln(a) \iff e^{\ln(x)} \cdot \ln(x) = \ln(a)$. Dacă se obține w ca o soluție a ecuației $\ln(a) = w \cdot e^w$ atunci $x = e^w$, care se exprimă formal astfel: $x = e^{W(\ln(a))}$. *Remarcă.* Dacă $y = w \cdot e^w$, sau scris parametrizat $y = w(y) \cdot e^{w(y)}$ pentru $y = \ln(z)$ atunci $\ln(z) = w(\ln(z)) \cdot e^{w(\ln(z))}$ are loc, și $e^{w(\ln(z))} = \ln(z)/w(\ln(z))$ astfel încât soluția ecuației $x^x = a$ se mai poate scrie ca: $x = e^{W(\ln(a))} = \ln(a)/W(\ln(a))$.

Alte exemple de utilizare a funcției W a lui Lambert pentru a obține soluțiile ecuațiilor transcendente includ ecuațiile în forma $x \cdot \log_b(x) = a$ (cu soluția $x = e^{W(a \cdot \ln(b))}$), valorile explicite ale curentului într-un circuit în care sunt în serie o diodă și o rezistență (relația între curent și tensiune în cazul diodei este dat de o lege exponențială [7] în timp ce la rezistor este o dependență liniară), încetitorul atomic Zeeman [8], curgerile și depunerile granulelor și suspensiilor, curgerea fluidelor vâscoase [9], și imagistica creierului [10].

În toate aceste cazuri, ca și în altele de asemeni, când este posibil obținerea unei soluții analitice sau, în cazul soluțiilor oferite prin intermediul funcției W a lui Lambert, obținerea unor soluții pseudo-analitice, aceste soluții analitice sunt de preferat în favoarea celor obținute prin **evaluare numerică**.

Rezolvarea ecuațiilor polinomiale

Un caz aparte de ecuații sunt cele în forma $y = a_n x^n + \dots + a_1 x + a_0$, numite ecuații polinomiale. Din fericire, astăzi există un algoritm care găsește toate rădăcinile oricărui polinom de orice grad [11], algoritm disponibil online [12] și care este implementat în majoritatea programelor profesionale care necesită rezolvarea de astfel de ecuații. Precizia în care acest algoritm rezolvă ecuațiile polinomiale este limitată numai de precizia de calcul a implementării. De exemplu o implementare C/C++ poate da o precizie de 10^{-14} , o implementare în FreePascal o precizie de 10^{-19} în timp ce o implementare în Fortran poate da o precizie de 10^{-23} .

Rezolvarea numeric a ecuațiilor transcendentale

Pentru ecuațiile la care literatura de specialitate nu ne pune la dispoziție metode deterministe care să ne conducă spre soluție, singura modalitate de abordare rimabilă este de a căuta soluțiile. Avantajul pe care îl are utilizarea calculatorului este imens: stocarea expresiei ecuației de rezolvat în calculator oferă practic posibilitatea evaluării rezultatului funcției asociate în orice moment. Fie astfel ecuația de rezolvat $y=f(x)$, în care y este o valoare care trebuie precizată de fiecare dată când facem apel la o rezolvare numerică. Se construiește funcția asociată $g(x) = f(x) - y$. Se poate evalua funcția $g(x)$ în orice valoare a domeniului său de definiție prin simpla sa implementare într-o rutină de calcul.

Rezolvarea numerică totdeauna presupune **cunoașterea domeniului** în care se află soluția (sau soluțiile) ecuației. Dacă nu avem nici o indicație asupra domeniului, o bună idee este să se încerce **reprezentarea grafică** a dependenței într-un program specializat (cum este MathCad). Odată identificat domeniul (fie acesta $[a,b]$) în care se află soluția sau soluțiile, următorul pas necesar este separarea soluțiilor, și anume găsirea unui subdomeniu (fie acesta $[c,d]$, $a < c, d < b$) în care există o singură soluție. Practic nu există metode standard care să ne ducă la acest subdomeniu, însă din nou o reprezentare grafică ne poate ajuta.

Condiția ca **într-un subdomeniu să existe cel puțin o soluție** este ca $g(c) \cdot g(d) < 0$. Condiția $g(c) \cdot g(d) < 0$ este **suficientă**, în sensul în care dacă $g(c) \cdot g(d) < 0$ atunci există cel puțin o soluție $g(z)=0$ în intervalul $[c,d]$ însă **nu este și necesară**, existența unui număr par de soluții în domeniul $[c,d]$ fiind când $g(c) \cdot g(d) > 0$ și totuși să existe soluții în acest interval.

Odată identificat un domeniu $[c,d]$ pentru care $g(c) \cdot g(d) < 0$, căutarea soluției ecuației $g(z)=0$ se poate face prin **căutări succesive**, și totdeauna un astfel de algoritm va produce cel puțin o soluție.

Cea simplă metodă de căutare succesivă este prin înjumătățirea intervalului (v. *Căutarea soluției $g(z)=0$ când $g(c) \cdot g(d) < 0$*).

```
Let a0=c; b0=d;
Repeat
    If(a0-b0<eps)then stop;//solution is any of a0 and b0
    If(b0-a0<eps)then stop; //solution is any of a0 and b0
    c0=(a0+b0)/2; if(g(c0)<eps)then stop;//solution is c0
    If(g(a0)*g(c0)<0)then b0=c0; else a0=c0;
Until(false);
```

Fig.8. Căutarea soluției $g(z)=0$ când $g(c) \cdot g(d) < 0$

Așa cum se observă în figura de mai sus (v. *Căutarea soluției $g(z)=0$ când $g(c) \cdot g(d) < 0$*) căutarea succesivă a soluției $g(z)=0$ impune definirea unei toleranțe (ϵ) a cărei valoare minimă este stabilită în funcție de **cerințele problemei de rezolvat** (o precizie de 4 cifre semnificative este foarte rar depășită de instrumentația de analiză fizică și chimică) și de **capacitatea procesorului matematic** al platformei de calcul (s-a arătat mai sus că limita de precizie poate merge până la 23 de cifre semnificative fără a face apel la calculul matematic în precizie arbitrară).

Rezolvarea sistemelor de ecuații liniare

Din fericire, și pentru acest caz sunt elaborate metode, unele dintre ele foarte eficiente, de rezolvare exactă. Diferențele între o metodă și o altă metodă în acest caz o reprezintă precizia cu care este oferită soluția. Pentru a putea rezolva un sistem de ecuații liniare, sunt necesare un număr mare de operații de înmulțire, împărțire, adunare și scădere, iar cea mai eficientă rezolvare are ordinul de complexitate polinomial, fiind de ordinul puterii a 3-a a numărului de ecuații, $O(n^3)$. Astfel, pentru un număr foarte mare de ecuații, sunt implicate și un număr mare de operații (pentru $n = 1000$, $n^3 = 10^9$), ceea ce face ca precizia maximă posibilă să scadă și ea (folosind o implementare C/C++, din precizia de 14 cifre semnificative, scăzând cele 9 datorate operațiilor aritmetice, ajungem la pragul de 5 cifre semnificative ca precizie maximă evaluată). Morala este că pentru cazurile în care se operează cu sisteme de ecuații foarte mari, utilizarea celor mai precise metode de rezolvare este obligatorie.

În figura următoare (v. *Procedura Gauss-Jordan pentru regresii liniare multiple*) este redat

procedura Gauss-Jordan pentru rezolvarea sistemelor de ecuații liniare, aplicat pentru modele de regresie liniară, când soluția este furnizată simultan cu calculul de varianță [13].

df = m+1	$y \sim = a_0 \cdot 1 + a_1 \cdot x_1 + \dots + a_m \cdot x_m$ pentru $(y_i, x_{1,i}, \dots, x_{m,i})_{1 \leq i \leq n}$								
i(linii)j(coloane)	-1	0	1	...	m	m+1	m+2	...	2m+1
0	y_i	n	$x_{1,i}$...	$x_{m,i}$	1	0	...	0
1	$y_i x_{1,i}$	$x_{1,i}$	$x_{1,i}^2$...	$x_{1,i} x_{m,i}$	0	1	...	0
...
m	$y_i x_{m,i}$	$x_{m,i}$	$x_{1,i} x_{m,i}$...	$x_{m,i} x_{m,i}$	0	0	...	1
Inițial	S_B	S_A				I_{m+1}			
Operații elementare pe linii									
Final	$A = S_A^{-1} S_B$	I_{m+1}				S_A^{-1}			
0	a_0	1	0	...	0	$(S_A^{-1})_{0,0}$
1	a_1	0	1	...	0	$(S_A^{-1})_{1,1}$
...
m	a_m	0	0	...	1	$(S_A^{-1})_{m,m}$
df = m	$y \sim = a_1 \cdot x_1 + \dots + a_m \cdot x_m$ pentru $(y_i, x_{1,i}, \dots, x_{m,i})_{1 \leq i \leq n}$								
i(linii)j(coloane)	-1	0	1	...	m	m+1	m+2	...	2m+1
0									
1	$y_i x_{1,i}$		$x_{1,i} x_{1,i}$...	$x_{1,i} x_{m,i}$		1	...	0
...
m	$y_i x_{m,i}$		$x_{1,i} x_{m,i}$...	$x_{m,i} x_{m,i}$		0	...	1
Inițial	S_B		S_A				I_m		
Operații elementare pe linii									
Final	$A = S_A^{-1} S_B$		I_m				S_A^{-1}		
0									
1	a_1		1	...	0		$(S_A^{-1})_{1,1}$
...
m	a_m		0	...	1		$(S_A^{-1})_{m,m}$
Coeficienții de varianță	$\hat{y}_i = a_0 + a_1 x_{1,i} + \dots + a_m x_{m,i}$					$s^2(a_i) = (S_A^{-1})_{i,i} \cdot \sigma^2 / (n - df)$			
Semnificația coeficienților	$t_i = t(a_i) = a_i / s(a_i)$					$p_i = p(t_i) = CDF_t(t_i, n - df)$			
Semnificația modelului	$r = \frac{n \cdot \sum y_i \hat{y}_i - \sum y_i \cdot \sum \hat{y}_i}{\sqrt{n \cdot \sum y_i^2 - \sum y_i \cdot \sum y_i} \cdot \sqrt{n \cdot \sum \hat{y}_i^2 - \sum \hat{y}_i \cdot \sum \hat{y}_i}}$ $F(r) = \frac{r^2}{1 - r^2} \cdot \frac{n - df}{m}, p_r = CDF_F(F(r), m, n - df)$								

Fig.9. Procedura Gauss-Jordan pentru regresii liniare multiple

Metode iterative

Pentru ecuații neliniare cu expresii complicate, și cu atât mai mult pentru ecuații transcendente, cel puțin o soluție poate rezulta prin aproximații succesive. Fie ecuația $f(x)=0$ pentru care c și m orădă în x .

O metodă iterativă totdeauna pornește de la o valoare inițială x_0 pe care c și m sînt o îmbunătățim astfel încât la final să obținem o bună aproximație $|f(x_0)| < \epsilon$ unde ϵ este o valoare arbitrar impusă de cerințele problemei de rezolvat.

Ecuația $f(x)=0$ se poate transforma astfel încât să se obțină o ecuație echivalentă cu aceasta în forma $x=g(x)$. Succesul metodei iterative depinde de modalitatea de exprimare a ecuației echivalente.

Exemplu. Fie $f(x)=x^2-2x-3$. În tabelul următor (v. *Aplicarea metodelor iterative*) sunt ilustrate mai multe modalități de exprimare a ecuației echivalente.

Ecua ie echivalent	$x = x^2 - x - 3$	$x = (2x + 3) / x$	$x = \pm\sqrt{2x + 3}$	$x = \pm\sqrt{2x + 3}$
Ecua ie iterativ	$x_{i+1} = x_i^2 - x_i - 3$	$x_{i+1} = (2x_i + 3) / x_i$	$x_{i+1} = \sqrt{2x_i + 3}$	$x_{i+1} = -\sqrt{2x_i + 3}$
i=0	1	1	9	9
i=1	5	-3	3.872983	-3.87298
i=2	2.6	9	2.178524	#NUM!
i=3	3.153846	69	1.164924	#NUM!
i=4	2.95122	4689	#NUM!	#NUM!
i=5	3.016529	21982029	#NUM!	#NUM!

Tab.9. Aplicarea metodelor iterative

A a cum se observ din tabelul anterior (v. *Aplicarea metodelor iterative*) nu orice alegere a metodei iterative conduce la convergen a c tre solu ie (f(x)=0).

Sisteme de ecua ii neliniare

Metoda aproxima iilor succesive poate fi o solu ie comod în anumite cazuri când se cere rezolvarea de sisteme de ecua ii neliniare.

Un exemplu tipic de aplicare este pentru maximizarea agrementului între observa ie i model. Fie exemplul ilustrat mai jos (v. *Observa ii la contingen a de doi factori multiplicativi*).

	Factor "B"	Nivel b ₁	Nivel b ₂	Nivel b ₃	Nivel b ₄
Factor "A"					
Nivel a ₁		Obs ₁₁	Obs ₁₂	Obs ₁₃	Obs ₁₄
Nivel a ₂		Obs ₂₁	Obs ₂₂	Obs ₂₃	Obs ₂₄
Nivel a ₃		Obs ₃₁	Obs ₃₂	Obs ₃₃	Obs ₃₄
Nivel a ₄		Obs ₄₁	Obs ₄₂	Obs ₄₃	Obs ₄₄

Tab.10. Observa ii la contingen a de doi factori multiplicativi

În cazul observa iilor la contingen a de doi factori multiplicativi (v. *Observa ii la contingen a de doi factori multiplicativi*) ipoteza este c Obs_{ij} ~ a_i·b_j i valoarea a teptat în urma observa iei este Exp_{ij}=a_i·b_j. Datorit unor factori aleatori necunoscu i, observa ia e afectat de erori, astfel încât aproape niciodat Exp_{ij}=Obs_{ij}. Dificultatea îns const în faptul c nu se cunosc valorile nivelelor factorilor (a₁, ..., a₄; b₁, ..., b₄) i nici tipul erorii (eroare propor ional cu valoarea observat , eroare în magnitudine uniform , etc.). Orice tentativ de a rezolva analitic sistemul (prin identificarea valorilor nivelelor) e sortit e ecului, deoarece sistemul are cel pu in o nedeterminare (putând fi astfel rezolvat doar cel mult parametric).

Avem îns posibilitatea s ob inem un set de valori ini iale pentru valorile a teptate cu formula:

$$Exp_{i,j} = \frac{\sum_{k=1}^r Obs_{i,k} \sum_{k=1}^c Obs_{k,j}}{\sum_{i=1}^r \sum_{j=1}^c Obs_{i,j}}$$

În acela i cadru al presupunerii naturale al efectului multiplicativ al celor doi factori asupra observabilei Obs din punct de vedere matematic se pot formula trei presupuneri cu privire la eroarea p tratic (Obs_{ij}-Exp_{ij})² produs de observa ie:

- ÷ m sur toarea este afectat de erori absolute întâmpl toare;
- ÷ m sur toarea este afectat de erori relative întâmpl toare;
- ÷ m sur toarea este afectat de erori întâmpl toare pe o scar intermediar între erori absolute i erori relative;

Prima dintre ipoteze (erori absolute întâmpl toare) conduce din punct de vedere matematic la minimizarea varian ei între model i observa ie (S²), a doua dintre ipoteze conduce la minimizarea p tratului coeficientului de varia ie (CV²) iar o solu ie (una din mai multe solu ii posibile) la cea de-a

treia dintre ipoteze (X^2) o reprezintă minimizarea statisticii X^2 (v. *Minimizarea diferitelor tipuri de erori pentru agrementul între observație și model*).

S^2	CV^2	X^2
$\sum_{i=1}^r \sum_{j=1}^c (Obs_{i,j} - a_i b_j)^2 = \min.$	$\sum_{i=1}^r \sum_{j=1}^c \frac{(Obs_{i,j} - a_i b_j)^2}{(a_i b_j)^2} = \min.$	$\sum_{i=1}^r \sum_{j=1}^c \frac{(Obs_{i,j} - a_i b_j)^2}{a_i b_j} = \min.$

Tab.11. Minimizarea diferitelor tipuri de erori pentru agrementul între observație și model

În relațiile de mai sus (v. *Minimizarea diferitelor tipuri de erori pentru agrementul între observație și model*) apar exprimați cei doi factori ("A" și "B") a căror independență se verifică prin intermediul efectului multiplicativ (a_i , $1 \leq i \leq r$ reprezintă contribuția primului factor la valoarea așteptată $E_{i,j}$ iar b_j , $1 \leq j \leq c$ reprezintă contribuția celui de-al doilea factor la valoarea așteptată $E_{i,j}$ și expresia valorii așteptate $E_{i,j}$ este dată, așa cum presupunerea naturală a fost făcută de produsul celor două contribuții: $E_{i,j} = a_i \cdot b_j$).

Minimizarea cantităților date de relațiile de mai sus în scopul determinării contribuțiilor factorilor A ($A = (a_i)_{1 \leq i \leq r}$) și B ($B = (b_j)_{1 \leq j \leq c}$) se face pe aceeași cale, dată generic de relația:

$$\left(\frac{\partial \cdot (a_i, b_j)}{\partial a_i} = 0 \right)_{1 \leq i \leq r} ; \left(\frac{\partial \cdot (a_i, b_j)}{\partial b_j} = 0 \right)_{1 \leq j \leq c}$$

unde expresia de derivat $\cdot (a_i, b_j)$ este una din expresiile date de S^2 , CV^2 și X^2 . În urma calculului (derivare) se poate obține condițiile de minimizare a diferitelor tipuri de erori pentru agrementul între observație sunt echivalente cu ecuațiile următoare (v. *Ecuații în minimizarea diferitelor tipuri de erori pentru agrementul între observație și model*).

S^2	CV^2	X^2	Domeniu
$a_i = \frac{\sum_{j=1}^c b_j O_{i,j}}{\sum_{j=1}^c b_j^2}$	$a_i = \frac{\sum_{j=1}^c \frac{O_{i,j}^2}{b_j^2}}{\sum_{j=1}^c \frac{O_{i,j}}{b_j}}$	$a_i^2 = \frac{\sum_{j=1}^c \frac{O_{i,j}^2}{b_j}}{\sum_{j=1}^c b_j}$	$i = 1..r$
$b_j = \frac{\sum_{i=1}^r a_i O_{i,j}}{\sum_{i=1}^r a_i^2}$	$b_j = \frac{\sum_{i=1}^r \frac{O_{i,j}^2}{a_i^2}}{\sum_{i=1}^r \frac{O_{i,j}}{a_i}}$	$b_j^2 = \frac{\sum_{i=1}^r \frac{O_{i,j}^2}{a_i}}{\sum_{i=1}^r a_i}$	$j = 1..c$

Tab.12. Ecuații în minimizarea diferitelor tipuri de erori pentru agrementul între observație și model

Se poate de asemenea arăta matematic că relațiile de mai sus (v. *Ecuații în minimizarea diferitelor tipuri de erori pentru agrementul între observație și model*) admit o infinitate de soluții și că familiile de soluții ale relațiilor se află în vecinătatea familiei de soluții date valorile inițiale ale valorilor așteptate ($Exp_{i,j} = a_i \cdot b_j$):

$$a_i \cdot b_j = \frac{\sum_{k=1}^r O_{i,k} \sum_{k=1}^c O_{k,j}}{\sum_{i=1}^r \sum_{j=1}^c O_{i,j}}$$

Calea directă de rezolvare a ecuațiilor de minimizare fără a face apel la ecuația de aproximare este ineficientă. De exemplu pentru $r=2, c=3$ substituțiile în relația pentru S^2 duc la:

$$\left(\frac{a_2}{a_1} \right)^2 + \frac{(Obs_{1,1}^2 + Obs_{1,2}^2 + Obs_{1,3}^2) - (Obs_{2,1}^2 + Obs_{2,2}^2 + Obs_{2,3}^2)}{(Obs_{1,1} Obs_{2,1} + Obs_{1,2} Obs_{2,2} + Obs_{1,3} Obs_{2,3})} \left(\frac{a_2}{a_1} \right) - 1 = 0$$

care este rezolvabil în (a_2/a_1) care dovedește că există o infinitate de soluții (pentru orice valoare nenulă a lui a_1 există o valoare a_2 care să verifice ecuația și gradul ecuației este dat de $\min(r,c)$). Ecuațiile ce se obțin pe calea substituției directe devin din ce în ce mai complicate cu creșterea lui r și c și cu coborârea dinspre relația (S^2) către relația (X^2).

Calea indirectă de rezolvare a ecuațiilor de minimizare este prin aproximații succesive fără când apel la soluția aproximativă oferită de estimarea inițială. Astfel, se folosește relația de estimare inițială pentru a obține prima aproximație (aproximația inițială) a soluției după care în fiecare succesiune de aproximații se înlocuiesc vechile valori ale aproximației în partea dreaptă a relațiilor pentru a obține

noile aproxima ii.

În acest caz metoda aproxima iilor succesive *converge rapid ctre soluia optimală*. Astfel pentru rela ia (S^2) trei itera ii sunt suficiente pentru a ob ine (vezi Tabelul 13) o valoare rezidual de 282.11735 i de la aceast itera ie încolo valoarea rezidual i schimb cifrele dincolo de a 5-a zecimal , în timp ce pentru rela ia (20) aceea i calitate a reprezent rii solu iei optimale este ob inut dup 4 itera ii.

Folosind datele din [14] redate în Tabelul 13, valorile sugerate de ecua iile de estimare ini ial pentru produsele (a_{ij})_{1 i 6;1 j 12} sunt redate în Tabelul 14, valorile ce rezult dup rezolvarea iterativ a rela iilor (18)-(20) sunt redate în Tabelele 15-17.

TV	UD	KK	KP	TP	ID	GS	AJ	BQ	ND	EP	AC	DY	Suma
DS	25.3	28	23.3	20	22.9	20.8	22.3	21.9	18.3	14.7	13.8	10	241.3
DC	26	27	24.4	19	20.6	24.4	16.8	20.9	20.3	15.6	11	11.8	237.8
DB	26.5	23.8	14.2	20	20.1	21.8	21.7	20.6	16	14.3	11.1	13.3	223.4
US	23	20.4	18.2	20.2	15.8	15.8	12.7	12.8	11.8	12.5	12.5	8.2	183.9
UC	18.5	17	20.8	18.1	17.5	14.4	19.6	13.7	13	12	12.7	8.3	185.6
UB	9.5	6.5	4.9	7.7	4.4	2.3	4.2	6.6	1.6	2.2	2.2	1.6	53.7
Suma	128.8	122.7	105.8	105	101.3	99.5	97.3	96.5	81	71.3	63.3	53.2	1125.7

Legend :

- ÷ TV: Tratament vs. Varietate
- ÷ UD, KK, KP, TP, ID, GS, AJ, BQ, ND, EP, AC, DY: variet i de cartofi (UD: Up to Date; KK: K of K; KP: Kerr's Pink; TP: Tinwald Perfection; ID: Iron Duke; GS: Great Scott; AJ: Ajax; BQ: British Queen; ND: Nithsdale; EP: Epicure; AC: Arran Comrade; DY: Duke of York)
- ÷ DS, DC, DB, US, UC, UB: tratamente (D* - cu fertilizant natural; U* - f r ; S - sol fertilizat cu sulfat; C - sol fertilizat cu cloruri; B - sol fertilizat cu baze)

Tab.13. Valori experimentale în tratamentul cartofilor

TV	UD	KK	KP	TP	ID	GS	AJ	BQ	ND	EP	AC	DY
DS	27.61	26.30	22.68	22.51	21.71	21.33	20.86	20.69	17.36	15.28	13.57	11.40
DC	27.21	25.92	22.35	22.18	21.40	21.02	20.55	20.39	17.11	15.06	13.37	11.24
DB	25.56	24.35	21.00	20.84	20.10	19.75	19.31	19.15	16.07	14.15	12.56	10.56
US	21.04	20.04	17.28	17.15	16.55	16.25	15.90	15.76	13.23	11.65	10.34	8.69
UC	21.24	20.23	17.44	17.31	16.70	16.41	16.04	15.91	13.35	11.76	10.44	8.77
UB	6.14	5.85	5.05	5.01	4.83	4.75	4.64	4.60	3.86	3.40	3.02	2.54

Tab.14. Valorile produselor (a_{ij})_{1 i 6;1 j 12}

TV	UD	KK	KP	TP	ID	GS	AJ	BQ	ND	EP	AC	DY
DS	27.07	26.42	22.64	21.85	21.85	21.94	20.94	20.63	17.93	15.48	13.54	11.61
DC	26.66	26.02	22.29	21.52	21.52	21.60	20.62	20.32	17.66	15.24	13.33	11.43
DB	24.91	24.32	20.83	20.11	20.11	20.19	19.27	18.99	16.50	14.25	12.46	10.69
US	20.64	20.15	17.26	16.66	16.66	16.73	15.96	15.73	13.67	11.80	10.32	8.85
UC	20.58	20.09	17.21	16.61	16.61	16.68	15.92	15.69	13.63	11.77	10.29	8.83
UB	6.29	6.14	5.26	5.08	5.08	5.10	4.86	4.79	4.17	3.60	3.14	2.70

Tab.15 Valorile optimizate ale produselor (a_{ij})_{1 i 6;1 j 12} folosind rela iile (S^2)

TV	UD	KK	KP	TP	ID	GS	AJ	BQ	ND	EP	AC	DY
DS	27.57	26.08	23.04	22.61	21.48	21.61	21.13	20.69	17.66	15.23	13.79	11.56
DC	27.38	25.9	22.88	22.45	21.34	21.46	20.99	20.55	17.54	15.13	13.69	11.48
DB	25.84	24.44	21.59	21.19	20.14	20.26	19.8	19.4	16.56	14.28	12.92	10.83
US	21.23	20.08	17.74	17.4	16.54	16.64	16.27	15.93	13.6	11.73	10.62	8.9
UC	21.47	20.31	17.94	17.61	16.73	16.83	16.46	16.12	13.76	11.86	10.74	9
UB	7.02	6.64	5.87	5.76	5.47	5.51	5.38	5.27	4.5	3.88	3.51	2.94

Tab.16. Valorile optimizate ale produselor (a_{ij})_{1 i 6;1 j 12} folosind rela iile (CV^2)

TV	UD	KK	KP	TP	ID	GS	AJ	BQ	ND	EP	AC	DY
DS	27.64	26.19	22.85	22.60	21.59	21.44	20.98	20.71	17.49	15.24	13.67	11.47
DC	27.35	25.91	22.61	22.36	21.36	21.22	20.76	20.50	17.30	15.08	13.52	11.35
DB	25.74	24.40	21.28	21.05	20.11	19.97	19.55	19.29	16.29	14.20	12.73	10.68
US	21.17	20.06	17.50	17.31	16.53	16.42	16.07	15.87	13.39	11.68	10.47	8.78
UC	21.40	20.28	17.69	17.50	16.71	16.60	16.25	16.04	13.54	11.80	10.58	8.88
UB	6.57	6.23	5.43	5.37	5.13	5.10	4.99	4.93	4.16	3.63	3.25	2.73

Tab.17. Valorile optimizate ale produselor $(a_i b_j)_{1 \leq i \leq 6; 1 \leq j \leq 12}$ folosind rela iile (X^2)

Tabelul 18 centralizeaz rezultatele ob inute pe cele 4 c i.

Cat	S^2				X^2				CV^2			
	eq(21)	eq(18)	eq(20)	eq(19)	eq(21)	eq(18)	eq(20)	eq(19)	eq(21)	eq(18)	eq(20)	eq(19)
DS	23.4	18.76	24.12	57.97	1.10	0.937	1.127	2.308	0.056	0.0515	0.0573	0.0971
DC	59.7	48.48	59.86	104.95	3.08	2.497	3.052	4.847	0.164	0.133	0.1611	0.2365
DB	69.8	66.77	71.47	95.21	3.78	3.596	3.796	4.803	0.221	0.2078	0.2167	0.2633
US	41.6	49.03	41.66	35.34	2.72	3.19	2.709	2.358	0.186	0.2158	0.183	0.1635
UC	57.6	59.01	56.53	82.16	3.46	3.66	3.339	4.367	0.218	0.2375	0.2065	0.2444
UB	37.5	40.1	37.13	28.26	7.89	8.295	7.659	5.956	1.751	1.8018	1.6696	1.3512
UD	30.3	26.3	28.2	78.9	2.66	2.35	2.15	3.58	0.335	0.293	0.235	0.232
KK	15.3	13.5	15.8	18.7	0.76	0.64	0.73	0.88	0.045	0.033	0.035	0.044
KP	63	62.7	64	67.5	3.11	3.15	3.13	3.19	0.155	0.162	0.159	0.155
TP	34.3	31.4	33.3	76.5	2.79	2.69	2.37	3.67	0.357	0.340	0.256	0.242
ID	3.4	3.9	4	4.5	0.21	0.27	0.28	0.26	0.017	0.028	0.029	0.021
GS	26.2	25.6	26.9	28.6	2.29	2.45	2.52	2.42	0.319	0.349	0.352	0.327
AJ	45	47	45.3	43.4	2.56	2.71	2.60	2.44	0.152	0.168	0.164	0.148
BQ	21.5	20.4	21	31.8	1.93	1.71	1.67	2.19	0.253	0.205	0.182	0.193
ND	18.3	17.9	19.1	20.5	2.13	2.29	2.35	2.27	0.393	0.424	0.427	0.403
EP	2.9	3.2	3.3	3.8	0.53	0.64	0.66	0.62	0.133	0.158	0.163	0.142
AC	18.2	18.8	18.7	19.3	1.76	1.87	1.84	1.83	0.209	0.232	0.233	0.221
DY	11.1	11.5	11.2	10.6	1.31	1.40	1.39	1.27	0.228	0.255	0.258	0.227
	289.5	282.2	290.8	404.1	22.04	22.17	21.69	24.62	2.596	2.647	2.493	2.355

Tabelul 18. Valori comparative pentru eroarea experimental întâmpl toare

Dup cum se observ î n Tabelul 18, fiecare dintre metodele definite de rela iile (S^2) - (X^2) î mbun t e te valoarea sumei obiectiv î n raport cu expresia definit de formula aproximativ de estimare ini ial i reprezint corec ii ale acesteia.

