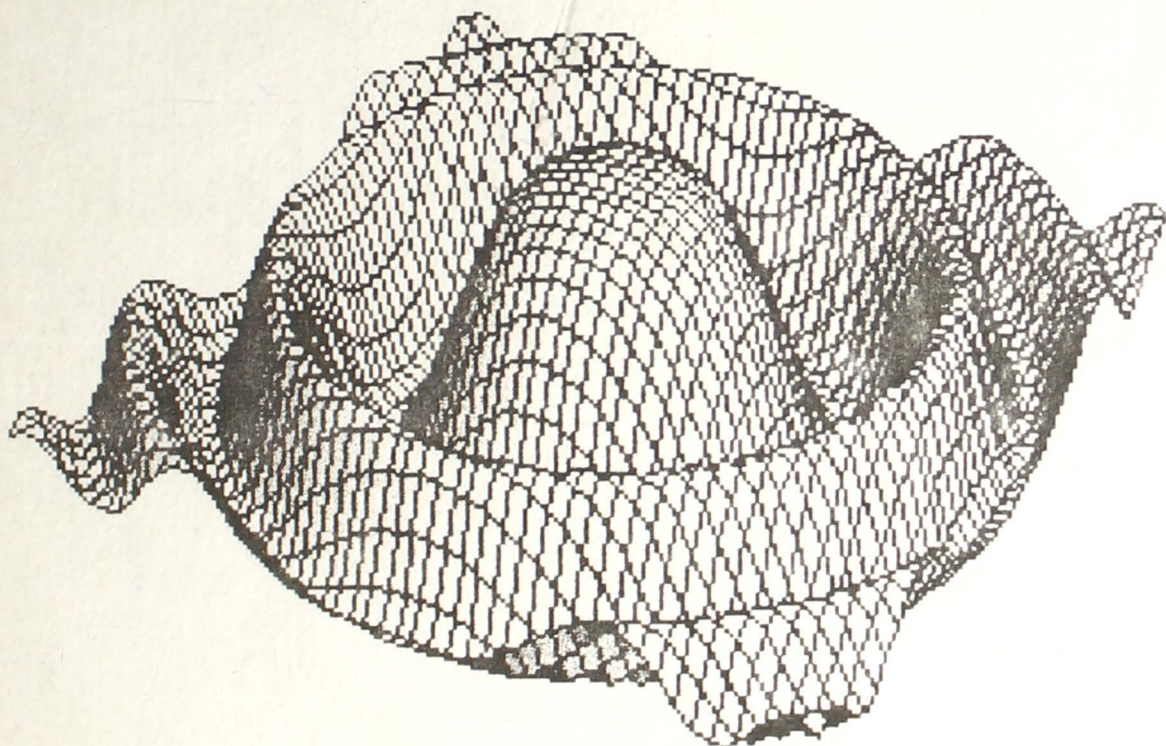


CAMELIA BALA

LUCIA MUTIHAC

CHEMOMETRIE

PRINCIPII ȘI APLICAȚII
ÎN CHIMIA ANALITICĂ



EDITURA UNIVERSITĂȚII DIN BUCUREȘTI

2000

<https://biblioteca-digitala.ro> / <https://unibuc.ro>

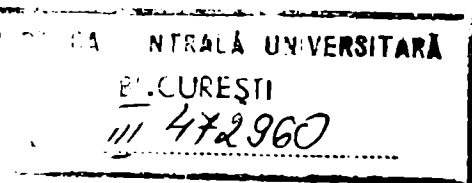
CAMELIA BALA

LUCIA MUTIAC

CHEMOMETRIE
PRINCIPII ȘI APLICAȚII
ÎN CHIMIA ANALITICĂ

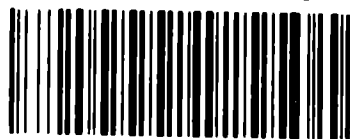
EDITURA UNIVERSITĂȚII DIN BUCUREȘTI
2000

Referenți științifici: Prof. dr. CONSTANTIN PĂTROESCU
Conf. dr. VICTOR DAVID



602/100

B.C.U. București



C20003844

Contribuția autorilor: Lucia MUTIHAC - cap. I, II, V, VI.
Camelia BALA - cap. III, IV.

© Editura Universității din București
Șos. Panduri, 90-92, București - 76235; Telefon/Fax 410.23.84

**Descrierea CIP a Bibliotecii Naționale
MUTIHAC, LUCIA**

Chemometrie: principii și aplicații în chimia analitică /

Lucia Mutihac, Camelia Bala - București

Editura Universității din București, 2000

144 p.; 23,5 cm.

Bibliogr.

ISBN 973-575-437-1

I. Bala, Camelia
51-7:543(075.8)

CUPRINS

Introducere	5
I. Metode chemometrice în chimia analitică	7
I.1. Chemometria și chimia analitică	7
I.2. Metode chemometrice	8
Bibliografie	10
II. Elemente de statistică matematică utilizate în chimia analitică	11
II.1. Generalități	11
II.2. Distribuții de probabilitate	13
II.2.1. Distribuția variabilelor aleatoare	13
II.2.2. Funcția de distribuție normală a variabilelor aleatoare	14
II.2.3. Funcția de distribuție χ^2	16
II.2.4. Funcția de distribuție t (Student)	17
II.2.5. Funcția de distribuție F	19
II.2.6. Distribuția binomială	20
II.3. Selecție și estimare	20
II.4. Intervale de încredere	21

II.4.1. Intervalul de încredere pentru media unei distribuții	22
II.4.2. Intervalul de încredere pentru dispersia de selecție	24
II.4.3. Intervalul de încredere pentru parametrii distribuției binomiale	25
II.5. Analiza de dispersie	25
Bibliografie	26
III. Concepte utilizate în evaluarea metodelor analitice	27
III.1. Tipuri de erori în chimia analitică	28
III.2. Precizia și acuratețea în chimia analitică	30
III.3. Sensibilitatea și limita de detecție	33
III.4. Selectivitatea și rezoluția	43
Bibliografie	48
IV. Analiza de regresie	50
IV.1. Principalele tipuri de regresii utilizate în analiza chimică	50
IV.2. Etapele analizei de regresie	54
IV.2.1. Definirea modelului	54
IV.2.2. Estimarea parametrilor modelului	56
IV.2.3. Validarea modelului	60
IV.2.3.1. Analiza de varianță-ANOVA	62

V. Procesarea semnalelor 70**V.1. Transformate Fourier 70****V.1.1. Transformata Fourier unidimensională (1D) 71****V.1.1.1. Proprietățile transformatelor
Fourier 72****V.1.1.2. Teorema de eșantionare 82****V.1.1.3. Transformata Fourier discretă 87****V.1.1.4. Convoluția 92****V.1.1.5. Corelația 99****V.1.1.6. Eșantionarea funcțiilor continue
unidimensionale 102****V.1.2. Transformata Fourier bidimensională (2D) 104****Bibliografie 106****VI. Aplicații ale rețelelor neuronale în chimie 108****VI.1. Introducere 108****VI.2. Noțiuni de teorie a modelării 109****VI.3. Modele de rețele neuronale 118****VI.3.1. Rețele neuronale formale cu instruire
supervizată (memorii asociative) 118**

VI.3.2. Rețele neuronale formale cu instruire nesupervizată (clasificatori vectoriali)	119
VI.4. Aplicații ale rețelelor neuronale	121
VI.4.1. Noțiuni fundamentale	121
VI.4.2. Clase de aplicații specifice rețelelor neuronale	123
VI.4.3. Abordarea modelării neuronale	126
VI.4.4. Aplicații ale rețelelor neuronale în chimie	128
VI.4.4.1. Reactivitatea legăturilor chimice	128
VI.4.4.2. Optimizarea proceselor HPLC	130
VI.4.4.3. Relația structură-activitate	131
VI.4.4.4. Caracteristicile generale ale structurii secundare a proteinelor	133
VI.4.4.5. Corelația spectru-structură	136
Bibliografie	139

Introducere

Chemometria este disciplina care utilizează metodele statistico-matematice în prelucrarea datelor analitice precum și în optimizarea metodelor de analiză. Alt obiectiv al chemometriei îl constituie adaptarea metodelor informaticii la cerințele necesare analiștilor. Alături de metodele statistico-matematice, chemometria pune la îndemâna chimiștilor metodele inteligenței artificiale în prelucrarea datelor chimice cum ar fi: rețelele neuronale și programarea simbolică.

Chimia analitică cuprinde un ansamblu de metode chimice, fizice și fizico-chimice, numite metode de analiză și metode statistice, matematice și informatice, numite metode chemometrice utilizate în procesarea și optimizarea informației chimice.

Lucrarea de față prezintă câteva aplicații semnificative ale metodelor chemometrice în procesarea datelor chimice. Astfel primul capitol cuprinde aspecte generale ale implicării chemometriei în chimia analitică. În capitolul al doilea sunt prezentate succint elemente de statistică matematică utilizate în chimia analitică. Metodele statistico-matematice vin în ajutorul chimiștilor atât în ceea ce privește perfecționarea unei metode de analiză existentă, cât și la elaborarea de noi metode de analiză.

În continuare sunt prezentate câteva dintre cele mai semnificative criterii utilizate în evaluarea metodelor analitice.

O atenție deosebită este acordată utilizării metodelor inteligenței artificiale în chimia analitică. Astfel în capitolele V și VI sunt prezentate aspecte ale procesării semnalelor, aplicații ale rețelelor neuronale atât

în recunoașterea de configurații cât și în modelarea, optimizarea, prelucrarea și interpretarea datelor experimentale.

Lucrarea de față concepută și realizată pe schema cursului predat studenților de la studii aprofundate, specializarea Chimie Analitică a Facultății de Chimie din Universitatea din București se adresează și chimiștilor analiști interesați în prelucrarea datelor chimice cu ajutorul metodelor matematice moderne.

Autorii

I. METODE CHEMOMETRICE ÎN CHIMIA ANALITICĂ

I.1. Chemometria și chimia analitică

Chimia analitică înglobează analiza chimică precum și ansamblul metodelor de interpretare a rezultatelor în urma cărora se iau decizii cu scopul de a rezolva probleme propuse ca subiecte de studiu. Chimia analitică este deci o metodologie care se bazează pe un ansamblu de metode chimice, fizice și fizico-chimice, numite *metode de analiză chimică* și metode informatice și statistice adaptate cerințelor specifice chimiștilor numite *metode chemometrice*. Ele au ca obiectiv rezolvarea problemelor prin intermediul procesării informației chimice. De asemenea, rolul lor este de a optimiza și îmbunătăți colecția de date utile, să se întrebe asupra conținutului lor, validității și semnificației lor. Un obiectiv important al chemometriei îl constituie aplicarea tehnicilor specifice tehnologiei informației în optimizarea proceselor analitice și evidențierea direcțiilor fundamentale de evoluție a metodelor computaționale. M. Otto [1] afirmă că dezvoltarea chemometriei este strâns legată de utilizarea calculatorului în chimie.

Chemometria este disciplina care are ca principal obiectiv introducerea și dezvoltarea metodelor matematice moderne în proiectarea sau selectarea măsurărilor și experimentelor optime precum și procesarea informației chimice. Pentru prima dată noțiunea de chemometrie a fost introdusă de suedezul Svante Wold și americanul Bruce R. Kowalski în 1972. În chimia analitică, chemometria oferă metodele matematice și statistice pertinente în scopul obținerii celor mai

statistico-matematice, chemometria pune la îndemâna chimiștilor metodele inteligenței artificiale în prelucrarea datelor chimice cum ar fi: rețelele neuronale și programarea simbolică. În general, rețelele neuronale sunt folosite în recunoașterea de configurații. Pe de altă parte, chemometria are un rol important în dezvoltarea altor metode cum ar fi metodele cromatografice de separare, automatizarea proceselor analitice sau în corelarea structură-activitate a compușilor chimici.

I.2. Metode chemometrice

Obiectivele metodelor chemometrice se pot defini astfel :

- permit o mai bună prelucrare a rezultatelor obținute de analiști în scopul extragerii informației celei mai pertinente posibil;
- creșterea calității măsurătorilor chimice prin optimizarea condițiilor de colectare ale acestora.

În 1975 s-a creat Chemometrics Society iar în fiecare an are loc un congres intitulat "Chemometrics in Analytical Chemistry". Există, de asemenea două reviste "Journal of Chemometrics" (Wiley) și "Journal of Intelligent Laboratory Systems" (Elsevier) unde sunt prezentate rezultatele semnificative obținute în acest domeniu.

Se disting 4 domenii semnificative de investigație referitoare la posibilitățile de aplicare a metodelor chemometrice și anume:

- a) achiziționarea semnalelor la ieșirea din aparatele de măsură, stocarea lor și aranjarea lor într-o formă corespunzătoare;

b) transformarea acestor semnale, expresia lor sub forma unor rezultate prin etalonare și validarea lor în termeni de precizie;

c) interpretarea rezultatelor în vederea rezolvării problemelor puse analiștilor;

d) gestionarea informațiilor din laboratoarele de analiză.

Dezvoltarea chemometriei face parte integrantă din evoluția chimiei. Chemometria a propus și propune în continuare soluții noi bazate pe un fundament matematic la problemele puse chimiștilor și în principal, analiștilor [2-6].

Chemometria își are un loc bine definit în chimia analitică atât în specificitatea instrumentelor de măsură utilizate, natura probelor precum și în modul de funcționare al laboratorului. În ceea ce privește calculatorul într-adevăr acesta rămâne util cercetării și dezvoltării, iar un obiectiv permanent îl constituie integrarea aparatelor la noile tehnici chemometrice.

Tehnologia informatică a început să pătrundă în chimia analitică în jurul anilor 1960, în particular în tot ceea ce înseamnă instrumente de laborator. În consecință, majoritatea aparatelor de măsură disponibile astăzi conțin microprocesoare. Acestea vin să înlocuiască componentele electronice cablate ale aparatelor crescând astfel fiabilitatea instrumentelor și îmbunătățind în același timp funcțiile de filtrare ale semnalului, calculul statistic și reprezentarea grafică. Chimia analitică este bogată în informații deci este logic ca dezvoltarea sa să beneficieze de metodele informaticii. Problema adaptării metodelor informaticii la cerințele necesare analiștilor constituie unul dintre obiectivele principale

Bibliografie

1. M. Otto, *Chemometrics, Statistics and Computer Application in Analytical Chemistry*, J.Wiley-VCH, Weinheim, **1999**.
2. D. L. Massart, A. Dijkstra și L. Kaufman, *Evaluation and Optimization of Laboratory Methods and Analytical Procedures*, Elsevier, Amsterdam, **1978**.
3. G. Kateman și J. Pijpers, *Quality Control in Analytical Chemistry*, John Wiley, New York, **1981**.
4. M. A. Sharaf, D. L. Illman și B. R. Kowalski, *Chemometrics*, John Wiley, New York, **1986**.
5. D. L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte și L. Kaufman, *Chemometrics : a textbook*, John Wiley, New York, **1988**.
6. M. Feinberg, *L'Actualité Chimique*, 3, 191, **1991**.