Integrarea numerică

O foarte larg categorie de probleme din fizic i chimie opereaz cu ecua ii î n care este necesar calculul unei integrale care nu poate fi evaluat formal, prin intermediul formulelor i metodelor de integrare cunoscute sub numele de *primitive*. Î n aceste situa ii se face apel la metodele de *integrare numerică*. Metodele de integrare numeric sunt metode aproximative de evaluare a valorii integralei definite. Rezultatul integr rii numerice poate fi evaluat aplicând un num r definit (î n preambulul integr rii) de pa i sau un num r de pa i care rezult din aplicarea î nse i a algoritmului numeric de integrare.

Semnifica ia fizic a integralei definite este aria subgraficului func iei de integrat. Î n acest sens, figura urm toare ilustreaz dou modalit i de aproximare î n care domeniul de integrare este î mpr it î n intervale de dimensiune egal .

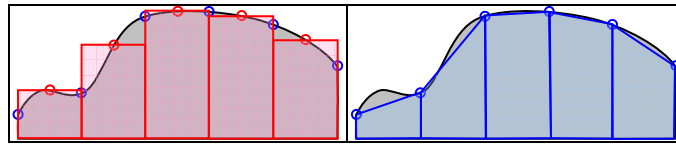


Fig.10. Aproximarea ariei prin dreptunghiuri și trapeze

În integrarea numerică este totdeauna un compromis între timpul de calcul (mai multe diviziuni necesită mai multe evaluări și deci mai mult timp) precizia dorită (față de care însuși criteriul de apreciere a preciziei este uneori foarte dificil de apreciat având în vedere că valoarea reală a ariei de sub graficul funcției este necunoscută), limita preciziei de calcul (eroarea cea mai mare se face în jurul valorii de 1 prin eroarea de evaluare a acestuia cu 0.9...99 și respectiv 1.0...01 și nu în jurul valorii de 0 unde modalitatea de stocare în virgulă flotant 1.2e-4 ne asigură o precizie foarte ridicată) și propagarea erorii din operații de calcul succesive (dacă $a=1.6$ și $b=3.0$ și calculăm a/b cu o precizie de 2 cifre obținem 0.53 pierzând astfel restul cifrelor de 3 ale numărului rațional $0.5(3)$ și calculele care urmează propagă și amplifică această pierdere).

Metodele de cuadratură (sau Newton-Cotes [15, 16]) aproximează funcția de integrat cu un polinom. Din acest punct de vedere:

- ÷ aproximarea ariei prin dreptunghiuri reprezintă aproximarea funcției cu un polinom constant; pentru $x_i \leq x \leq x_{i+1}$ $f(x) = c_i$ ($c_i = f(x_i)$, $c_i = f(x_{i+1})$ sau $c_i = f(\xi_i)$, $\xi_i = (x_i + x_{i+1})/2$)
- ÷ aproximarea ariei prin trapeze reprezintă aproximarea funcției cu o funcție liniară; pentru $x_i \leq x \leq x_{i+1}$ $f(x) = \alpha x + \beta$, $\alpha = (f(x_{i+1}) - f(x_i)) / (x_{i+1} - x_i)$, $\beta = (f(x_i) \cdot x_{i+1} - f(x_{i+1}) \cdot x_i) / (x_{i+1} - x_i)$, relații ce rezultă din condițiile $f_i(x_i) = \alpha x_i + \beta = f(x_i)$ și $f_i(x_{i+1}) = \alpha x_{i+1} + \beta = f(x_{i+1})$

Este de asemenea posibilă aproximarea prin polinoame de ordinul 2, 3 și superior. În acest sens este exemplificat în continuare pentru polinoamele de ordinul 2 (regula 1/3 a lui Simpson [17]), existând însuși regula 3/8 a lui Simpson de interpolare cu polinoame de ordinul 3 (regula 3/8 a lui Simpson) și formule polinomiale de grad egal cu numărul de intervale.

Regula 1/3 a lui Simpson de integrare numerică

Fie un ir de sub-intervale în care se evaluează funcția $f(x)$: $a = x_0, x_1, \dots, x_n = b$, de lungimi egale: $x_{i+1} - x_i = (b-a)/n$ (pentru $0 \leq i < n-1$). Pe o pereche de intervale oarecare consecutive $[x_{i-1}, x_i]$ și $[x_i, x_{i+1}]$ expresia funcției se aproximează cu o funcție parabolică $g(x) = a \cdot x^2 + b \cdot x + c$. Se pot identifica coeficienții a, b și c direct în expresia integrată:

$$\int_{x_{i-1}}^{x_{i+1}} g(x) dx = \int_{x_{i-1}}^{x_{i+1}} (ax^2 + bx + c) dx = \frac{a(x_{i+1} - x_{i-1})^3}{3} + \frac{b(x_{i+1} - x_{i-1})^2}{2} + \frac{c(x_{i+1} - x_{i-1})}{1}$$

dacă se folosește valoarea funcției în punctul x_i :

$$g(x_i) = g\left(\frac{x_{i-1} + x_{i+1}}{2}\right) = a\left(\frac{x_{i-1} + x_{i+1}}{2}\right)^2 + b\left(\frac{x_{i-1} + x_{i+1}}{2}\right) + c = \frac{a(x_{i-1} + x_{i+1})^2 + 2b(x_{i-1} + x_{i+1}) + 4c}{4}$$

Sucesiunea de calcule este următoarea (unde $G(x)$ este primitiva lui $g(x)$):

$$\begin{aligned} \int_{x_{i-1}}^{x_{i+1}} g(x) dx &= \frac{x_{i+1} - x_{i-1}}{6} \left(2a(x_{i+1}^2 + x_{i+1}x_{i-1} + x_{i-1}^2) + 3b(x_{i+1} + x_{i-1}) + 6c \right) \\ &= \frac{x_{i+1} - x_{i-1}}{6} \left(a(x_{i+1}^2 + 2x_{i+1}x_{i-1} + x_{i-1}^2) + 2b(x_{i+1} + x_{i-1}) + 4c + a(x_{i+1}^2 + x_{i-1}^2) + b(x_{i+1} + x_{i-1}) + 2c \right) \\ &= \frac{x_{i+1} - x_{i-1}}{6} (4g(x_i) + g(x_{i+1}) + g(x_{i-1})) \end{aligned}$$

Exprimând valorile funcției g cu ajutorul funcției f în punctele x_{i-1}, x_i, x_{i+1} rezultă că aproximarea integralei funcției f cu integralele funcțiilor g se realizează aadar cu ajutorul formulelor:

$$\int_{x_{i-1}}^{x_{i+1}} f(x) dx \cong \frac{x_{i+1} - x_{i-1}}{6} (f(x_{i-1}) + 4f(x_i) + f(x_{i+1}))$$

Însumând integralele de pe fiecare domeniu $[x_{i-1}, x_{i+1}]$ se obține o expresie în care capetele

$[x_0, x_1]$ și $[x_{n-1}, x_n]$ sunt luate o singură dată iar celelalte intervale $[x_i, x_{i+1}]$, pentru $0 < i < n-2$ sunt toate luate de două ori, adică :

$$\sum_{i=1}^{n-1} \int_{x_{i-1}}^{x_{i+1}} f(x) dx = 2 \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x) dx - \int_{x_0}^{x_1} f(x) dx - \int_{x_{n-1}}^{x_n} f(x) dx = 2 \int_a^b f(x) dx - \int_{x_0}^{x_1} f(x) dx - \int_{x_{n-1}}^{x_n} f(x) dx$$

Expresia integralei funcției f pe întreg domeniul $[a, b]$ se obține ca:

$$2 \int_a^b f(x) dx = \sum_{i=1}^{n-1} \int_{x_{i-1}}^{x_{i+1}} f(x) dx - \int_{x_0}^{x_1} f(x) dx - \int_{x_{n-1}}^{x_n} f(x) dx$$

Relația de mai sus arată că evaluarea completă a integralei funcției f mai necesită o aproximare, și anume a integralelor intervalelor de la capete ($[x_0, x_1]$ și $[x_{n-1}, x_n]$), care se poate face de exemplu cu metoda trapezelor:

$$\int_a^b f(x) dx \cong \frac{1}{2} \sum_{i=1}^{n-1} \frac{x_{i+1} - x_{i-1}}{6} (f(x_{i-1}) + 4f(x_i) + f(x_{i+1})) - \frac{1}{2} \int_{x_0}^{x_1} f(x) dx - \frac{1}{2} \int_{x_{n-1}}^{x_n} f(x) dx$$

$$\int_{x_0}^{x_1} f(x) dx \cong \frac{1}{2} \frac{f(x_1) + f(x_0)}{2} (x_1 - x_0), \quad \int_{x_{n-1}}^{x_n} f(x) dx \cong \frac{1}{2} \frac{f(x_n) + f(x_{n-1})}{2} (x_n - x_{n-1})$$

înând cont că $x_{i+1} - x_i = (b-a)/n$,

$$\int_a^b f(x) dx \cong \frac{b-a}{6n} \left(\sum_{i=1}^{n-1} f(x_{i-1}) + 4 \sum_{i=1}^{n-1} f(x_i) + \sum_{i=1}^{n-1} f(x_{i+1}) \right) - \frac{1}{4} \frac{b-a}{n} (f(x_0) + f(x_1) + f(x_{n-1}) + f(x_n))$$

în care se înțelege seama acum că :

$$\sum_{i=1}^{n-1} f(x_{i-1}) = \sum_{i=0}^{n-2} f(x_i) = \sum_{i=0}^n f(x_i) - f(x_n) - f(x_{n-1})$$

$$\sum_{i=1}^{n-1} f(x_i) = \sum_{i=0}^n f(x_i) - f(x_0) - f(x_n)$$

$$\sum_{i=1}^{n-1} f(x_{i+1}) = \sum_{i=2}^n f(x_i) = \sum_{i=0}^n f(x_i) - f(x_1) - f(x_0)$$

$$\sum_{i=0}^n f(x_i) \stackrel{\text{def}}{=} S, \quad \int_a^b f(x) dx \stackrel{\text{def}}{=} I$$

de unde:

$$I \cong \frac{b-a}{n} \left(\frac{6S - f(x_n) - f(x_{n-1}) - 4f(x_0) - 4f(x_n) - f(x_1) - f(x_0)}{6} - \frac{f(x_0) + f(x_1) + f(x_{n-1}) + f(x_n)}{4} \right)$$

$$I \cong \frac{(b-a)S}{n} - \frac{(b-a)}{12n} (13f(x_0) + 5f(x_1) + 5f(x_{n-1}) + 13f(x_n))$$

Elemente de teoria probabilităților

Metode statistice pentru populații finite și infinite

Frecvent operăm cu date provenite din măsurători repetate ale aceluiași fenomen. Aproape niciodată însă, două măsurători independente asupra aceleiași observabile, produce același rezultat. Din acest punct de vedere, odată încheiată observația fenomenului fizic, se pune problema prelucrării și interpretării informației din spațiul informațional și un alt tip de experiment trebuie derulat aici: **experimentul statistic**.

Într-un experiment statistic operăm cu variabile aleatoare - o **variabilă aleatoare** fiind asociat unei măsurabile ale cărei valori sunt colectate în spațiul informațional. Variabila aleatoare are o **serie de valori** care corespund rezultatelor măsurătorilor înregistrate. Totdeauna în spațiul informațional vom poseda serii finite de valori în timp ce în spațiul fizic s-ar putea repeta experimentul de un număr infinit de ori. Din acest motiv, în statistică se diferențiază **eșantionul** - seria de valori colectată în spațiul informațional, de **populație** - întreaga mulțime infinită de valori care ar putea fi aduse din spațiul fizic printr-un proces continuu de măsurare.

Din punctul de vedere al reprezentării prin valori în spațiul informațional, avem **variabile aleatoare discrete**, ale căror valori apar în unei mulțimi finite (înzestrate sau nu cu o relație de ordine) sau infinite dar atunci înzestrate cu o relație de ordine și numărăbile (cum este mulțimea numerelor naturale, sau a celor întregi, sau a celor raționale) și **variabile aleatoare continue**, ale căror valori apar în mulțimii numerelor reale (de puterea continuului) - și asta chiar dacă, în spațiul informațional le stocăm tot folosind o mulțime de valori discrete și ordonate (folosind o anumită precizie de exprimare a numerelor). O altă clasificare a tipului variabilelor aleatoare este în **cantitative** - dacă scala de măsurare a fenomenului observat permite operația de adăugare ("+") și **calitative** - dacă nu e permis (nu are sens).

Distribuții statistice în urma unor măsurători repetate

Așa cum s-a anticipat la începutul secțiunii anterioare, o problemă la care statistica are o soluție de exprimare a rezultatului este cu privire la rezultatul urmând o serie de măsurători repetate. În tabelul următor (v. *Măsurători statistice pentru caracterizarea variabilelor cantitative*) sunt prezentate măsurătorile statistice ce rezultă din aplicarea pe populații finite respectiv pe infinite, pentru cazul variabilelor cantitative.

Măsurătorie	Referință	Expresie	Interpretare
Suma valorilor	Un șir de numere	(\cdot)	-
Numărul de valori		$ \cdot $	-
Valoarea medie		$E(\cdot) = (\cdot)/ \cdot $	Valoarea așteptată
Moment central de ordin $k, k > 1$		$E_k(\cdot) = E((X - E(X))^k)$	-
Media caracteristicii X	O populație	$\mu = \mu(X) = E(X)$	Tendința centrală
Media observabilei Y	Un eșantion	$m = m(Y) = E(Y)$	
Estimatorul mediei caracteristicii X	O populație	$M(Y) = m(Y)$	
Varianța caracteristicii X	O populație	$\text{Var}(X) = E((X - \mu)^2)$	Împrețierea
Deviația standard a caracteristicii X		$s = s(X) = \sqrt{\text{Var}(X)}$	Dispersia
Varianța observabilei Y	Un eșantion	$\text{var} = \text{var}(Y) = E((Y - E(Y))^2)$	Împrețierea
Deviația standard a observabilei Y		$s = s(Y) = \sqrt{\text{var}(Y)}$	Dispersia
Estimatorul varianței caracteristicii X	O populație	$\text{VAR}(Y) = \frac{ Y }{ Y - 1} \text{var}(Y)$	Împrețierea
Estimatorul deviației standard a caracteristicii X		$S = S(Y) = \sqrt{\frac{ Y }{ Y - 1}} s(Y)$	Dispersia

Tabelul 19. Metode statistice pentru caracterizarea variabilelor cantitative

Un caz foarte frecvent în m sur torile repetate este când valorile se distribuie 'normal' (vezi mai jos **distribuția Gauss**), i din acest punct de vedere, este esen ial caracterizarea dep rt rii de la normalitate (v. *Statistici pentru caracterizarea depărtării de normalitate a variabilelor cantitative*).

Simbol i m sur	Refer	Expresie	M rimi ce intervin
g_1 , Asimetria caracteristicii X	O popula ie	$g_1 = \mu_3/\mu_2^{3/2}$	$\mu_k = E_k(X), k>1$
b_2 , Boltirea caracteristicii X		$b_2 = \mu_4/\mu_2^2$	
g_2 , Excesul de boltire al caracteristicii X		$g_2 = b_2 - 3$	
g_1 , Asimetria observabilei Y	Un e antion	$g_1 = m_3/m_2^{3/2}$	$m_k = E_k(Y), k>1$
b_2 , Boltirea observabilei Y		$b_2 = m_4/m_2^2$	
g_2 , Excesul de boltire al observabilei Y		$g_2 = b_2 - 3$	
Estimatorul asimetriei caracteristicii X	O popula ie	$G_1 = \frac{\sqrt{n_Y(n_Y-1)}}{(n_Y-2)} M_3/M_2^{3/2}$	$n_Y = Y $
Estimatorul boltirii caracteristicii X		$B_2 = \frac{(n_Y-1)(n_Y+1)}{(n_Y-2)(n_Y-3)} M_4/M_2^2$	$M_k = \frac{n_Y}{n_Y-1} E_k(Y),$
Estimatorul excesului de boltire a caracteristicii X		$G_2 = B_2 - 3 \cdot \frac{(n_Y-1)^2}{(n_Y-2)(n_Y-3)}$	$k>1$

Tabelul 20. Statistici pentru caracterizarea dep rt rii de normalitate a variabilelor cantitative

M rime i nota ie	Valoare
Media mediei, $\mu_{\bar{Y}}$	$\mu_{\bar{Y}} = \mu(m(Y)) = \mu(X)$
Varian a mediei, $\sigma_{\bar{Y}}^2$	$\sigma_{\bar{Y}}^2 = \sigma^2(m(Y)) = \sigma^2(X)/n_Y$
Media varian ei, $\mu(s^2)$	$\mu(s^2) = \mu(s^2(Y)) = \sigma^2(X)(n_Y-1)/n_Y$
Varian a varian ei, $\sigma^2(s^2)$	$\sigma^2(s^2) = \sigma^2(s^2(Y)) = \frac{(n_Y-1)^2}{n_Y^3} \mu_4(X) - \frac{(n_Y-1)(n_Y-3)}{n_Y^3} \mu_2^2(X)$

Tabelul 21. Medii i varian e ale mediei i varian ei observabilei Y ce rezult din distribu ia de e antionare din popula ia cu caracteristica X

M rime i nota ie	Aproximare
Media mediei, $\mu_{\bar{Y}}$	$\mu_{\bar{Y}} \cong m(Y)$
Varian a mediei, $\sigma_{\bar{Y}}^2$	$\sigma_{\bar{Y}}^2 \cong s^2(Y)/(n_Y-1)$
Media varian ei, $\mu(s^2)$	$\mu(s^2) \cong s^2(Y)$
Varian a varian ei, $\sigma^2(s^2)$	$\sigma^2(s^2) \cong \frac{(n_Y-1)}{n_Y^2} m_4(Y) - \frac{(n_Y-3)}{n_Y(n_Y-1)} m_2^2(Y)$

Tabelul 22. Valori aproximative pentru mediile i varian ele mediei i varian ei observabilei Y în ipotezele teoremei limit centrale

Un alt caz foarte important de distribu ie este **distribuția de eșantionare**. Extragerea repetat de e antioane (de volum dat) dintr-o popula ie face ca valorile ob inute s surmeze o distribu ie, numit distribu ia de e antionare. Tabelul 21 prezint rezultatele care se ob in pentru varian a m rimilor statistice prin extragerea repetat de e antioane dintr-o popula ie.

Când valorile parametrilor statistici ai popula iei nu sunt cunoscute, dar se poate face presupunerea c distribu ia popula iei se comport suficient de bine [18], ace tia pot fi aproxima i cu ajutorul estimatorilor acestora (Tabelul 19). Formulele de calcul aproximativ ale mediei i varian ei pentru medie i varian sunt redade în Tabelul 22.

Dac se pot asuma ipoteze cu privire la distribu ia caracteristicii X în popula ie, atunci se pot ob ine formule de calcul pentru parametrii statistici (ai popula iei) i folosind rela iile din Tabelul 19

estimatorii parametrilor statistici ai populaiei din m sur torile (statisticile) efectuate asupra e antionului.

Legi de distribu ie i statistici ale acestora

Tabelele 23-41 dau expresiile unor m rimi statistice (valabile pentru popula ie) în timp ce expresiile pentru estimatori se pot ob ine din [Tabelul 19](#).

M rime statistic	Expresie de calcul
Suport	$k \in \{a, a+1, \dots, b-1, b\}$
Minim; Maxim	a; b
Func ia de probabilitate	$1/(b-a+1)$
Func ia de reparti ie	$([k]-a+1)/(b-a+1)$
Media i mediana; varian a	$(a+b)/2; ((b-a+1)^2-1)/12$
Asimetria; excesul de boltire	$0; -\frac{6((b-a+1)^2+1)}{5((b-a+1)^2-1)}$

Tabelul 23. M rimi statistice ale distribu iei discrete uniforme

M rime statistic	Expresie de calcul
Suport	$k \in \{0,1\}; p \in (0,1)$
Minim; Maxim	0; 1
Func ia de probabilitate	$(1-p), k=0$ $p, k=1$
Func ia de reparti ie	$(1-p), k \in [0,1)$ $1, 1-k$
Media; varian a	$p; p(1-p)$
Asimetria; excesul de boltire	$0; (6p^2-6p+1)/(p(1-p))$

Tabelul 24. M rimi statistice ale distribu iei discrete Bernoulli

M rime statistic	Expresie de calcul
Suport	$k \in \{0, \dots, n\}; p \in (0,1)$
Minim; Maxim	0; n
Func ia de probabilitate	$\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$
Func ia de reparti ie	$\sum_{i=0}^k \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$
Media; varian a	$np; np(1-p)$
Asimetria; excesul de boltire	$(1-2p)/\sqrt{np(1-p)}; \frac{1-6p(1-p)}{np(1-p)}$

Tabelul 25. M rimi statistice ale distribu iei discrete binomiale

M rime statistic	Expresie de calcul
Suport	$k = 0, 1, \dots; 0$
Minim; Maxim	0;
Func ia de probabilitate	$e^{-\lambda} \lambda^k / k!$
Func ia de reparti ie	$\sum_{i=0}^k e^{-\lambda} \lambda^i / i!$
Media; varian a	;
Asimetria; excesul de boltire	$1/\sqrt{\lambda}; 1/\lambda$

Tabelul 26. M rimi statistice ale distribu iei discrete Poisson

M rime statistic	Expresie de calcul
Suport; Minim; Maxim	$x \in [a, b]$; a; b
Func ia de probabilitate	$1/(b-a)$
Func ia de reparti ie	$(x-a)/(b-a)$
Media i mediana; varian a	$(a+b)/2$; $(b-a)^2/12$
Asimetria; excesul de boltire	0; -6/5

Tabelul 27. M rimi statistice ale distribu iei continue uniforme

M rime statistic	Expresie de calcul
Suport	$x \in (- ,)$; $x_0 \in (- ,)$; $\gamma \in (0,)$
Minim; Maxim	- ;
Func ia de probabilitate; Func ia de reparti ie	$\frac{1}{\gamma\pi \left(1 + \left(\frac{x-x_0}{\gamma}\right)^2\right)}$; $\frac{1}{\pi} \arctan\left(\frac{x-x_0}{\gamma}\right) + \frac{1}{2}$
Mediana i moda	x_0

Tabelul 28. M rimi statistice ale distribu iei continue Cauchy-Lorentz

M rime statistic	Expresie de calcul
Suport	$x \in (- ,)$; $\gamma \in (0,)$
Minim; Maxim	- ;
Func ia de probabilitate	$\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}$, $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$
Func ia de reparti ie	$\frac{1}{2} + x\Gamma\left(\frac{v+1}{2}\right) \cdot \sum_{n \geq 0} \frac{(-x^2/v)^n}{n!} \prod_{i=0}^{n-1} \frac{(1+2i)(v+1+2i)}{2(3+2i)}$
Media; mediana; moda; varian a	0 (> 1); 0; 0; $v/(v-2)$, > 2
Asimetria; excesul de boltire	0, > 3 ; $6/(v-4)$, > 4

Tabelul 29. M rimi statistice ale distribu iei continue Student t

M rime statistic	Expresie de calcul
Suport	$x \in [0,)$; $d_1, d_2 \in (0,)$
Minim; Maxim	0;
Func ia de probabilitate	$\frac{\Gamma((d_1 + d_2)/2)}{\Gamma(d_1/2)\Gamma(d_2/2)} \frac{(d_1)^{d_1/2} (d_2)^{d_2/2} x^{d_1/2-1}}{(d_1 x + d_2)^{(d_1+d_2)/2}}$, $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$
Func ia de reparti ie	$IB\left(\frac{d_1 x}{d_1 x + d_2}, \frac{d_1}{2}, \frac{d_2}{2}\right) / IB\left(1, \frac{d_1}{2}, \frac{d_2}{2}\right)$, $IB(z, a, b) = \int_0^z t^{a-1} (1-t)^{b-1} dt$
Media; moda	$d_2/(d_2 - 2)$, $d_2 > 2$; $d_2(d_1 - 2)/d_1(d_2 + 2)$, $d_1 > 2$
Varian a; asimetria	$\frac{2d_2^2(d_1 + d_2 - 2)}{d_1(d_2 - 2)^2(d_2 - 4)}$, $d_2 > 4$; $\frac{(2d_1 + d_2 - 2)\sqrt{8(d_2 - 4)}}{(d_2 - 6)\sqrt{d_1(d_1 + d_2 - 2)}}$, $d_2 > 6$
Excesul de boltire	$\frac{3d_2^3 + (5d_1 - 8)d_2^2 + (5d_1^2 - 32d_1 + 20)d_2 - 22d_1^2 + 44d_1 - 16}{d_1(d_2 - 6)(d_2 - 8)(d_1 + d_2 - 2)/12}$, $d_2 > 8$

Tabelul 30. M rimi statistice ale distribu iei continue Fisher-Snedecor F

M rime statistic	Expresie de calcul
Suport	$x \in [0, \infty); d \in (0, \infty)$
Minim; Maxim	0;
Func ia de probabilitate	$(1/2)^{d/2} x^{d/2-1} e^{-x/2} / \Gamma(d/2), \Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$
Func ia de reparti ie	$\int_0^{x/2} t^{d/2-1} e^{-t} dt / \Gamma(d/2)$
Media; mediana; moda; varian a	$d; \cong d - 2/3; d - 2, d > 2; 2d$
asimetria; excesul de boltire	$\sqrt{8/d}; 12/d$

Tabelul 31. M rimi statistice ale distribu iei continue ²

M rime statistic	Expresie de calcul
Suport	$x \in [0, \infty); \lambda \in (0, \infty)$
Minim; Maxim; Func ia de probabilitate; Func ia de reparti ie	0; ; $\lambda e^{-\lambda x}; 1 - e^{-\lambda x}$
Media; mediana; moda; varian a; asimetria; excesul de boltire	$1/\lambda; \ln(2)/\lambda; 0; 1/\lambda^2; 2; 6$

Tabelul 32. M rimi statistice ale distribu iei continue exponen iale

M rime statistic	Expresie de calcul
Suport	$x \in [0, \infty); k \in (0, \infty)$
Minim; Maxim	0;
Func ia de probabilitate; func ia de reparti ie	$kx^{k-1} e^{-(x/\lambda)^k} / \lambda^k; 1 - e^{-(x/\lambda)^k}$
Media; mediana; moda	$\mu = \lambda \Gamma(1 + 1/k); \lambda (\ln(2))^{1/k}; \lambda ((k-1)/k)^{1/k}, k > 1$
Varian a; asimetria	$\sigma^2 = \lambda^2 \Gamma(1 + 2/k) - \mu^2; \gamma_1 = (\Gamma(1 + 3/k) \lambda^3 - 3\mu \sigma^2 - \mu^3) / \sigma^3$
Excesul de boltire	$\gamma_2 = (\lambda^4 \Gamma(1 + 4/k) - 4\gamma_1 \sigma^3 \mu - 6\mu^2 \sigma^2 - \mu^4) / \sigma^4$

Tabelul 33. M rimi statistice ale distribu iei continue Weibull

M rime statistic	Expresie de calcul
Suport	$x \in [0, \infty); \mu \in (-\infty, \infty); \sigma \in (0, \infty)$
Minim; Maxim	0;
Func ia de probabilitate	$e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} / (x\sigma\sqrt{2\pi})$
Func ia de reparti ie	$\frac{1 + \operatorname{erf}\left(\frac{(\ln(x)-\mu)}{\sigma\sqrt{2}}\right)}{2}; \operatorname{erf}(z) = 2 \int_0^z e^{-t^2} dt / \sqrt{\pi}$
Media; mediana; moda; varian a	$e^{\mu+\sigma^2/2}; e^\mu; e^{\mu-\sigma^2}; (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$
Asimetria; excesul de boltire	$(e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}; e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 6$

Tabelul 34. M rimi statistice ale distribu iei continue Log-normale

M rime statistic	Expresie de calcul
Suport	$b \in (0, \infty); \mu, x \in (-\infty, \infty)$
Minim; Maxim	- ;
Func ia de probabilitate	$e^{- x-\mu /b} / 2b$
Func ia de reparti ie	$e^{(x-\mu)/b} / 2, x < \mu$ $1 - e^{-(x-\mu)/b} / 2, \mu \leq x$
Media; mediana; moda; varian a	$\mu; \mu; \mu; 2b^2$
Asimetria; excesul de boltire	0; 3

Tabelul 35. M rimi statistice ale distribu iei continue Laplace (dublu exponen ial)

M rime statistic	Expresie de calcul
Suport	$\mu, \gamma \in (0, \infty); x \in (\mu, \infty)$
Minim; Maxim	μ ;
Func ia de probabilitate	$\frac{\sqrt{\frac{x-\mu}{\beta}} + \sqrt{\frac{\beta}{x-\mu}}}{2\gamma(x-\mu)} N_{0,1}\left(\left(\sqrt{\frac{x-\mu}{\beta}} - \sqrt{\frac{\beta}{x-\mu}}\right)/\gamma\right)$
Func ia de probabilitate standard	$\frac{\sqrt{x} + \sqrt{1/x}}{2\gamma(x-\mu)} N_{0,1}\left((\sqrt{x} - \sqrt{1/x})/\gamma\right), N_{0,1}(z) = \int_{-\infty}^z \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$
Func ia de reparti ie standard	$N_{0,1}\left((\sqrt{x} - \sqrt{1/x})/\gamma\right)$
Media; varian a (standard)	$1 + \gamma^2/2; \gamma\sqrt{1 + 5\gamma^2/4}$

Tabelul 36. M rimi statistice ale distribu iei continue Birnbaum-Saunders (a vie ii obosite)

M rime statistic	Expresie de calcul
Suport	$k, \theta \in (0, \infty); x \in [0, \infty)$
Minim; Maxim	0;
Func ia de probabilitate	$x^{k-1} e^{-x/\theta} \theta^{-k} / \Gamma(k), \Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$
Func ia de reparti ie	$\frac{\int_0^{x/\theta} t^{k-1} e^{-t} dt}{\int_0^{\infty} t^{k-1} e^{-t} dt}$
Media; moda; varian a	$k; (k-1)^{-1/k}, k > 1; k^{-2}$
Asimetria; excesul de boltire	$2/\sqrt{k}; 6/k$

Tabelul 37. M rimi statistice ale distribu iei continue Gamma

M rime statistic	Expresie de calcul
Suport	$\mu \in (0, \infty); x \in (-\infty, \infty)$
Minim; Maxim	- ;
Func ia de probabilitate	$\exp(-\exp(-(x-\mu)/\beta))/\beta \exp(-(x-\mu)/\beta)/\beta$
Func ia de reparti ie	$\exp(-\exp(-(x-\mu)/\beta))$
Media; mediana; moda; varian a	$\mu + \beta; \mu - \beta \cdot \ln(\ln(2)); \mu; \beta^2/6$
Asimetria; excesul de boltire	$\frac{12\sqrt{6}\zeta(3)}{\pi^3} \cong 1.14; 12/5$

Tabelul 38. M rimi statistice ale distribu iei continue Gumbel (log-Weibull)

M rime statistic	Expresie de calcul
Suport	$\alpha, \beta \in (0, \infty); x \in [0, 1]$
Minim; Maxim	0; 1
Func ia de probabilitate	$x^{\alpha-1} (1-x)^{\beta-1} / IB(1, \alpha, \beta); IB(z, a, b) = \int_0^z t^{a-1} (1-t)^{b-1} dt$
Func ia de reparti ie	$IB(x, \alpha, \beta) / IB(1, \alpha, \beta)$
Media; moda; varian a	$\frac{\alpha}{\alpha+\beta}; \frac{\alpha-1}{\alpha+\beta-2}, \alpha > 1; \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Asimetria; excesul de boltire	$\frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}; \frac{\alpha^3 - (2\beta-1)\alpha^2 - 2\alpha\beta(\beta+2) + (\beta+1)\beta^2}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)/6}$

Tabelul 39. M rimi statistice ale distribu iei continue Beta

M rime statistic	Expresie de calcul
Suport	$\in (0, \infty); \mu, \sigma \in (-\infty, \infty)$
Minim; Maxim	- ;
Func ia de probabilitate	$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)/(\sigma\sqrt{2\pi})$
Func ia de reparti ie	$\left(1 + \operatorname{erf}\left(\frac{(x-\mu)/(\sigma\sqrt{2})}{1}\right)\right)/2$; $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$
Media; moda; varian a	$\mu; \mu; \mu; \sigma^2$
Asimetria; excesul de boltire	0; 0

Tabelul 40. M rimi statistice ale distribu iei continue Gauss (normale)

M rime	Popula ie (finit) de volum n_X	E antion de volum n_Y	Estimator
Media	$\mu_{\bar{X}} = \mu; \sigma_{\bar{X}}^2 = \sigma^2/n_X$	$\mu_{\bar{Y}} = m; \sigma_{\bar{Y}}^2 = s^2/n_Y$	$m; s^2/(n_Y-1)$
Varian a	$\frac{(n_X-1)^2/n_X}{n_X^3 \mu_4^{-1} - n_X^3(n_X-3)^{-1}}$	$\frac{(n_Y-1)s^2/n_Y}{n_Y^3 m_4^{-1} - n_Y^3(n_Y-3)^{-1}}$	$\frac{s^2}{n_Y-1} \frac{(n_Y-1)m_4 - (n_Y-3)m_2^2}{n_Y^2(n_Y-1)}$ $= \frac{2\sigma^4(n_Y-1)}{n_Y^2} \cong \frac{2s^4}{n_Y-1}$
Var σ_1^2	$\frac{6n_X(n_X-1)}{(n_X-2)(n_X+1)(n_X+3)}$	$\frac{6n_Y(n_Y-1)}{(n_Y-2)(n_Y+1)(n_Y+3)}$	$c_4^2 \cdot \operatorname{var}(g_1)$ $c_4 - \text{Vezi Tabelul 29}$
Var σ_2^2	$\frac{24n_X(n_X-1)^2(n_X-3)^{-1}}{(n_X-2)(n_X+3)(n_X+5)}$	$\frac{24n_Y(n_Y-1)^2(n_Y-3)^{-1}}{(n_Y-2)(n_Y+3)(n_Y+5)}$	$c_4^2 \cdot \operatorname{var}(g_2)$ $c_4 - \text{Vezi Tabelul 29}$

Tabelul 41. Alte m rimi statistice ale distribu iei continue Gauss (normale)

Agrementele între observa ie i model: statistici

Fie $Y = (Y_1, \dots, Y_n)$ i $X = (X_1, \dots, X_n)$ serie de observa ii pereche i obiectivul s fie g sirea unei func ii $f(x, a_1, \dots, a_m)$, pentru care $Y = f(X)$ este cea mai bun solu ie posibil a aproximare $\sim Y$. Atingerea acestui obiectiv presupune g sirea expresiei func iei f i a valorilor parametrilor a_1, \dots, a_m . Sub ipoteza de acord între observa ie i model, expresia func iei f se presupune a fi cunoscut (sau cel pu in ar trebui, atunci când se desf oar o c utare dup un anumit set de expresii alternative).

Astfel, ar mas ob inerea valorilor parametrilor a_1, \dots, a_m . Pentru a avea o solu ie unic pentru valorile parametrilor a_1, \dots, a_m , cel pu in se cere ca $m \leq n$ s fie asigurat . O serie de alternative sunt disponibile pentru aproximarea $Y \sim f(X)$, iar cele considerate cele mai importante sunt exemplificate în continuare.

Minimizarea erorii de acord (minimizarea dezacordului)

În această ipotez o serie de alternative sunt disponibile (ecua ia de mai jos, cu diferite alegeri pentru p i q):

$$S(p, q) = \sum_{i=1}^n |Y_i - f(X_i)|^p / f^q(X_i) = \min., q = 0, 1, p/2, p$$

În cazul în care seria Y reprezint frecven a (nenul) a observa iilor distincte X , atunci $f(x)$ ar trebui s fie o func ie pozitiv de asemenea, i modul num r torului nu mai este necesar.

Distribu ia Gauss [19] a valorilor termenilor sumei sunt traduse în $p = 2$, i distribu ia Laplace [20] apare când $p = 1$ (vezi i [21]). Func ia densitate de probabilitate (PDF) a unor reprezentan i ai familiei care con ine distribu iile Gauss i Laplace standard ($\mu = 0$ i $\sigma = 1$), sunt exemplificate în fig. 11. Minimizarea erorii de acord pentru valori p diferite ofer solu ii diferite pentru parametrii, i cum se poate observa în figura 1, sunt asociate cu diferite forme de eroare.

$$GL(0;0.5) = \sqrt{15/2} \cong 2.739$$

$$GL(0;1.0) = \sqrt{1/2} \cong 0.707$$

$$GL(0;2.0) = 1/\sqrt{2\pi} \cong 0.399$$

$$GL(0;3.0) = \dots \cong 0.342$$

$$GL(0;4.0) = \Gamma^2(3/4) \cdot 2^{1/4} \cdot \pi^{-3/2} \cong 0.321$$

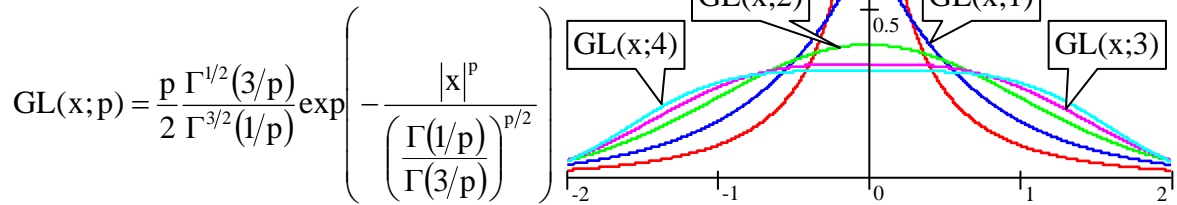


Fig.11. Distribu ia Gauss-Laplace

Dou cazuri particulare sunt de obicei utilizate pentru a estima parametrii necunoscu i ai unei distribu ie atunci c ând $p = 2$ (vezi [22] i [23]). Abordarea cea mai general de ob inere a parametrilor de distribu ie este de a ghici valorile sau s aplice o procedur care reduce în fiecare etap cantitatea dat de ecua ia (1) pân c ând cantitatea redus este mult mai mic decât cea r mas .

Utilizarea momentelor

Sub ipoteza c $Y \sim f(X)$ ar trebui s fie chiar mai precis (a doua ipotez fiind caracterul aleatoriu al erorii, cu o medie de zero) i aproximarea $X_i^k Y_i \sim X_i^k \cdot f(x_i)$ este folosit pentru $k = 0$. Metoda momentelor d greutate maxim la primele momente, astfel, o solu ie a_1, \dots, a_m a $Y \sim f(X)$ poate proveni direct din ecua ia $X_i^k Y_i = X_i^k \cdot f(x_i)$. Modul cel mai convenabil pentru cazul general este repetarea c ut rii parametrilor a_1, \dots, a_m începând de la anumite valori ini iale):

$$\sum_{i=1}^n X_i^k Y_i \sim \sum_{i=1}^n X_i^k f(X_i), k = 0, 1, \dots$$

Utilizarea momentelor centrale

Se înt re te aproximarea $X_i Y_i \sim X_i \cdot f(X_i)$ i $(Y_i - \bar{Y})^k \sim (X_i - \bar{X})^k \cdot f(X_i)$ pentru $k \geq 2$. Metoda momentelor d greutate maxim primelor momente centrale, deci o solu ie a_1, \dots, a_m a $Y \sim f(X)$ poate proveni din ecua ia:

$$\sum_{i=1}^n X_i Y_i \sim \sum_{i=1}^n X_i f(X_i) \text{ and } \sum_{i=1}^n (Y_i - \bar{Y})^k \sim \sum_{i=1}^n (X_i - \bar{X})^k f(X_i), k = 2, 3 \dots$$

Folosind statisticile referitoare la popula ie

O u oar modificare a metodei anterioare poate beneficia de disponibilitatea expresiei pentru media, devia ia standard, asimetria i excesul asimetriei popula iei în func ie de parametrii distribu iei pentru un num r mare de distribu ii bine cunoscute.

Estimare pe baza șansei maxime

Principiul ansei maxime este c o estimare rezonabil pentru un parametru este cea care maximizeaz probabilitatea (P în ecua ia de mai jos) de observare a datelor experimentale [24], iar setul mai probabil al parametrilor a_1, \dots, a_m va face P maxim (i implicit MLE este maxim).

$$MLE = \log_2(P) = \sum_{i=1}^n \log_2(f(X_i))$$

Statistica Benford

Testul Benford folose te distribu ia Z (normal) pentru a verifica ipoteza c un ir de numere urmeaz distribu ia Benford, frecven ele dup care se distribuie o anume cifr a fiec rui num r din ir.

Un ir de numere urmeaz distribu ia Benford dac probabilitatea de distribu ie a unei cifre (d_i)

a numerelor ($d=d_0d_1\dots$) reprezentate în baza de numera ie b (uzual baza 10) urmeaz legea (Benford):

$$\begin{aligned}
 p(d_0) &= \log_b \left(1 + \frac{1}{d_0} \right), d_0 = 1..(b-1); \\
 p(d_1) &= \sum_{k=1}^{b-1} \log_b \left(1 + \frac{1}{k \cdot b + d_1} \right), d_1 = 0..(b-1) \\
 p(d_2) &= \sum_{j=1}^{b-1} \sum_{k=0}^{b-1} \log_b \left(1 + \frac{1}{j \cdot b^2 + k \cdot b + d_2} \right), d_2 = 0..(b-1) \\
 &\dots
 \end{aligned}$$

Fig.12. Statistica Benford

Ipoteza acestei legi de distribu ie este c valorile m sur torilor rezultate din observa ie sunt frecvent distribuite logaritmic i astfel logaritmul setului de m sur tori este distribuit uniform. Legea de distribu ie este numit dup fizicianul Frank BENFORD care a formulat-o intuitiv în 1938 [25], dar demonstra ia acesteia a fost dat mult mai târziu [26].

Acest rezultat intuitiv de num rare a apari iilor a fost g sit aplicându-se la o mare varietate de seturi de date incluzând facturile la electricitate, adresele de str zi, pre urile ac iunilor, numerele popula iei, ratele de deces, lungimile râurilor, constante fizice i matematice i procesele descrise de legi putere (care sunt foarte comune în natur). Este foarte important de tiut c rezultatul (odat observat într-o baz de numera ie) are loc independent de baza de numera ie în care se exprim numerele, chiar dac propor iile de reprezentare se schimb . De aici, **acest rezultat poate fi folosit pentru a verifica datele în suspiciunea de alterare (mistificare) a acestora prin compararea frecvențelor teoretice cu cele observate pentru prima cifră a acestora.**

Statistica Jarque-Bera

Testul Jarque-Bera [27, 28] calculeaz i atribuie probabilitatea statistic ca valorile unui e antion ce provine din popula ie normal distribuit s î i abat simultan asimetria i excesul de boltire de la valorile teoretice corespunz toare distribu iei normale.

Statistica Jarque-Bera se calculeaz cu rela ia:

$$JB = n \frac{g_1^2 + g_2^2 / 4}{6}$$

în care g_1 este asimetria, g_2 este excesul de boltire i n este volumul e antionului.

Statistica JB are o distribu ie asimptotic c tre $^2(df=2)$.

g_1 , Asimetria observabilei Y	Un e antion	$g_1 = m_3/m_2^{3/2}$	$m_k = E_k(Y), k>1$
b_2 , Boltirea observabilei Y		$b_2 = m_4/m_2^2$	
g_2 , Excesul de boltire al observabilei Y		$g_2 = b_2 - 3$	

Statistica Kolmogorov-Smirnov

Testul Kolmogorov-Smirnov [29] poate fi folosit pentru verificarea ipotezei c un e antion de date urmeaz o anumit lege de distribu ie (redat în continuare), precum i pentru compararea legilor de distribu ie ale popula iilor din care provin dou e antioane [30].

Statistica Kolmogorov-Smirnov verific dac observa iile independente $X=(X_i)_{1 \leq i \leq n}$ provin dintr-o popula ie ce urmeaz legea de distribu ie dat de func ia cumulativ de probabilitate $F(x)$ prin calcularea maximumului diferen ei absolute între $F_i(x)$ i func ia cumulativ de probabilitate observat $F_0(x)$ în toate punctele observa iei:

$$D = \max_{1 \leq i \leq n} |F_i(X_i) - F_0(X_i)| \quad (\text{K-S Stat})$$

Distribuția Kolmogorov

Legea de distribuție Kolmogorov se obține pentru variabila aleatoare K dată de:

$K = \max_{0 \leq t \leq 1} B(t) ,$ <p>unde B este puntea Brownian condiționat de:</p> $B(0) = B(1) = 0$ $M(B(t)) = 0$ $\text{Var}(B(t)) = t(t-1)$ $P(K \leq x) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 x^2} = \frac{\sqrt{2\pi}}{x} \sum_{i=1}^{\infty} e^{-(2i-1)^2 \pi^2 / 8x^2}$	(K-S Dist)
--	------------

Testul Kolmogorov-Smirnov

Ipoteza testului este că următoarea convergență are loc în distribuție:

$D\sqrt{n} \rightarrow \sup_{t \in [0,1]} B(F(t)) $ <p>Ipoteza se respinge la nivelul de semnificație dacă :</p> $D\sqrt{n} > K_{\alpha}, \text{ unde } K : P(K \leq K_{\alpha}) = 1 - \alpha$	(K-S Test)
---	------------

Pentru compararea a două distribuții observate:

$D = \max_{1 \leq i \leq \max(n,m)} F_{o1}(X_i) - F_{o2}(X_i) $ <p>Ipoteza se respinge la nivelul de semnificație dacă :</p> $D\sqrt{\frac{mn}{m+n}} > K_{\alpha}$	(K-S Test)
---	------------

Statistica Anderson-Darling

Testul Anderson-Darling [31] verifică dacă este o evidență statistică că un eantion s nu provine dintr-o funcție de probabilitate dată.

Statistica Anderson verifică dacă asupra observațiilor distincte ordonate crescător $(X_i)_{1 \leq i \leq n}$, $X_i < X_{i+1}$ se poate respinge ipoteza că provine dintr-o distribuție dată de funcția cumulativă de probabilitate F calculând valoarea A dată de relația:

$$A^2 = -n - \sum_{k=1}^n \frac{2k-1}{n} (\ln(F(Y_k)) + \ln(1 - F(Y_{n+1-k})))$$

O aplicație de interes înșă reprezintă testul Anderson-Darling pentru mai multe eantioane asupra cărora se poate verifica proveniența din aceeași populație, caz în care legea de distribuție a populației nu mai trebuie să fie specificată [32,33]. Formulele de calcul și interpretarea testului pentru compararea de eantioane se găsesc la adresa [34].

În cazul comparării unei legi de distribuție discrete cunoscute cu legea de distribuție observată eantion varianța statisticii A^2 se calculează cu formula ([35], n - numărul de observații din eantion, =3.1415926535897932384626434...):

$$\text{Var}(A^2) = \frac{2(\pi^2 - 9)}{3} + \frac{10 - \pi^2}{n}$$

În cazul verificării ipotezei de normalitate, este posibil să se aproximeze probabilitatea de observație asociată valorii statisticii A^2 [36]. Se aplică corecția de volum al eantionului:

$$A^2_c = A^2(1 + 0.75/n + 2.25/n^2)$$

$$p = \begin{cases} 1 - \exp(-13.436 + 101.14 \cdot x - 223.73 \cdot x^2), & x < 0.2 \\ 1 - \exp(-8.318 + 42.796 \cdot x - 59.938 \cdot x^2), & 0.2 \leq x < 0.34 \\ \exp(0.9177 - 4.279 \cdot x - 1.38 \cdot x^2), & 0.34 \leq x < 0.6 \\ \exp(1.2937 - 5.709 \cdot x + 0.0186 \cdot x^2), & x \leq 0.6 \end{cases}$$

Statistica Pearson-Fisher Chi Square

Experimentul varianțelor ce conduce la distribuția χ^2

Distribuția χ^2 a fost descoperită de Karl PEARSON [37] în urma încercării de a explica varianța observată a numerelor care provin din distribuția normală.

Astfel, dacă se consideră distribuția normală standard $N(0,1)$ și variabila întâmplătoare X ce urmează această distribuție (Figura 13), probabilitatea (dp) de a extrage valorile $-x$ și x din $N(0,1)$ sunt ambele egale și egale cu diferențiala funcției de densitate de probabilitate a distribuției normale ($PDF_{N(0,1)}$).

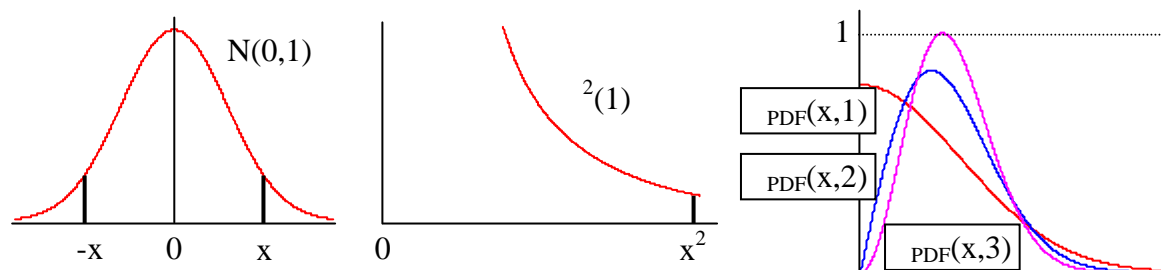


Figura 13. Funcțiile de densitate de probabilitate (PDFs) pentru $N(0,1)$, $\chi^2(1)$ și $\chi^2(k)$

$$PDF_{N(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (1)$$

Distribuția normală standard are media 0; astfel, pentru a exprima probabilitatea de observație pentru deviația x^2 trebuie adunate două probabilități (pentru $-x$ și x) date de relația (1):

$$dp(x^2) = 2 \cdot dPDF_{N(0,1)}(x) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \quad (2)$$

Pentru a reconstitui PDF pentru x^2 trebuie să efectuăm o schimbare de variabilă $x^2 = t$; atunci $x = \sqrt{t}$:

$$dp(t) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{t}{2}\right) d\sqrt{t} = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{t}{2}\right) \frac{1}{2\sqrt{t}} dt \quad (3)$$

Este ușor de verificat că (3) este un caz particular al lui (4) când $k = 1$:

$$\chi^2_{PDF}(t, k) = \frac{1}{2^{k/2} \Gamma(k/2)} t^{k/2-1} \exp\left(-\frac{t}{2}\right) \quad (4)$$

Procedura descrisă mai sus corespunde pentru distribuția Chi Square cu un grad de libertate (extragerea lui X din distribuția normală). Dacă sunt extrase mai multe valori (k valori) din distribuția normală atunci se obține distribuția Chi Square cu k grade de libertate, și demonstrația că ecuația (4) este adevărată poate fi găsită în [38].

Calea directă de la distribuția normală la distribuția χ^2 nu este reversibilă (Figura 13); astfel, definind variabila φ în relația (5) - ce reprezintă o expresie modificată a coeficientului de asociere definit de LIEBETRAU [39]:

$$\varphi = \varphi(X^2, k) = \sqrt{\frac{X^2}{k}} \quad (5)$$

obținerea distribuției lui φ se poate obține pe o cale similară cu cea descrisă mai sus; notând $u = \varphi$ în (5) și substituind $t = X^2 = ku^2$ în (4) se obține ($du^2 = 2u \cdot du$):

$$d\chi^2(ku^2, k) = \frac{1}{2^{k/2} \Gamma(k/2)} (ku^2)^{k/2-1} \exp\left(-\frac{ku^2}{2}\right) d(ku^2) \quad (6)$$

După rearanjarea termenilor:

$$d\varphi_{PDF}(u, k) = \frac{2u^{k-1}}{\Gamma(k/2)} \left(\frac{k}{2}\right)^{k/2} \exp\left(-\frac{u^2}{2/k}\right) du \quad (7)$$

Pornind de la densitatea de probabilitate (PDF) a distribuției Gamma:

$$\Gamma_{\text{PDF}}(x; a, b, c) = \frac{cx^{ca-1}}{b^{ca}\Gamma(a)} \exp\left(-\left(x/b\right)^c\right) \quad (8)$$

este u or de verificat c :

$$\varphi_{\text{PDF}}(x, k) = \Gamma_{\text{PDF}}\left(x, \frac{k}{2}, \sqrt{\frac{2}{k}}, 2\right) \quad (9)$$

Rela ia (9) demonstreaz c distribu ia lui $\sqrt{\frac{X^2}{k}}$ este un caz particular al distribu iei Gamma (Figura 13).

Testul χ^2 ca măsură a independenței, omogenității și asocierii în distribuție

Distribu ia χ^2 are 3 aplica ii imediate:

- ÷ Testul Chi Square pentru verificarea independen ei
 - testeaz asocierea între dou variabile cu valori grupate pe categorii;
 - se poate aplica dac au loc dou condi ii:
 - nici una din valorile a teptate nu este mai mic decât 1;
 - nu mai mult de 20% din valorile a teptate nu sunt mai mici de 5;
 - ipotezele de lucru sunt: nu exist nici o asociere între cele dou variabile (ipoteza nul) i este o asociere între cele dou variabile (ipoteza contrar);
 - Când statistica Chi Square (X^2) este mai mare decât valoarea func iei cumulative de probabilitate a distribu iei Chi Square (χ^2) pentru num rul de grade de libertate egal cu num rul de cazuri minus unu i pentru riscul de a fi în eroare (nivelul de semnifica ie) ales, atunci exist o diferen semnificativ de la ipoteza lipsei de asociere i cele dou variabile sunt asociate;
- ÷ Testul Chi Square pentru verificarea omogenit ii
 - testeaz dac mai multe popula ii sunt similare (sau omogene sau egale) în anumite caracteristici (acele caracteristici care sunt incluse în testare);
 - ipotezele de lucru sunt: popula iile sunt similare (sau omogene sau egale) în caracteristica supus observa iei (ipoteza nul) i popula iile sunt diferite în caracteristic (ipoteza contrar);
 - uzual caracteristica supus observa iei este un moment central (ex. valoare medie, varian);
- ÷ Testul Chi Square pentru verificarea asocierii în distribu ie
 - testeaz dac un model teoretic poate fi asociat observa iilor;
 - ipotezele de lucru sunt: datele observate urmeaz distribu ia dat de modelul teoretic (ipoteza nul) i datele observate nu provin dintr-o popula ie ce urmeaz modelul teoretic (ipoteza contrar);

Probleme frecvente în aplicarea testului χ^2 ca măsură a asocierii în distribuție

Testul χ^2 , propus ca m sur a dep rt rii întâmpl toare între observa ie i modelul teoretic de Karl PEARSON [37] a fost corectat în interpretare de Ronald FISHER prin reducerea num rului de grade de libertate corespunz tor cu o unitate (datorit estim rii frecven ei teoretice din frecven a observat , [40]), i cu num rul parametrilor necunoscu i ai distribu iei teoretice estima i din observa ii din m suri ale tendin ei centrale ([41]).

Testarea agrementului între observa ie i ipotez se realizeaz prin divizarea observa iilor într-un num r definit de intervale (n), pentru care se calculeaz expresia X^2 (unde s este num rul de parametri ai distribu iei teoretice estima i din momente centrale, O_i este frecven a experimental observat în clasa de frecven i , E_i este frecven a a teptat calculat din legea de distribu ie teoretic pentru clasa de frecven i , X^2 este valoarea statisticii chi square iar χ^2 este valoarea parametrului statistic chi square din distribu ia cu acela i nume):

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \approx \chi^2(n-s-1) \quad (10)$$

Pe baza distribuției teoretice χ^2 se calculează probabilitatea de respingere a ipotezei de agrement. Uzual ipoteza de agrement este acceptată dacă probabilitatea de respingere a ipotezei de agrement ($P_{CDF}(\chi^2, n-s-1)$) este mai mică de 5%.

În ciuda faptului că testul χ^2 este cea mai cunoscută statistică pentru verificarea agrementului între observațiile și ipoteza, testarea independenței și a omogenității, definirea cadrului de aplicare al acesteia este dintre cele mai complexe [42].