II. ELEMENTE DE STATISTICĂ MATEMATICĂ UTILIZATE ÎN CHIMIA ANALITICĂ

II.1. Generalități

Statistica este o ramură a matematicii aplicate care cuprinde un grup de metode cu ajutorul cărora se pot obține datele cele mai probabile privind fenomenele de masă, dependente de o multitudine de factori [1-4].

În chimia analitică, statistica matematică pune la dispoziție atât metode de apreciere și analiză a stabilității sistemelor analitice cât și metode de ridicare a performanțelor de stabilitate ale acestora.

Sistemele analitice sunt sisteme informaționale, rezultatele funcționării lor fiind informațiile obținute asupra compoziției cantitative (concentrația unui constituent) și calitative (prezența de grupări funcționale) a materialului de cercetat.

În analiza chimică apar frecvent de rezolvat probleme de statistică aplicată, ca de exemplu :

- stabilirea legilor statistice de repartiție a datelor;
- realizarea de selecții și estimări de parametri;
- evaluarea intervalului de încredere pentru mărimile estimate;
- verificări de ipoteze.

Metodele statisticii matematice își găsesc aplicații în următoarele direcții:

- obținerea, prelucrarea și prezentarea rezultatelor determinărilor analitice,

- evaluarea caracteristicilor de performanță ale instrumentelor și metodelor de analiză;
- elaborarea de noi metode de analiză;
- optimizarea proceselor analitice.

Statistica matematică are utilizări specifice în domeniul analizei în urme. Problemele limitei de detecție și a limitei de determinare în analiza chimică sunt rezolvate cu ajutorul metodelor statistice. Teoria statistică a detecției semnalelor permite definirea riguroasă a limitei de detecție și totodată, pune la dispoziție metode de estimare corectă a acesteia.

În cazul rezultatelor unor măsurători, ca de exemplu : spectrele, cromatogramele, polarogramele de diverse tipuri, pentru prelucrarea lor se utilizează metode speciale ale statisticii matematice, și anume : netezirea semnalelor, derivarea semnalelor, autocorelația și intercorelația semnalelor, descompunerea în componente a semnalelor, filtrarea semnalelor. De asemenea, pentru calibrarea instrumentelor analitice sunt necesare metode matematice de estimare a funcțiilor de calibrare liniare sau neliniare, de una sau mai multe variabile.

În vederea selecționării unei anumite metode de analiză din mai multe posibile, se utilizează testele statistice privind compararea caracteristicilor de performanță ale metodelor.

Metodele statistico-matematice sunt prezente atât în perfecționarea unor metode de analiză existente, cât și la elaborarea unor noi metode de analiză.

II.2. Distribuții de probabilitate

În general distribuțiile de probabilitate sunt utilizate în prelucrarea rezultatelor experimentale în scopul obținerii unei cantități maxime de informație. Acest aspect se întâlnește în estimarea parametrilor, evaluarea intervalelor de încredere în legătură cu parametrii estimați și în verificarea ipotezelor în legătură cu parametrii estimați.

II.2.1. Distribuția variabilelor aleatoare

Repetând un experiment analitic de determinare, rezultatele măsurătorilor alcătuiesc o mulțime ce este descrisă de către o funcție de distribuție $F(x)$. Această funcție de distribuție măsoară probabilitatea ca variabila aleatoare x să fie mai mică decât o anumită valoare. Probabilitatea ca variabila x să fie mai mică decât un nivel x_1 este:

$$P(x < x_1) = F(x_1) \quad (2.1)$$

Funcția de distribuție are o valoare cuprinsă în domeniul $[0,1]$, adică:

$$0 \leq F(x) \leq 1 \quad (2.2)$$

Prin derivarea funcției de distribuție $F(x)$ se obține funcția densitate de probabilitate corespunzătoare $p(x)$:

$$p(x) = dF(x)/dx = F'(x) \quad (2.3)$$

Probabilitatea ca variabila aleatoare să ia o valoare situată în domeniul $[x_1, x_2]$ este dată de integrala:

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} p(x) dx \quad (2.4)$$

Caracteristicile cele mai importante ale unei funcții de repartiție sunt valoarea medie m (momentul de ordinul unu) și dispersia σ^2 (momentul centrat de ordinul doi):

$$m = \int_{-\infty}^{+\infty} xp(x)dx = M(x) \quad (2.5)$$

$$\sigma^2 = \int_{-\infty}^{+\infty} |(x - M(x))|^2 p(x)dx \quad (2.6)$$

II.2.2. Funcția de distribuție normală a variabilelor aleatoare

Funcția de distribuție normală (Gauss) poate fi considerată model pentru cercetarea probabilistică a procesului de măsură. Funcția de distribuție normală, fundamentală în statistica matematică și în teoria erorilor, are densitatea de probabilitate conform expresiei:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad -\infty < x < +\infty \quad (2.7)$$

Funcția de distribuție normală, notată prin $N(m,\sigma)$, prezintă următoarele proprietăți importante:

- admite un maxim pentru $x = m$;
- este simetrică în raport cu m ;
- modificarea parametrului m determină deplasarea curbei normale de-a lungul abscisei, fără a-i afecta forma;
- modificarea lui σ duce la lățirea sau îngustarea curbei, fără ca media (m) să fie afectată;
- curba are două puncte de inflexiune, corespunzătoare valorilor $x = \pm \sigma$.

Există două teoreme care prezintă o mare importanță pentru aplicațiile distribuției normale:

1. Media a n valori de selecție aparținând unei funcții de distribuție normală $N(m, \sigma)$ are la rândul ei o funcție de distribuție normală

$$N\left(m, \frac{\sigma}{\sqrt{n}}\right).$$

2. Media a n valori dintr-o selecție, aparținând unei distribuții nenormale, pentru n număr mare are o funcție de densitate de probabilitate normală sau apropiată de aceasta.

Dacă se consideră o variabilă aleatoare x , caracterizată de funcția de distribuție $N(m, \sigma)$, atunci variabila aleatoare z , numită și abaterea normată :

$$z = \frac{x - m}{\sigma} \quad (2.8)$$

are o funcție de distribuție $N(0,1)$, cu o funcție de densitate de probabilitate normală standard sub forma:

$$p(z) = \frac{1}{\sqrt{2\pi}} \exp \frac{-z^2}{2}, \quad -\infty < z < +\infty, \text{ cu } m = 0 \text{ și}$$

$$\sigma^2 = 1 \quad (2.9)$$

Funcția de probabilitate a distribuției normale standard are forma:

$$P(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz \quad (2.10)$$

valorile acestora se găsesc în tabel.

Simetria funcției de densitate de probabilitate normală față de $z = 0$ impune egalitatea:

$$p(-z) = 1-p(z) \quad (2.11)$$

Acceptarea distribuției normale ca model în studiul fluctuației determinărilor, a măsurătorilor analitice în general, este justificată atât din punct de vedere practic cât și teoretic. Eroarea întâmplătoare a unei determinări este rezultatul tuturor erorilor întâmplătoare elementare care intervin în procesul de analiză. În condițiile unui sistem analitic bine conceput și cu o execuție corectă a analizelor, rezultatele experimentale trebuie să aibă o distribuție normală sau apropiată de aceasta. Abaterile pronunțate ale distribuției rezultatelor analitice de la o distribuție normală pune în evidență nereguli ale sistemului analitic studiat.

II.2.3. Funcția de distribuție χ^2

Dacă x_1, x_2, \dots, x_n sunt variabile aleatoare independente cu o funcție de distribuție normală $N(0,1)$, variabila aleatoare χ^2 prezentată mai jos:

$$\chi^2 = \sum_{i=1}^n x_i^2, \quad ; 0 < \chi^2 < +\infty \quad (2.12)$$

va avea o distribuție χ^2 cu ν grade de libertate și o funcție de densitate de probabilitate:

$$p(\chi^2) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}}} (\chi^2)^{\left(\frac{\nu}{2}\right)-1} e^{-\frac{\chi^2}{2}}, \quad \nu > 0; 0 \leq \chi^2 \leq \infty \quad (2.13)$$

Media acestei distribuții este egală cu ν iar dispersia egală cu 2ν

Funcția de probabilitate a distribuției χ^2 este:

$$P(\chi^2) = \int_0^{\chi^2} \frac{1}{\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}}} (\chi^2)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}} d(\chi^2) \quad (2.14)$$

unde valorile pentru variabila χ^2 precum și valorile pentru $P(\chi^2)$ sunt tabelate.

Se poate demonstra că, efectuând o selecție x_1, x_2, \dots, x_n asupra unei variabile aleatoare cu o funcție de repartiție $N(m, \sigma)$, dispersia de selecție corespunzătoare

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \text{ aparține mulțimii de valori dată de relația:}$$

$$s^2 \in \frac{\chi^2 \sigma^2}{n-1} \quad (2.15)$$

unde $\chi^2 =$ valorile funcției dedistribuție $F(\chi^2)$ corespunzătoare la $\nu = n-1$ grade de libertate.

Deci se poate deduce funcția de distribuție pentru dispersia de selecție cu ajutorul funcțiilor dedistribuție $F(\chi^2)$. Această precizare va fi utilizată la evaluarea unui interval de încredere pentru dispersie. Relația (15) permite de asemenea ca, prin luarea în considerare a distribuției variabilei F să permită elaborarea unui test de semnificație pentru compararea a două dispersii de selecție.

II.2.4. Funcția de distribuție t (Student)

Variabila aleatoare t a funcției de distribuție Student reprezintă raportul a două variabile aleatoare, și anume z , variabila normală

standard, cu o funcție de distribuție normală $N(0,1)$ și variabila aleatoare

$\sqrt{\frac{\chi^2}{\nu}}$, independentă de prima:

$$t = \frac{z}{\sqrt{\frac{\chi^2}{\nu}}} = \frac{z}{\frac{s}{\sigma}} = \frac{x-m}{\sigma} = \frac{1}{\frac{s}{\sigma}} = \frac{x-m}{s} \quad (2.16)$$

Funcția densitate de probabilitate pentru variabila t are expresia următoare:

$$p(t) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad -\infty < t < +\infty \quad (2.17)$$

unde ν reprezintă numărul gradelor de libertate asociat cu s^2 . Funcția de probabilitate corespunzătoare funcției densitate de probabilitate pentru variabila t are expresia:

$$P(t) = \int_{-\infty}^t p(t) dt \quad (2.18)$$

Când numărul gradelor de libertate $\nu \rightarrow \infty$, distribuția t tinde către o distribuție normală standard (valorile variabilei t se află tabelate).

În general funcția de distribuție t se folosește pentru evaluarea intervalului de încredere pentru media unei distribuții normale și de asemenea pentru verificarea ipotezelor statistice pe care le implică analiza chimică pentru selecții mici ($n < 30$).

II.2.5. Funcția de distribuție F

Fiind date două variabile aleatoare χ_1^2 și χ_2^2 independente, cu ν_1 și ν_2 grade de libertate, variabila

$$F = \frac{\nu_2 \chi_1^2}{\nu_1 \chi_2^2} \quad (2.19)$$

are funcția de densitate de probabilitate dată de expresia:

$$p(F) = \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} F^{\left(\frac{\nu_1}{2}-1\right)} \left(1 + \frac{\nu_1}{\nu_2} F\right)^{-\frac{\nu_1 + \nu_2}{2}}, \quad 0 < F < \infty$$

(2.20)

Funcția de distribuție F, ale cărei valori sunt tabelate, se utilizează la verificarea ipotezelor de semnificație a două dispersii de selecție. Dacă se consideră selecțiile $x_{11}, x_{12}, \dots, x_{1n}$ și $x_{21}, x_{22}, \dots, x_{2n}$ efectuate asupra aceleiași distribuții normale sau a două distribuții normale cu aceeași dispersie, variabila:

$$F = \frac{s_1^2}{s_2^2} = \frac{\left(\frac{\chi_1^2}{\nu_1}\right)}{\left(\frac{\chi_2^2}{\nu_2}\right)} \quad (2.21)$$

are o distribuție F cu $\nu_1 = n_1 - 1$ și $\nu_2 = n_2 - 1$ grade de libertate, iar $s_i^2 =$

$$\frac{\chi^2 \sigma^2}{\nu_i}$$

II.2.6. Distribuția binomială

Dacă un eveniment are o probabilitate de apariție p , probabilitatea ca el să apară de x ori în urma a n experimente independente, este evaluată de funcția:

$$p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad (2.22)$$

unde $x = 0, 1, 2, \dots, n$

Pentru $0,3 \leq p \leq 0,70$ și $n \geq 30$, distribuția binomială este exprimată bine de către o distribuție normală cu $m = np$ și dispersia $\sigma^2 = np(1-p)$.

II.3. Selecție și estimare

Noțiunea de selecție este strâns legată de cea de estimare, ambele prezentând o mare importanță pentru analiza chimică. Un exemplu tipic de selecție analitică îl constituie mulțimea rezultatelor determinărilor analitice x_1, x_2, \dots, x_n efectuate pe un material, în scopul evaluării conținutului unui component. În absența erorilor sistematice,

media acestei mulțimi de selecție ($\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$) este cea mai bună estimare

a conținutului adevărat al materialului. Un alt exemplu de estimare îl constituie estimarea preciziei (împrăștierii) rezultatelor, cu ajutorul abaterii standard de selecție:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.23)$$

În procesele analitice sunt: numeroase exemple de estimare, cum ar fi : estimarea funcțiilor de calibrare, estimarea parametrilor distribuțiilor rezultatelor și erorilor de analiză, estimarea performanțelor proceselor analitice, estimarea funcției de răspuns a procesului analitic etc.

Generalizat, problema estimării parametrilor poate fi prezentată astfel: se consideră cazul în care între variabilele de interes $X = |x_1, x_2, \dots, x_n|$ și răspunsul Y există o dependență liniară, de tipul:

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m + e \quad (2.24)$$

Pentru a estima parametri $B = |b_1, b_2, \dots, b_m|$ se utilizează o mulțime de n perechi de date ($X(i) = |x_1(i), x_2(i), \dots, x_n(i)|, y(i)$), generate în acord cu relația:

$$Y(i) = BX(i) + e(i), \quad i = 1, 2, \dots, n \quad (2.25)$$

unde $e(i)$ este zgomotul cu o valoare de așteptare egală cu zero.

Se notează cu \hat{B} valorile de estimare pentru parametrii iar regula de estimare prin:

$$\hat{B} = G(x; y) \quad (2.26)$$

Cei mai importanți estimatori sunt: estimatori Bayes, estimatorul verosimilității maxime, estimatorul Markov și estimatorul celor mai mici pătrate [xx].

II.4. Intervale de încredere

Intervalul statistic de încredere a fost utilizat pentru prima dată de către Laplace, cu scopul de a mări gradul de semnificație al estimării parametrilor distribuției binomiale. În continuare Neymann este cel care

a dezvoltat teoria și terminologia intervalului de încredere. Noțiunea de estimare impune noțiunea de intervalului de încredere; deci problema intervalului de încredere apare ori de câte ori este pusă problema de estimare.

Dacă se consideră o variabilă aleatoare cu o funcție de densitate de probabilitate ipotetică $p(x,b)$, parametrul b va fi estimat în baza unei selecții: $x(1), x(2), \dots, x(n)$ prin intermediul unui estimator:

$$\hat{b} = G[x(1), x(2), \dots, x(n)] \quad (2.27)$$

Valoarea de estimație \hat{b} este la rândul ei o variabilă aleatoare a cărei funcție de densitate de probabilitate o notăm cu $p(\hat{b}, b)$. Probabilitatea P ca variabila \hat{b} să se găsească într-un interval $b_1 < \hat{b} < b_2$ este dată de integrala:

$$P = P(b_1 < \hat{b} < b_2) = \int_{b_1}^{b_2} p(\hat{b}, b) d\hat{b} \quad (2.28)$$

Parametrul P este numit și coeficient de încredere. De obicei, în statistică problema se pune invers, adică: dacă se dă o probabilitate P , să se stabilească un interval de încredere pentru parametrul existent.

II.4.1. Intervalul de încredere pentru media unei distribuții

Există două situații distincte:

1. *Dispersia distribuției normale este cunoscută.*

În acest caz, pentru estimarea mediei \bar{x} se realizează o selecție $x(1), x(2), \dots, x(n)$ referitoare la o variabilă x , distribuită $N(m, \sigma)$. La

rândul ei, media va avea o distribuție $N(m, s/\sqrt{n})$. Variabila normală standard $z = (\bar{x} - m)\sqrt{n}/s$ are o distribuție normală $N(0,1)$. Pentru o probabilitate de încredere P , mediei m a distribuției îi corespunde un interval de încredere simetric, conform expresiei: $\bar{x} \pm z_{(1-\alpha/2)} s/\sqrt{n}$, în care $\alpha = 1-P$ iar $z_{(1-\alpha/2)}$ se va deduce din condiția:

$$P[z_{(1-\alpha/2)}] = (1+p)/2 \quad (2. 29)$$

Exemplu. Să se evalueze un interval de încredere simetric pentru medie, ce corespunde unei probabilități de încredere $P \geq 0,95$.

Deoarece $\bar{x} = 0,41$, $\sigma^2 = 0,000625$, $n = 9$ determinări analitice iar $z_{0,975} = 1,96$ (din anexă), intervalul de încredere pentru medie este:

$$0,41 - 0,016 < m < 0,41 + 0,016$$

2. Dispersia distribuției normale este necunoscută.

Acest caz al evaluării intervalului de încredere pentru media unei distribuții normale cu dispersie necunoscută este foarte des utilizat în domeniul chimiei analitice.

Pentru evaluarea intervalului de încredere simetric, corespunzător unei probabilități de încredere P , în cazul unei distribuții $N(m, \sigma)$ la care nu se cunoaște nici media nici dispersia, se ia în considerare variabila $t = (\bar{x} - m)/(s/\sqrt{n})$ cu o distribuție t cu $\nu = n-1$ grade de libertate. Expresia intervalului de încredere este:

$$\bar{x} \pm t_{(1-\alpha/2; \nu)} s/\sqrt{n} \quad (2. 30),$$

unde $\alpha = 1-P$ iar $t_{(1-\alpha/2; \nu)}$ se deduce din ecuația: $P[t_{(1-\alpha/2; \nu)}] = (1+P)/2$.

În același mod se poate calcula intervalul de încredere pentru diferența a două medii (m_1, m_2) corespunzătoare variabilelor aleatoare x și y cu distribuțiile $N(m_1, \sigma)$ și $N(m_2, \sigma)$.

Ca aplicație în chimia analitică, acest caz se utilizează în special la evaluarea intervalului de încredere pentru diferența dintre mediile rezultatelor obținute în două laboratoare diferite, prin aceeași metodă.

II.4.2. Intervalul de încredere pentru dispersia de selecție

Pentru estimarea dispersiei unei variabile aleatoare cu o distribuție $N(m, \sigma)$ se consideră o selecție $x(1), x(2), \dots, x(n)$, din care se evaluează dispersia de selecție s^2 . Variabila $\chi^2 = s^2(n-1)/\sigma^2$ are o distribuție cu $\nu = n-1$ grade de libertate. Intervalul de încredere pentru dispersie, pentru o probabilitate P are următoarea expresie:

$$\frac{(n-1)s^2}{c_{(1-\alpha/2; \nu)}^2} < t^2 < \frac{(n-1)s^2}{c_{(1-\alpha/2; \nu)}^2} \quad (2.31)$$

Dacă se utilizează volume de selecție mari ($n \geq 30$), abaterea standard de selecție s are o distribuție $N(s, s/\sqrt{2n})$. Deci, se poate admite că variabila:

$$z = \frac{s - s}{s/\sqrt{2n}} \quad (2.32)$$

are o distribuție $N(0,1)$.

În concluzie, pentru o probabilitate de încredere corespunzătoare P , se poate evalua un interval de încredere pentru abaterea standard a unei distribuții normale, pentru volume de selecție mari, cu expresia:

$$s \pm z_{(1-\alpha/2; \nu)} s/\sqrt{2n} \quad (2.33)$$

II.4.3. Intervalul de încredere pentru parametrii distribuției binomiale

Se consideră că prin repetarea unor determinări de n ori, fenomenul care ne interesează s-a manifestat de m ori. Probabilitatea p de manifestare a evenimentului este estimată de raportul $\hat{p} = m/n$. Ținând cont de distribuția binomială, variabila:

$$z = (\hat{p} - p) / [\hat{p}(1 - \hat{p})/n]^{1/2} \quad (2.34)$$

are aproximativ distribuția $N(0,1)$ și deci intervalul de încredere pentru parametrul p , corespunzător unei probabilități de încredere P , are următoarea expresie:

$$\hat{p} \pm z_{(1-\alpha/2;v)} [\hat{p}(1 - \hat{p})/n]^{1/2} \quad (2.35)$$

II.5. Analiza de dispersie

Analiza de dispersie, introdusă în statistica matematică de către Fisher, are aplicații în evaluarea și optimizarea proceselor analitice cu aplicații în chimie, biologie, biotehnologie.

Analiza de dispersie este o tehnică statistică utilizată la evaluarea unor factori care distorsionează veridicitatea rezultatelor. Din punct de vedere al metodologiei, se pot menționa analizele de dispersie monofactorială, bifactorială și trifactorială.

Studiile realizate în literatura de specialitate [5,6] în domeniul analizei de dispersie prezintă exemple de utilizare a analizei de dispersie monofactorială și bifactorială privind descompunerea dispersiei rezultatelor de analiză în două componente (între zile și în cursul

aceleiași zile de lucru) sau trei componente. Analiza de dispersie este utilizată de asemeni în studiile efectuate între laboratoare.

În continuare se va prezenta un model de utilizare a analizei de dispersie trifactoriale pentru evaluarea unei metode de analiză conform unui plan experimental. Pentru testarea metodei de analiză se iau în considerare trei factori: laboratorul, compoziția materialului de analizat și timpul. Există deci r laboratoare, p materiale cu un conținut constant în componentul de determinat și diferit de restul componentilor și u zile de funcționare, deci timp; la fiecare moment, fiecare material este analizat de n ori. În cazul acestui experiment dispersia totală este descompusă în componente pe factori: laboratorul, compoziția materialului și timpul.

În analiza de dispersie trifactorială observațiile aparțin unei rețele tridimensionale formată din r , p , u celule, în fiecare celulă având n observații. Notațiile r , p și u reprezintă numărul de nivele la care se situează cei trei factori studiați și n , numărul de observații la un anumit nivel.

BIBLIOGRAFIE

1. C. Liteanu și I. Rîcă, *Teoria și metodologia statistică a analizei urmelor*, Editura Scrisul Românesc, Craiova, 1979.
2. L. Mutihac și V. David, *Chemometrie*, Editura Universității din București, 1997.
3. D.L. Massart, B.G.M. Vandeginste, S.N. Deming și Y. Michotte și L. Kaufman, *Chemometrics : a textbook*, Ed. Elsevier, Amsterdam, 1988.
4. D. Ceașescu, *Utilizarea statisticii matematice în chimia analitică*, Ed. Tehnică, București, 1982.
5. J.H. Riddich, R. Flora și Q.L. van Meter, *Clin. Chem.*, 18, 250, 1972.
6. G. Wernimont, *Mat. Res. Stds.*, 9-10, 8, 1969.

III. CONCEPTE UTILIZATE ÎN EVALUAREA METODELOR ANALITICE

În chimia analitică pentru evaluarea metodelor se folosesc o serie de criterii de performanță, care sunt enumerate în tabelul 3.1.

Tabelul 3.1. Criterii de performanță pentru metodele analitice.

Criteriu	Definiție
Acuratețea	Abilitatea de a obține date corecte
Precizia	Reproductibilitatea măsurărilor repetate
Sensibilitatea	Mărimea semnalului analitic generat în urma unor modificări cu o unitate a concentrației analitului
Limita de detecție	Cantitate minimă ce poate fi detectată pentru un anumit nivel de încredere
Domeniu de linearitate	Domeniul în care semnalul analitic este proporțional cu concentrația de analit din probă
Selectivitatea	Abilitatea de a face distincție între analitul de interes și alți compuși prezenți în probă
Viteza de analiză	Rapiditatea măsurărilor combinată cu timpul necesar pentru pregătirea probei
Robustețea	Durabilitatea instrumentului analitic sau, în cazul unei metode, se referă la ușurința cu care se obțin rezultate corecte
Portabilitatea	Posibilitatea de a utiliza instrumentul analitic pentru analize directe, în teren. În cazul unei metode se referă la posibilitatea acesteia de a fi utilizată cu aceleași rezultate în laboratoare diferite.
Numărul de probe	Nr. de probe care pot fi analizate într-o perioadă determinată de timp
Cost	Investiția inițială, costul reactivilor și al consumabililor, costul pregătirii personalului calificat
Ecologic – (“Greenness”)	Metoda nu conduce la obținerea și/sau utilizarea unei cantități importante de compuși toxici

Orice chimist analist care realizează un experiment și dorește să conceapă și să optimizeze o metodă de analiză trebuie să investigheze factorii care influențează semnalul analitic și să optimizeze performanțele metodei analitice. Dintre factorii tipici în chimia analitică putem aminti: pH-ul, tăria ionică, solventul, concentrația reactivilor, compoziția amestecului de reacție, etc.

În principal, pentru optimizarea unui experiment se disting două direcții. Prima, are în vedere selectarea și testarea celor mai importanți factori, și optimizarea acestora. Aici intervine experiența cercetătorului iar succesul alegerii va depinde de nivelul cunoștințelor acestuia în domeniu. De asemenea abordarea nesistematică a problemei nu garantează descoperirea soluției optime. Cea de a doua direcție are în vedere investigarea sistematică a factorilor și optimizarea caracteristicilor de performanță [1].

III.1. Tipuri de erori în chimia analitică

Rezultatele numerice obținute în urma determinărilor din chimia analitică sunt însoțite de erori iar estimarea mărimi acestor erori este importantă pentru validarea rezultatelor. Astfel scopul determinărilor în chimia analitică îl constituie estimarea valorii adevărate.

Eroarea de determinare analitică este dată de diferența dintre rezultatul determinării și valoarea adevărată a determinării. Dacă se notează cu X_0 valoarea adevărată a conținutului componentului de determinat, iar cu X rezultatul analizei eroarea de determinare se poate calcula, astfel:

$$e = X - X_0 \quad (3.1.)$$

iar eroarea relativă de determinare are următoarea expresie:

$$e = \frac{(X - X_0)}{X_0} \quad (3.2.)$$

Eroarea relativă de determinare poate fi exprimată și procentual, în raport cu valoarea adevărată a conținutului:

$$e \% = \frac{(X - X_0)}{X_0} \cdot 100 \quad (3.3.)$$

În general, tipurile de erori de determinare pot fi clasificate în:

- *erori întâmplătoare*, și
- *erori sistematice*.

Dacă același experiment este repetat de mai multe ori, fiecare determinare individuală va varia în jurul valorii medii \bar{X} . Erorile întâmplătoare se pot evidenția atunci când la repetarea unui experiment de determinare se obține de fiecare dată alt rezultat, datorită unor factori necunoscuți sau ce nu pot fi controlați.

Eroarea întâmplătoare este dată de diferența dintre rezultatele obținute și valoarea medie a acestora la efectuarea repetată a aceluiași experiment [2].

Spre deosebire de erorile întâmplătoare, **erorile sistematice** sunt acele erori care prezintă un caracter constant, într-un singură direcție, distorsionând rezultatelor determinărilor cu o cantitate constantă față de valoarea adevărată. De asemenea, la clasificarea erorilor sistematice, este necesar să fie luat în considerare parametrul timp. Astfel, deosebim erori sistematice variabile în timp și erori sistematice constante în timp.

Liteanu și Rîcă [3] au introdus, în clasificare, **erorile grosolane** care sunt definite ca fiind erorile întâmplătoare, intermitente care provoacă o abatere mare a rezultatelor de la valoarea adevărată. Ele se deosebesc semnificativ de restul rezultatelor și pot fi identificate cu ușurință.

III.2. Precizia și acuratețea în chimia analitică

Evaluarea cantitativă a erorii întâmplătoare prezentă într-un set de date este dată de **precizie sau reproductibilitate**. **Acuratețea**, care este o măsură a deviației față de valoarea adevărată, este afectată de eroarea sistematică. Acuratețea este definită ca deviația valorii medii a rezultatelor față de valoarea adevărată.

Precizia este dată de împrăștierea rezultatelor unei mulțimi de determinări în jurul mediei lor și este evaluată de abaterea standard. Dacă X_1, X_2, \dots, X_n este mulțimea rezultatelor pentru aceeași probă, n este numărul rezultatelor și \bar{X} este media rezultatelor, **abaterea standard** este:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \quad (3.4)$$

Atunci când sunt disponibile un număr mic de rezultate, abaterea standard calculată poate fi subestimată, deoarece valoarea medie este considerată valoare adevărată, iar media este calculată din același volum mic de date. În acest caz se folosește **abaterea standard modificată**:

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} \quad (3.5)$$

Aceasta este abaterea standard folosită în practica statistică pentru estimarea abaterii standard a întregii populații [4].