O serie de probleme la compararea unei distribuții observate cu o distribuție teoretică apar în calcularea statisticii χ^2 și în aplicarea testului χ^2 .

O primă problemă este alegerea numărului de clase de frecvență și există mai multe soluții, dintre care două sunt:

- ÷ calcularea prin rotunjire a numărului de clase de frecvență din entropia Hartley [43] a observației vs. așteptată: $\log_2(2N)$, unde N este numărul de observații (EasyFit [44] folosește această procedură);
- ÷ calcularea numărului de clase de frecvență odată cu lățimea clasei folosind histograma ca estimator al densității [45] și alegerea pe baza acesteia a criteriului optimal pentru lățimea clasei (Dataplot [46] generează automat clasele de frecvență folosind această regulă: lățimea clasei de frecvență este $0.3 \cdot s$ unde s este deviația standard a ecartului; limitele inferioară și superioară sunt date de medie $\pm 6 \cdot s$ și clasele de frecvență observate marginale sunt omise);

O a doua problemă este lățimea claselor de frecvență; și aici există cel puțin două abordări:

- ÷ Datele pot fi grupate în clase de frecvență de probabilitate (teoretică sau observată) egal;
- ÷ Datele pot fi grupate în intervale de lățime egal;

Prima abordare (probabilitate egală) este mai frecvent adoptată deoarece este o soluție mai bună pentru observații foarte grupate.

O altă problemă este numărul de observații din interiorul fiecărei clase de frecvență. Fiecare clasă de frecvență trebuie să conțină cel puțin 5 observații, astfel încât în practică clase de frecvență alăturate se reunesc pentru a satisface această impunere.

Măsuri ale tendinței centrale

Fie X un ir de n valori X_1, X_2, \dots, X_n . Pentru calculul valorii medii, următorii indicatori sunt cei mai folosiți:

- ÷ Media Aritmetică, $AM(X)$, dată de:

$$AM(X) = \frac{\sum_{i=1}^n X_i}{n}$$

- ÷ Media Geometrică, $GM(X)$, obținută din expresia (de notat că pentru n par, expresia pentru GM poate fi nedeterminată, când produsul X_i este negativ):

$$GM(X) = \sqrt[n]{\prod_{i=1}^n X_i} = \exp(AM(\ln(X)))$$

- ÷ Media Armonică, $HM(X)$, dată de:

$$HM(X) = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}} = 1 / AM(1/X)$$

- ÷ Media lui Euler, $EM(X)$, calculată ca:

$$EM(X) = \sqrt{\frac{\sum_{i=1}^n X_i^2}{n}} = \sqrt{AM(X^2)}$$

- ÷ Valoarea mediană, $m(X)$, este numărul (π a.î. X_π este ir ordonat):

$$m(X) = \begin{cases} \left(X_{\pi(\frac{n}{2})} + X_{\pi(\frac{n}{2}+1)} \right) / 2, & \text{pt. } n \text{ par} \\ X_{\pi(\frac{n+1}{2})}, & \text{pt. } n \text{ impar} \end{cases}$$

÷ Valorile la mod sunt numerele date de:

$$\tilde{X} = \{X_i \mid f_i = \sup\{f_j, 1 \leq j \leq n\}, f_j \text{ frecventa aparitiei lui } X_j\}$$

Binary and multinomial variables analysis: binomial confidence intervals

The origins of the mathematical study of natural phenomena are found in the fundamental work [47] of Isaac Newton [1643-1727].

Even if first studies about binomial expressions were made by Euclid [48], the mathematical basis of the binomial distribution study was put by Jacob Bernoulli [1654-1705], of which studies of especially significance for the theory of probabilities [49] was published 8 years later after his death by his nephew, Nicolaus Bernoulli. In *Doctrinam de Permutationibus & Combinationibus* section of this fundamental work he demonstrates the Newton binomial series expansion. Later, Abraham De Moivre [1667-1754] put the basis of approximated calculus for, using the normal distribution for binomial distribution approximation [50]. Later, Johann Carl Friedrich Gauss [1777-1855] with work [51] put the basis of mathematical statistics. Abraham Wald [1902-1950, born in Cluj] do his contributions on binomial confidence intervals approximation, elaborated and published the confidence interval that carry his name now [52].

Nowadays, the most prolific researcher on confidence intervals domain is Allan Agresti, which it was named the Statistician of the Year for 2003 by American Statistical Association, and at the prize ceremony (October 14, 2003) it spoken about Binomial Confidence Intervals [53, 54, 55, 56].

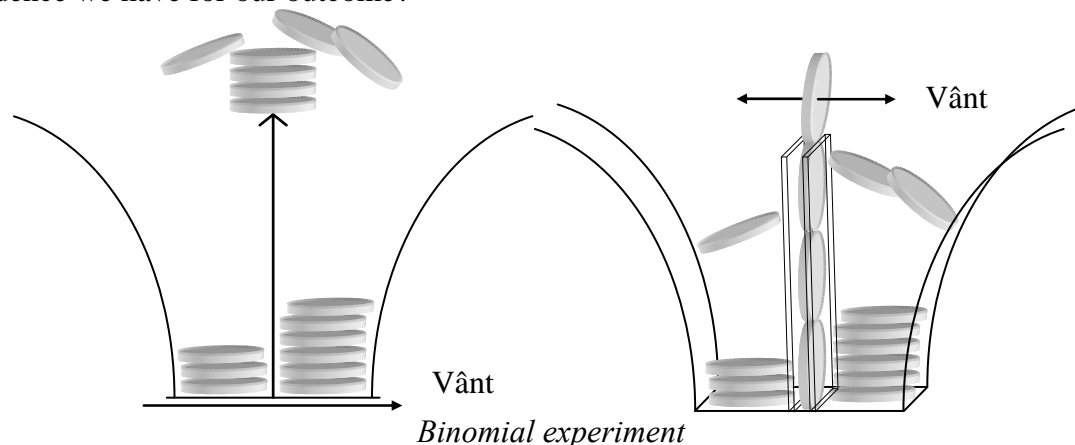
Binary observations always generate binomial distribution. A broad range of experiments are subject of binomials. Thus, law of binomial distribution are proved at heterometric bands of tetrameric enzyme in [57], the stoichiometry of the donor and acceptor chromophores implied in enzymatic ligand/receptor interactions in [58], translocation and exfoliation of type I restriction endonucleases in [59], biotinidase activity on neonatal thyroid hormone stimulator in [60], the parasite induced mortality at fish in [61], the occupancy/activity for proteins at multiple nonspecific sites containing replication in [62].

Finally, let us to give a reference to a very good essay about the frame of binomial distribution model applied to the natural phenomena [63].

(binomial design and model)

Let us give now a close look at the binomial experiment (see *Binomial experiment*). In a binary experiment, two values may come from observation, and may be encoded as 0 and 1 in the informational space (let's say if the coin goes to the left, then is recorded 0 or it goes to the right, and then is recorded 1. In a binomial experiment, we count the number of 0's and 1's from a series of n repetitions of binary experiments (let's say from n fallen coins in our *Binomial experiment*).

If from the total time t of running the experiment, wind has t_1 time the direction from left to right and t_2 the direction from right to left, the expectance is to found about $100 \cdot t_1 / (t_1 + t_2) \%$ coins in the right basket and $100 \cdot t_2 / (t_1 + t_2) \%$ coins in the left basket, but the truth is that we will see a such proportion only if we will spend enough time counting the coins. Going further, we may want to estimate the wind direction ratio from counting the coins, but then what level of confidence we have for our outcome?



The previous experiment is a construction meant to estimate a ratio between two real variables (times of wind having a certain direction) in which the estimation is given from a repeated binary experiment. As longer time we will spend on counting fallen coins as better accuracy will have on estimating times ratio. This *experiment* proves that is no difference in the quality of the measurement encoded by binary values than the measurement encoded by the real values (e.g. any real number can be encoded as a succession of 0's and 1's as in our experiment. More than that, counting the coins can be even more accurate than any other instrument if the time goes to infinity.

Binomial distribution

Coins fallen	In the right basket (probability)	In the left basket (probability)	Explanation from observation of the right basket
1	0 (p)	1 (p)	0 = 0
	1 (1-p)	0 (1-p)	1 = 1
2	0 (p ²)	2 (p ²)	0 = 0 + 0
	1 (p·(1-p)·2)	1 (p·(1-p)·2)	1 = 0 + 1 = 1 + 0
	2 ((1-p) ²)	0 ((1-p) ²)	2 = 1 + 1
...
n	0 (p ⁿ)	n (p ⁿ)	0 = 0 + 0 + ... + 0
	1 (p ⁿ⁻¹ ·(1-p)·(n))	n-1 (p ⁿ⁻¹ ·(1-p)·(n))	1 = 1 + 0 + ... + 0 = ... = 0 + ... + 0 + 1

	k	n-k	?

	n-1 (p·(1-p) ⁿ⁻¹ ·(n))	1 (p·(1-p) ⁿ⁻¹ ·(n))	n-1 = 0 + 1 + ... + 1 = ... = 1 + ... + 1 + 0
n ((1-p) ⁿ)	0 ((1-p) ⁿ)	n = 1 + 1 + ... + 1	

?: The probability of falling k from n comes from the Newton's binomial expansion (and may be verified by induction):

$$(p + (1-p))^n = \sum_{k=0}^n \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Let's go back to the binomial experiment. The probability to see a certain number of coins in the right basket are related with wind speed ratio (or true proportion in the population) and the total number of fallen coins (see *Binomial distribution*).

According to [64] in general if p is a population parameter, T_1 a sufficient statistic in estimating p , and T_2 is any other statistic, the sampling distribution of simultaneous values of T_1 and T_2 must be such that for any given value of T_1 , the distribution of T_2 does not involve p (if $f(p, T_1, T_2)dT_1dT_2$ - probability that T_1 and T_2 to fall in ranges dT_1 and dT_2 - it separates as follows: $f(p, T_1, T_2) = f_1(p, T_1) \cdot f_2(T_1, T_2)$).

The probability of drawing in order any particular sample x_1, x_2, \dots, x_n (of 0's and 1's) is:
 $p^{x_1}(1-p)^{1-x_1} \cdot p^{x_2}(1-p)^{1-x_2} \cdot \dots \cdot p^{x_n}(1-p)^{1-x_n} = p^{\sum x_i} (1-p)^{n-\sum x_i}$

This quantity can be divided in two factors:

$$p^{\sum x_i} (1-p)^{n-\sum x_i} = \frac{n!}{(\sum x_i)!(n-\sum x_i)!} p^{\sum x_i} (1-p)^{n-\sum x_i} \cdot \frac{(\sum x_i)!(n-\sum x_i)!}{n!}$$

of which the first represent the probability that the actual total $\sum x_i$ should have been scored, and the second the probability, given this total, that the partition of it among the n observations should be actually observed. In the latter factor, p , the parameter sought, does not appear. Now when the mean $\sum x_i$ is known, any further information which the sample has to give depends on the observed partition ($x_i/(n-x_i)$), but the probability of any particular distribution is wholly independent of the value of p .

The parameter p can be estimated from the mean of the observed sample by using the Maximum Likelihood Estimation (MLE) method [65]:

$$\text{MLE} = \ln\left(\frac{n!}{(\sum x_i)!(n - \sum x_i)!} p^{\sum x_i} (1-p)^{n - \sum x_i}\right) = \max. \text{ then } \frac{\partial \text{MLE}}{\partial p} = \frac{\partial \text{MLE}}{\partial n} = 0$$

The parameters n and p are independent, and then from $\text{MLE}/p = 0$ it results:

$$\frac{\partial}{\partial p} \ln\left(\frac{n!}{(\sum x_i)!(n - \sum x_i)!}\right) + \frac{\partial}{\partial p} \ln(p^{\sum x_i}) + \frac{\partial}{\partial p} \ln((1-p)^{n - \sum x_i}) = 0$$

This previous equation gives a relationship between estimators under assumption of the maximum likelihood:

$$\frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = 0, \text{ from which } \hat{p} \cdot \hat{n} = \sum x_i$$

Solution of the second derivative ($\text{MLE}/n = 0$) has no analytical close form and can be computed only numerically [66]. Thus, if we assume that we draw always n objects, then $T_1 = x_i$ is a sufficient statistics.

More, if we obtain an estimate of p, \hat{p} , from $\hat{p} \cdot n = \sum x_i$, then (as long is the solution of MLE) then:

$$\text{var}(\hat{p}) = \left(-\left(\frac{\partial^2 \text{MLE}}{\partial p^2}\right)\right)^{-1} = \left(\frac{\sum x_i}{p^2} + \frac{n - \sum x_i}{(1-p)^2}\right)^{-1} =_{\hat{p} \cdot n = \sum x_i} \left(\frac{\hat{p} \cdot n}{p^2} + \frac{n - \hat{p} \cdot n}{(1-p)^2}\right)^{-1} =_{p=\hat{p}} \frac{p(1-p)}{n}$$

Confidence for a binomial proportion is relatively simple to be expressed as long as we enumerate entire probability space of the binomial distribution.

A variable (Y) confined to the whole domain of values (from 0 to n) is binomial distributed if the probability of its taking any particular value j is:

$$P_B(Y = j) = \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j}$$

When we already conducted a previous experiment in which we seen i objects on the right from m objects in total, then we can use i/m as an estimate for p and the probability of Y taking j values on the right become:

$$P_B(Y = j, n, i/m) = \frac{n!}{j!(n-j)!} \left(\frac{i}{m}\right)^j \left(1 - \frac{i}{m}\right)^{n-j}$$

(binomial confidence intervals)

In order to collect the probabilities given by the previous relation and to express a 95% (or other threshold) confidence interval we must solve a combinatorial problem since $P_B(j)$ function is not monotone nor continuous (see $P_B(j, n, 7/10)$ binomial probability table).

$P_B(j, n, 7/10)$ binomial probability

j	0	1	2	3	4	5	6	7	8	9	10
$P_B(j)$	$5.9 \cdot 10^{-6}$	$1.4 \cdot 10^{-4}$	$1.4 \cdot 10^{-3}$	$9.0 \cdot 10^{-3}$	$3.7 \cdot 10^{-2}$	$1.0 \cdot 10^{-1}$	$2.0 \cdot 10^{-1}$	$2.6 \cdot 10^{-1}$	$2.3 \cdot 10^{-1}$	$1.2 \cdot 10^{-1}$	$2.8 \cdot 10^{-2}$
OC	1	2	3	4	6	7	9	11	10	8	5
$_{i \text{ OC}}$	$5.9 \cdot 10^{-6}$	$1.4 \cdot 10^{-4}$	$1.6 \cdot 10^{-3}$	$1.1 \cdot 10^{-2}$	$7.6 \cdot 10^{-2}$	$1.8 \cdot 10^{-1}$	$5 \cdot 10^{-1}$	$1.0 \cdot 10^0$	$7.3 \cdot 10^{-1}$	$3.0 \cdot 10^{-1}$	$3.9 \cdot 10^{-2}$
j	0	1	2	3	4	5	6	7	8	9	10
OC: Order by chance (from unlikely to likely); ^{95%} CI(j, 10, 7/10) = [4, 9]; Err([4, 9]) = 3.9% (<5%)											

It is a fact that we cannot express a mathematical formula indicating how many numbers from left and how many from right (see $P_B(j, n, 7/10)$ binomial probability table) should be excluded in general when we express a confidence interval from a binomial experiment. For this reason usually expressed confidence intervals uses approximating formulas. Following table (see *Binomial and Binomial-like confidence intervals****) lists the most known of it.

*Binomial and Binomial-like confidence intervals****

Group	Name	Method	Acronym*	Refs
Normality	Wald	Classic	Wald_N	[52], [67], [68], [69]

Group	Name	Method	Acronym*	Refs
approximation	Agresti-Coull	Continuity corrected	Wald_C	[⁷⁰]
		Classic	A_C_N	[⁵³]
		Continuity corrected	A_C_C	[⁷⁰]
		Continuity corrected	A_C_D	-**
	Wilson	Classic	Wilson_N	[⁷¹]
		Continuity corrected	Wilson_C	[⁶⁸]
Harmonic approximation	ArcSine	Classic	ArcS_N	[⁷²]
		Continuity corrected	ArcS_C	[⁷⁰]
		Continuity corrected	ArcS_D	[⁷⁰]
		Continuity corrected	ArcS_E	-
Log-normality approximation	Logit	Classic	Logit_N	[⁷³]
		Continuity corrected	Logit_C	[⁷⁴]
Binomial approximation	Bayes (Fisher)	Classic	BetaC11	[⁷⁵]
	Clopper-Pearson	Classic	BetaC01	[⁷⁶], [⁷⁷]
	Jeffreys	Classic	BetaCJ0	[⁷⁸]
	BetaC00	Continuity corrected	BetaC00	-
	BetaC10	Continuity corrected	BetaC10	-
	BetaCJ1	Continuity corrected	BetaCJ1	-
	BetaCJ2	Continuity corrected	BetaCJ2	-
	BetaCJA	Continuity corrected	BetaCJA	-
Obtained from optimization	Blyth-Still-Cassella	Probabilistic optimization	B_S_C	[⁷⁹], [⁸⁰]
	OptiBin	Numerical optimization	OptiBin	[⁸¹]
	ComB	Combinatorial enumeration	ComB	[⁸²], [⁸³]

* According to http://l.academicdirect.org/Statistics/confidence_intervals/;
** New corrections
** English translation of Table 1 p. 42 from [⁸⁴]

(contingency of binomials)

A more complex problem occurs when expressions containing proportions requires confidence intervals. This is the common case when ratios or differences of proportions express a relative or absolute measure of comparison between two populations, or two treatments on populations. This sort of expressions is often used in medical and biological studies [⁸⁴].

The general shape of a study involving two binomial proportions is given by a 2X2 contingency table (see *2X2 contingency*):

2X2 contingency

Experiment 1\2	Left	Right
Up	a	b
Down	c	d

Depending on the type of the experiment, a list of six binomial expressions involving two binomial variables may comprise the currently reported in the literature [⁸⁴] (see *Expressions involving two binomials*) - where the place of X and Y in the contingency table may vary (for example one arrangement is when X = a, m = a+b, Y = c, n = c+d and other one is when X = a, m = a+c, Y = b, n = b+d).

Expressions involving two binomials

Function	f3	f4	f5	f6	f7	f8
Expression	$\frac{X(n - Y)}{Y(m - X)}$	$\frac{Y}{n} - \frac{X}{m}$	$\left \frac{Y}{n} - \frac{X}{m} \right $	$1 / \left \frac{Y}{n} - \frac{X}{m} \right $	$\frac{Xn}{Ym}$	$\left 1 - \frac{Xn}{Ym} \right $

(multinomial distribution)

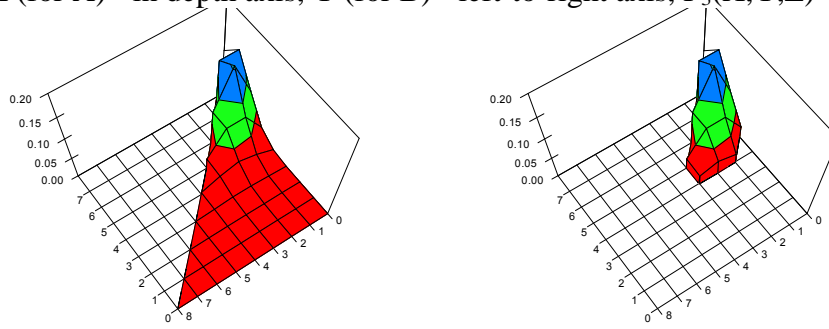
If the observations are split into the informational space in more than two categories then knowing as a fact that we observe n objects which falls in k classes with the probabilities p_1, p_2, \dots, p_k ($1 = p_1 + p_2 + \dots + p_k$) then the probability $P_M(x_1, \dots, x_n; n; p_1, \dots, p_k)$ to observe a certain distribution of the objects ($n = x_1 + x_2 + \dots + x_k$) in the classes (1, ..., k) is:

$$P_M(x_i; n; p_i) = n! \prod_{j=1}^k p_j^{x_j} / x_j!$$

We have many multinomial distributions when we discuss about the distribution of living organisms in a certain area, for example distribution of bacteria species belonging on a class of bacteria in a Petri plate after treatment with an antibiotic. Also for these cases we want to express with a certain level of confidence, the effect of the treatment, for example.

Our knowledge about expressing confidence for expressions containing binomial variables can be extent to cover trinomial (see *Trinomial(6A,1B,1C) - distribution and 95% coverage confidence interval*) or multinomial variables as well [85].

Trinomial(6A,1B,1C) - distribution and 95% coverage confidence interval
Depicted: X (for A) - in depth axis, Y (for B) - left-to-right axis, $P_3(X,Y,Z)$ - vertical axis



Ordinal variables analysis: ranks statistic

In not at all few cases the measurement or observation has the purpose to arrange a number of individuals (such as molecules) in order according to some quality which they possess to a varying degree (such as a molecular structure descriptor or compound chromatographic elution time). The arrangement in which every member of the whole has a rank are called ranking.

Ranked material can arise in many different ways [⁸⁶]:

- ÷ Purely arrangement of objects considered only by reference to their position in space or time;
- ÷ According to some quality which we cannot measure on an objective scale. "If A scratches B when the two are rubbed together" gives a scale of "hardness" then this is such kind of scale and, by transitivity, we may only rank the hardness;
- ÷ According to some measurable or countable quality. We may rank molecules according to melting point or to the number of molecular formula isomers. It may not always be necessary to carry out the actual measurements in such cases, as for instance, if we arrange the molecules by the number of atoms standing in place of the number of isomers;
- ÷ According to some quality which we believe to be measurable but cannot measure for practical or theoretical reasons; the electronegativity we cannot measure, but we can conduct practical or theoretical experiments to see near to are located the binding electrons.

We can always rank a series of quantitative measurements according to their position on the scale. We may replace then the values with their ranks. Disadvantage of ranks use in the place of the values is because (and when) the information regarding the closeness between individuals is omitted. Compensating accuracy lose, the advantage are found in generality, because with ranks we gain the independence of ranks on stretching the scale of measurement - ranking is invariant under stretching the scale of measurement.

(order statistics)

If $X (X_1, \dots, X_n)$ is a random sample coming from a continuous distribution f_X (with $F_X = \text{CDF}(f_X)$) then $U = F_X(X)$ comes from the standard uniform distribution ($U_1 = F_X(X_1), \dots, U_n = F_X(X_n)$). We will note through $U_{(k)}$ the k -th value in the (ascending) ordered series of U .

By splitting the $[0,1]$ interval in three parts, and counting the number of cumulative probabilities $U_{(k)}$ ($1 \leq k \leq n$) falling in these three intervals we actually count the number of objects in a multinomial distribution with three categories (see *Probability of order statistic falling in an interval*).

If we have two random samples coming from (two) continuous distribution functions, and $U_{(i)} = F_X(X_{(i)})$ and $V_{(j)} = F_Y(Y_{(j)})$ the associated order statistics, then the joint probability density function of the two order statistics $U_{(i)} < U_{(j)}$ are constructed from a multinomial distribution with five categories (see *Probability of joint order statistics falling in an interval*).

When distribution functions are known (we known or we derive what distribution function has X and Y) the above obtained relations may allow us to obtain the PDF of the order statistics ($u = F_X(x)$, $v = F_Y(y)$, $du = f_X(x)dx$, $dv = f_Y(y)dy$) - see *Probability distribution function of order statistics*.

Probability of order statistic falling in an interval

Interval	[0,u]	(u,u+du]	(u+du,1]
Probability	u	du	1-u-du
Objects falling in	k-1	1	n-k
$P_3(U_{(k)}; n; u) = \frac{n!}{(k-1)!1!(n-k)!} u^{k-1} \cdot du \cdot (1-u-du)^{n-k} \cong \frac{n!}{(k-1)!(n-k)!} u^{k-1} (1-u)^{n-k} du$			

Probability of joint order statistics falling in an interval

Interval	[0,u]	(u,u+du]	(u+du,v]	(v,v+dv]	(v+dv,1]
Probability	u	du	v-u	dv	1-v-dv
Objects falling in	i-1	1	j-i-1	1	n-j
$P_5(U_{(i)} < U_{(j)}; n; u) = \frac{n!}{(i-1)!1!(j-i-1)!1!(n-j)!} u^{i-1} \cdot du \cdot (v-u)^{j-i-1} \cdot dv \cdot (1-v-dv)^{n-j}$					

Probability distribution function of order statistics

Expression	Meaning
$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} (F_X(x))^{k-1} (1-F_X(x))^{n-k} f_X(x)$	Probability to see the k-th value in the ordered (ranked) list of observations
$f_{X_{(i)}, X_{(j)}}(x, y) = n! \frac{(F_X(x))^{i-1}}{(i-1)!} \frac{(F_X(y) - F_X(x))^{j-i-1}}{(j-i-1)!} \frac{(1-F_X(y))^{n-j}}{(n-j)!} dx dy$	Probability to see the i-th and j-th values in the ordered (ranked) list of observations

(Fisher-Yates correlation coefficient) If $(X_i, Y_i)_{1 \leq i \leq n}$ are already sorted by X values and $U_i = \text{rank}(X_i)$ and $V_i = \text{rank}(Y_i)$ then (and if $U = (1, 2, \dots, n)$ - meaning that there are no tied X values - the Fisher-Yates correlation coefficient are defined by $r_{FY} = \frac{\sum_{1 \leq i \leq n} \xi(i|n)\xi(v_i|n)}{\sum_{1 \leq i \leq n} \xi^2(i|n)}$ ($\xi(i|n)$ is normal order statistic - the expected value of the i-th largest standardized deviate in a sample of size n from a normal population):

$$r_{FY} = \frac{\sum_{1 \leq i \leq n} \xi(i|n)\xi(v_i|n)}{\sum_{1 \leq i \leq n} \xi^2(i|n)}$$

(generalized correlation coefficient)

For n pairs of values (x_i, y_i) we may construct two matrices $(a_{i,j})$ and $(b_{i,j})$ for defining scores and possessing two properties (see *Scores for correlation coefficients*) - $a_{i,i} = b_{i,i} = 0$, $a_{i,j} = a_{j,i}$, $b_{i,j} = b_{j,i}$.

Following table gives the generalized correlation coefficient [86] and three particular cases of it (See *Pearson, Spearman and Kendall correlation coefficients*).

Scores for correlation coefficients

$(a_{i,j})$					$(b_{i,j})$				
x_i vs. y_j	y_1	y_2	...	y_n	y_j vs. x_i	x_1	x_2	...	x_n
x_1	0	$a_{1,2}$...	$a_{1,n}$	y_1	0	$b_{1,2}$...	$b_{1,n}$
x_2	$-a_{1,2}$	0	...	$a_{2,n}$	y_2	$-b_{1,2}$	0	...	$b_{2,n}$
...	0	0	...
x_n	$-a_{1,n}$	$-a_{2,n}$...	0	y_n	$-b_{1,n}$	$-b_{2,n}$...	0

Pearson, Spearman and Kendall correlation coefficients

Generalized formula	Pearson [87]	Spearman [88]	Kendall [89]
$Q = \frac{\sum a_{i,j} b_{i,j}}{\sqrt{\sum a_{i,j}^2} \sqrt{\sum b_{i,j}^2}}$	$a_{i,j} = x_j - x_i$	$a_{i,j} = \text{rank}(x_j) - \text{rank}(x_i)$	$a_{i,j} = \text{sign}(x_j - x_i)$
	$b_{i,j} = y_j - y_i$	$b_{i,j} = \text{rank}(y_j) - \text{rank}(y_i)$	$b_{i,j} = \text{sign}(y_j - y_i)$

A correlation coefficient should (and these three it) follow a list of rules:

- ÷ If the agreement is perfect (if $x_i = y_i$, $1 \leq i \leq n$) then Q should be 1 and if the disagreement is perfect (if $x_i = y_{n-i}$, $1 \leq i \leq n$) then Q should be -1; indeed, $(\sum a_{i,j} b_{i,j})^2 = (\sum a_{i,j})^2 (\sum b_{i,j})^2$ - being Cauchy- Bunyakovsky-Schwarz inequality [90, 91, 92];
- ÷ Increasing values from -1 to 1 should reflect increasing agreement between (x_i) and (y_i) ; this

should be made by construction of $(a_{i,j})$ and $(b_{i,j})$ matrices.

(Jackknife method) Suppose we are interested in estimating some parameter θ using $\hat{\theta} = f(x_1, \dots, x_n)$ where (x_1, \dots, x_n) is a sample of n independent observations with cumulative distribution function $F(\cdot, X)$. Assuming that $\hat{\theta}$ is a good estimate of θ , then following steps (see *Jackknife algorithm*) gives us so called jackknife estimate of first order (then the bias is of $O(1/n)$ order) [93]:

Jackknife algorithm

```

÷ For each i = 1..n
    ○ Remove  $x_i$  and compute  $\hat{\theta}_{-i} = f(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$  from reminders;
    ○ Compute pseudo value  $\hat{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{-i}$ ;
÷ EndFor

```

In order to remove high order bias we must proceed further to remove all pairs of size 2, 3, and so on [94].