Dacă numărul de rezultate disponibile este mai mare, valoarea abaterii standard modificate, s , se apropie de valoarea abaterii standard, σ . Când numărul de rezultate este mai mare de 20 se poate utiliza ecuația 3.4.

De asemenea, poate fi determinat **coeficientul de variație (CV)** sau **abaterea relativă standard (RSD)**.

$$RSD = \frac{s}{\bar{X}} \quad \text{sau} \quad \%RSD = \frac{s}{\bar{X}} \cdot 100 \quad (3.6.)$$

Evaluarea preciziei se face cu ajutorul abaterii standard dar cu specificarea obligatorie a condițiilor și modului de obținere a mulțimii de rezultate [5-7].

Diferența dintre precizie și acuratețe poate fi ilustrată de către următorul exemplu: pentru analiza, prin spectrometrie de absorbție atomică, a unei probe ce conține 10 $\mu\text{g/kg}$ plumb, s-au făcut cinci determinări repetate, în patru laboratoare diferite, iar rezultatele sunt prezentate în tabelul 3.2. și figura 3.1.

Rezultatele obținute în laboratorul A prezintă o variație mică, dar nu sunt apropiate de valoarea adevărată. Astfel, seria de determinări realizate în acest laborator prezintă o precizie mare și acuratețe mică. Acest fapt semnifică că în laboratorul A, rezultatele sunt afectate de o sursă neidentificabilă de erori și deci se impune examinarea atentă a

metodelor de calibrare, a standardelor folosite. Rezultatele obținute de laboratorul D sunt caracterizate printr-o variație mică și o acuratețe bună.

Tabelul 3.2. Rezultatelor obținute, în patru laboratoare diferite, la analiza unei probe ce conține 10 $\mu\text{g}/\text{kg}$ plumb.

Lab	Analiza					Media	s
	1	2	3	4	5		
A	8,99	8,67	9,03	8,92	8,77	8,87	0,152
B	9,80	7,99	8,01	7,55	10,07	8,68	1,16
C	8,43	9,88	11,21	10,77	9,66	9,99	1,08
D	9,87	9,79	10,04	10,12	9,99	9,96	0,132

Laboratoarele B și C au furnizat rezultate ce se caracterizează printr-o precizie mică; dar acuratețea este mai bună pentru un rezultatele dintr-un laborator de cât în celălalt. Obținerea unei bune acurateți cu o precizie mică, pentru un număr mic de analize repetate pentru aceeași probă, reprezintă de obicei un accident.

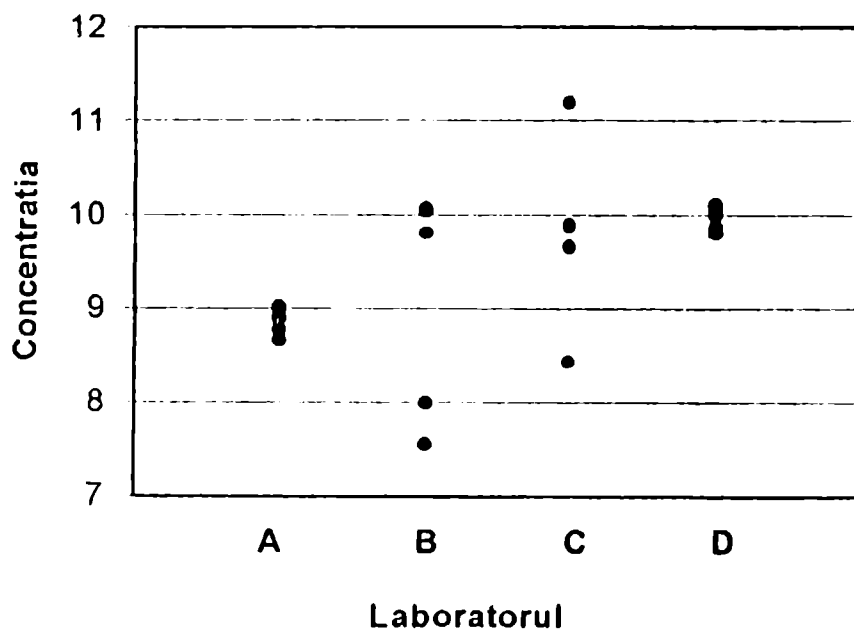


Figura 3.1. Seturi de cinci determinări pentru aceeași probă, obținute de patru laboratoare diferite. Valoarea adevărată pentru concentrația probei este 10 $\mu\text{g}/\text{kg}$ plumb.

III.3. Sensibilitatea și limita de detecție

Sensibilitatea, S, a unei metode analitice este definită ca fiind derivata funcției de răspuns $y = f(x)$ în raport cu concentrația sau cantitatea, x a analitului de determinat.

$$S = \frac{d y}{d x} \quad (3.7)$$

În cazul unei funcții de răspuns liniare,

$$y = a + b x \quad (3.8)$$

sensibilitatea este egală cu panta funcției:

$$S = b = \frac{\Delta y}{\Delta x} \quad (3.9)$$

iar pentru o funcție de răspuns neliniară, sensibilitatea se modifică cu valoarea conținutului de determinat. Pentru concentrații mici de analit, precizia determinărilor scade, iar în acest caz, sensibilitatea poate fi considerată o funcție de precizie. Se definește astfel, **sensibilitatea analitică** pentru o anumită concentrație de analit, ca fiind raportul dintre sensibilitate, b , (panta dreptei în cazul unei curbe de calibrare liniare) și abaterea standard a semnalului, σ .

$$S_s = \frac{b}{\sigma} \quad (3.10)$$

În figura 3.2. sunt prezentate curbele de calibrare obținute în cazul determinării fierului utilizând spectrometria de absorbție moleculară în vizibil, cu următorii reactivi: 2-piridin aldoximă, *o*-fenantrolină și acidul 2,6-piridin carboxilic.

Din calculul sensibilității se poate observa că determinarea fierului cu 2-piridin aldoximă se realizează cu o sensibilitate mult mai mare ($dA/dc = 0,20$) decât în cazul în care se folosesc drept reactivi *o*-fenantrolină ($dA/dc = 0,14$) sau acidul 2,6-piridin carboxilic ($dA/dc = 0,025$). Determinările au fost realizate în cuve cu grosimea stratului de soluție de 1 cm.

Sensibilitatea unei metode analitice se poate modifica adeseori. Astfel, dacă semnalului analitic obținut este amplificat, de exemplu de la milivolți la volți, conform definiției sensibilității, rezultă o creștere a acesteia de 1000 de ori. În același mod poate fi crescută sensibilitatea determinărilor fotometrice prin creșterea lungimii drumului optic.

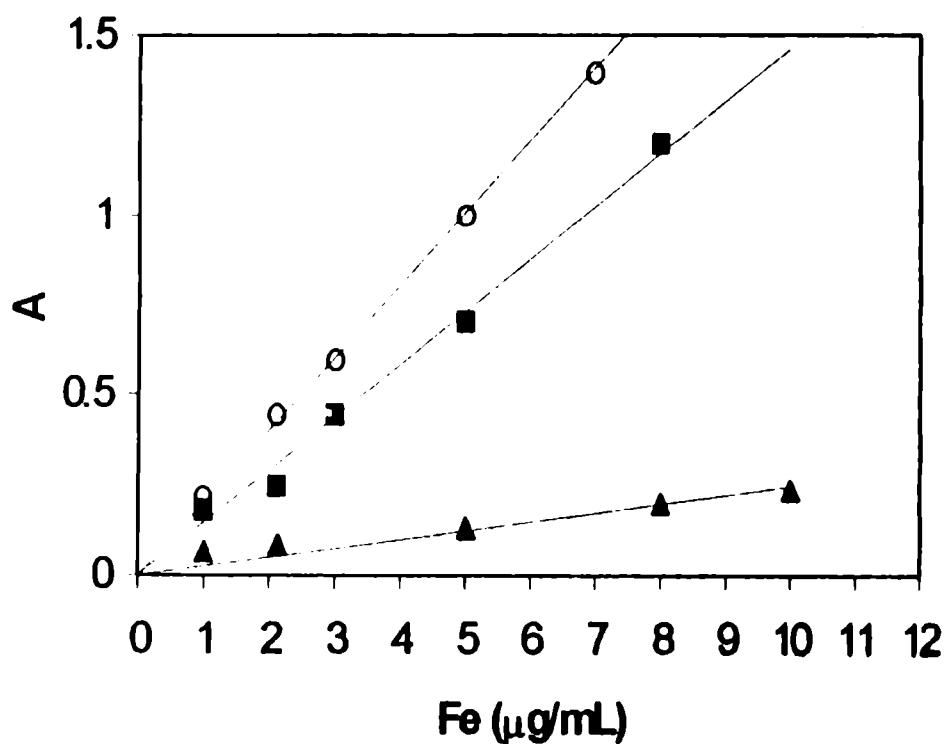


Figura 3.2. Curbele de calibrare obținute la determinarea fierului cu 2-piridină aldoximă (□), *o*-fenantrolină (□) și acidul 2,6-piridin carboxilic (▲).

Sensibilitatea unei determinări analitice este constantă pentru un anumit domeniu de concentrații care trebuie specificat. De asemenea, determinarea sensibilității este influențată de modelul de liniarizare ales de către chimistul analist. Limita inferioară a domeniului de concentrații, în care sensibilitatea poate fi considerată constantă, este dat de limita de detecție, iar limita superioară corespunde concentrației de analit pentru care sensibilitatea se modifică. În general o astfel de modificare se produce gradat, iar limita superioară poate fi determinată numai în cazul în care se precizează semnificația curbei de calibrare liniară. O definiție, general acceptată, precizează că limita superioară a domeniului de linearitate, între semnalul analitic și concentrația analitului de determinat, corespunde concentrației de analit pentru care valoarea semnalului analitic diferă cu o valoare precizată (în general 3%) de valoarea semnalului analitic rezultată din sensibilitate [8].

Conceptul de domeniu de linearitate este ilustrat în figura 3.3. În general domeniul de linearitate se exprimă prin număr de decade cuprinse între limita inferioară și cea superioară.

Una dintre cele mai importante preocupări ale chimiștilor analiști se referă la detectarea și determinarea analiților în volume mici de probă (microanaliza), determinarea concentrațiilor sau maselor foarte mici de analiți în volume mari de probă (analiza de urme) sau chiar determinarea unor concentrații mici de analiți prezenți în cantități mici de probă.

Conform IUPAC [9], **limita de detecție** “este dată de concentrația sau cantitatea de analit care conduce la cel mai mic semnal analitic, x , care poate fi măsurat cu o anumită *incertitudine* pentru o anumită procedură analitică”. Morrison [10] definește limita de detecție

astfel: "concentrația cea mai mică de analit care poate fi detectată cu *certitudine*, într-un proces analitic". Pentru a putea defini termenii de "*amunită incertitudine*" și "*deteție cu certitudine*" trebuie să se țină seama de erorile întâmplătoare asociate metodelor analitice. Aceste două abordări deși definesc noțiunea de limită de dectție, nu sunt complete atunci când ne referim la **limita de dectție** ca criteriu de performanță pentru metodele analitice.

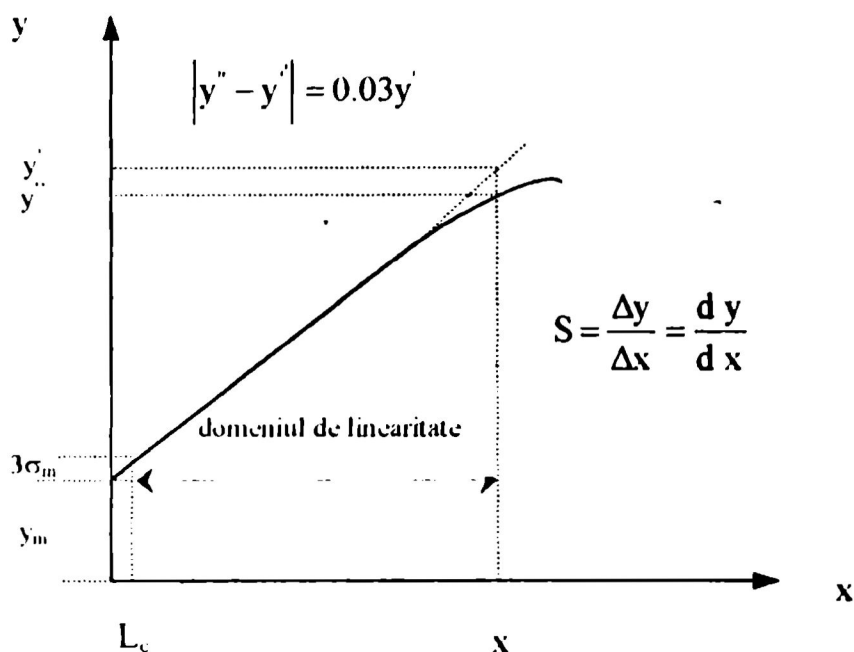


Figura 3.3. Ilustrarea conceptului de domeniu de linearitate; L_c -limită de decizie, y_m -semnalul martorului; σ_m -adaterea standard pentru martor.

Chimistul analist trebuie să accepte faptul că erorile întâmplătoare nu pot fi evitate și în consecință dectția (sau determinare) unui analit este limitată. De obicei, semnalul analitic măsurat, este determinat prin diferența a două măsurători, de ex. măsurarea unei probe necunoscute, y , și măsurarea probei martor, y_m . Această problemă poate fi formulată în

două moduri: dacă y diferă semnificativ față de y_m , sau dacă $y - y_m$ diferă semnificativ față de zero. Răspunsul este dat de statistica matematică. Distribuția normală a semnalului analitic al probei este prezentată în figura 3.4. Se folosesc următoarele notații pentru proba martor: σ_m - abaterea standard și y_m - valoarea adevărată a semnalului analitic.

Probabilitatea de a măsura semnale analitice $y_m > L_c$ este

$$\alpha = \int_{L_c}^{\infty} P(y_m) dy_m \quad (3.11)$$

unde $P(y_m)$ reprezintă funcția de distribuție corespunzătoare semnalului analitic y_m . Dacă un semnal analitic mai mare decât limita de decizie, L_c , este interpretat prin “prezență analitului de interes”, atunci o fracție α din măsurătoarea realizată pentru martor va fi interpretată greșit.

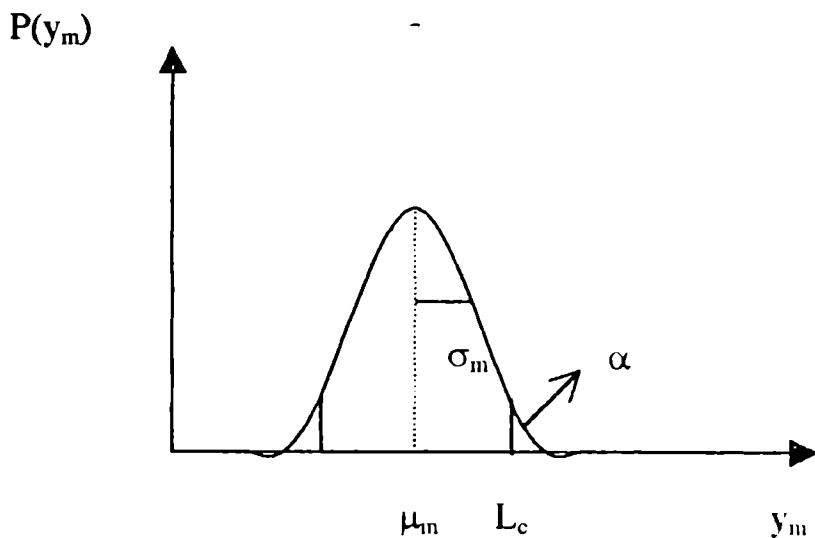


Figura 3.4. Distribuția normală a y_m .

Semnalul analitic corespunzător **limitei de decizie** poate fi exprimat astfel:

$$L_c = y_m + k_c \sigma_m \quad (3.12)$$

Pentru o distribuție normală a semnalului analitic al probei martor, o valoare $k_c = 3$, conduce la o probabilitate $1 - \alpha = 99,87\%$. Aceasta înseamnă că valoarea adevărată a semnalul probei martor poate fi determinată cu un factor de risc, $\alpha = 0,13\%$ (factorul de risc se referă la probabilitatea prezenței analitului de interes în proba martor).

Deci, limita de decizie, L_c , reprezintă cel mai mic semnal analitic (concentrație) care poate fi atins prin metoda analitică aleasă. Pentru $k_c = 3$, limita de decizie, L_c , este egală cu limita de detecție așa cum a fost definită de IUPAC [9].

În principiu limita de decizie, L_c , nu poate fi considerată un criteriu de performanță pentru metodele analitice. Conform definiției limitei de decizie, probabilitatea de a nu detecta analitul de interes atunci când concentrația acestuia generează un semnal corespunzător L_c , este de 50%. Funcțiile de distribuție pentru martor, y_m și probă, y sunt reprezentate în figura 3.5. pentru un număr mare de determinări, repetate, pentru aceeași probă cu concentrația corespunzătoare limitei de decizie, L_c . Se poate observa că probabilitatea de a nu detecta analitul de interes atunci când este prezent în probă la o concentrație corespunzătoare L_c , este de 50%.

Abaterea standard pentru martor, σ_m și pentru probă σ sunt considerate egale (valabil de obicei pentru concentrații mici). Măsurarea unui semnal analitic mai mare decât semnalul corespunzător L_c , poate fi interpretată ca fiind datorată prezenței analitului de interes în probă cu o certitudine mai mare de 50%. Un semnal analitic mai mic decât L_c poate fi datorat absenței analitului de interes, certitudinea interpretării fiind mai mică de 50%. β este dat de relația:

$$\beta = \int_{-\infty}^{L_c} P(y) dy \quad (3.13)$$

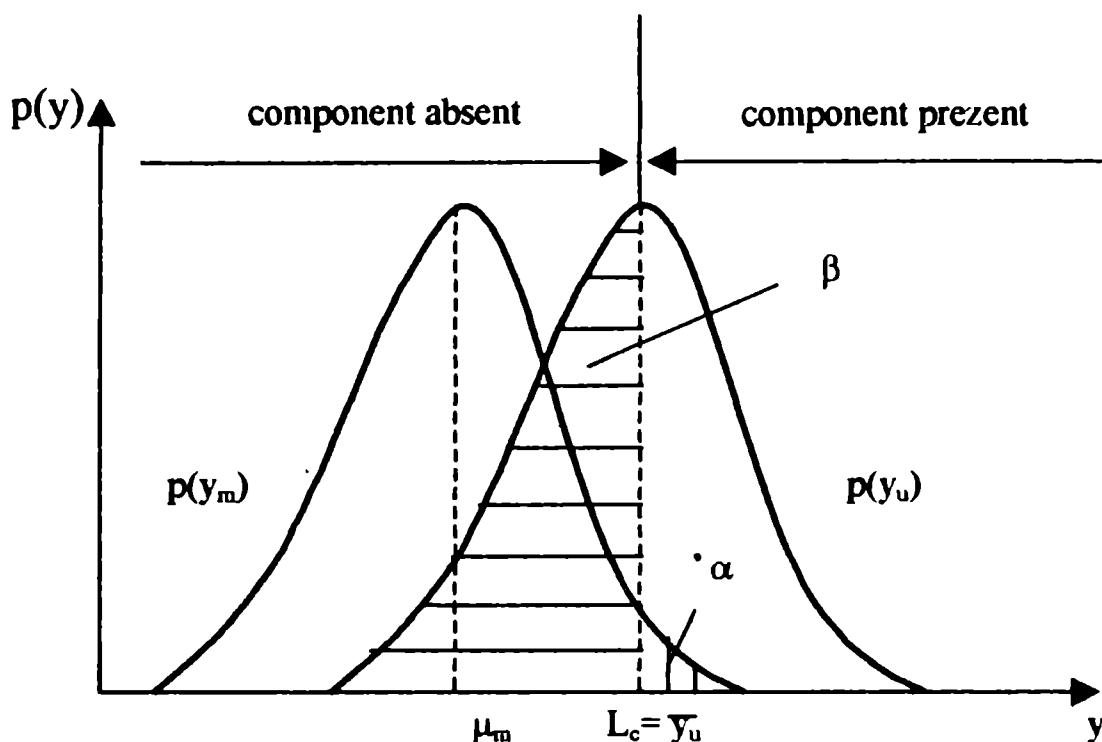


Figura 3.5. Ilustrarea grafică a limitei de decizie.

În vederea caracterizării unei metode analitice este necesar să se specifice limita de detecție, L_d , care corespunde concentrației care conduce la obținerea unui semnal analitic ce poate fi distins de semnalul analitic corespunzător probei martor. Cu alte cuvinte, **limita de detecție**, trebuie definită astfel încât valorile corespunzătoare pentru α sau β să fie echilibrate. În figura 3.6. este reprezentat cazul în care $\alpha = \beta$.

Semnalul analitic corespunzător **limitei de detecție** poate fi exprimat astfel:

$$L_d = y_m + k_d \cdot \sigma_m = L_c + k'_d \cdot \sigma_m \quad (3.14)$$

Considerându-se că deviațiile standard corespunzătoare funcțiilor de distribuție $P(y_m)$ și $P(y)$ sunt egale. Dacă concentrația analitului de interes prezent în probă este egală cu limita de detecție (pentru $k_d=6$ sau $k'_d=3$),

acesta poate fi determinat cu o certitudine de 99,87%. Concentrațiile mai mici pot fi detectate cu un nivel de încredere mai mic.

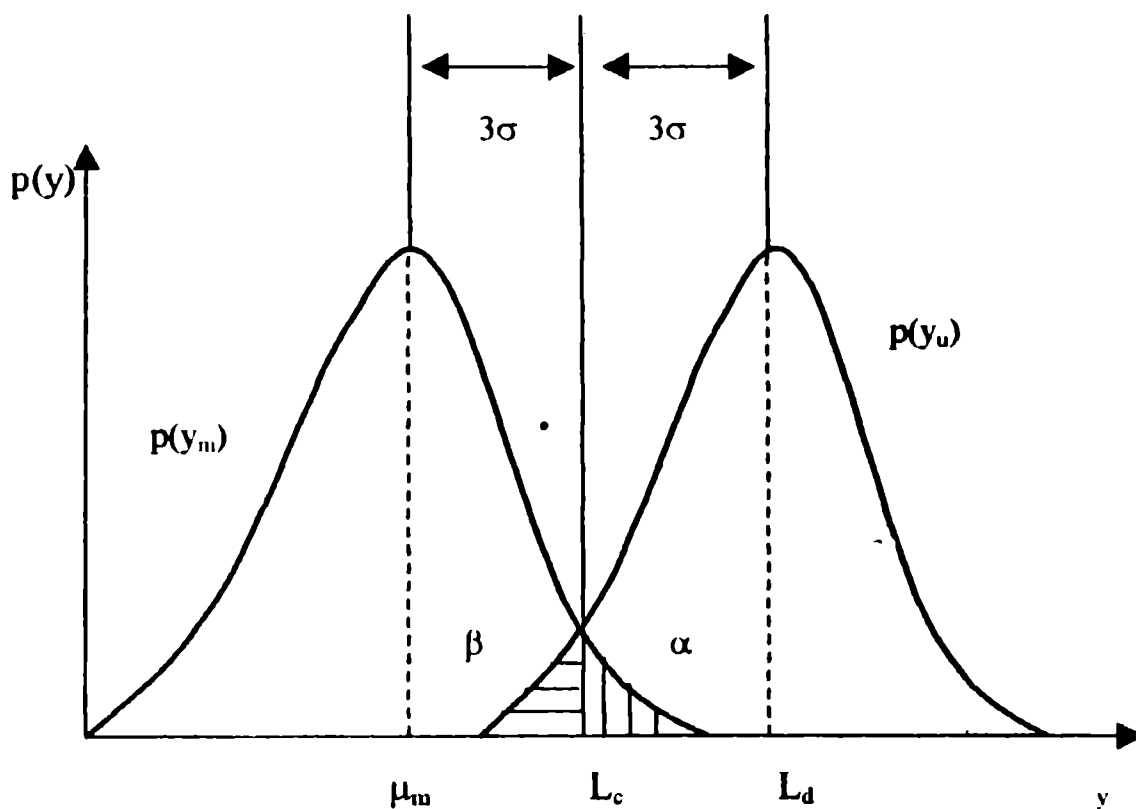


Figura 3.6. Ilustrarea grafică a limitei de detecție.

Limita de determinare, L_q , poate fi definită ca limita pentru care o metodă analitică dată este suficient de precisă pentru a permite măsurarea valorii conținutului de determinat pentru un analitul de interes. Cu alte cuvinte, limita de determinare este concentrația care poate fi determinată cu o abatere relativă standard cu valoare cunoscută și maximă.

Semnalul analitic corespunzător **limitei de determinare** poate fi exprimat astfel:

$$L_q = y_m + k_q \cdot \sigma_m \quad (3.15)$$

unde $k_q = 20$ dacă valoarea maximă admisă pentru abaterea relativă standard este 5% sau $k_q = 10$ dacă valoarea maximă admisă pentru abaterea relativă standard este 10%. Abaterea relativă standard pentru măsurători la nivelul limitei de determinare, L_q , este dată de raportul $1/k_q$.

Definițiile limitei de decizie, a limitei de detecție și a limitei de determinare sunt ilustrate grafic în figura 3.7.

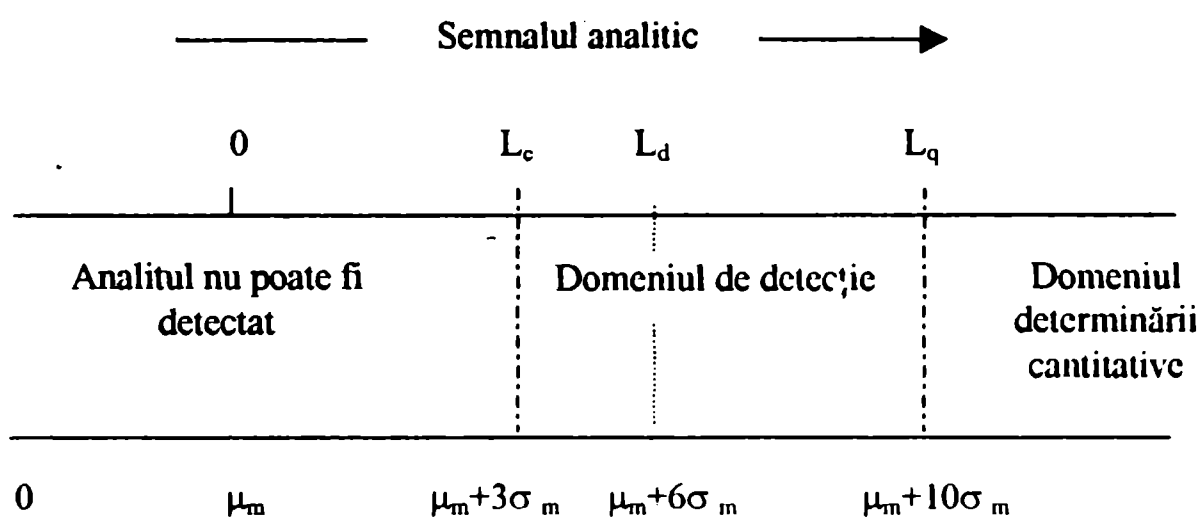


Figura 3.7. Ilustrarea grafică a limitei de decizie, limitei de detecție și limitei de determinare.

Adeseori în literatură nu există un consens în definirea, interpretarea și determinarea limitei de detecție. Din acest motiv atunci când se stabilesc caracteristicile de performanță pentru o anumită metodă analitică, chimiștii analiști trebuie să precizeze modalitatea de definire a limitei de detecție.

Exemplu

În urma analizei prin spectrometrie de absorbție atomică a mercurului dintr-o serie de soluții standard s-a obținut o curbă de calibrare cu următoarea ecuație:

$$y = 1,15 C_{\text{Hg}} + 0.025 \quad (3.16)$$

unde y este semnalul analitic iar C_{Hg} este concentrația de Hg exprimată în $\mu\text{g/mL}$. Soluția martor a fost măsurată de 10 ori, obținându-se o valoare medie a semnalului analitic de 0.28 cu o abatere standard egală cu 0.008. Pentru soluția cu o concentrație de mercur de $1,0 \mu\text{g/mL}$ abaterea standard este 0.01.

Să se determine sensibilitatea metodei, sensibilitatea analitică pentru $C_{\text{Hg}}=1,0 \mu\text{g/mL}$ și limita de detecție ($k = 3$).

- Sensibilitatea este dată de panta curbei de calibrare, iar în acest caz este:

$$S = 1,15 \text{ unitate semnal analitic}/(\mu\text{g/mL}).$$

- Sensibilitatea analitică pentru $C_{\text{Hg}}=1,0 \mu\text{g/mL}$ se determină cu ecuația 3.10.

$$S_a = \frac{1,15}{0.01} = 115 \text{ mL} / \mu\text{g}$$

- Limita de detecție

Semnalul analitic minim măsurabil este:

$$y = y_m + 3 \sigma_m$$

$y = 0,28 + 3 (0.008) = 0,304$ iar concentrația corespunzătoare limitei de detecție va fi:

$$C'_{\text{Hg}} = \frac{0,304 - 0,28}{1,15} = 0,021 \mu\text{g/mL}$$

Exercițiu

La analiza conținutului de carbon total în apă s-au obținut următoarele

date:

Concentrația ($\mu\text{g/mL}$)	Nr. de determinări	Media semnalului analitic	Abaterea standard
0,0	20	0,03	0,008
6,0	10	0,45	0,0084
10,0	7	0,71	0,0072
19,0	5	1,28	0,015

- Să se determine sensibilitatea metodei.
- Care este limita de detecție a acestei metode ($k=3$)?
- Calculați abaterea relativă standard pentru fiecare set de determinări.