For groups up to size k , the bias is removed to order $O(1/n^k)$ and the formula for the k -th order jackknife is obtained ($\hat{\theta}_{(j)}$ the mean of the estimates based on removing groups of size j). Based on computed values, Jackknife estimates for the parameter and its variance are obtained (see *Jackknife estimates*).

Jackknife estimates

Estimates using		$\text{Var}_{\text{est}}(\cdot)$
first order jackknife	$J_n^1(\theta) = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i$	$\text{SE}^2(J_n^1(\theta)) = \frac{1}{n(n-1)} \sum_{i=1}^n (\hat{\theta}_i - J_n^1(\theta))^2$
second order jackknife	$J_n^2(\theta) = \frac{1}{C_2^n} \sum_{1 \leq i < j \leq n} \hat{\theta}_{i,j}$	$\text{SE}^2(J_n^2(\theta)) = \frac{1}{C_2^n(C_2^n - 1)} \sum_{1 \leq i < j \leq n} (\hat{\theta}_{i,j} - J_n^2(\theta))^2$
k -th order jackknife	$J_n^k(\theta) = \frac{1}{C_k^n} \sum_{1 \leq i_1 < \dots < i_k \leq n} \hat{\theta}_{i_1, \dots, i_k}$	$\text{SE}^2(J_n^k(\theta)) = \frac{1}{C_k^n(C_k^n - 1)} \sum_{1 \leq i_1 < \dots < i_k \leq n} (\hat{\theta}_{i_1, \dots, i_k} - J_n^k(\theta))^2$
combined till k -th order	$\tilde{J}_n^k(\theta) = \frac{1}{k!} \sum_{j=0}^k (-1)^j \binom{k}{j} (n-j)^k J_n^j(\theta)$	probability from p_1, p_2, \dots, p_k using [95]: "combining independent tests of significance"

(Bootstrap method) Suppose we are interested in estimating some parameter θ using $\hat{\theta} = f(x_1, \dots, x_n)$ where (x_1, \dots, x_n) is a sample of n independent observations with cumulative distribution function $F(\cdot, X)$. Assuming that $\hat{\theta}$ is a good estimate of θ , then a series of steps gives (- see *Bootstrap algorithm*) us the bootstrap estimates [96].

Bootstrap algorithm

```

÷ Construct empirical probability distribution  $\hat{F}$  giving weight of  $1/n$  to each  $x_i$  (this is bootstrap population):
    Value
    Probability  $\begin{pmatrix} x_1 & \dots & x_n \\ 1/n & 1/n & 1/n \end{pmatrix}$ 
÷ For each i = 1..N (50 N 200):
    ○ Extract a sample (bootstrap sample) of same size  $n$  (as the initial sample) from the bootstrap population using discrete uniform distribution (Random(1..n));
    ○ Calculate the estimate  $\hat{\theta}_{(i)}$  of  $\theta$  for the sample;
÷ EndFor

```

Following table gives the bootstrap estimates (see *Bootstrap estimates*).

Bootstrap estimates

	$\text{Var}_{\text{est}}(\)$
$B_n(\theta) = \frac{1}{N} \sum_{i=1}^N \hat{\theta}_{(i)}$	$\text{SE}^2(B_n(\theta)) = \frac{1}{N(N-1)} \sum_{i=1}^N (B_n(\theta) - \hat{\theta}_{(i)})^2$

(tied values and fractional ranking)

Coming back to our example relating number of atoms with the number of isomers, following table gives the number of isomers for first X alkanes (see Ranking the isomers of alkanes):

Ranking the isomers of alkanes

MF	CH ₄	C ₂ H ₆	C ₃ H ₈	C ₄ H ₁₀	C ₅ H ₁₂	C ₆ H ₁₄	C ₇ H ₁₆	C ₈ H ₁₈	C ₉ H ₂₀	C ₁₀ H ₂₂	C ₁₁ H ₂₄	C ₁₂ H ₂₆	C ₁₃ H ₂₈	C ₁₄ H ₃₀
NC	1	2	3	4	5	6	7	8	9	10	11	12	13	14
IS	1	1	1	2	3	5	9	18	35	75	159	355	802	1858
r _{NC}	1	2	3	4	5	6	7	8	9	10	11	12	13	14
r _{IS}	2	2	2	4	5	6	7	8	9	10	11	12	13	14

MF: Molecular formula; NC: Number of C atoms; IS: Number of structural isomers

A lot of particular cases do not provide necessary information to distinguish between different objects, such as in our example of ranking alkanes isomers. Even more, we wish to measure the degree of correspondence between these two things and we may involve ranks since the relationship is clear to not be linear.

When ranking are conducted using a given property only (as number of structural isomers), then we are not able to distinguish between the objects (as between methane, ethane and propane) and we should assign the same rank. We can do this keeping constant the sum of ranks by averaging all equal rank values (as in: Ranking the isomers of alkanes). This sort of ranking may provide fractional ranks.

(Spearman ρ)

For given set of *n* objects with two measured characteristics *X* and *Y* ranked as (q_i)_{1 i n} and (r_i)_{1 i n} using fractional ranking following formula defines Spearman correlation coefficient [88], where diff_i = q_i-r_i (see Isomers of alkanes - Spearman correlation measures):

$$\rho = 1 - \frac{6 \sum_{i=1}^n \text{diff}_i^2}{(n^3 - n)}$$

Isomers of alkanes - Spearman correlation measures

MF	CH ₄	C ₂ H ₆	C ₃ H ₈	C ₄ H ₁₀	C ₅ H ₁₂	C ₆ H ₁₄	C ₇ H ₁₆	C ₈ H ₁₈	C ₉ H ₂₀	C ₁₀ H ₂₂	C ₁₁ H ₂₄	C ₁₂ H ₂₆	C ₁₃ H ₂₈	C ₁₄ H ₃₀
NC	1	2	3	4	5	6	7	8	9	10	11	12	13	14
IS	1	1	1	2	3	5	9	18	35	75	159	355	802	1858
r _{NC}	1	2	3	4	5	6	7	8	9	10	11	12	13	14
r _{IS}	2	2	2	4	5	6	7	8	9	10	11	12	13	14
diff	-1	0	1	0	0	0	0	0	0	0	0	0	0	0
diff ²	1	0	1	0	0	0	0	0	0	0	0	0	0	0

_idiff_i² = 2; T = 0; U = 3³-3 = 24; _a = 2718/2730; _b = 2706/2730

MF: Molecular formula; NC: Number of C atoms; IS: Number of structural isomers

When are sets of ties in the rankings then we must count the total number of their pairs for *X* (T = $\sum_i (t_i^3 - t_i)$) and for *Y* (U = $\sum_j (u_j^3 - u_j)$). Two formulas are derived in the presence of ties T and U [86]:

$$\rho_a = \frac{(n^3 - n) - 6 \sum_{i=1}^n \text{diff}_i^2 - \frac{T+U}{2}}{(n^3 - n)}, \quad \rho_b = \frac{(n^3 - n) - 6 \sum_{i=1}^n \text{diff}_i^2 - \frac{T+U}{2}}{\sqrt{(n^3 - n) - T} \sqrt{(n^3 - n) - U}}$$

The distribution of ρ_a is approximately normal. Two approximate statistics are available:

$$\div \text{var}(\rho_a) = \frac{1.060}{n-3}, z_{\rho_a} = \sqrt{\frac{n-3}{1.06}} z_r, z_r = \frac{1}{2} \ln \frac{1+r}{1-r} \text{ as gives in [97];}$$

$$\div t = r \sqrt{\frac{n-2}{1-r^2}} \text{ as given in [98];}$$

(Kendall τ)

For given set of n objects with two measured characteristics X and Y ranked as $(q_i)_{1 \leq i \leq n}$ and $(r_i)_{1 \leq i \leq n}$ the number C of concordant and D of discordant pairs using fractional ranking must be computed (see *XY concordance contingency*).

XY concordance contingency

C=cnn+cpp D=dpn+dnp		(Y _j -Y _i)		
		<0	=0	>0
(X _j -X _i)	<0	cnn	tny	dnp
	=0	txn	txy	txp
	>0	dpn	tpy	cpp

Then following formula defines Kendall τ_a and τ_b correlation coefficients [99], where τ_b is the correction for ties:

$$\tau_a = \frac{2(C-D)}{(n^2-n)}, \tau_b = \frac{2(C-D)}{\sqrt{(n^2-n) - \sum_i t_i(t_i-1)} \sqrt{(n^2-n) - \sum_j u_j(u_j-1)}}, \tau_c = \frac{2(C-D)}{n^2 \frac{\min(\text{rows}, \text{cols})-1}{\min(\text{rows}, \text{cols})}}$$

The sampling distribution of τ_a has an expected value of zero. The distribution of τ_a cannot be expressed in a close form. May be calculated exactly for small samples and it is common to use an approximation to the normal distribution, for larger samples, with zero mean and variance given in the next formula and z_{τ_a} statistic approximates the distribution of τ_a [86]:

$$\text{var}(\tau_a) \cong \frac{2(2n+5)}{9n(n-1)}, z_{\tau_a} = \frac{\tau_a}{\sqrt{\text{var}(\tau_a)}} = \frac{2(C-D)}{n(n-1)} \frac{\sqrt{9n(n-1)}}{\sqrt{2(2n+5)}} = \frac{3(C-D)}{\sqrt{n(n-1)(2n+5)/2}}$$

or [97]: $\text{var}(\tau_a) \cong \frac{0.437}{n-4}$

The following statistic z_{τ_b} provides an approximation for the τ_b distribution:

$$z_{\tau_b} = \frac{C-D}{s_{CD}}, s_{CD}^2 = \frac{v_0 - v_t - v_u}{18} + v_1 + v_2$$

$$v_0 = n(n-1)(2n+5), v_t = \sum_i t_i(t_i-1)(2t_i+5), v_u = \sum_j u_j(u_j-1)(2u_j+5)$$

$$v_1 = \frac{(\sum_i t_i(t_i-1))(\sum_j u_j(u_j-1))}{2n(n-1)}, v_2 = \frac{(\sum_i t_i(t_i-1)(t_i-2))(\sum_j u_j(u_j-1)(u_j-2))}{9n(n-1)(n-2)}$$

When we assess the correlation between the properties of more than 10 objects in the expression of z_{τ_b} we should decrease the absolute value of $2(C-D)$ with one unit since the frequency of it is only in paired points. For example [86] when $n = 9$ and $2(C-D) = 20$ the exact probability is 4.4%, the not corrected one is 3.7% and the corrected one is 4.8%, and thus with continuity corrected expression we obtain a better approximation when $n \geq 10$.

(Goodman and Kruskal's γ) By using same notation as for Kendall τ , gamma correlation coefficient is given by [100]:

$$\gamma = \frac{C-D}{C+D}$$

The coefficient approximately follows a Student t distribution with $(n-1)$ degrees of freedom [101]:

$$t_\gamma \cong \gamma \sqrt{\frac{C+D}{n(1-\gamma^2)}}$$

(Hoeffding D) A more general measure of dependence was developed [102] (where $D_1 = \sum_i (Q_i - 1)(Q_i - 2)$, $D_2 = \sum_i (R_i - 1)(R_i - 2)(S_i - 1)(S_i - 2)$, $D_3 = \sum_i (R_i - 2)(S_i - 2)(Q_i - 1)$, $R_i = \text{Rank}(X_i)$, $S_i = \text{Rank}(Y_i)$, $Q_i = 1 + |\{(x_j, y_j) \text{ s.th. } x_j < x_i, y_j < y_i\}|$):

$$D = 30 \frac{(n-2)(n-3)D_1 + D_2 - 2(n-2)D_3}{n(n-1)(n-2)(n-3)(n-4)}$$

It was slightly changed in definition [103] to point out the calculation behind (see *Partition on θ^2 and Hoeffding D*):

Partition on θ^2 and Hoeffding D

$X_i \setminus Y_i$	$y_i < Y_i$	$y_i > Y_i$
$x_i < X_i$	a_i	b_i
$x_i > X_i$	c_i	d_i
$B_n = \sum_i (a_i d_i - b_i c_i)^2 / n^4$		

(other statistics) There are many statistics developed to measure in certain manners the agreement between the ranks in data series. A brief list of other methods is given in the next table (see *Other ordinal correlation measures*).

Other ordinal correlation measures

Expression	Ref
$d = (C - D) / (C + D + T_Y)$	Sommer's d [104]
$\tau_K = 4 \cdot \text{cnn} \cdot (n(n-1))^{-1} - 1$	Kendall's τ_K [105]
$\rho_F = 1 - \frac{4}{n^2} \sum_{1 \leq i, j \leq n} \text{rank}(x_i) - \text{rank}(y_i) $	Spearman's footrule ρ_F [105]

(ordered contingencies)

Let us take a general two-way table representing the contingency of two variables $(X_i)_{1 \leq i \leq R}$ and $(Y_j)_{1 \leq j \leq C}$ (see *R\C ordered contingency*).

R\C ordered contingency

$R \setminus C$	Y_1	...	Y_j	...	Y_C
X_1	$n_{1,1}$...	$n_{1,j}$...	$n_{1,C}$
...
X_i	$n_{i,1}$...	$n_{i,j}$...	$n_{i,C}$
...
X_R	$n_{R,1}$...	$n_{R,j}$...	$n_{R,C}$
$X_1 < \dots < X_i < \dots < X_R$					
$Y_1 < \dots < Y_j < \dots < Y_C$					

In order to give computing formula for rank correlation measures (see *Computing rank correlation measures for multinomial contingencies*) a series of notations should be introduced (see *Notations for computing rank correlation measures for ordered contingencies*).

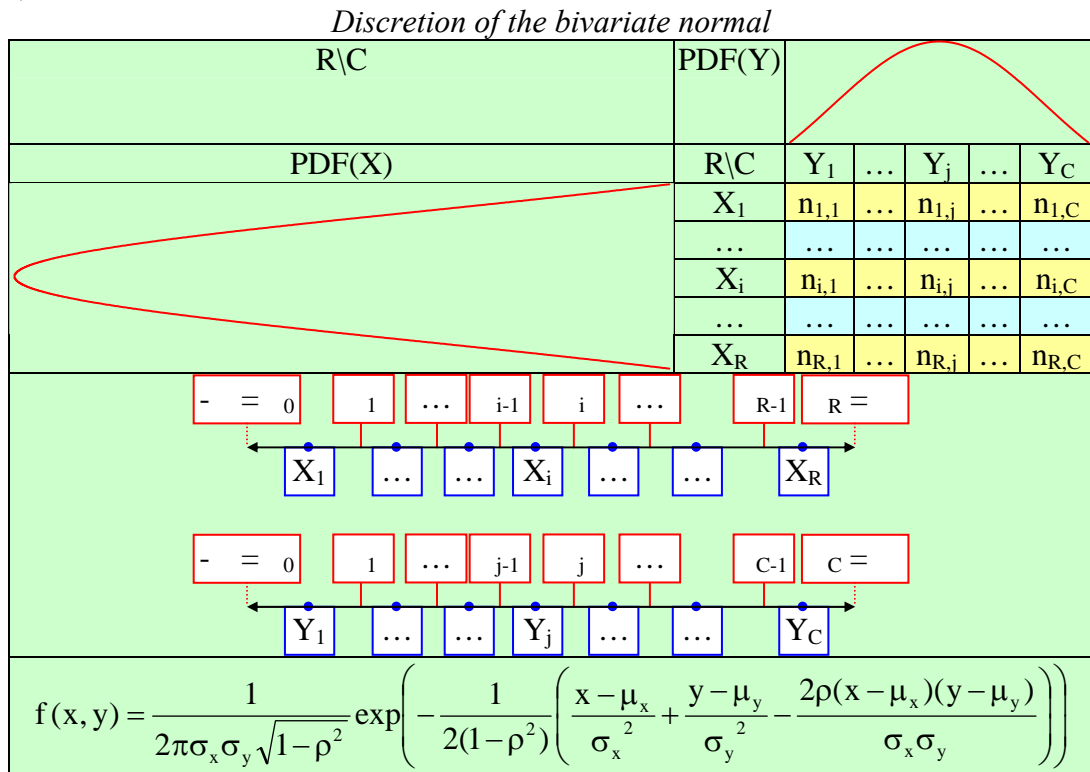
Notations for computing rank correlation measures for ordered contingencies

Label	Formula	Meaning
$n_{i,\cdot}$; $n_{\cdot,j}$; n	$\sum_j n_{i,j}$; $\sum_i n_{i,j}$; $\sum_i \sum_j n_{i,j}$	row, column, and overall totals
R_i ; C_j	i or R_i ; j or C_j	score for row i and for column j
\bar{R} ; \bar{C}	$\sum_i n_{i,R_i} / n$; $\sum_j n_{\cdot,j} C_j / n$	average row and column score
R_i ; C_j	$\sum_{k < i} n_{k,\cdot} + \frac{n_{i,\cdot} + 1}{2}$; $\sum_{l < j} n_{\cdot,l} + \frac{n_{\cdot,j} + 1}{2}$	rank score of row i and column j
$C_{i,j}$; $D_{i,j}$	$\sum_{k > i} \sum_{l > j} n_{k,l} + \sum_{k < i} \sum_{l < j} n_{k,l}$; $\sum_{k > i} \sum_{l < j} n_{k,l} + \sum_{k < i} \sum_{l > j} n_{k,l}$	concordances and discordances
C ; D	$\sum_i \sum_j n_{i,j} C_{i,j}$; $\sum_i \sum_j n_{i,j} D_{i,j}$	twice number of concordances twice number of discordances

Computing rank correlation measures for ordered contingencies

Measure	Its expression
	$(C - D)/(C + D)$
Asymptotic variance	$\frac{16}{(C + D)^4} \sum_i \sum_j n_{i,j} (D \cdot C_{i,j} - C \cdot D_{i,j})^2$
Variance to be different from 0	$\frac{4}{(C + D)^2} \left(\sum_i \sum_j n_{i,j} (C_{i,j} - D_{i,j})^2 - \frac{(C - D)^2}{n} \right)$
b	$\frac{C - D}{w}$, $w = \sqrt{w_r w_c}$, $w_r = n^2 - \sum_i n_{i,\cdot}^2$, $w_c = n^2 - \sum_j n_{\cdot,j}^2$ $d_{i,j} = C_{i,j} - D_{i,j}$, $v_{i,j} = n_{i,\cdot} w_c - n_{\cdot,j} w_r$
Asymptotic variance	$\frac{1}{w^4} \left(\sum_i \sum_j n_{i,j} (2w d_{i,j} + t_b \cdot v_{i,j})^2 - n^3 t_b^2 (w_r + w_c)^2 \right)$
Variance to be different from 0	$\frac{4}{w_r w_c} \left(\sum_i \sum_j n_{i,j} (C_{i,j} - D_{i,j})^2 - \frac{(C - D)^2}{n} \right)$
d_{cR}	$\frac{C - D}{w_r}$, $w_r = n^2 - \sum_i n_{i,\cdot}^2$, $d_{i,j} = C_{i,j} - D_{i,j}$
Asymptotic variance	$\frac{4}{w_r^4} \left(\sum_i \sum_j n_{i,j} (w_r d_{i,j} - (C - D)(n - n_{i,\cdot}))^2 \right)$
Variance to be different from 0	$\frac{4}{w_r^2} \left(\sum_i \sum_j n_{i,j} (C_{i,j} - D_{i,j})^2 - \frac{(C - D)^2}{n} \right)$
r	ss_{rc}/w , $w = \sqrt{ss_r ss_c}$ $ss_r = \sum_i \sum_j n_{i,j} (R_i - \bar{R})^2$, $ss_c = \sum_i \sum_j n_{i,j} (C_j - \bar{C})^2$ $ss_{rc} = \sum_i \sum_j n_{i,j} (R_i - \bar{R})(C_j - \bar{C})$, $b_{i,j} = (R_i - \bar{R})^2 ss_c + (C_j - \bar{C})^2 ss_r$
Asymptotic variance	$\frac{1}{w^4} \left(\sum_i \sum_j n_{i,j} \left(w(R_i - \bar{R})(C_j - \bar{C}) - \frac{b_{i,j} ss_{rc}}{2w} \right)^2 \right)$
Variance to be different from 0	$\frac{1}{w^2} \left(\sum_i \sum_j n_{i,j} (R_i - \bar{R})(C_j - \bar{C}) - \frac{ss_{rc}^2}{n} \right)$
s	$\frac{v}{w}$, $v = \sum_i \sum_j n_{i,j} R_{0i} C_{0j}$, $w = \frac{1}{12} \sqrt{FG}$, $F = n^3 - \sum_i n_{i,\cdot}^3$, $G = n^3 - \sum_j n_{\cdot,j}^3$, $R_{0i} = R_i - n/2$, $C_{0j} = C_j - n/2$ $\bar{z} = \sum_i \sum_j n_{i,j} z_{i,j}$, $z_{i,j} = w v_{i,j} - v w_{i,j}$, $w_{i,j} = \frac{-n}{96w} (F n_{\cdot,j}^2 + G n_{i,\cdot}^2)$ $v_{i,j} = n \left(R_{0i} C_{0j} + \frac{1}{2} \sum_l n_{i,l} C_{0l} \frac{1}{2} \sum_k n_{k,j} R_{0k} + \sum_l \sum_{k>l} n_{k,l} C_{0l} + \sum_k \sum_{l>j} n_{k,l} R_{0k} \right)$
Asymptotic variance	$\frac{1}{n^2 w^4} \sum_i \sum_j n_{i,j} (z_{i,j} - \bar{z})^2$
Variance to be different from 0	$\frac{1}{n^2 w^2} \sum_i \sum_j n_{i,j} (v_{i,j} - \bar{v})^2$, $\bar{v} = \sum_i \sum_j n_{i,j} v_{i,j} / n$

(polychoric correlation) Assuming for two variables that comes from normal distribution, and the contingency counts the number of observations for a given set of discrete values of X (X_1, \dots, X_R) and Y (Y_1, \dots, Y_C) then we may estimate the correlation between the population of X and Y by using the bivariate normal distribution (see *Discretion of the bivariate normal*).



The values of μ_X and σ_X may come from MLE on $(X_i)_{i=1}^R$ values under assumption of normality and the values of μ_Y and σ_Y may come from MLE on $(Y_j)_{j=1}^C$ values under assumption of normality and are assumed known with known data series.

The joint probability of X_i and Y_j falling in $(\alpha_{i-1}, \alpha_i] \times (\beta_{j-1}, \beta_j]$ is estimated by the observed count, from which the likelihood of all $(X_i)_{i=1}^R$ and $(Y_j)_{j=1}^C$ falling in the designed intervals $(\alpha_{i-1}, \alpha_i] \times (\beta_{j-1}, \beta_j]$ can be derived. Maximum likelihood is obtained when derivative of log of likelihood is null (MLE method, [65]), giving us gives a system of equations from which is possible to estimate (numerically) of unknown "thresholds" $(\alpha_i)_{i=1}^{R-1}$ and $(\beta_j)_{j=1}^{C-1}$ as well as correlation coefficient ρ .

$$n_{i,j} \sim \int_{\alpha_{i-1}}^{\alpha_i} \int_{\beta_{j-1}}^{\beta_j} f(x, y) dx dy, \quad P(\alpha; \beta; \rho) = \prod_{i=1}^R \prod_{j=1}^C \int_{\alpha_{i-1}}^{\alpha_i} \int_{\beta_{j-1}}^{\beta_j} f(x, y) dx dy$$

$$\ln(P(\alpha; \beta; \rho)) = \sum_{i=1}^R \sum_{j=1}^C \int_{\alpha_{i-1}}^{\alpha_i} \int_{\beta_{j-1}}^{\beta_j} f(x, y) dx dy, \quad \frac{\partial(\ln(P(\alpha; \beta; \rho)))}{\partial \alpha} = \frac{\partial(\ln(P(\alpha; \beta; \rho)))}{\partial \beta} = \frac{\partial(\ln(P(\alpha; \beta; \rho)))}{\partial \rho} = 0$$

Linear regression analysis: linear models

The simplest association model is linear association. The model assumes that exist a relationship between two paired characteristics expressed by a straight line. We may start expressing this association by using the implicit equation of a straight line: $ax + by + c = 0$.

If $a = 0$ then the equation of the line reduces to $by + c = 0$. If further $c \neq 0$ gives a relationship which defines the mean of the Y associated characteristic but no relationship with X. Similarly if $b = 0$ then the equation of the line reduces to $ax + c = 0$ and if further $c \neq 0$ gives a relationship which defines the mean of the X associated characteristic but no relationship with Y. The remained case, if $c = 0$ defines a degenerated linear model in which is no intercept between the characteristics X and Y.

Which expression of the linear equation we should use is a matter of experimental error treatment. Going further, if a linear model defines the relationship between the X and Y characteristics, then if we take samples $(x_i, y_i)_{i=1}^n$ of these two (X and Y) characteristics then we should still see the relationship in terms of a experimental error.

We may make errors measuring both characteristics, and thus if we express as (\hat{x}_i, \hat{y}_i) the expected (or true) values each corresponding to a pair of measurements (x_i, y_i) then:

÷ The linear model is given by: $a\hat{x}_i + b\hat{y}_i + c = 0, i = 1..n$;

÷ The experimental errors are given by: $\varepsilon_i = x_i - \hat{x}_i$ and $\eta_i = y_i - \hat{y}_i$;

In order to be able to find the linear model (e.g. it's coefficients a, b and c) we must make further assumptions on experimental errors. Thus, at least we should know the expected value of the experimental error. If \hat{x}_i is the expected value of x_i then the expected value of ε_i , namely $\hat{\varepsilon}_i$ is 0: $\hat{\varepsilon}_i = 0$, but this fact happens only if we not made systematically errors on measurement of the characteristic X. Similarly, if \hat{y}_i is the expected value of y_i then the expected value of η_i , namely $\hat{\eta}_i$ is 0: $\hat{\eta}_i = 0$, but this fact happens only if we not made systematically errors on measurement of the characteristic Y. More than that, under presence of systematically errors on measurement of X or Y is no way to obtain the relationship between X and Y excepting the situation when we know the (expected) value of the systematic error (but then we can systematically extract it from the values to obtain new series of data which do not posses systematically errors). Thus, first assumptions were made, namely:

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n (x_i - \hat{x}_i)}{n} = 0 \text{ (no systematically error on X); } \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} = 0 \text{ (idem for Y)}$$

Even more strong assumptions are generally accepted and are required in order to solve (or obtain) the relationship, namely:

$$\frac{\sum_{i=1}^n (x_i - \hat{x}_i)}{n} \cong 0 \text{ and } \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \cong 0$$

Please note that these two relationships above do not imply that the absolute or other sort of modulus based differences are null to. In general:

$$\frac{\sum_{i=1}^n |x_i - \hat{x}_i|^\alpha}{n} \geq 0 \text{ and } \frac{\sum_{i=1}^n |y_i - \hat{y}_i|^\beta}{n} \geq 0$$

We should know more about the distribution of the error. A common assumption is to expect that an error ε_i (or η_i) to occur in equal probability as an error $-\varepsilon_i$ (or $-\eta_i$), and then the distribution of the experimental error is symmetrical.

A choice is possible here coming from the experiment: to give different weight to the errors (and then we have a weighted regression). Usually the weight are function of the observable and/or expectance ($w_i = f(x_i, \hat{x}_i)$, $w_i = g(y_i, \hat{y}_i)$). The reason of giving weight to the

errors is to normalize (e.g. the distribution of the errors to become normal or at least to have a known distribution).

Here comes another assumption regarding the experimental error (ϵ_i and η_i) that we must take in order to obtain estimations of the population parameters: that follows a known distribution.

We may make inferences about the distribution (see for example *Distribution of errors when binary responses are recorded*).

Distribution of errors when binary responses are recorded

x_i	$p_i = p(x_i)$	\hat{x}_i	$q_i = p(\hat{x}_i)$	$x_i - \hat{x}_i$	$ x_i - \hat{x}_i ^\alpha$ ($\alpha > 0$)	$p_i q_i$
0	p	0	q	0	0	pq
0	p	1	(1-q)	-1	1	p(1-q)
1	(1-p)	0	q	1	1	(1-p)q
1	(1-p)	1	(1-q)	0	0	(1-p)(1-q)

$x_i - \hat{x}_i$	$p(x_i - \hat{x}_i)$	$n = n_{-1} + n_0 + n_1$ (observation errors)
-1	$p_{-1} = p(1-q)$	$P_3(n_{-1}, n_0, n_1; n; \{p(1-q), (1-p)q, 2pq\}) =$ $\frac{n!(p(1-q))^{n_{-1}} ((1-p)q)^{n_0} (1-p-q+2pq)^{n_1}}{n_{-1}! n_0! n_1!} =$ $\frac{n! 2^{n_0} \cdot p^{n-n_1} q^{n-n_1} (1-p)^{n_1} (1-q)^{n_1} (1-p-q+2pq)^{n_0}}{n_{-1}! n_0! n_1!}$
0	$p_0 = 1-p-q+2pq$	
1	$p_1 = q(1-p)$	
$ x_i - \hat{x}_i ^\alpha$	$p(x_i - \hat{x}_i ^\alpha)$	$n = n_0 + n_1$ (observation errors)
0	$p_0 = 1-p-q+2pq$	$P_2(n_0, n_1; n; \{1-p-q-2pq, p+q-2pq\}) =$ $\frac{n!(1-p-q+2pq)^{n_0} (p+q-2pq)^{n_1}}{n_0! n_1!} =$ Binomial($1-p-q+2pq; n$)
1	$p_1 = p+q-2pq$	

The common continuous symmetrical distribution assumed to be followed by the experimental error is the Gauss-Laplace distribution [106]:

$$GL(x; \mu, \sigma, p) = \frac{p}{2\sigma} \frac{\Gamma^{1/2}(3/p)}{\Gamma^{3/2}(1/p)} \exp\left(-\frac{|x - \mu|^p}{\sigma} \left/ \left(\frac{\Gamma(1/p)}{\Gamma(3/p)}\right)^{p/2}\right.\right)$$

When $p = 1$ the Laplace distribution occurs [107] (often referred in physical sciences) and when $p = 2$ the Gauss distribution occurs [108] (often referred in biological sciences).

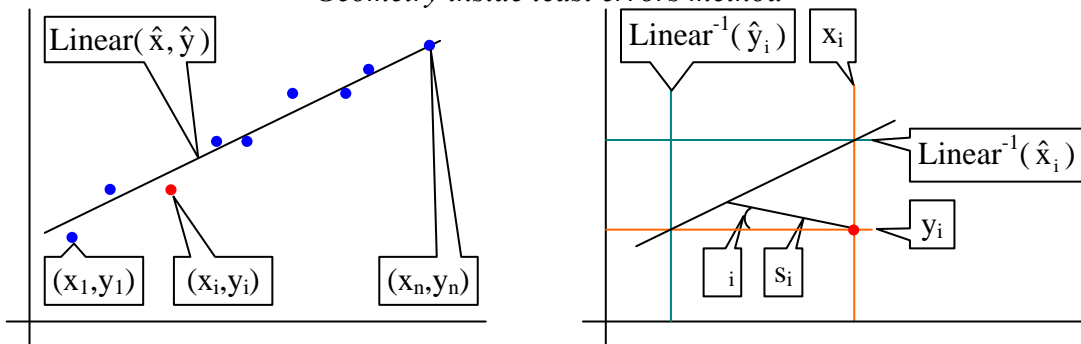
The relationship of Gauss-Laplace distribution with the distribution of the experimental error is immediate (see *Gauss-Laplace and least errors method*).

Gauss-Laplace and least errors method

Errors distribution assumption	Regression model	Least errors method
$\left(\frac{ x_i - \hat{x}_i ^\alpha}{\hat{x}_i}\right)_{1 \leq i \leq n} \sim GL(\epsilon; 0, \sigma_x, \alpha)$	$x \sim \hat{x} = b \cdot y + c$	$\sum_{i=1}^n \epsilon_i^\alpha = \sum_{i=1}^n x_i - \hat{x}_i ^\alpha = \min.$
$\left(\frac{ y_i - \hat{y}_i ^\beta}{\hat{y}_i}\right)_{1 \leq i \leq n} \sim GL(\eta; 0, \sigma_y, \beta)$	$y \sim \hat{y} = a \cdot x + c$	$\sum_{i=1}^n \eta_i^\beta = \sum_{i=1}^n y_i - \hat{y}_i ^\beta = \min.$
$\left(\frac{ x_i - \hat{x}_i ^\gamma}{\hat{x}_i}\right)_{1 \leq i \leq n} \sim GL(\epsilon; 0, \sigma_x, \gamma)$ $\left(\frac{ y_i - \hat{y}_i ^\gamma}{\hat{y}_i}\right)_{1 \leq i \leq n} \sim GL(\eta; 0, \sqrt{1 - \sigma_x^2}, \gamma)$	only when $c = 1$: $1 \sim \hat{1} = a \cdot x + b \cdot y$	$\sum_{i=1}^n 0_i^\gamma = \sum_{i=1}^n ax_i + by_i - 1 ^\gamma = \min.$

The geometrical interpretation of the least errors method is depicted below (see *Geometry inside least errors method*). Subject to minimization is $S(\hat{x}, \hat{y}) = \sum_{i=1}^n s_i$. Different values of s_i and give different assumptions on errors, different linear models and different methods of calculation [109] (see *Calculations in linear regression - least errors method*).

Geometry inside least errors method



Calculations in linear regression - least errors method

i		Regression model	S min.	Method	Further assumption
0	2	$x \sim \hat{x} = b \cdot y + c$	$\sum_{i=1}^n (x_i - b y_i - c)^2$	analytical from $\frac{\partial S}{\partial b} = \frac{\partial S}{\partial c} = 0$	measurement of $(y_i)_{1 \leq i \leq n}$ without errors
	1		$\sum_{i=1}^n x_i - b y_i - c $	numerical from $\frac{\partial S}{\partial b} = \frac{\partial S}{\partial c} = 0$	
/2	2	$y \sim \hat{y} = a \cdot x + c$	$\sum_{i=1}^n (y_i - a x_i - c)^2$	analytical from $\frac{\partial S}{\partial a} = \frac{\partial S}{\partial c} = 0$	measurement of $(x_i)_{1 \leq i \leq n}$ without errors
	1		$\sum_{i=1}^n y_i - a x_i - c $	numerical from $\frac{\partial S}{\partial a} = \frac{\partial S}{\partial c} = 0$	
s.th. $s_i \perp$ on $\text{Linear}(\hat{x}, \hat{y})$	2	$x \sim \hat{x} = b \cdot y + c$	see [109]	analytical from $\frac{\partial S}{\partial b} = \frac{\partial S}{\partial c} = 0$	$i = (x_i); \quad i = (y_i)$ $i / i = x_i^2 / y_i^2$
	1			numerical from $\frac{\partial S}{\partial b} = \frac{\partial S}{\partial c} = 0$	$i = (x_i); \quad i = (y_i)$ $i / i = x_i / y_i $
	2	$y \sim \hat{y} = a \cdot x + c$		analytical from $\frac{\partial S}{\partial a} = \frac{\partial S}{\partial c} = 0$	$i = (x_i); \quad i = (y_i)$ $i / i = x_i^2 / y_i^2$
	1			numerical from $\frac{\partial S}{\partial a} = \frac{\partial S}{\partial c} = 0$	$i = (x_i); \quad i = (y_i)$ $i / i = x_i / y_i $
s.th. s_i is median	2	$x \sim \hat{x} = b \cdot y + c$
	1	
	2	$y \sim \hat{y} = a \cdot x + c$
	1	
	2	$1 \sim a \cdot x + b \cdot y$
	1	

The multiple regression models are generalizations of simple linear regression. We may conduct the same rationalizations (as in Gauss-Laplace and least errors method or in Calculations

in linear regression - least errors method) in order to derive representatives of the family of multiple regression models. We will take only the simplest and in the same time the most common case:

$$y \sim = a_0 \cdot 1 + a_1 \cdot x_1 + \dots + a_m \cdot x_m \text{ for } (y_i, x_{1,i}, \dots, x_{m,i})_{1 \leq i \leq n} \text{ and at least } m < n$$

Deriving of $\sum_i (a_0 + a_1 x_{1,i} + \dots + a_m x_{m,i} - y_i)^2 = \min$. conduct to the following system of equations:

$$S_A = \begin{pmatrix} 0 & 1 & \dots & m \\ \sum_i 1 \cdot 1 & \sum_i x_{1,i} \cdot 1 & \dots & \sum_i x_{m,i} \cdot 1 \\ \sum_i 1 \cdot x_{1,i} & \sum_i x_{1,i} \cdot x_{1,i} & \dots & \sum_i x_{n,i} \cdot x_{1,i} \\ \dots & \dots & \dots & \dots \\ \sum_i 1 \cdot x_{n,i} & \sum_i x_{1,i} \cdot x_{n,i} & \dots & \sum_i x_{n,i} \cdot x_{m,i} \end{pmatrix} \begin{matrix} 0 \\ 1 \\ \dots \\ m \end{matrix}, S_B = \begin{pmatrix} 0 \\ \sum_i y_i \cdot 1 \\ \sum_i y_i \cdot x_{1,i} \\ \dots \\ \sum_i y_i \cdot x_{m,i} \end{pmatrix} \begin{matrix} 0 \\ 1 \\ \dots \\ m \end{matrix}, S_A \times A = S_B$$

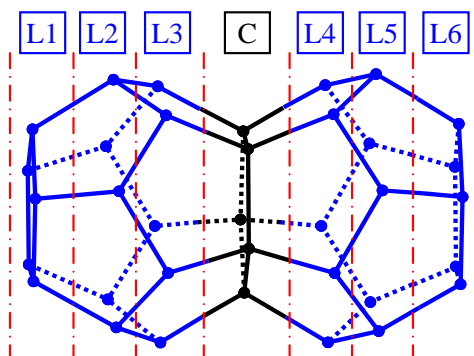
from which the solution is immediate: $S_A^{-1} \times S_A \times A = S_A^{-1} \times S_B \quad A = S_A^{-1} \times S_B$.

The Gauss-Jordan numeric method [110] exploits the feature of elementary row operations (multiplication/division and addition/subtraction keep the matrix equality unchanged) to obtain simultaneous with the coefficients their variance (see *Gauss-Jordan procedure for linear multiple regression*).

Aplica ii

Problem : Nivele de reprezentare

Se dore te s se realizeze un experiment în care s fie colectate (în spa iul informa ional) observa ii cu privire la structura de mai jos, în care pe fiecare din cele 6 nivele pot fi prezen i atomi de carbon, azot i bor.



Câte informa ii distincte se vor reg si, maxim, în spa iul informa ional, tiind c structura este rigid i fixat într-un ansamblu mai mare (cazul a) i respectiv mobil (cazul b).

Solu ie

Straturi: 6 (L1, L2, L3, L4, L5, L6); Nivele: 3 (B, N, C); Num r total de combina ii: 729; Structuri distincte: 378

Studiu de caz: Propriet i din Spartan '10

MolVol: Molecular volume Å³; SurfA: Surface area Å²; Ovality: Ovality dimensionless 1.234; HOMO: Highest Occupied Molecular Orbital Energy eV; LUMO: Lowest Unoccupied Molecular Orbital Energy eV; Estimated polarizability: *10⁻³⁰ m³; Electronegativity= -HOMO+LUMO/2; Hardness= -HOMO-LUMO/2

Property	Distinct	Grp.	Zero	Property	Distinct	Grp.	Zero
DipoleT_0	376	352		Lumo-Homo_0	353	375	
DipoleT_1	377	351		Lumo-Homo_1	331	397	
DipoleT_2	377	351		Lumo-Homo_2	372	356	
EnergyHF_0	131	510	87	MolVol_0	175	512	41
EnergyHF_1	139	507	82	MolVol_1	167	530	31
EnergyHF_2	346	382		MolVol_2	372	356	
HOMO_0	337	389	2	Ovality_0	46	516	166
HOMO_1	320	408		Ovality_1	51	590	87
HOMO_2	367	361		Ovality_2	56	645	27
LUMO_0	318	410		Polariz_0	349	379	
LUMO_1	300	428		Polariz_1	316	412	
LUMO_2	367	361		Polariz_2	371	357	
Lumo+Homo_0	355	373		SurfA_0	221	477	30
Lumo+Homo_1	324	404		SurfA_1	202	520	6
Lumo+Homo_2	356	370	2	SurfA_2	368	360	

Problem : Precizie i exactitate

S se analizeze sub aspectul preciziei urm toarele m sur tori ale vitezei vântului în raport cu afirma ia c în acea zi nu s-au înregistrat deplas ri de aer.

Moment	Viteza	Direc ia	Durata
9 ³⁰	2.8ms ⁻¹	NV	1 min.
10 ³⁰	1.0ms ⁻¹	E	2 min.
11 ³⁰	1.4ms ⁻¹	NE	1 min.
12 ³⁰	3ms ⁻¹	S	1 min.

Ce observa ie lipse te pentru ca s se sus in afirma ia?

Soluție

Se alc tuie te un tabel de forma celui de mai jos:

v	t	dir	dx	dy	sdx	sdy
2.8	60	NV	-118.8	118.8	-118.8	118.8
1	120	E	120.0	0.0	1.2	118.8
1.4	60	NE	59.4	59.4	60.6	178.2
3	60	S	0.0	-180.0	60.6	-1.8

Problem : Separare maxim

În cromatografie se folosesc o serie de m suri pentru a caracteriza separarea compu ilor dintr-un amestec:

- ÷ $RF(i, e) = l(i, e)/l(e)$, în care i este un compus separat, e faza mobil , $l(i, e)$ distan a de migrare a lui i în e , $l(e)$ distan a de migrare a lui e i ce conduce la *mulțimea factorilor de retardare* (RF)
- ÷ $RFO(i, e) = l(\pi(i), e)/l(e)$, unde $\pi(i)$ este o permutare ce transform lista într-o list ordonat (cresc tor sau descresc tor) i ce conduce la *șirul ordonat al factorilor de retardare* (RFO)
- ÷ $nc(e) = \text{Count}(\{i \mid l(\pi(i+1), e) - l(\pi(i), e) > (w(\pi(i+1), e) + w(\pi(i), e))/8)\})$ ce conduce la *numărul de componenți observați la o distanță (unul de altul) de cel puțin 1σ* (nc)
- ÷ $mnc = \max\{nc(e) \mid e \in E\}$, în care E mul imea tuturor fazelor mobile posibile i ce conduce la *numărul de componenți existenți în amestec* (mnc)
- ÷ $RSM(i, j, e) = 2 \cdot (l(i, e) - l(j, e)) / (w(i, e) + w(j, e))$, unde $w(i, e)$ este l rgimea spotului i iar $w(j, e)$ este l rgimea spotului j , i ce conduce la *matricea rezoluțiilor* (RSM)
- ÷ $RSO(i, e) = 2 \cdot (l(\pi(i+1), e) - l(\pi(i), e)) / (w(\pi(i+1), e) + w(\pi(i), e))$, unde din nou π este permutarea ce ordoneaz (cresc tor) irul factorilor de retardare i ce conduce la *șirul rezoluțiilor spoturilor adiacente* (RSO)
- ÷ $RSS(e) = \sum_{i=1}^{nc(e)} RSO(i, e)$ ce conduce la *suma rezoluțiilor* (RSS)
- ÷ $QN_{\text{eff}}(e) = \sqrt{N_{\text{eff}}(e)} = 4 \cdot l(e) \cdot nc(e) / \sum_{i=1}^{nc(e)} w(e, i)$ ce conduce la *radical din numărul efectiv de talere* (QN_{eff}); ajustând condi iile experimentale date idealul de separare se ob ine când $4 \cdot RSS(e) / QN_{\text{eff}}(e)$
- ÷ $RSP(e) = 25 \cdot RSS(e) / QN_{\text{eff}}(e)$ ce d *rezoluția împărțită la numărul de talere* (RSP); idealul de separare se ob ine când $RSP(e) = 100$
- ÷ $RSA(e) = RSS(e) / nc(e)$ ce d *rezoluția medie a separării* (RSA)
- ÷ $RRP(e) = \prod_{i=1}^{nc(e)} RSO(i, e) / \sum_{i=1}^{nc(e)} RSO(i, e)$ care este *produsul rezoluțiilor relative* (RRP)
- ÷ $RSH(e, p) = \left(\sum_{i=1}^{nc(e)} (RSO(i, e))^p / nc(e) \right)^{1/p}$ unde p este o valoare real arbitrar i ce conduce la *media Hölder a rezoluțiilor* (RSH); când $p=2$ $RSH(e, 2)$ este un mai bun descriptor al separ rii decât RSA
- ÷ $RFD(e) = \sqrt{\sum_{i=1}^{nc(e)} (RFO(i+1, e) - RFO(i, e) - 1/mnc)^2} / \sqrt{nc(e) \cdot (nc(e) + 1)}$, unde $1/mnc$ este diferen a teoretic între 2 factori de retardare i ce conduce la un indice de separare medie exprimat ca devia ie între ideal ($1/mnc$) i observat ($RFO(i+1, e) - RFO(e)$) - *deviația factorilor de retardare* (RFD)
- ÷ $IEne(e) = mnc^2 - \sum_{i=1}^{mnc} n_i^2$, unde n_i este num rul de spoturi în al i -lea interval echidistant i ce

conduce la *energia informațională asociată separării* (IEne)

$$\div \text{IEnt}(e) = \sum_{i=1}^{mnc} n_i \cdot \log(n_i), \text{ unde } n_i \text{ este numrul de spoturi în al } i\text{-lea interval echidistant } i \text{ ce}$$

conduce la *entropia informațională asociată separării* (IEnt)

$$\div \text{QF}(e) = \min_{1 \leq i, j \leq nc(e)} \text{RSM}(i, j, e) = \min_{1 \leq i \leq nc(e)} \text{RSO}(i, e) \text{ care conduce la } \textit{factorul de calitate a separării}$$

(QF) dat de cea mai defavorabil separare

S se calculeze i apoi s se analizeze propriet ile (opera iile matematice permise) de ace ti parametrii.

Soluție

Imaginând procesul de separare drept un proces în care apari ia migr rii spoturilor este o problem de durat , din punct de vedere experimental este de interes ce se întâmpl cu valorile acestor parametrii când se detecteaz apari ia unui nou spot (eventual la începutul sau la sfâr itul irului de spoturi deja existent).

Fie e un eluent (acela i). Vom simplifica nota iile f când referire mereu la acela i eluent.

RF. Fie $\{RF(i) | 1 \leq i \leq nc\}$ mul imea factorilor de retardare. Apari ia unui nou factor face ca la mul imea factorilor de retardare s se adauge un element. Din acest punct de vedere, mul imea factorilor de retardare poate fi v zut ca o mul ime de caracteristici. Asocierea între compusul chimic i factorul de retardare al acestuia e unic (pe calea direct , adic un compus are un singur factor de retardare în condi iile experimentale date; reciproca îns nu este adev rat , putând exista 2 compu i cu acela i factor de retardare) i poate indica ulterior absen a compusului dintr-un amestec prin absen a valorii factorului de retardare din mul imea valorilor m surate pentru amestec.

RFO. irul ordonat al factorilor de retardare nu este o mul ime standard, fiind echipat i cu o rela ie de ordine. Din acest punct de vedere, irul ordonat al factorilor de retardare necesit o opera ie simpl de ad ugare la început sau la sfâr it doar dac valoarea ad ugat este fie mai mic , fie mai mare decât toate valorile deja existente în ir; în caz contrar, ad ugarea necesit i identificarea pozi iei în care trebuie f cut inserarea.

nc. Numrul de componen i observa i î i modific valoarea la apari ia unui nou spot numai dac este îndeplinit condi ia ca distan a dintre spoturile adiacente s fie cel pu in $1/8$ din l rgimea acestora. Din acest punct de vedere, ad ugarea unui nou compus în amestec poate avea ca efect unul din urm toarele:

\div $nc - nc - 1$ când noul spot apare între 2 spoturi deja existente, dar apari ia acestuia face ca cele 2 spoturi existente ini ial împreun cu cel de-al treilea nou spot s nu mai poat fi distinse unul de cel alt;

\div $nc - nc$ când noul spot apare în vecin tatea altuia dar suficient de departe de oricare alt spot

\div $nc - nc + 1$ când noul spot apare suficient de departe de oricare alt spot

RSM. Matricea rezolu iilor sufer o opera iune de ad ugare a unei linii i a unei coloane pentru fiecare nou compus ad ugat în amestec. Liniile i coloanele existente ini ial în această matrice r mân cu valori neschimbate.

RSO. irul rezolu iilor ordonate nu se poate ob ine din ordonarea matricei RSM (de exemplu ordonând o linie prin permutare de coloane). Matricea rezolu iilor ordonate se ob ine din diferen ele între valorile din irul ordonat al factorilor de retardare (RFO - RSO). Din acest punct de vedere, rezultatul ob inut prin ad ugarea unui nou compus în amestec poate produce apari ia sau dispari ia unui element în irul rezolu iilor ordonate sau s se p streze numrul acestora (vezi nc) necesitând totodat i schimbarea valorilor adiacente valorii inserate (dac numrul elementelor r mâne acela i sau cre te cu o unitate).

Problemă: Metoda adaosului standard

Se consider o prob sub form de solu ie în care compu ii dizolva i sunt cunoscu i dar sunt în cantit i necunoscute. De asemenea sunt disponibili ace ti compu i prepara i separat sub form de solu ii apoase. Care este modalitatea de determinare a cantit ilor necunoscute, tiind c identificarea se realizeaz cu ajutorul unui detector care produce un semnal propor ional cu concentra ia (fie aceasta molar) de compus identificat.

Soluție

Fie x i y cantit ile în grame de compu i cunoscu i C_1 i C_2 dizolva i în 100 ml H_2O (în proba

necunoscut) și fie a și b cantitățile în grame de compuși cunoscuți dizolvați în 100 ml H₂O (în etaloanele preparate din acești compuși). Fie masele molare M₁ (în g, pentru C₁) și M₂ (în g, pentru C₂). În acest caz avem la dispoziție 3 soluții: S_U (soluția necunoscută), S₁ (din C₁) și S₂ (din C₂):

Soluție	S _U	S ₁	S ₂
Volum	100 ml	100 ml	100 ml
Substanț dizolvat	x gC ₁ , y gC ₂	a gC ₁	b gC ₂

Imaginem o serie de 3 experimente cu acești compuși în care folosim câte 25 ml din soluția cu cantități dizolvate necunoscute (S_U) și câte un volum (deocamdată neprecizat) din soluțiile etalon. Un calcul simplu de compoziția amestecurilor:

Experiment	S _U	S ₁	S ₂	V(H ₂ O)	m(C ₁)	m(C ₂)
1	25 ml	V ₁₁ ml	V ₁₂ ml	25+V ₁₁ +V ₁₂	25·x/100+V ₁₁ ·a/100	25·y/100+V ₁₂ ·b/100
2	25 ml	V ₂₁ ml	V ₂₂ ml	25+V ₂₁ +V ₂₂	25·x/100+V ₂₁ ·a/100	25·y/100+V ₂₂ ·b/100
3	25 ml	V ₃₁ ml	V ₃₂ ml	25+V ₃₁ +V ₃₂	25·x/100+V ₃₁ ·a/100	25·y/100+V ₃₂ ·b/100

Detectorul va înregistra semnale. În acest moment avem 2 soluții de implementare: când detecția se realizează după separare, caz în care pentru fiecare compus identificat avem câte un semnal în fiecare experiment (fie R₁₁ și R₁₂ semnalele din primul experiment și asemeni R₂₁, R₂₂, R₃₁ și R₃₂) sau detecția se realizează direct asupra amestecului, caz în care avem câte un singur semnal din fiecare experiment (fie Z₁ semnalul înregistrat în primul experiment corespunzător detecției, Z₂ semnalul înregistrat în al doilea experiment și Z₃ semnalul înregistrat în al treilea experiment) proporționale cu concentrațiile molare, deci este necesar să evaluăm aceste concentrații:

Experiment	n(C ₁)/V(H ₂ O)	n(C ₂)/V(H ₂ O)
1	(25·x/100+V ₁₁ ·a/100)/M ₁ /(25+V ₁₁ +V ₁₂)	(25·y/100+V ₁₂ ·b/100)/M ₂ /(25+V ₁₁ +V ₁₂)
2	(25·x/100+V ₂₁ ·a/100)/M ₁ /(25+V ₂₁ +V ₂₂)	(25·y/100+V ₂₂ ·b/100)/M ₂ /(25+V ₂₁ +V ₂₂)
3	(25·x/100+V ₃₁ ·a/100)/M ₁ /(25+V ₃₁ +V ₃₂)	(25·y/100+V ₃₂ ·b/100)/M ₂ /(25+V ₃₁ +V ₃₂)
Semnal	Separat pentru C ₁	Separat pentru C ₂
1	R ₁₁	R ₁₂
2	R ₂₁	R ₂₂
3	R ₃₁	R ₃₂
Semnal	Împreună pentru C ₁ și C ₂ în amestec	
1	Z ₁	
2	Z ₂	
3	Z ₃	

Se exprimă raportul între concentrație și semnal care pentru un anumit compus este același de la un experiment la altul. Rezolvăm în continuare pe cele două cazuri:

Cazul semnalului separat

$$\div \text{ Pentru } C_1: \frac{x/4 + a \cdot V_{11}/100}{M_1 \cdot (25 + V_{11} + V_{12}) \cdot R_{11}} = \frac{x/4 + a \cdot V_{21}/100}{M_1 \cdot (25 + V_{21} + V_{22}) \cdot R_{21}} = \frac{x/4 + a \cdot V_{31}/100}{M_1 \cdot (25 + V_{31} + V_{32}) \cdot R_{31}}$$

$$\div \text{ Pentru } C_2: \frac{y/4 + b \cdot V_{12}/100}{M_2 \cdot (25 + V_{11} + V_{12}) \cdot R_{12}} = \frac{y/4 + b \cdot V_{22}/100}{M_2 \cdot (25 + V_{21} + V_{22}) \cdot R_{22}} = \frac{y/4 + b \cdot V_{32}/100}{M_2 \cdot (25 + V_{31} + V_{32}) \cdot R_{32}}$$

Simplificând:

$$\div \text{ Pentru } C_1: \frac{x \cdot 25 + a \cdot V_{11}}{(25 + V_{11} + V_{12}) \cdot R_{11}} = \frac{x \cdot 25 + a \cdot V_{21}}{(25 + V_{21} + V_{22}) \cdot R_{21}} = \frac{x \cdot 25 + a \cdot V_{31}}{(25 + V_{31} + V_{32}) \cdot R_{31}}$$

$$\div \text{ Pentru } C_2: \frac{y \cdot 25 + b \cdot V_{12}}{(25 + V_{11} + V_{12}) \cdot R_{12}} = \frac{y \cdot 25 + b \cdot V_{22}}{(25 + V_{21} + V_{22}) \cdot R_{22}} = \frac{y \cdot 25 + b \cdot V_{32}}{(25 + V_{31} + V_{32}) \cdot R_{32}}$$

Pentru fiecare dintre cei doi compuși, avem exact 2 soluții (nu neapărat identice) corespunzătoare celor 2 egalități. În fapt, problema poate fi rezolvată și mai corect minimizând eroarea de observare (tem de cas!).

În continuare, fie aceste două egalități următoarele:

$$x \cdot \frac{25}{(25 + V_{11} + V_{12}) \cdot R_{11}} + \frac{a \cdot V_{11}}{(25 + V_{11} + V_{12}) \cdot R_{11}} = x \cdot \frac{25}{(25 + V_{21} + V_{22}) \cdot R_{21}} + \frac{a \cdot V_{21}}{(25 + V_{21} + V_{22}) \cdot R_{21}}$$

$$x \cdot \frac{25}{(25 + V_{11} + V_{12}) \cdot R_{11}} + \frac{a \cdot V_{11}}{(25 + V_{11} + V_{12}) \cdot R_{11}} = x \cdot \frac{25}{(25 + V_{31} + V_{32}) \cdot R_{31}} + \frac{a \cdot V_{31}}{(25 + V_{31} + V_{32}) \cdot R_{31}}$$

De unde:

$$x = a \cdot \frac{\frac{V_{21}/R_{21}}{25 + V_{21} + V_{22}} - \frac{V_{11}/R_{11}}{25 + V_{11} + V_{12}}}{\frac{25/R_{11}}{25 + V_{11} + V_{12}} - \frac{25/R_{21}}{25 + V_{21} + V_{22}}} \quad \text{i/sau} \quad x = a \cdot \frac{\frac{V_{31}/R_{31}}{25 + V_{31} + V_{32}} - \frac{V_{11}/R_{11}}{25 + V_{11} + V_{12}}}{\frac{25/R_{11}}{25 + V_{11} + V_{12}} - \frac{25/R_{31}}{25 + V_{31} + V_{32}}}$$

În mod similar rezultăm expresiile pentru y .