III.4. Selectivitatea și rezoluția

Selectivitatea în chimia analitică se referă la posibilitatea unei metode de a detecta sau determina unul sau mai mulți componenți din probă, independent de prezența altor componenți. În general, fiecare probă este un amestec de mai mulți componenți e conține: (1) *analizii* de interes, care sunt componenții ce vor fi determinați cantitativ și (2) *matricea*, reprezentată de toți ceilalți componenți prezenți în probă, care nu prezintă interes din punctul de vedere al determinării cantitative, dar care pot influența măsurătorile.

Răspunsul sistemului pentru determinarea conținutului unui analit poate fi influențat de:

(a) Matricea și/sau alți analiți prezenți în proba de analizat. În chimia analitică acest efect este cunoscut sub denumirea de *efect de matrice*;

b) Prezența în matrice a anumitor elemente care pot contribui la răspunsul analitic, fără să influențeze sensibilitatea detectorului pentru analitul de interes. Acest fenomen este cunoscut sub numele de *interferență*. Exemple tipice în acest sens îl constituie detectorii pentru gaz cromatografe și spectrofotometrele UV-vis. Când două componente din amestec sunt eluate împreună, semnalul de răspuns al detectorului (de ex., detector cu ionizare în flacără) este dat de suma răspunsurilor generate de cele două componente.

În spectrometria UV-vis mai mulți compuși absorb radiația la aceeași lungime de undă iar atunci când aceștia sunt prezenți în aceeași probă, absorbanta măsurată este dată de suma absorbanțelor fiecărui component în parte.

c) Efectul de matrice și de interferențe.

Fenomenul de *interferență* se produce atunci când semnalul elementului de determinat nu poate fi separat de semnalele altor componente din proba de analizat. În acest caz se folosește **rezoluția** ca măsură a gradului de suprapunere a două semnale experimentale. În figura 3.8. este reprezentată grafic definirea rezoluției.

Gradul de suprapunere al celor două picuri este dat de raportul:

$$R_s = 2 \frac{\Delta x}{x_1 + x_2} \quad (3.17)$$

iar gradul de suprapunere a două semnale, de orice formă, este evaluat prin integrala:

$$A = \int_{-\infty}^{+\infty} y_1(x)y_2(x)dx \quad (3.18)$$

în care $y_1(x)$ și $y_2(x)$ reprezintă funcțiile celor două semnale. Când cele două semnale sunt separate, A este egal cu zero, iar atunci când se suprapun ($A > 0$), valoarea parametrului A crește odată cu creșterea gradului de suprapunere.

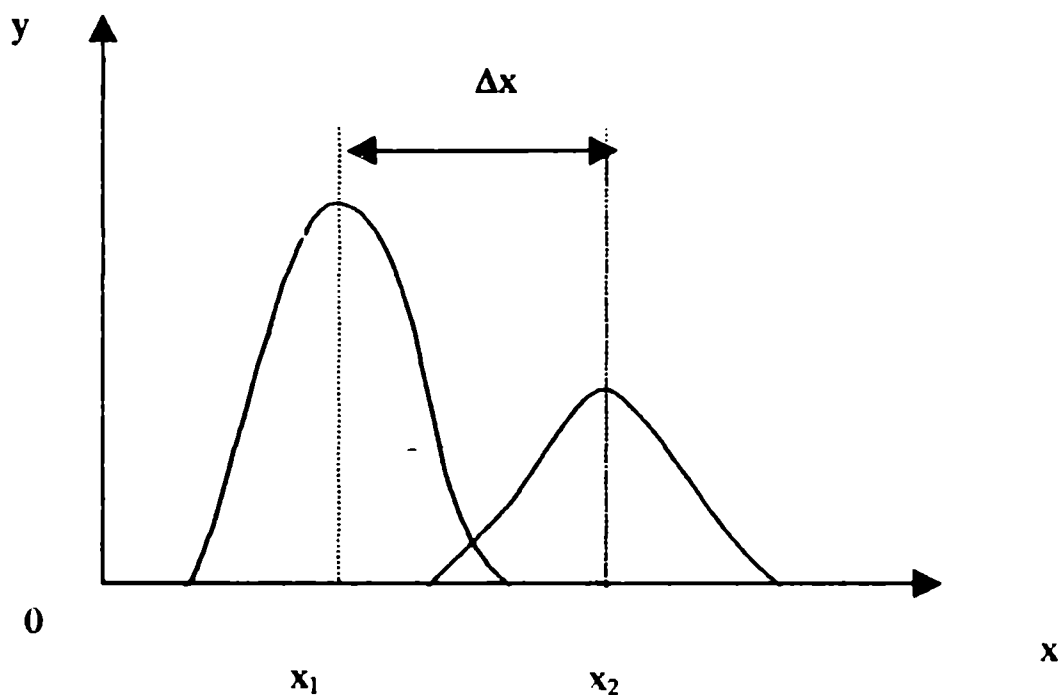


Figura 3.8. Reprezentarea schematică a noțiunii de rezoluție.

Efectul de matrice și interferențele reprezintă o problemă ce este abordată permanent de către chimiștii analiști; ambele efecte determinând complexitatea procedurilor analitice. Cu cât efectul de matrice și interferențele sunt diminuate cu atât mai mult se pot obține rezultate corecte și deci se sporește calitatea determinărilor printr-o anumită procedură analitică.

În prezent, în literatura de specialitate nu se constată o uniformitate în descrierea termenilor de selectivitate, specificitate, interferențe și efect de matrice [10-14]. IUPAC [9] recomandă definirea unei interferențe pentru a procedură analitică ca fiind cea care cauzează o eroare sistematică predeterminată pentru rezultatul analitic.

Următorul exemplu poate fi de folos pentru înțelegerea termenilor de specific și selectiv. Dacă un reactiv formează un complex colorat numai cu un singur analit, acesta este *specific* pentru acel analit. De asemenea dacă reactivul formează complecși colorați cu mai mulți componenți ai probei, dar fiecare complex are culoare diferită, procedura analitică de complexare este *selectivă*. Deci o procedură analitică selectivă conține “senzori” specifici pentru unul dintre analiți.

De asemenea când se vorbește de specificitate și selectivitate trebuie să se țină cont de toți analiții prezenți în probă. O metodă analitică ce are la bază tehnica spectrometriei de absorbție atomică poate fi specifică “în totalitate” pentru determinări de Ca în probe care nu conțin Al și PO_4^{3-} . Atunci când Al și PO_4^{3-} sunt prezenți sensibilitatea metodei pentru determinarea Ca scade.

Selectivitatea poate fi evaluată cu ajutorul funcției de răspuns. În cazul unei funcții de răspuns liniare în raport cu conținutul componentelor se poate scrie următoarea relație:

$$y = a_1c_1 + a_2c_2 + \dots + a_nc_n \quad (3.19)$$

unde prin c_1 s-a notat conținutul componentului de analizat, iar prin c_2, c_3, \dots, c_n conținuturile celorlalți componenți din probă. Metoda este selectivă dacă a_1 este singurul coeficient diferit de zero.

Dacă se consideră cazul unui sistem analitic capabil de a determina m componenți, sistemul de m ecuații se poate scrie:

$$\begin{aligned}
 y_1 &= a_{11}c_1 + a_{21}c_2 + \dots + a_{n1}c_n \\
 y_2 &= a_{12}c_1 + a_{22}c_2 + \dots + a_{n2}c_n \\
 &\dots\dots\dots \\
 y_m &= a_{1m}c_1 + a_{2m}c_2 + \dots + a_{nm}c_n
 \end{aligned}
 \tag{3.20}$$

În cazul în care funcția de răspuns este liniară în raport cu conținutul componentelor, matricea corespunzătoare acestui sistem care ne conduce la informația referitoare la selectivitatea sistemului are următoarea formă:

$$\begin{vmatrix}
 a_{11} & a_{21} & \dots & a_{n1} \\
 a_{12} & a_{22} & \dots & a_{n2} \\
 \dots & \dots & \dots & \dots \\
 a_{1m} & a_{2m} & \dots & a_{nm}
 \end{vmatrix}
 \tag{3.21}$$

Metoda este selectivă pentru toți cei m componenți de determinat dacă numai coeficienții de tip a_{11} sunt diferiți de zero. Dacă funcția de răspuns nu este liniară în raport cu conținutul componentelor, funcția de răspuns nu mai corespunde formei din ecuația 3.19.

Metoda cea mai potrivită pentru evaluarea selectivității este metoda experimentului factorial planificat însoțită de analiza de regresie. Evaluarea selectivității unei metode nu este o problemă simplă. Ea presupune corelarea cunoștințelor matematice și statistice și adaptarea acestora la informația chimică cerută.

Bibliografie

1. B.B. Kebbekus și S. Mitra, *Performance of analytical methods: figures of merit* în “*Environmental Chemical Analysis*”, Ed. Blackie Academic & Professional, New York, 1998, pp.24-29.
2. L. Mutihac și V. David, *Chemometrie*, Ed. Universității București, 1997, pp.13-21.
3. C. Liteanu și I. Rîcă, *Optimizarea proceselor analitice*, Ed. Academiei Române, București, 1985.
4. D. Ceașescu, *Utilizarea statisticii matematice în chimia analitică*, Ed. Tehnică, București, 1982.
5. C. Liteanu și I. Rîcă, *Teoria și metodologia statistică a analizei urmelor*, Ed. Scrisul Românesc, Craiova, 1979.
6. D.L. Massart, A. Dijkstra și L. Kaufman, *Evaluation and Optimization of Laboratory Methods and Analytical Procedures*, Elsevier Science Publishers, Amsterdam, 1978.
7. D.L. Massart,, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, *Chemometrics: A Textbook*, Elsevier Science Publishers, Amsterdam, 1987.
8. O. Matthias, *Chemometrics. Statistics and Computer Application in Analytical Chemistry*, Ed. Wiley-VCH, Weinheim, 1999, pp.81-100.
9. IUPAC, *Spectrochim. Acta*, 33 B, 1978, 242.
10. R. Belcher, *Talanta*, 23, 1976, 1883.
11. R. Belcher și D. Betteridge, *Talanta*, 13, 1966, 535.

12. H. Kaiser, *Z. Analyt. Chem.*, 260, 1972, 252.
13. A.I. Wilson, *Talanta*, 21, 1974, 1109.
14. L. Pszonicki, *Anal. Chem*, 25, 1980, 439.

IV. ANALIZA DE REGRESIE

IV.1. Principalele tipuri de regresii utilizate în analiza chimică

Analiza de regresie reprezintă un instrument matematic utilizat frecvent în analiza chimică și care permite modelarea datelor experimentale afectate de erori. Liteanu și Rîcă [1] au arătat că cele mai numeroase aplicații ale analizei de regresie, în analiza chimică, se întâlnesc la evaluarea funcțiilor de etalonare a sistemelor analitice, la stabilirea condițiilor optime de lucru, la evidențierea și corectarea fenomenelor de interferență și la evaluarea performanțelor calitative ale sistemelor analitice. Analiza de regresie este utilizată de chimistul analist ori de câte ori este necesară găsirea unui model matematic care să descrie cât mai bine setul de date experimentale de care dispune.

Răspunsul unui sistem analitic, notat cu z , este influențat de un număr mare de factori experimentali cum ar fi: concentrația analiților și a interferenților din proba analizată, mediu de reacție, condițiile de operare etc. Deoarece este de dorit ca modelarea răspunsului să se facă în funcție de un număr limitat de factori, se variază sistematic factorii experimentali considerați la elaborarea modelului – variabilele independente – și se determină valoarea răspunsului – variabila dependentă.

Astfel valoarea răspunsului obținut în experimentul numărul i va fi dată de o relație de forma:

$$y_i = f(x_{j_i}; b_k) + e_i \quad (4.1)$$

unde x_{ji} reprezintă valorile obținute în experimentul numărul i a factorilor x_j ($j = 1, m$); b_k – o constantă denumit parametrul modelului ($k = 0, m$); $f(x_{ji}; b_k)$ – reprezintă componenta deterministă a valorii răspunsului și e_i – componenta aleatoare sau stohastică care modelează variabilitatea datelor analitice ca urmare a variațiilor condițiilor experimentale [2].

Principalele tipuri de regresii întâlnite în analiza chimică sunt:

- *Regresie monovariabilă liniară*

$$y = b_0 + b_1 x \quad (4.2)$$

- *Regresie monovariabilă polinomială*

$$y = b_0 + \sum_{k=1}^m b_k x^k \quad (4.3)$$

- *Regresie monovariabilă neliniară*

$$y = f_{neliniara}(x) \quad (4.4)$$

- *Regresie multivariabilă liniară*

$$y = b_0 + \sum_{i=1}^m b_i x_i \quad (4.5)$$

Regresia monovariabilă presupune existența unei singure variabile independente, independent care este relația de dependență dintre aceasta și variabila dependentă, în timp ce regresia multivariabilă presupune că variabila dependentă este o funcție de mai multe variabile independente [3]. În funcție de forma matematică a modelului utilizat, se poate face distincție între metodele de regresie liniare și cele neliniare. În cazul regresiei liniare, variabila dependentă poate fi reprezentată printr-o combinație liniară de operatori care sunt funcție de variabila independentă (sau variabila independentă).

În cazul regresiei monovariabile, astfel de operatori sunt x^c ($c = \text{const.}$), $\log x$, $1/x$, $\sin x$, etc.

În cazul analizei din chimia analitică se utilizează adeseori regresia liniară propriu zisă pentru obținerea curbei de calibrare atunci când variabila independentă semnifică concentrația componentului analizat din etaloane iar b_0 și b_1 – ordonata la origine a dreptei de calibrare și, respectiv, panta acesteia.

În cazul analizei bazate pe metodele spectrometrice [3] se folosește regresia multivariabilă pentru determinarea simultană a mai multor componenți prezenți în soluția de analizat (în acest caz se presupune că $b_0 = 0$).

Ecuțiile utilizate în cazul metodelor neliniare pot fi liniarizate, în anumite condiții, prin dezvoltări în serie. Dacă avem o funcție ce depinde de mai mulți parametrii, b_1, b_2, \dots, b_m , de forma:

$$y = f(x, b_1, \dots, b_m) \quad (4.6)$$

aceasta poate fi aproximată în jurul unui punct din spațiul parametrilor, b_0 , definit prin vectorul linie

$$b_0^T = (b_{10}, b_{20}, \dots, b_{m0}) \quad (4.7)$$

printr-o dezvoltare în serie Taylor din care se rețin termenii de ordinul 0 și 1, conform ecuației:

$$y = f[x, b_{10}, \dots, b_{m0}] + \left(\frac{\delta f}{\delta b_1} \right)_{b=b_0} \cdot \Delta b_1 + \left(\frac{\delta f}{\delta b_2} \right)_{b=b_0} \cdot \Delta b_2 + \dots + \left(\frac{\delta f}{\delta b_m} \right)_{b=b_0} \cdot \Delta b_m \quad (4.8)$$

cu condiția ca Δb_k ($k=1, m$) să fie suficient de mici. Utilizând această proprietate, regresia neliniară poate fi efectuată în același mod ca regresia liniară, în cazul în care inițial se cunoaște sau se poate determina o estimare suficient de bună a valorilor parametrilor modelului.

În analiza chimică se utilizează frecvent ca modele monoparametrice neliniare următoarele funcții:

- funcțiile exponențiale

$$y = b_1 \exp(b_2 x) \quad (4.9)$$

care descriu funcțiile de răspuns ale instrumentelor analitice după aplicare unui semnal în formă de treaptă;

- funcțiile putere

$$y = b_1 x^{b_2} \quad (4.10)$$

care descriu intensitatea liniilor spectrale în spectrografia de emisie;

- funcțiile logaritmice

$$y = b_1 + b_2 \log x \quad (4.11)$$

care descriu răspunsul electrozilor ioni selectivi;

- funcțiile de tip Gauss

$$y = b_1 \exp\left[-b_2 (x - b_3)^2\right] \quad (4.12)$$

care descriu forma picurilor cromatografice;

- funcțiile de tip Lorentz

$$y = \frac{b_1}{1 + b_2 (x - b_3)^2}$$

care descriu forma picurilor de rezonanță magnetică nucleară.

IV.2. Etapele analizei de regresie

Analiza de regresie presupune parcurgerea următoarelor etape: definirea modelului, estimarea parametrilor modelului (regresia propriu zisă), validarea modelului și calculul intervalelor de încredere ale parametrilor.

IV.2.1. Definirea modelului

În această etapă se are în vedere explicitarea relației 4.1. și stabilirea numărului de parametri ai modelului (constantele b_k). Adeseori modelul matematic adoptat este sugerat de anumite legi fizice care se respectă în cazul studiat, de exemplu, legea Lambert-Beer în spectrometrie, ecuația Nernst în potențimetrie, etc.

Pentru identificarea tipurilor de interferențe care afectează o anumită metodă de analiză se compară rezultatele obținute cu metoda studiată, y_i , cu rezultatele obținute, pentru aceleași probe, cu o metodă de referință considerată exactă, x_i . În astfel de cazuri este foarte probabil să fie respectat un model liniar de forma:

$$y = b_0 + b_1 x \quad (4.13)$$

Valorile estimate ale parametrilor furnizează următoarele informații:

- (i) dacă $b_0 = 0$ și $b_1 = 1$, metoda studiată este exactă;

(ii) dacă $b_0 \neq 0$ și $b_1 = 1$, metoda de analiză este afectată de o eroare sistematică de tip aditiv și pentru obținerea rezultatului corect, b_0 trebuie scăzut din rezultatul determinat;

(iii) dacă $b_0 = 0$ și $b_1 \neq 1$, metoda este afectată de o eroare de tip multiplicativ și pentru obținerea rezultatului corect se poate aplica metoda adărilor standard;

(iv) dacă $b_0 \neq 0$ și $b_1 \neq 1$, metoda este afectată atât de erori de tip aditiv cât și multiplicativ.

Atunci când nu se cunoaște relația dintre factorii de optimizat și răspuns pentru optimizarea parametrilor de lucru se folosește modelarea empirică. Pentru a putea utiliza regresia monovariabilă liniară, se presupun următoarele relații între perechile de date experimentale, (y_1, x_1) , (y_2, x_2) ,....., (y_n, x_n) măsurate în condiții reale:

$$\begin{aligned} y_1 &= b_0 + b_1 x_1 + e_1 \\ y_2 &= b_0 + b_1 x_2 + e_2 \\ &\dots\dots\dots \\ y_n &= b_0 + b_1 x_n + e_n \end{aligned} \tag{4.14}$$

Dacă se notează cu y , b , e și X vectorii coloană ai măsurătorilor, parametrilor, erorilor și respectiv matricea variabilelor independente, sistemul de mai sus poate fi scris în formă matriceală [4].

$$y = X b + e \tag{4.15}$$

unde:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ - \\ - \\ y_n \end{pmatrix}; \quad b = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}; \quad e = \begin{pmatrix} e_1 \\ e_2 \\ - \\ - \\ y_n \end{pmatrix}; \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ - & - \\ - & - \\ 1 & x_n \end{pmatrix} \quad (4.16)$$

Transpusa vectorilor coloană sau matricelor se reprezintă adăugând litera **T** în dreapta simbolului. De exemplu, dacă **z** reprezintă vectorul coloană cu **n** elemente definit prin ecuația 4.16, atunci **y^T** reprezintă următorul vector linie:

$$y^T = (y_1 \quad y_2 \quad \dots \quad y_n) \quad (4.17)$$

iar

$$X^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \quad (4.18)$$

Aplicarea algebrei matriceale în analiza de regresie prezintă avantajul că ecuațiile și soluțiile se scriu într-o formă generală aplicabilă tuturor problemelor și ușor de transpus în limbajul calculatoarelor [5].

IV.2.2. Estimarea parametrilor modelului

Estimarea parametrilor modelului se referă la suprapunerea modelului peste perechile de date experimentale și calcularea valorilor necunoscute ale parametrilor. Aceasta presupune minimalizarea unui

criteriu matematic a căruia expresie este funcție de natura distribuției erorilor în setul de date. Dacă modelul considerat aproximează suficient de bine dependența exprimată prin relația 4.1. prin funcția f pe tot domeniul de variație al factorilor, variațiile $(y_i - \hat{y}_i)$, (\hat{y}_i este estimarea mărimii y_i) reprezintă o estimare a erorilor experimentale, e_i .

Criteriul matematic, R , utilizat pentru estimarea corectă a valorilor parametrilor modelului este dat de suma pătratelor abaterilor [6]:

$$R = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.19)$$

cu condiția ca erorile experimentale de determinare a răspunsului să reprezinte variabile aleatoare care să îndeplinească următoarele condiții;

- (i) să fie distribuite normal cu media zero;
- (ii) să fie independente;
- (iii) să aibă aceeași varianță.

Criteriul matematic, R , definit anterior reprezintă norma vectorului reziduurilor, notată $|e|$ și calculată, în notațiile matriceale cu relația:

$$|e| = e^T e = (y - \hat{y})^T (y - \hat{y}) = (y - Xb)^T (y - Xb) \quad (4.20)$$

În practică se îndeplinesc în cvasitotalitate proprietățile i) și ii) ale erorilor experimentale menționate anterior. Proprietatea iii), denumită *homoscedasticitate*, nu este de obicei respectată într-un domeniu larg de

variație al factorilor. În acest caz pentru fiecare pereche de valori a factorilor, $x_i^T = (x_{1i}, x_{2i}, \dots, x_{mi})$, se repetă experimentul analitic și se măsoară valorile răspunsului y_{ii} după care se calculează dispersiile și se testează egalitatea acestora [7].

Estimarea parametrilor modelului cel mai des utilizate se numesc *regresii prin metoda celor mai mici pătrate*.

În cazul multor tehnici analitice, incluzând măsurătorile radiometrice și spectrometrice, varianța erorilor experimentale este dependentă de valoarea răspunsului, y_i , proprietatea aceasta numindu-se *heteroscedasticitate* [7]. Aceasta se întâmplă datorită tipurilor specifice de zgomot care afectează măsurătorile. Astfel, în determinările radiometrice, bazate pe numărarea particulelor care ajung la detector, predomină un tip de zgomot supus distribuției lui Poisson. În astfel de cazuri, dispersia erorilor experimentale este proporțională cu rădăcina pătrată a valorilor semnalelor chiar în prezența altor surse de erori (în afara celor de numărare datorate detectorului) cum ar fi: erori de pipetare, variații ale timpilor de reacție sau a temperaturii etc.

În tehnicile spectrometrice, cum ar fi absorbția atomică, zgomotul datorat fluctuațiilor intensității sursei de lumină este proporțional cu intensitatea semnalelor. În unele tehnici analitice poate să predomină zgomotul datorat componentelor electronice ale instrumentului de măsură care poate avea varianță constantă sau, uneori, dependentă de valoarea semnalelor.

În cazul heteroscedasticității nu se mai poate aplica metoda celor mai mici pătrate și, pentru efectuarea regresiei, se apelează la una din următoarele alternative: aplicarea metodei celor mai mici pătrate

ponderate sau transformarea variabilelor pentru obținerea homoscedasticității.

Când se aplică metoda celor mai mici pătrate ponderată trebuie să se cunoască dispersiile pentru valorile răspunsului, y_i , pentru fiecare punct, x_i , din domeniul factorilor.

În acest caz criteriul de optimizare, R , are în acest caz următoare expresie:

$$R = \sum_{i=1}^n (w_i e_i)^2 = \sum [w_i (\bar{y}_i - \hat{y})]^2 \quad (4.21)$$

unde:

w_i reprezintă factorii de pondere ai reziduurilor:

$$w_i = \frac{1}{s_i} \quad (4.22)$$

\bar{y}_i reprezintă media măsurătorile individuale ale variabilei dependente obținute la aceleași valori ale variabilelor independente:

$$\bar{y}_i = \sum_{l=1}^{n_i} y_{il} \quad (4.23)$$

Ideea de bază a regresiei ponderate prin metoda celor mai mici pătrate este de a atribui cea mai mare importanță datelor măsurate ce cea mai mare precizie, și deci o problemă o va constitui estimarea corectă a factorilor de pondere, fiind necesar un număr mare de repetări ale experimentelor [8].

IV.2.3. Validarea modelului

Validarea modelului se realizează prin testarea concordanței dintre modelul matematic obținut anterior și perechile de date experimentale. Pentru validarea modelului se pot utiliza metode grafice sau metodele statisticii matematice. Metodele grafice pornesc de la premiza că modelul este corect, dacă diferențele dintre datele experimentale și cele calculate pe baza modelului sunt variabile aleatoare cu caracteristici apropiate de cele ale erorilor experimentale care afectează măsurarea valorilor răspunsului. De obicei sunt preferate metodele statistice deoarece sunt mai obiective.

Natura distribuției reziduurilor se poate testa fie grafic, prin reprezentarea frecvențelor de apariție a acestora, fie statistic [7]. Reziduurile calculate pot fi reprezentate grafic în funcție de valorile estimate, sau determinate experimental, ale răspunsului (y_i sau \hat{y}_i), sau de valorile factorilor și se analizează forma benzii în care acestea sunt dispersate.

Cazurile cele mai întâlnite sunt (figura 4.1.):

1. Reziduurile se distribuie aleator într-o bandă cu media zero și cu lărgime constantă, situație care poate apărea în caz de homoscedasticitate și dacă modelul este corect;
2. Reziduurile se distribuie aleator într-o bandă cu media zero și cu lărgime variabilă; această situație poate descrie un caz de heteroscedasticitate, dacă modelul este corect;

3. Reziduurile se distribuie aleator într-o bandă a cărei media, în unele porțiuni, se depărtează sistematic de la valoarea zero, caz în care modelul va fi invalidat;
4. Majoritatea reziduurilor sunt distribuite într-o bandă de forma celor din figura 4.1. –(1) și (2)- cu excepția unora care se situează mult în afara limitelor benzii. Perechile de date care conduc la valori foarte mari ale reziduurilor pot fi eliminate din setul de date inițial, după care se recalculează valorile parametrilor.

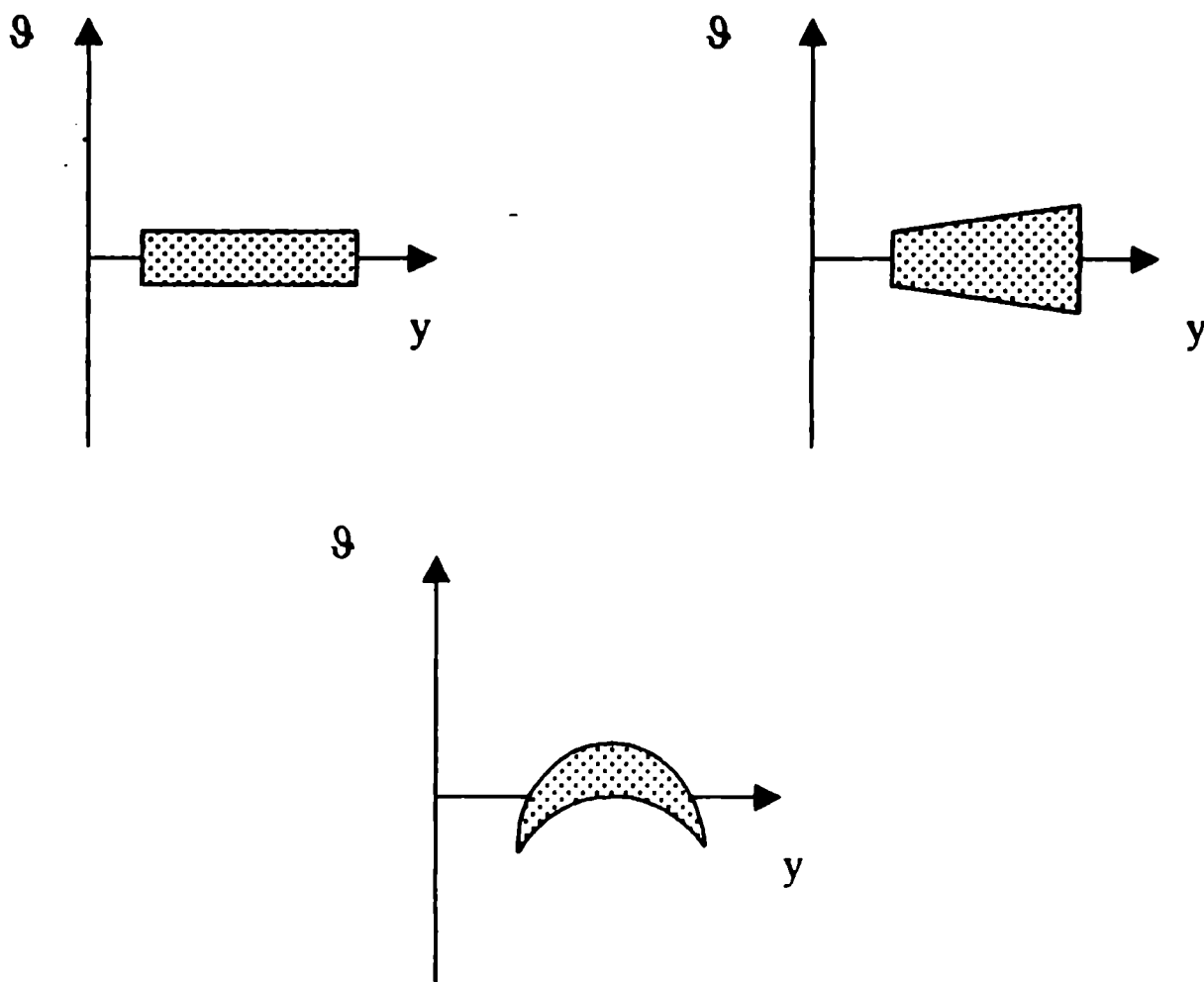


Figura 4.1. Forme de distribuție a reziduurilor; exemple.