Cazul amestecului

În cazul măsurării directe pe amestec, se suprapune semnalul de la primul compus cu semnalul de la al doilea compus, însă în mod necesar cu aceeași intensitate:

$$\div \text{ Experiment 1: } \frac{x/4 + a \cdot V_{11}/100}{M_1 \cdot (25 + V_{11} + V_{12})} \cdot \alpha + \frac{y/4 + b \cdot V_{12}/100}{M_2 \cdot (25 + V_{11} + V_{12})} \cdot \beta = Z_1$$

$$\div \text{ Experiment 2: } \frac{x/4 + a \cdot V_{21}/100}{M_1 \cdot (25 + V_{21} + V_{22})} \cdot \alpha + \frac{y/4 + b \cdot V_{22}/100}{M_2 \cdot (25 + V_{21} + V_{22})} \cdot \beta = Z_2$$

$$\div \text{ Experiment 3: } \frac{x/4 + a \cdot V_{31}/100}{M_1 \cdot (25 + V_{31} + V_{32})} \cdot \alpha + \frac{y/4 + b \cdot V_{32}/100}{M_2 \cdot (25 + V_{31} + V_{32})} \cdot \beta = Z_3$$

Ecuațiile de mai sus formează un sistem de 3 ecuații cu 4 necunoscute (x , y , α , β). Este evident că pentru a rezolva acest sistem este nevoie de încă o ecuație, deci de încă un experiment:

$$\text{Experiment 4: } \frac{x/4 + a \cdot V_{41}/100}{M_1 \cdot (25 + V_{41} + V_{42})} \cdot \alpha + \frac{y/4 + b \cdot V_{42}/100}{M_2 \cdot (25 + V_{41} + V_{42})} \cdot \beta = Z_4$$

Sistemul de ecuații rezultat se poate scrie prin intermediul valorilor cunoscute astfel (unde c_{11} , ... c_{45} sunt valori cunoscute):

$$x \cdot \alpha \cdot c_{11} + \alpha \cdot c_{12} + y \cdot \beta \cdot c_{13} + \beta \cdot c_{14} = c_{15}$$

$$x \cdot \alpha \cdot c_{21} + \alpha \cdot c_{22} + y \cdot \beta \cdot c_{23} + \beta \cdot c_{24} = c_{25}$$

$$x \cdot \alpha \cdot c_{31} + \alpha \cdot c_{32} + y \cdot \beta \cdot c_{33} + \beta \cdot c_{34} = c_{35}$$

$$x \cdot \alpha \cdot c_{41} + \alpha \cdot c_{42} + y \cdot \beta \cdot c_{43} + \beta \cdot c_{44} = c_{45}$$

Chiar dacă pare complicat, în esență este un sistem de ecuații ușor de rezolvat. Fie când substituim $x = \gamma$, $y = \delta$, sistemul devine un sistem liniar omogen în necunoscutele γ , δ , α , β :

$$\gamma \cdot c_{11} + \alpha \cdot c_{12} + \delta \cdot c_{13} + \beta \cdot c_{14} = c_{15}$$

$$\gamma \cdot c_{21} + \alpha \cdot c_{22} + \delta \cdot c_{23} + \beta \cdot c_{24} = c_{25}$$

$$\gamma \cdot c_{31} + \alpha \cdot c_{32} + \delta \cdot c_{33} + \beta \cdot c_{34} = c_{35}$$

$$\gamma \cdot c_{41} + \alpha \cdot c_{42} + \delta \cdot c_{43} + \beta \cdot c_{44} = c_{45}$$

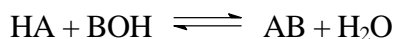
Odată obținut soluția acestui sistem (γ , δ , α , β) valorile necunoscutele x și y sunt imediate: $x = \gamma / \alpha$ și $y = \delta / \beta$.

Problemă: Calculul pH-ului în funcție de volumul de soluție adăugată

Să se exprime variația pH-ului în titrarea în soluție apoasă (H_2O) a unui acid slab (HA) cu o bază slabă (BOH).

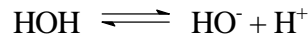
Soluție

Modelul reacției de titrare pornește de la scrierea ecuației reacției chimice a titrării unui acid slab HA cu o bază slabă BOH, pentru care echilibrul este caracterizat de constanta de solubilitate a sării obișnuite:

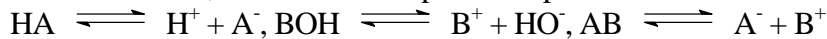


Dacă reacția se desfășoară în apă, trebuie să considerăm influența asupra pH-ului din disocierea

moleculilor de apă :



Procesele de disociere ale acidului, bazei și sării în apă sunt exprimate de relațiile:



Titrarea începe cu adăugarea unei mici cantități de bază în acid. În acest moment sunt prezente în soluție speciile: H^+ , HO^- , HA și A^- . Din ecuația de disociere a acidului și apei rezultă :

$$[\text{H}^+] \cdot [\text{A}^-] = K_a \cdot [\text{HA}], [\text{H}^+] \cdot [\text{HO}^-] = K_w$$

unde $[\cdot]$ este exprimată concentrația molară ($[\text{H}^+]$ este concentrația molară), K_a este constanta de aciditate iar K_w este constanta de disociere a apei la temperatura considerată. Dacă se aplică bilanțul de masă pentru acid și sare, rezultă că C_a , concentrația analitică a acidului și respectiv C_s , concentrația analitică a sării sunt date de:

$$C_a = [\text{HA}] + [\text{H}^+] - [\text{HO}^-], C_s = [\text{A}^-] - [\text{H}^+] + [\text{HO}^-]$$

După substituțiile corespunzătoare în ecuațiile de mai sus, se obține o ecuație de gradul 3 a pH-ului (ecuația Brønsted-Lowry $[\text{H}^+]^{111}$, $[\text{H}^+]^{112}$), $[\text{H}^+] = x$:

$$x^3 + (K_a + C_s) \cdot x^2 - (K_w + C_x \cdot K_a) \cdot x - K_w \cdot K_a = 0$$

Ecuația de mai sus admite o soluție unică în intervalul (0,1), înțelegând seama că :

$$C_s = C_b \cdot V_x / (V_a + V_x), C_x = (C_a \cdot V_a - C_b \cdot V_x) / (V_a + V_x)$$

unde C_b este concentrația analitică a bazei, V_x este volumul de bază adăugat, C_x concentrația analitică a acidului după adăugare iar V_a este volumul inițial de acid. Substituind în ecuația Brønsted, aceasta poate fi rezolvată numeric.

La punctul de echivalență, se pornește modelul de la același punct inițial și se consideră toate echilibrurile menționate. La hidroliză mică, $C_s = [\text{A}^-] = [\text{B}^+]$ și atunci $[\text{H}^+] = x$:

$$x = \sqrt{\frac{K_w \cdot K_a \cdot (K_b + C_s)}{K_b \cdot (K_a + C_s)}}$$

După punctul de titrare, prin deduceri similare se obține că $[\text{H}^+] = x$:

$$x^3 + (K_w / K_b + C_x) \cdot x^2 - (K_w + C_s \cdot K_w / K_b) \cdot x - K_w^2 / K_b = 0$$

unde expresiile lui C_x și C_s sunt:

$$C_x = (C_b \cdot V_x - C_a \cdot V_a) / (V_a + V_x), C_s = C_a \cdot V_a / (V_a + V_x)$$

Un program care implementează rezolvarea acestor probleme este disponibil online la adresa de Internet: <http://l.academicdirect.org/Education/Training/titration/v1.1/>.

Problemă : Modelarea procesului de acțiune enzimatic Michaelis-Menten

- ÷ Prima dată a fost observat la invertază $[\text{H}^{13}]$;
- ÷ Ecuația: $\text{S} + \text{E} \rightleftharpoons \text{C} \rightleftharpoons \text{P} + \text{E}$, (S - substrat, E - enzimă, C - complex, P - produs (concentrații: s, e, c, p));
- ÷ ipoteză : enzima (E) nu suferă modificări în ceea ce privește cantitatea sa totală (și astfel nici concentrația) în timp ($e + c = \text{constant}$);
- ÷ modelul cinetic permite exprimarea evoluției sistemului către echilibru; presupune scrierea ecuațiilor de viteză de reacție pentru fiecare reacție elementară și aplicarea principiului conservării numărului de atomi;

Soluție: v. Algoritm pentru simularea cineticii Michaelis-Menten

1. Scrierea reacțiilor elementare; scrierea ecuațiilor de viteză				
(1): $\text{S} + \text{E} \xrightarrow{k_1} \text{C}$, $v_{(1)} = k_1 \cdot s \cdot e$	(2): $\text{C} \xrightarrow{k_2} \text{S} + \text{E}$, $v_{(2)} = k_2 \cdot c$	(3): $\text{C} \xrightarrow{k_3} \text{P} + \text{E}$, $v_{(3)} = k_3 \cdot c$		
2. Aplicarea principiului conservării numărului de atomi				
(S): $\dot{s} = v_{(2)} - v_{(1)}$	(E): $\dot{e} = v_{(2)} + v_{(3)} - v_{(1)}$	(C): $\dot{c} = v_{(1)} - v_{(2)} - v_{(3)}$	(P): $\dot{p} = v_{(3)}$	
3. Presupuneri și notații				
$s(0) = s_0$	$e(0) = e_0$	$c(0) = 0$	$p(0) = 0$	$e = e_0 - c$
4. Ecuații de rezolvat				
$\dot{s} = k_2 c - k_1 s (e_0 - c)$		$\dot{c} = k_1 s (e_0 - c) - (k_2 + k_3) c$		
5. Diferențe abordări				

aproxima ia "QSSA"	Briggs & Haldane, [114]	$\dot{c} = 0 \Rightarrow$	$c = \frac{e_0 s}{\kappa + s}; -\dot{s} = \dot{p} = \frac{k_3 e_0 s}{\kappa + s}; \kappa = \frac{k_2 + k_3}{k_1}$
aproxima ia "EA"	Henri [115]	$\dot{s} = 0 \Rightarrow$	$c = \frac{e_0 s}{\kappa + s}; \dot{p} = \frac{k_3 e_0 s}{\kappa + s}; \kappa = \frac{k_2}{k_1}$
Cazul general	ecua ie implicit nu are solu ie analitic !	phases space	$\frac{dy}{dx} = b \frac{x - y - xy}{-x + ay + xy}$
	substitu ii în ecua ia explicit $t = k_1 e_0 \tau$, timp ini ial	$0 < a < 1$ $b > 0$	$a = \frac{k_2}{k_2 + k_3}; b = \frac{k_2 + k_3}{k_1 e_0}; x = \frac{k_1 s}{k_2 + k_3}; y = \frac{c}{e_0}$ $\dot{x} = -x + ay + xy; \dot{y} = b(x - y - xy);$

6. Rezolvare numeric (i=1..n)

$x_0 = 3$ $y_0 = 0$	$x_{i+1} = x_i + (-x_i + ax_i + x_i y_i)$ $y_{i+1} = y_i + b(x_i - y_i - x_i y_i)$	$= 10^{-2}$ $n = 3000$	$a \in \{\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}\}$	$b \in \{\frac{250}{25}, \frac{50}{25}, \frac{10}{25}, \frac{4}{25}\}$
------------------------	---	---------------------------	--	--

7. Foaie de calcul Excel:

	A	B	C	D	E	F
1	$x_0 = 3$		i	x_i		y_i
2	$y_0 = 0$		=0	=B1		=B2
3	=1.0e-2		=D2+1	=E2+\$B\$3*(-E2+\$B\$4*E2+E2*F2)		=F2+\$B\$5*\$B\$3*(E2-F2-E2*F2)
4	a=0.2	
5	b=10	

Algoritm pentru simularea cineticii Michaelis-Menten

Problem : Modelarea mecanismului Lindemann - Hinshelwood al complexului activat

Prima oar comunicat de Lindemann [116] i ulterior analizat de Hinshelwood [117] furnizeaz un mecanism de ac iune enzimatic . Se modeleaz mecanismul $R + R \rightleftharpoons R^* + R \rightleftharpoons P$.

Solu ie: v. Algoritm pentru simularea mecanismului Lindemann - Hinshelwood

1. Scrierea reac iilor elementare; scrierea ecua iilor de viteza

(1): $R + R \rightarrow R^* + R, v_{(1)} = k_1[R]^2$	$R^* + R \rightarrow R + R, v_{(2)} = k_2[R][R^*]$	$R^* \rightarrow P, v_{(3)} = k_3[R^*]$
--	--	---

2. Aplicarea principiului conserv rii num rului de atomi (necunoscute $[R] = x; [R^*] = y; [P] = z$):

(R): $\dot{x} = -v_{(1)} + v_{(2)}$	(R*): $\dot{y} = v_{(1)} - v_{(2)} - v_{(3)}$	(P): $\dot{z} = v_{(3)}$
-------------------------------------	---	--------------------------

3. Presupuneri i nota ii

$r(0) = r_0$	$r^*(0) = 0$	$p(0) = 0$	$k_1 = a$	$k_2 = b$	$k_3 = c$
--------------	--------------	------------	-----------	-----------	-----------

4. Ecua ii de rezolvat

$\dot{x} = -ax^2 + bxy$	$\dot{y} = ax^2 - bxy - cy$	$\dot{z} = cy$
-------------------------	-----------------------------	----------------

5. Abordare gre it

C utarea unei solu ii analitice este f r de succes.

6. Rezolvare numeric (i=1..n)

$x_0 = 3$ $y_0 = 0$ $z_0 = 0$	$x_{i+1} = x_i + (-ax_i^2 + bx_i y_i)$ $y_{i+1} = y_i + (ax_i^2 - bx_i y_i - cy_i)$ $z_{i+1} = z_i + cy_i$	$= 10^{-2}$ $n = 3000$	$a = 10^{-1}$ $b = 10^{-2}$ $c = 10^{-3}$
-------------------------------------	--	---------------------------	---

7. Foaie de calcul Excel:

	A	B	C	D	E	F	G
1	$x_0 = 1$		i	x_i		y_i	z_i
2	$y_0 = 0$		=0	=B1		=B2	=B3
3	$z_0 = 0$		=D2+1	=E2+(-B\$1*E2^2+B\$2*E2*F2)*B\$4		=F2+(B\$1*E2^2-B\$2*E2*F2-B\$3*F2)*B\$4	=G2+B\$3*F2*B\$4
4	=1e-2	

5	a=	1e-1
6	b=	1e-2
7	c=	1e-3

Algoritm pentru simularea mecanismului Lindemann - Hinshelwood

Problem : Simularea cineticii autocatalizei

Autocataliza este un caz special de cinetic chimic în care reacțiile au loc doar în prezența atât reactanților cât și a produsilor de reacție și au o ecuație generală de reacție dată de: $R \rightarrow P$, cu $v_{(1)} = k \cdot [R] \cdot [P]$.

Soluție: v. Algoritm pentru simularea autocatalizei.

1. Scrierea reacțiilor elementare; scrierea ecuațiilor de viteză					
(1): $R \rightarrow P, v_{(1)} = k \cdot [R] \cdot [P]$					
2. Aplicarea principiului conservării numărului de atomi (necunoscute $[R] = x; [R^*] = y; [P] = z$):					
(R): $\dot{r} = -v_{(1)} = -k_1 r p$			(P): $\dot{p} = v_{(1)} = k_1 r p$		
3. Presupuneri și notații					
$[R] = r$	$[P] = p$	$k_1 = a$	$r + p = r_0 + p_0 = b$	$r(0) = r_0$	$p(0) = p_0$
4. Ecuații de rezolvat					
$\dot{p} = a(b - p)p$			$\dot{r} = -ar(b - r)$		
5. Abordări - existența soluției analitice					
$\dot{p} = a(b - p)p \Rightarrow \frac{dp}{p(b - p)} = a dt \Rightarrow \frac{1}{b} \ln \frac{p}{b - p} = at + c \Rightarrow \frac{p}{b - p} = e^{b(k_1 t + c)} \Rightarrow$ $p = \frac{b}{1 + e^{-b(k_1 t + c)}} = \frac{b}{1 + e^{-bk_1 t} e^{-bc}}$					
6. Constantele "b" & "c" - din valorile inițiale ale concentrației (la momentul $t = 0$).					
$\frac{1}{b} \ln \frac{p(0)}{b - p(0)} = a \cdot 0 + c \Rightarrow \frac{1}{b} \ln \frac{p_0}{r_0} = c; bc = \ln \frac{p_0}{r_0}; -bc = \ln \frac{r_0}{p_0}; e^{-bc} = \frac{r_0}{p_0}$					
7. Soluția analitică și interpretarea ei					
$p = p(t) = p_0 \frac{r_0 + p_0}{r_0 e^{-(r_0 + p_0)k_1 t} + p_0}; r = r_0 + p_0 - p$			dac $p_0 = 0$ atunci $p = 0$ & astfel nu evoluează dac $r_0 = 0$ atunci $p = p_0$ & astfel nu evoluează		
8. Desen în cazul $p_0 \neq 0$ și $r_0 \neq 0$ (aplicație numerică, folosind MathCad [118])					
dac $p_0 = 0.1; r_0 = 0.9; k_1 = 0.2$ atunci $p(t) = \frac{1}{9e^{-0.2t} + 1}$ & $r(t) = 1 - \frac{1}{9e^{-0.2t} + 1}$					

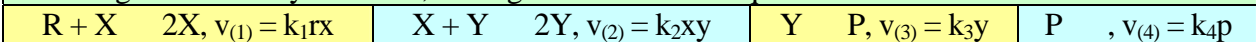
Algoritm pentru simularea autocatalizei

(Lotka - Volterra mechanism)

Proposed by Lotka for the first time in [119] as a complex mechanism in homogeneous phase with damped oscillations. Later [120] a slightly changing of the mechanism was proposed when no damping occurs in the oscillations. Volterra made a statistical analysis of involving these equations [121]. By using same notations ($[R]=r, [X]=x, [Y]=y, [P]=p$) following algorithm is for the simulation of the Lotka - Volterra mechanism, in which the concentration of the reactant are kept constant by the

experimental design (such as adding substance or having an equilibrium with a nonmiscible phase containing it) - see *Algorithm for simulation of Lotka - Volterra kinetics*.

1. Writing of elementary reactions; writing of reaction rates equations



2. Equations to solve

$\dot{x} = v_{(1)} - v_{(2)} = k_1 r x - k_2 xy$	$\dot{y} = v_{(2)} - v_{(3)} = k_2 xy - k_3 y$	$\dot{p} = v_{(3)} + v_{(4)} = k_3 y + k_4 p$
--	--	---

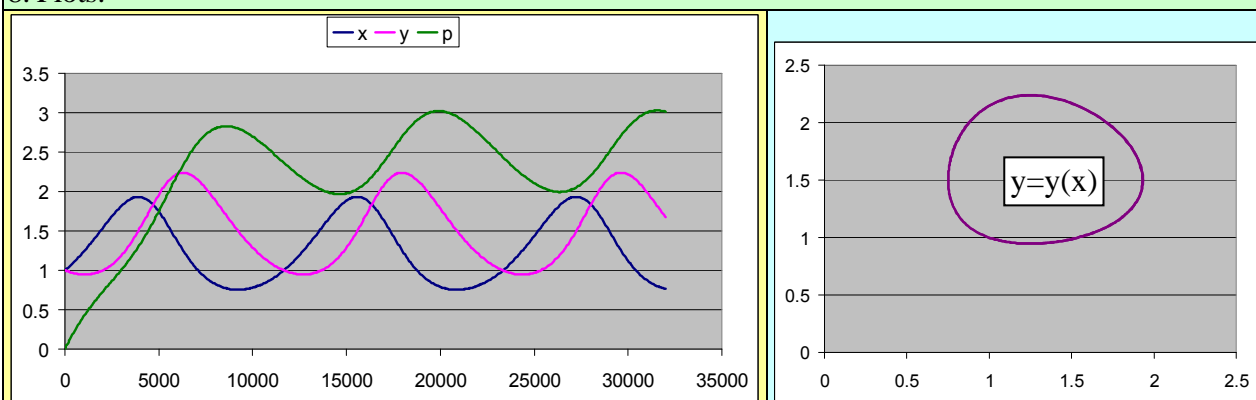
3. Solving numerically (i=1..n)

$x_0 = 3$ $y_0 = 0$ $z_0 = 0$	$x_{i+1} = x_i(1+(k_1 r - k_2 y_i) \Delta t)$ $y_{i+1} = y_i(1+(k_2 x_i - k_3) \Delta t)$ $p_{i+1} = p_i + (k_3 y_i + k_4 p_i) \Delta t$	$\Delta t = 10^{-2}$ $n = 5 \cdot 10^5$	$k_1 = 3$ $k_2 = 4$ $k_3 = 5$	$k_4 = 3$ $r = 1$
-------------------------------------	--	--	-------------------------------------	----------------------

7. Excel sheet:

	A	B	C	D	E	F	G
1	x0=	1		i	xi	yi	pi
2	y0=	1		=0	=B1	=B2	=B3
3	p0=	0		=D2+1	=D2*(1+(B\$5*B\$4 - B\$6*E2)*B\$9)	=E2*(1+(B\$6*D2 - B\$7)*B\$9)	= F2+(B\$7*E2 - B\$8*F2)*B\$9
4	r=	2	
5	k1=	3	
6	k2=	4	
7	k3=	5	
8	k4=	3	
9		=1e-4	

8. Plots:



Remarks: the $y=y(x)$ equation is almost impossible to be extracted analytically, but we may extract it approximately (numerically). Thus, for the simulated mechanism above, the $y=y(x)$ equation is:

$$(x-1.32)^2 + 0.824 \cdot (y-1.57)^2 = 0.35 \pm 0.05$$

Algorithm for simulation of Lotka - Volterra kinetics

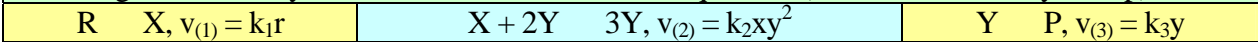
(brusselator mechanism) - autocatalytic

A group of researchers from Bruxelles proposed for the first time a model in which phases space converges to a limit cycle [122]. Many authors modified it and studied systems working under this sort of mechanisms [123, 124, 125]. A simplified variant is given below (as in previous cases the concentration of reactant, r , are kept constant, see *Algorithm for simulation of brusselator kinetics*).

The equations of the system (see *Algorithm for simulation of brusselator kinetics*) does not goes all the time to a limit cycle independent of the values of the flow constants and/or initial concentrations. Trying to solve this system is full of surprises: for most of the combinations the equations provide a system going to a equilibrium point; there are values for which the system evolves with damped oscillations to the equilibrium; undamped oscillations have a significant weight between the general solutions of the system - fact proved by the nature system itself in which are numerous living systems basing their existence on such kind of oscillations. Heart periodical pulses are due to

such kind of process. Importance if this type of processes is big. This is the reason for which in 1997, Ilya PRIGOGINE received the Nobel prize in Chemistry for their theoretical studies on dissipative systems (see *Algorithm for simulation of Lotka - Volterra kinetics*).

1. Writing of elementary reactions and of reaction rates equations ($[R]=r, [X]=x, [Y]=y, [P]=p$):



2. Applying of atoms conservation principle (unknowns $[X] = x; [Y] = y$):

(X): $\dot{x} = v_{(1)} - v_{(2)} = k_1r - k_2xy^2$	(Y): $\dot{y} = v_{(2)} - v_{(3)} = k_2xy^2 - k_3y$	(P): $\dot{p} = k_3y$
---	---	-----------------------

3. Approach (simplifying $r = 1, k_1 = 1$ and $k_3 = 1$):

$\dot{x} = 1 - k_2xy^2$	$\dot{y} = k_2xy^2 - y$
-------------------------	-------------------------

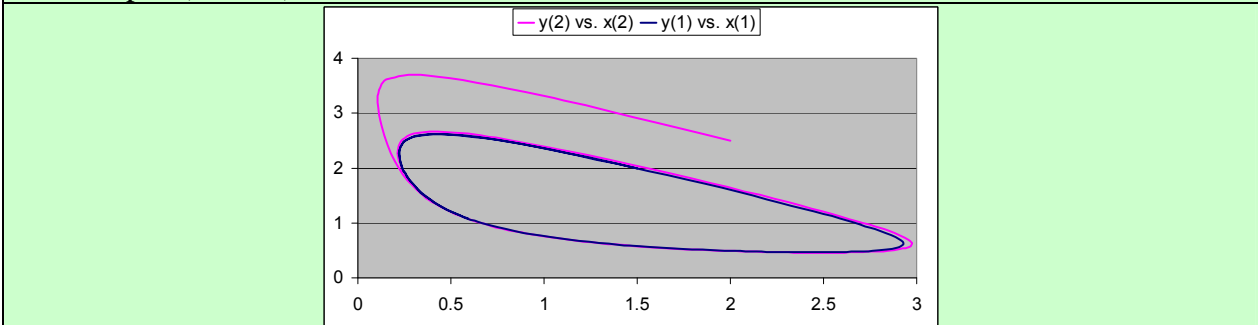
4. Solving numerically (i=1..n):

$k_2 = 0.88$	$x_{n+1} = x_n + (1 - k_2x_ny_n^2) \cdot 10^{-2}$ $y_{n+1} = y_n + (k_2x_ny_n^2 - y_n) \cdot 10^{-2}$	$n = 10000$	Case 1 $x_0 = 1.5$ $y_0 = 2$	Case 2 $x_0 = 2$ $y_0 = 2.5$
--------------	--	-------------	------------------------------------	------------------------------------

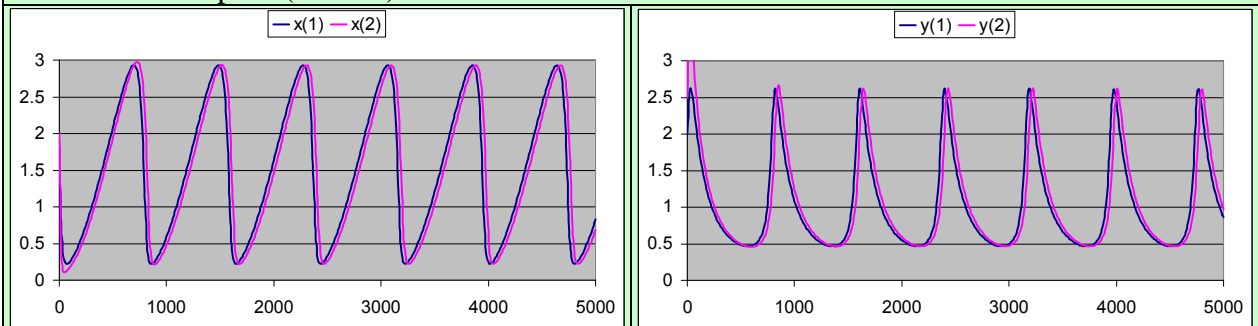
5. Excel sheet:

	A	B	C	D	E	F	G	H
1	k2=	0.88			Case 1			Case 2
2	=	1e-2	i	xi		yi	xi	yi
3	Case 1		=0	=B4		=B5	=B7	=B8
4	x0=	1.5	=D2+1	=D3+(1-B\$1*D3*E3^2)*B\$2		=E3+(B\$1*D3*E3^2-E3)*B\$2
5	y0=	2
6	Case 2	
7	x0=	2
8	y0=	2.5

6. Phases plot (n=1000):



7. Intermediaries plots (n=5000):



8. Simulation analysis

Not for any values the attractor appears. For a given k_2 (such is 0.88) exists minimum values of x_0 and y_0 (x_{0-min}, y_{0-min}) from which periodical oscillations occurs and the system tends to the attractor.

Algorithm for simulation of Lotka - Volterra kinetics

There is a fundamental difference between brusselator and Lotka-Volterra mechanisms evolution: if the L-V evolves oscillating around the initial values of the intermediaries, the brusselator converges (in time) to same equation, independent of the initial values of the intermediaries concentrations.

(oregonator mechanism)

Initiated by a group from Oregon (USA), implies 18 elementary stages and 21 different chemical species [126]. A simplified variant is given below (see *Algorithm for simulation of oregonator kinetics*).

1. Writing of elementary reactions and of reaction rates equations ($[X]=x$, and idem for the rest of):

$A + Y \xrightarrow{k_1} X$ $v_{(1)} = k_1 ay$	$X + Y \xrightarrow{k_2} P$ $v_{(2)} = k_2 xy$	$A + X \xrightarrow{k_3} 2X + Z$ $v_{(3)} = k_3 ax$	$2X \xrightarrow{k_4} Q$ $v_{(4)} = k_4 x^2$	$Z \xrightarrow{k_5} Y$ $v_{(5)} = k_5 z$
---	---	--	---	--

2. Applying of atoms conservation principle (unknowns $[X] = x$; $[Y] = y$, $[Z] = z$):

(X): $\dot{x} = k_1 ay - k_2 xy + k_3 ax - 2k_4 x^2$	(Y): $\dot{y} = -k_1 ay - k_2 xy + k_5 z$	(Z): $\dot{z} = k_3 ax - k_5 z$
--	---	---------------------------------

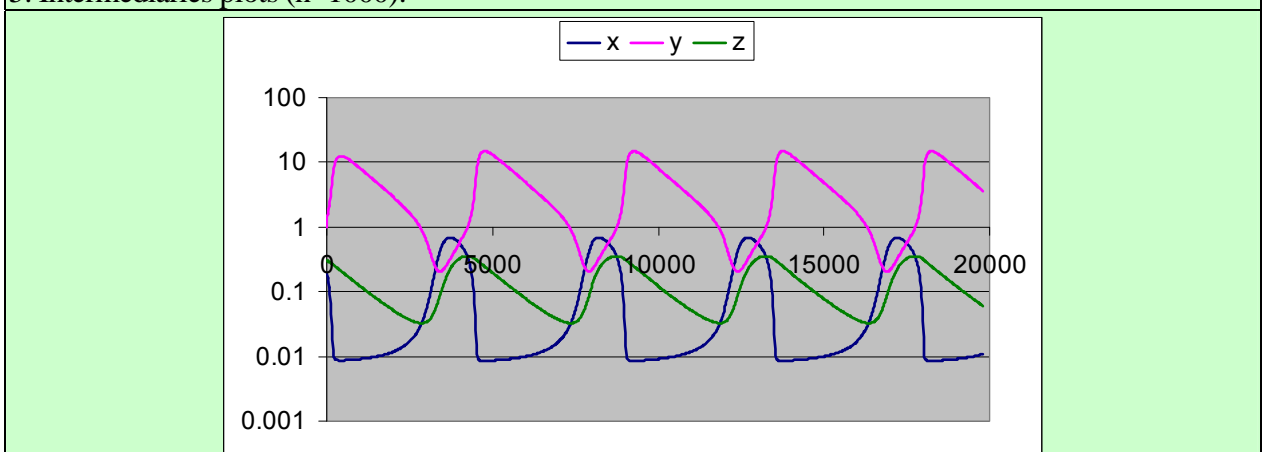
3. Solving numerically (i=1..n, after a long series of substitutions and rescaling):

$x_{n+1} = x_n + (qy_n - x_n y_n + x_n(1 - x_n)) /$ $y_{n+1} = y_n + (-qy_n - x_n y_n + fz_n) /$ $z_{n+1} = z_n + (x_n - y_n)$	$x_0 = 0.2$ $y_0 = 1$ $z_0 = 0.3$	$= 8e-3$ $= 1e-1$	$q = 2e-3$ $f = 1$	$= 1e-3$ $n = 19800$
--	---	----------------------	-----------------------	-------------------------

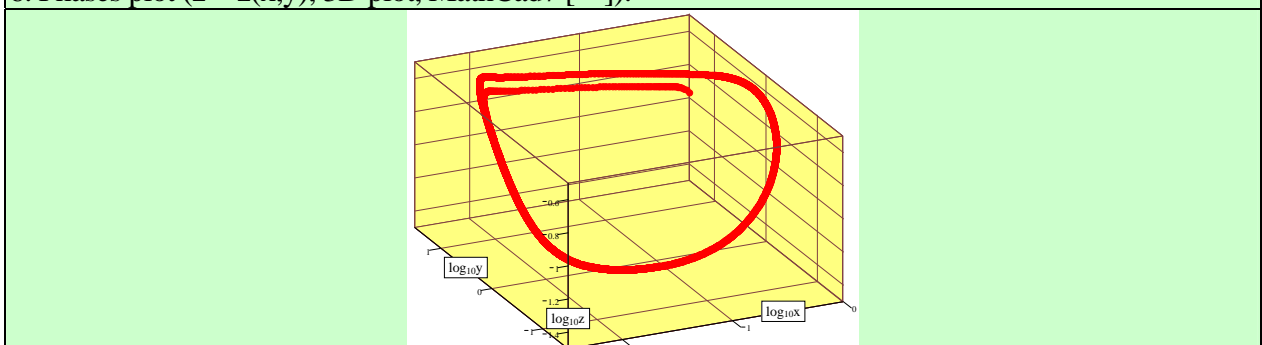
4. Excel sheet:

	A	B	C	D	E	F	F	
1	q=	8e-3	i	xi		yi	zi	
2	=	1e-1	=0	=B6		=B7	=B8	
3	=	2e-3	=D2+1	=IF(D2+(B\$1*E2-D2*E2	+D2*(1-D2))*B\$5/B\$2>0,	=IF(E2+(-B\$1*E2-D2*E2+	+B\$4*F2)*B\$5/B\$3>0,	=IF(F2+
				D2+(B\$1*E2-D2*E2+	D2*(1-D2))*B\$5/B\$2,0)	E2+(-B\$1*E2-D2*E2+	F2+(D2-F2)*B\$5,0)	
				D2*(1-D2))*B\$5/B\$2,0)		B\$4*F2)*B\$5/B\$3,0)		
4	f=	1	
5	=	1e-3	
6	x0=	0.2	
7	y0=	1	
8	z0=	0.3	

5. Intermediaries plots (n=1000):



6. Phases plot ($z = z(x,y)$); 3D plot, MathCad7 [118]:



Algorithm for simulation of oregonator kinetics

Referin e

¹ George BOOLE, 1854. An Investigation of the Laws of Thought. (Reprinted 2003 as Laws of Thought. New York: Prometheus Books. ISBN 1-59102-089-1), p. 430.

² Ronald A. FISHER, 1922. On the interpretation of χ^2 from contingency tables, and the calculation of P. Journal of the Royal Statistical Society 85(1):87-94. DOI:10.2307/2340521

³ Wiki, 2012. List of unsolved problems in physics (<http://en.wikipedia.org/w/index.php?oldid=469618085>, version of January 5, 2012): http://en.wikipedia.org/wiki/List_of_unsolved_problems_in_physics

⁴ Werner HEISENBERG, 1927. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, Zeitschrift für Physik 43(3-4):172-198, doi: 10.1007/BF01397280

⁵ Schrödinger E, 1926. An Undulatory Theory of the Mechanics of Atoms and Molecules. Physical Review 28(6): 1049-1070.

⁶ Barry N. TAYLOR, Ambler THOMPSON (Eds.), 2008. The international system of units (SI). NIST Special Publication 330. Gaithersburg: National Institute of Standards and Technology.

⁷ Carl WAGNER, 1931. On the theory of the rectifier effect (In German). Physikalische Zeitschrift 32:641-645.

⁸ William D. PHILLIPS, Harold J. METCALF, 1982. Laser Deceleration of an Atomic Beam. Physical Review Letters 48(9):596-599.

⁹ R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, D.E. Knuth, 1996. On the Lambert W function. Advances in Computational Mathematics 5(1): 329-359.

¹⁰ Roberto C. SOTERO, Iturria-Medina YASSER, 2011. From blood oxygenation level dependent (BOLD) signals to brain temperature maps. Bulletin of Mathematical Biology 73(11):2731-2747.

¹¹ M.A. Jenkins, 1975. Algorithm 493: Zeros of a Real Polynomial [C2]. ACM Transactions on Mathematical Software 1(2):178-189.

¹² <http://www.netlib.org/toms/493>

¹³ William H. PRESS, Saul A. TEUKOLSKY, William T. VETTERLING, Brian P. FLANNERY, 1992. Gauss-Jordan Elimination (§2.1) and Gaussian Elimination with Backsubstitution (§2.2) In: Numerical Recipes in FORTRAN: The Art of Scientific Computing (2nd ed.) Cambridge: Cambridge University Press p. 27-32 (§2.1) and 33-34 (§2.2).

¹⁴ Ronald A. FISHER, 1923. Studies in Crop Variation. II. The Manurial Response of Different Potato Varieties. Journal of Agricultural Science 13:311-320.

¹⁵ Isaac NEWTON, 1711. Analysis through a series of quantities, integrals, and differentials: with a list of lines of the third order (in Latin). London: William Jones.

¹⁶ Robert COTES, 1722. Harmony of measures (in Latin). Cambridge: Robert Smith.

¹⁷ Thomas SIMPSON, 1823. Doctrine and Application of Fluxions. Edinburgh: Bell and Bradfute.

¹⁸ Teorema Limit Central

÷ Cronologia contribu iilor majore:

- Abraham DE MOIVRE. 1733. Approximatio ad Summam Terminorum Binomii $(a+b)^n$ in Seriem expansi. In: The Doctrine of Chance: or The Method of Calculating the Probability of Events in Play (Abraham DE MOIVRE). W. Pearform 1738: 235-243.
- Joseph L. LAGRANGE. 1776. Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations; dans lequel on examine les avantages de cette méthode par le calcul des probabilités; et où l'on résoud différents problèmes relat ifs à cette matière. Miscellanea Taurinensia 5:167-232.
- Pierre S. LAPLACE. 1812. Théorie Analytique des Probabilités. Courcier, 465 p.
- Aleksandr M. LIAPUNOV. 1901. Nouvelle forme du théoreme sur la limite des probabilités. Mémoires de l'Académie Impériale des Sciences de St. Pétersbourg 12(5):1-24.

÷ Enun ul teoremei (fie $(X_n)_{n \geq 1}$ variabile independente $i \exists > 0$ a.î. $\mu_{2+}(X_n) < \infty$):

$$\text{○ dac } \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \mu_{2+\delta}(X_k)}{\left(\sum_{k=1}^n \sigma_k^2\right)^{(2+\delta)/2}} = 0 \text{ atunci } \frac{\sum_{i=1}^n (X_n - \mu_1(X_n))}{\sqrt{\sum_{k=1}^n \sigma_k^2}} \xrightarrow[n \rightarrow \infty]{} N(0,1)$$

¹⁹ Carl F. GAUSS, 1809. Theoria Motus Corporum Coelestium. Perthes et Besser, Hamburg. Translated, 1857, as Theory of Motion of the Heavenly Bodies Moving about the Sun in Conic Sections, trans. C. H. Davis. Little, Brown; Boston. Reprinted, 1963, Dover, New York.

²⁰ Pierre S. LAPLACE, 1812. Théorie Analytique des Probabilités. Paris: Courcier.

²¹ Ronald A. FISHER, 1920. A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error, and by the Mean Square Error. Monthly Notices of the Royal Astronomical Society 80:758-770.

²² Fisher, R. A. and Mackenzie W. A., 1923. Studies in Crop Variation. II. The manurial response of different potato varieties. Journal of Agricultural Science 13:311-320.

- ²³ R. A. Fisher, 1924. The Conditions Under Which χ^2 Measures the Discrepancy Between Observation and Hypothesis. *Journal of the Royal Statistical Society* 87:442-450.
- ²⁴ R. A. Fisher, 1912. On an Absolute Criterion for Fitting Frequency Curves. *Messenger of Mathematics* 41:155-160.
- ²⁵ BENFORD Frank. 1938. The law of anomalous numbers. *Proceedings of the American Philosophical Society* 78(4):551-572.
- ²⁶ HILL Theodore P. 1995. Base invariance implies Benford's Law. *Proceedings of the American Mathematical Society* 123(3):887-895.
- ²⁷ Carlos M JARQUE, Anil K BERA. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ Lett* 6(3):255-259.
- ²⁸ Carlos M JARQUE, Anil K BERA. 1981. Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Econ Lett* 7(4):313-318.
- ²⁹ KOLMOGOROV Andrey. 1941. Confidence Limits for an Unknown Distribution Function. *The Annals of Mathematical Statistics* 12(4):461-463.
- ³⁰ SMIRNOV Nikolay V. 1948. Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics* 19(2):279-281.
- ³¹ Theodore W ANDERSON, Donald A DARLING. 1952. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics* 23(2):193-212.
- ³² Fritz W SCHOLZ, Michael A STEPHENS. 1987. K-sample Anderson-Darling Tests. *Journal of the American Statistical Association* 82(399):918-924.
- ³³ Department of Defense Handbook. 2002. *Composite Materials Handbook. Volume 1. Polymer Matrix Composites Guidelines for Characterization of Structural Materials. Chapter 8. Statistical Methods. 8.3.2.2 The k-sample Anderson-Darling test MIL-HDBK-17-1F:8-17.*
- ³⁴ Lorentz JÄNTSCHI. 2009. <http://1.academicdirect.org/Statistics/tests/kAD/>, k-sample Anderson-Darling.
- ³⁵ Fritz W SCHOLZ, Michael A STEPHENS. 1986. K-Sample Anderson-Darling Tests of Fit, for Continuous and Discrete Cases. Technical Report. University of Washington. GN-22:81.
- ³⁶ A. Trujillo-Ortiz, R. Hernandez-Walls, K. Barba-Rojo, A. Castro-Perez. 2007. AnDartest:Anderson-Darling test for assessing normality of a sample data. <http://mathworks.com/matlabcentral/fileexchange/14807>
- ³⁷ PEARSON Karl. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine* 5th Ser 50:157-175.
- ³⁸ Ronald A. FISHER, 1935. The Mathematical Distributions Used in the Common Tests of Significance. *Econometrica* 3:353-365.
- ³⁹ LIEBETRAU Albert M. 1983. Measures of association. Newbury Park, CA: Sage Publications. *Quantitative Applications in the Social Sciences* 32:1-96 (p.13).
- ⁴⁰ FISHER Ronald A. 1922. On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society* 85:87-94.
- ⁴¹ FISHER Ronald A. 1924. The Conditions Under Which χ^2 Measures the Discrepancy Between Observation and Hypothesis. *Journal of the Royal Statistical Society* 87:442-450.
- ⁴² SNEDECOR George W. and COCHRAN William G. 1989. *Statistical Methods, Eighth Edition*, Iowa State University Press.
- ⁴³ HARTLEY Ralph V L. 1928. Transmission of Information. *Bell System Technical Journal* 1928:535-563.
- ⁴⁴ Software. 2008. EasyFit v.5. MathWave Technologies. <http://www.mathwave.com>
- ⁴⁵ SCOTT David. 1992. *Multivariate Density Estimation*. John Wiley, Chapter 3.
- ⁴⁶ Software. 2005. Dataplot. National Institute for Standards and Technology. <http://www.itl.nist.gov/div898/software/dataplot.html>
- ⁴⁷ Isaac NEWTON, 1687. *Philosophiæ naturalis principia mathematica*. London: Juffu Societatis Regiæ, full text: <http://www.ntnu.no/ub/spesialsamlingene/ebok/02a019654.html>
- ⁴⁸ Julian L. COOLIDGE, 1949. The Story of the Binomial Theorem. *The American Mathematical Monthly* 56(3):147-157. full text: <http://www.jstor.org/stable/2305028>
- ⁴⁹ Jacobi BERNOULLI, 1713. *Ars Conjectandi*. Basel: Thurnisius.
- ⁵⁰ Abraham De MOIVRE, 1738. Approximatio ad Summam Terminorum Binomii $(a+b)^n$ in Seriem expansi (presented privately to some friends in 1733), p. 235-243 In: *The Doctrine of Chance: or The Method of Calculating the Probability of Events in Play* (2nd ed.), London: W. Pearson.
- ⁵¹ Johann C. F. GAUSS, 1823. *Theoria combinationum observationum erroribus minnimis obnoxiae. Pars prior, et Pars posterior*. *Comment. Societ. R. Sci. Göttingensis Recentiores* 5:33-90.
- ⁵² Abraham WALD, 1939. Contributions to the Theory of Statistical Estimation and Testing Hypothesis. *The Annals of Mathematical Statistics* 10(4):299-326.
- ⁵³ Allan AGRESTI, Brent A. COULL, 1998. Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. *American Statistician* 52(2):119-126.
- ⁵⁴ Allan AGRESTI, Yongyi MIN, 2001. On small-sample confidence intervals for parameters in discrete distributions. *Biometrics* 57(3):963-971.

- ⁵⁵ Allan AGRESTI, 2001. Exact inference for categorical data: Recent advances and continuing controversies. *Statistics in Medicine* 20(17-18):2709-2722.
- ⁵⁶ Allan AGRESTI, 2002. Dealing with discreteness: Making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Statistical Methods in Medical Research* 12(1):3-21.
- ⁵⁷ Wolfgang K. ENGEL, 1973. Onset of synthesis of mitochondrial enzymes during mouse development. Synchronous activation of parental alleles at the gene locus for the M form of NADP dependent malate dehydrogenase. *Humangenetik* 20(2):133-140.
- ⁵⁸ David L. MEADOWS, Jerome S. SCHULTZ, 1991. A molecular model for singlet/singlet energy transfer of monovalent ligand/receptor interactions. *Biotechnology and Bioengineering* 37(11):1066-1075.
- ⁵⁹ Mark D. SZCZELKUN, 2002. Kinetic models of translocation, head-on collision, and DNA cleavage by type I restriction endonucleases. *Biochemistry* 41(6):2067-2074.
- ⁶⁰ Tijen TANYALÇIN, François J. M. EYSKENS, Eddy PHILIPS, Marc LEFEVERE, Benal BÜYÜKGEBİZ, 2002. A marked difference between two populations under mass screening of neonatal TSH and biotinidase activity. *Accreditation and Quality Assurance* 7(11):498-506.
- ⁶¹ Enrique A. OSSET, Mercedes FERNÁNDEZ, Juan A. RAGA, Aneta KOSTADINOVA, 2005. Mediterranean *Diploodus annularis* (Teleostei: Sparidae) and its brain parasite: Unforeseen outcome. *Parasitology International* 54(3):201-206.
- ⁶² Clarke R. CONANT, Marc R. Van GILST, Stephen E. WEITZEL, William A. REES, Peter H. von HIPPEL, 2005. A Quantitative Description of the Binding States and In Vitro Function of Antitermination Protein N of Bacteriophage? *Journal of Molecular Biology* 348(5):1039-1057.
- ⁶³ Matthew A. CARLTON, William D. STANSFIELD, 2005. Making babies by the flip of a coin? *American Statistician* 59(2):180-182.
- ⁶⁴ Ronald A. FISHER, 1925. Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society* 22:700-725.
- ⁶⁵ Ronald A. FISHER, 1912. On an Absolute Criterion for Fitting Frequency Curves. *Messenger of Mathematics* 41:155-160.
- ⁶⁶ Evan COOCH, Gary WHITE, 2011. Program MARK. A gentle introduction. Internet: E-book (22 Dec. 2011, 927p.).
- ⁶⁷ Bernard ROSNER, 1995. Hypothesis Testing: Categorical Data (p. 345-442) In: *Fundamentals of Biostatistics* (4th ed.). Belmont: Duxbury Press.
- ⁶⁸ Robert G. NEWCOMBE, 1998. Two-sided confidence intervals for the single proportion; comparison of seven methods. *Statistics in Medicine* 17(8):857-872.
- ⁶⁹ Ana M. PIRES, 200X. Confidence intervals for a binomial proportion: comparison of methods and software evaluation (web only). URL: http://www.math.ist.utl.pt/~apires/AP_COMPSTAT02.pdf
- ⁷⁰ Lawrence D. BROWN, Tony T. CAI, Anirban DasGUPTA, 2001. Interval estimation for a binomial proportion. *Statistical Science* 16(2):101-133.
- ⁷¹ Edwin B. WILSON, 1927. Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association* 22(158):209-212.
- ⁷² J. R. ANDERSON, Leslie E. BERNSTEIN, Malcolm C. PIKE, 1982. Approximate Confidence Intervals for Probabilities of Survival and Quantiles in Life-Table Analysis. *Biometrics* 38(2):407-416.
- ⁷³ Barnet WOOLF, 1955. On estimating the relation between blood group and disease. *Annals of Human Genetics* 19:251-253.
- ⁷⁴ John J. GART, 1966. Alternative analyses of contingency tables. *Journal of Royal Statistical Society B* 28:164-179.
- ⁷⁵ Ronald A. FISHER, 1956. *Statistical Methods for Scientific Inference*. Edinburgh: Oliver and Boyd.
- ⁷⁶ C. J. Clopper, Egon S. PEARSON, 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26:404-413.
- ⁷⁷ Allan AGRESTI, 2003. Dealing with discreteness: making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Statistical Methods in Medical Research* 12:3-21.
- ⁷⁸ Harold JEFFREYS. *Theory of Probability* (3rd Ed). Clarendon Press, Oxford, 1961. <http://lccn.loc.gov/62000074>
- ⁷⁹ Colin R. BLYTH, Harold A. STILL, 1983. Binomial confidence intervals. *Journal of the American Statistical Association* 78(381):108-116.
- ⁸⁰ George CASELLA, 1986. Refining binomial confidence intervals. *The Canadian Journal of Statistics* 14(2):113-129.
- ⁸¹ Sorana D. BOLBOAC , Lorentz JÄNTSCHI, 2008. Optimized Confidence Intervals for Binomial Distributed Samples. *International Journal of Pure and Applied Mathematics* 47(1):1-8.
- ⁸² Lorentz JÄNTSCHI, Sorana D. BOLBOAC , 2007. How to Assess Dose-Response Study Outcome: a Statistical Approach. *Recent Advances in Synthesis & Chemical Biology VI*:P36.
- ⁸³ Lorentz JÄNTSCHI, Sorana D. BOLBOAC , 2010. Exact Probabilities on Confidence Limits for Binomial Samples: Applied to the Difference between Two Proportions, *TheScientificWorldJOURNAL* 10(5):865-878.
- ⁸⁴ Sorana D. BOLBOAC , Lorentz JÄNTSCHI, 2008. Assessment of Confidence Intervals used in Medical Studies (in Romanian), Cluj-Napoca: AcademicPres & AcademicDirect, p. 234 (+p. 10 intro), 2008.

- ⁸⁵ Lorentz JÄNTSCHI, 2011. Sampling distribution in biodiversity analysis - Project proposal PN-II-ID-PCE-2011-3-0103, not funded (I was found "not eligible", see UEFISCDI.ro website).
- ⁸⁶ Maurice G. KENDALL, 1962. Rank Correlation Methods (3rd Ed). London: C. Griffin & Co.
- ⁸⁷ Karl F. R. S. PEARSON, 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine 5th Ser 50(302):157-175. doi: 10.1080/14786440009463897
- ⁸⁸ Charles SPEARMAN, 1904. The proof and measurement of association between two things. Amer. J. Psychol. 15:72-101.
- ⁸⁹ Maurice G. KENDALL, Babington B. SMITH, 1939. The Problem of m Rankings. The Annals of Mathematical Statistics 10(3):275-287.
- ⁹⁰ Augustin-Louis CAUCHY, 1821. Oeuvres 2. III, p. 373
- ⁹¹ Viktor BUNYAKOVSKY, 1859. Sur quelques inégalités concernant les intégrales aux différences finis (in French). Mem. Acad. Sci. St. Petersburg 7(1):9.1-9.18.
- ⁹² Hermann A. SCHWARZ, 1888. 15. Einführung der Constante c In: Über ein Flächen kleinsten Flächeninhalts betreffendes Problem der Variationsrechnung (in German). Acta Societatis Scientiarum Fennicae XV:344-345.
- ⁹³ Maurice H. QUENOUILLE, 1949. Approximate tests of correlation in time-series. J R Stat Soc B 11(1):68-84.
- ⁹⁴ Maurice H. QUENOUILLE, 1956. Notes on bias in estimation. Biometrika 43(3-4):353-360.
- ⁹⁵ Ronald A. FISHER, 1948. Combining independent tests of significance. The American Statistician 2(5):30.
- ⁹⁶ Bradley EFRON, 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics 7(1):1-26.
- ⁹⁷ Edgar C. FIELLER, Herman O. HARTLEY, Egon S. PEARSON, 1957. Tests for rank correlation coefficients. I. Biometrika 44:470-481.
- ⁹⁸ Maurice G. KENDALL, Alan STUART, 1973. §31.19 & §31.21, In: The Advanced Theory of Statistics, Volume 2: Inference and Relationship. London: C. Griffin. <http://lccn.loc.gov/77360686>
- ⁹⁹ Maurice G. KENDALL, 1938. A New Measure of Rank Correlation. Biometrika 30(1-2):81-89.
- ¹⁰⁰ Leo A. GOODMAN, William H. KRUSKAL, 1954. Measures of Association for Cross Classifications. Journal of the American Statistical Association 49(268):732-764.
- ¹⁰¹ Leo A. GOODMAN, William H. KRUSKAL, 1972. Measures of Association for Cross Classifications IV: Simplification of Asymptotic Variances. Journal of the American Statistical Association 67(338):415-421.
- ¹⁰² Wassily Hoeffding, 1948. A Non-parametric Test of Independence. Annals of Mathematical Statistics, 19(4):546-557.
- ¹⁰³ Julius R. BLUM, James KIEFER, Murray ROSENBLATT, 1961. Distribution Free Tests of Independence Based on the Sample Distribution Function. Ann Math Statist 32(2):485-498.
- ¹⁰⁴ Robert H. SOMERS, 1962. A new asymmetric measure of association for ordinal variables. American Sociological Review 27(6):799-811.
- ¹⁰⁵ Maurice G. KENDALL, 1990. Rank Correlation Methods. Oxford: Oxford University Press.
- ¹⁰⁶ Lorentz JÄNTSCHI, Sorana D. BOLBOAC , 2009. Observation vs. Observable: Maximum Likelihood Estimations according to the Assumption of Generalized Gauss and Laplace Distributions. Leonardo Electronic Journal of Practices and Technologies 8(15):81-104.
- ¹⁰⁷ Pierre-Simon LAPLACE, 1814. Théorie analytique des probabilité (in French). Paris: Courcier.
- ¹⁰⁸ Carl Friedrich GAUSS. 1809. Theoria motus corporum coelestium in sectionibus conicis solem ambientium. Paris: Treuttel & Würtz. London: R.H. Evans.
- ¹⁰⁹ Lorentz JÄNTSCHI, Sorana D. BOLBOAC , 2007. The Jungle of Linear Regression Revisited. Leonardo Electronic Journal of Practices and Technologies 6(10):169-187.
- ¹¹⁰ William H. PRESS, Saul A. TEUKOLSKY, William T. VETTERLING, Brian P. FLANNERY, 1992. Gauss-Jordan Elimination (§2.1) and Gaussian Elimination with Backsubstitution (§2.2) In: Numerical Recipes in FORTRAN: The Art of Scientific Computing (2nd ed.) Cambridge: Cambridge University Press p. 27-32 (§2.1) and 33-34 (§2.2).
- ¹¹¹ Bronsted J.N., 1923. Some remarks on the concept of acids and bases (In German). Recueil des Travaux Chimiques des Pays-Bas 42(8):718-728, DOI: 10.1002/recl.19230420815
- ¹¹² Lowry T.M., 1923. The Uniqueness of Hydrogen. Chemistry and Industry 42(3):43-47. DOI: 10.1002/jctb.5000420302
- ¹¹³ Leonor MICHAELIS, Maud L. MENTEN, 1913. Die Kinetik der Invertinwirkung. Biochem Z 49:333-369.
- ¹¹⁴ George E. BRIGGS, John B.S. HALDANE, 1925. A note on the kinematics of enzyme action. Biochem J 19(2):338-339.
- ¹¹⁵ Victor HENRI, 1903. Lois Générales de l'Action des Diastases. Paris: Hermann.
- ¹¹⁶ Frederick A. LINDEMANN, Svante ARRHENIUS, Irving LANGMUIR, N. R. DHAR, J. PERRIN, W. C. McC. LEWIS, 1922. Discussion on "the radiation theory of chemical action". Transactions of the Faraday Society, 17:598-606. doi: 10.1039/TF9221700598
- ¹¹⁷ Cyril N. HINSHELWOOD, 1926. On the Theory of Unimolecular Reaction. Proc. R. Soc. Lond. A 113:230-233. DOI:10.1098/rspa.1926.0149
- ¹¹⁸ MathSoft, 1997. MathCad7 Professional (software). MathSoft. URL: <http://www.mathsoft.com>

-
- ¹¹⁹ Alfred J. LOTKA, 1909. Contribution to the Theory of Periodic Reactions. *The Journal of Physical Chemistry* 14(3):271-274. doi: 10.1021/j150111a004
- ¹²⁰ Alfred J. LOTKA, 1909. Undamped oscillations derived from the law of mass action. *Journal of the American Chemical Society* 42(8):1595-1599. doi: 10.1021/ja01453a010
- ¹²¹ Vito VOLTERRA, 1926. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem. Acad. Lincei Roma* 2:31-113.
- ¹²² Ilya PRIGOGINE, Grégoire NICOLIS, 1967. On Symmetry-Breaking Instabilities in Dissipative Systems. *J Chem Phys* 46(9):3542-3550.
- ¹²³ Richard M. NOYES, 1976. Oscillations in chemical systems. XII. Applicability to closed systems of models with two and three variables. *The Journal of Chemical Physics* 64(4):1266-1269.
- ¹²⁴ G. B. COOK, Peter GRAY, D. G. KNAPP, Stephen K. SCOTT, 1989. Bimolecular routes to cubic autocatalysis. *The Journal of Physical Chemistry* 93(7):2749-2755.
- ¹²⁵ Daniel T. GILLESPIE, 1977. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* 81(25):2340-2361.
- ¹²⁶ Richard J. FIELD, Richard M. NOYES, 1974. Oscillations in Chemical Systems IV. Limit cycle behavior in a model of a real chemical reaction, *J. Chem. Phys.* 60:1877-1884. DOI:10.1063/1.1681288