IV.2.3.1. Analiza de varianță – ANOVA

Analiza de varianță – ANOVA – este cea mai riguroasă tehnică statistică de analiză a măsurătorilor, aplicată pentru validarea sau invalidarea unui model, și depinde de mai mulți factori care acționează simultan în vederea estimării importanței și efectelor fiecăruia [9].

Analiza de varianță permite să se determine ce parte a variabilității unei populații este cauzată de diverși factori sistematici și, respectiv de întâmplare.

În tabelul 4.1. sunt prezentate datele necesare pentru analiza de varianță cu un singur factor.

Tabelul 4.1. Datele necesare pentru analiza de varianță cu un singur factor.

Nr. de determinări	Grupul de măsurători			
	1	2	q
1	y_{11}	y_{12}	..	y_{1q}
2	y_{21}	y_{21}		y_{2q}
	$y_{n_1 1}$	$y_{n_2 1}$..	$y_{n_q 1}$
Media	\bar{y}_1	\bar{y}_2	..	\bar{y}_{total}

În vederea testării diferențelor repetate dintre grupurile de măsurători se presupune că fiecare măsurătoare, y_{ij} este descrisă de media globală a valorilor determinate, \bar{y}_{total} , media valorilor din fiecare grup, \bar{y}_j , și eroare întâmplătoare reziduală, e_{ij} , conform relației:

$$y_{ij} = \bar{y}_{total} + (\bar{y}_j - \bar{y}_{total}) + e_{ij} \quad (4.24)$$

Varianța totală, exprimată ca suma de pătrate a variațiilor față de medie, este dată de varianța *dintre* diferite grupuri și *în* cadrul aceluiași grup. Aceasta înseamnă că suma de pătrate totală corectată, SS_{cor}^2 , se obține din suma de pătrate corespunzătoare variației valorilor *dintre* grupuri, SS_{fact}^2 , și suma de pătrate reziduală corespunzătoare variației valorilor *în* grup., SS_R^2 :

$$SS_{cor}^2 = SS_{fact}^2 + SS_R^2 \quad (4.25)$$

unde:
$$SS_{cor}^2 = \sum_{j=1}^q \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{total})^2 \quad (4.26)$$

$$SS_{fact}^2 = \sum_{j=1}^q n_j (\bar{y}_j - \bar{y}_{total})^2 \quad \text{și} \quad \bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij} \quad (4.27)$$

$$SS_R^2 = \sum_{j=1}^q \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 \quad (4.28)$$

$$\hat{y}_{total} = \frac{1}{n} \sum_{j=1}^q \sum_{i=1}^{n_j} y_{ij} \quad (4.29)$$

q - numărul de grupuri;

n_j = numărul de măsurători repetate pentru grupul j ;

n = numărul total de măsurători, ex. $n = \sum_{j=1}^q n_j$

Pentru validarea sau invalidarea modelului se utilizează testul statistic F, valoarea calculată fiind dată de relația:

$$F = \frac{\frac{SS_{fact}^2}{(q-1)}}{\frac{SS_R^2}{(n-q)}} \quad (4.30)$$

Dacă $F < F_{n, N-n}^\alpha$, unde n și $N - n$ reprezintă gradele de libertate corespunzătoare, iar α este nivelul de semnificație al testului, atunci se poate admite, cu probabilitatea $P = 1 - \alpha$, că modelul de regresie analizat exprimă corect relația deterministă dintre variabilele dependente și independente [10].

Spre exemplificare se va discuta analiza de varianță – ANOVA în cazul analizei de regresie liniară:

$$y = b_0 + b_1 x \quad (4.31)$$

sau

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix} \cdot \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \quad (4.32)$$

iar valoarea parametrului \mathbf{b} este dată de relația:

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (4.33)$$

În tabelul 4.2. se prezintă semnificația termenilor sumelor de pătrate (SS) pentru analiza ANOVA completă în regresia liniară.

Tabelul 4.2. Calculul sumelor de pătrate (SS) în analiza ANOVA completă pentru cazul regresiei liniare.

Suma de pătrate	Matricea de operare	Formula de calcul	Grade de libertate
SS_T , Totală	$y^T y$	$\sum_{i=1}^n y_i^2$	n
SS_M , Medie	$\bar{y}^T \bar{y}$	$n\bar{y}^2$	1
SS_{cor} , Corectată față de medie	$(y - \bar{y})^T (y - \bar{y})$	$\sum_{i=1}^n (y_i - \bar{y})^2$	$n - 1$
SS_{fac} , Factorială	$(\hat{y} - \bar{y})^T (\hat{y} - \bar{y})$	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$p - 1$
SS_R , Reziduală	$(y - \hat{y})^T (y - \hat{y})$	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$n - p$
SS_{lof} , "lack-of-fit"	$(j - \hat{y})^T (j - \hat{y})$	$\sum_{i=1}^n (\bar{y}_i - \hat{y}_i)^2$	$f - p$
SS_{pe} , erorilor pur experimentale	$(y - j)^T (y - j)$	$\sum_{i=1}^n (y_i - \bar{y}_i)^2$	$n - f$

j – acest vector media valorilor pentru fiecare punct i ; f – numărul diferitelor determinări a variabilelor independente; n – numărul de măsurători; p –

numărul de parametri; $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ - media totală.

Exemplu – ANOVA în regresia liniară.

În tabelul 4.3. sunt prezentate perechile de date $x - y$ pentru 4 măsurători care urmează să fie prelucrate prin analiza de regresie liniară.

Tabelul 4.3.

Nr.	x	y
1	0	0,3
2	1	2,2
3	2	3
4	2	4

Introducerea acestor date în modelul de regresie liniară dat de ecuația 4.32 conduce la relația:

$$\begin{pmatrix} 0,3 \\ 2,2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \quad (4.34)$$

Parametrul b se calculează cu ecuația 4.33:

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = b(X^T X)^{-1} X^T y = \begin{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix}^{-1} \\ \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 \end{pmatrix} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0,3 \\ 2,2 \\ 3 \\ 4 \end{pmatrix} = \quad (4.35)$$

$$= \begin{pmatrix} 4 & 5 \\ 5 & 9 \end{pmatrix}^{-1} \begin{pmatrix} 9,5 \\ 16,2 \end{pmatrix} = \begin{pmatrix} 0,818 & -0,454 \\ -0,454 & 0,363 \end{pmatrix} \begin{pmatrix} 9,5 \\ 16,2 \end{pmatrix} = \begin{pmatrix} 0,41 \\ 1,57 \end{pmatrix}$$

Astfel, ecuația modelului de regresie liniară propus va fi:

$$y = 0,41 + 1,57 x$$

Aplicând analiza ANOVA pentru acest model se pot calcula sumele de pătrate din tabelul 4.2.

Varianța valorilor y rezultă din *suma totală de pătrate*, SS_T :

$$SS_T = y^T y = (0,3 \ 2,2 \ 3 \ 4) \begin{pmatrix} 0,3 \\ 2,2 \\ 3 \\ 4 \end{pmatrix} = \sum_{i=1}^n y_i^2 = 0,3^2 + 2,2^2 + 3^2 + 4^2 = 29,93$$

Suma pătratelor mediei, SS_M se calculează din media totală valorilor y folosind următoarea formulă:

$$SS_M = n\bar{y} = 4 \cdot 2,375^2 = 22,56$$

Asemănător se calculează și **suma pătratelor corectată față de medie**:

$$SS_{cor} = \sum_{i=1}^n (y_i - \bar{y})^2 = (0,3 - 2,375)^2 + (2,2 - 2,375)^2 + (3 - 2,375)^2 + (4 - 2,375)^2 = 7,37$$

Influența factorilor este reflectată de către **suma de pătrate** SS_{fac} :

$$SS_{fac} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = (0,41 - 2,375)^2 + (1,98 - 2,375)^2 + (3,55 - 2,375)^2 + (3,55 - 2,375)^2 = 6,80$$

Diferența dintre valorile y observate și cele estimate este dată de **suma de pătrate reziduală**, SS_R :

$$SS_R = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = (0,3 - 0,41)^2 + (2,2 - 1,98)^2 + (3 - 3,55)^2 + (4 - 3,55)^2 = 0,57$$

Variația mediei grupurilor față de modelul de regresie este dată de

SS_{lof} .

$$SS_{lof} = \sum_{i=1}^n (\bar{y}_i - \hat{y}_i)^2 = (0,3 - 0,41)^2 + (2,2 - 1,98)^2 + (3,5 - 3,55)^2 + (3,5 - 3,55)^2 = 0,07$$

Suma de pătrate datorată erorilor experimentale este:

$$SS_{pe} = \sum_{i=1}^n (y_i - \bar{y}_i)^2 = (0,3 - 0,3)^2 + (2,2 - 2,2)^2 + (3 - 3,5)^2 + (4 - 3,5)^2 = 0,5$$

Rezultatele analizei de varianță ANOVĂ sunt prezentate în tabelul 4.4.

Tabelul 4.4. Tabelul ANOVA pentru regresia liniară a datelor din tabelul 4.3.

Sursa de varianță	SS	df	MSS	Testul F	p
SS_T , Totală	25,93	4	7,48		
SS_M , Medie	22,56	1	22,56		
SS_{cor} , Corectată față de medie	7,37	3	2,47		
SS_{fac} , Factorială	6,80	1	6,80	24,06	0,0391
SS_R , Reziduală	0,57	2	0,283		
SS_{lof} , "lack-of-fit"	0,07	1	0,07	0,14	0,779
SS_{pe} , erorilor pur experimentale	0,50	1	0,5		

SS - suma de pătrate; MSS - media sumei de pătrate; df - gradele de libertate.

IV.2.4. Calculul intervalelor de încredere

Informația necesară determinării intervalelor de încredere a parametrilor este conținută în matricea de varianță – covarianță, care poate fi exprimată în funcție de media sumelor de pătrate corespunzătoare erorilor pur experimentale, astfel:

$$C = MSS_{pe} (X^T X)^{-1} \quad (4.36)$$

unde $MSS_{pe} = \frac{SS_{pe}}{n - f}$

Pentru cazul în care SS_{lof} are valori mici se folosește varianța reziduală, s_R^2 , în locul varianței datorate erorilor experimentale:

$$s_R^2 = MSS_R = \frac{SS_R}{n - p} \quad (4.37)$$

iar matricea varianță – covarianță va avea expresia:

$$C = s_R^2 (X^T X)^{-1} \quad (4.38)$$

Diagonala matricei varianță – covarianță constă din varianța parametrilor estimați iar ceilalți termeni sunt dați de covarianța parametrilor.

În cazul modelului de regresie liniară, cu parametrii b_0 și b_1 , matricea corespunzătoare este:

$$C = \begin{pmatrix} s_{b_0}^2 & s_{b_0, b_1}^2 \\ s_{b_1, b_0}^2 & s_{b_1}^2 \end{pmatrix} \quad (4.37)$$

Bibliografie

1. C. Liteanu și I. Rîcă, *Teoria și metodologia statistică a analizei urmelor*, Ed. Scisul Românesc, Craiova, 1979.
2. D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte și L. Kaufman, *Chemometrics: A Textbook*, Elsevier Science Publishers, Amsterdam, 1987.
3. R. Shiavi, *Introduction to Applied Statistical Analysis*, Aksen Publishers, Boston, 1991.
4. M. Otto, *Chemometrics. Statistical and Computer Application in Analytical Chemistry*, Ed. Wiley-VCH, Weinheim, 1999.
5. R. Henrion și G. Henrion, *Multivariate data analysis*, Springer, Berlin, 1994.
6. M. Draner și H. Smith, *Applied Regression Analysis*, 2nd Edition, Wiley, New York, 1981.
7. L. Mutihac și V. David, *Chemometrie*, Ed. Universității din București, 1997.
8. S. Weisberg, *Applied Linear Regression*, 2nd Edition, Wiley, New York, 1985.
9. D. Ceașescu, *Utilizarea statisticii matematice în chimia analitică*, Ed. Tehnică, București, 1982.
10. E.V. Thomas, *Anal. Chem.*, 66, 795A, 1994.

V. PROCESAREA SEMNALELOR

V.1. TRANSFORMATE FOURIER

Un proces fizic poate fi descris fie în *domeniul spațial (temporal)*, de exemplu prin valorile unei mărimi funcție de coordonatele spațiale $f(x,y)$, sau la fel de bine în *domeniul frecvenței*, unde procesul este specificat prin amplitudinea sa $F(u,v)$, care este în general o funcție complexă ce conține și faza. Din considerente de calcul, este convenabil de conceput funcțiile $f(x,y)$ și $F(u,v)$ ca reprezentări echivalente ale aceleiași mărimi [1-12].

Transformata Fourier transpune o secvență de date din domeniul spațial în domeniul frecvenței, în timp ce transformata Fourier inversă realizează operația inversă [13, 14].

Exemplu

Codul de mai jos scris în **MATHEMATICA** realizează transformarea Fourier a unor date numerice care corespund unui puls dreptunghiular

```
In[1]:={-1, -1, -1, -1, 1, 1, 1, 1}
Out[1]={-1, -1, -1, -1, 1, 1, 1, 1}
```

Transformata Fourier a datelor de mai sus implică numere complexe

```
In[2]:=Fourier[%]
Out[2]={0.+0. I, -0.707107-1.70711 I, 0. +0.
I, -0.707107-0.292893 I,
```

0. +0. I, -0.707107+0.292893 I, 0. +0. I, -
0.707107+1.70711 I}

Transformarea Fourier inversă a secvenței de date obținută mai sus conduce la

In[3]:=InverseFourier[%]

Out[3]={-1.+0. I, -1.-1.57009 10⁻¹⁶ I, -1.+0. I,
-1.+1.57009 10⁻¹⁶ I,
1. +0. I, 1. +1.57009 10⁻¹⁶ I, 1.+0. I, 1.-
1.57009 10⁻¹⁶ I}

care după trunchiere prin utilizarea instrucțiunii Chop produce exact secvența datelor de intrare

In[4]:=Chop[%]

Out[4]={-1., -1., -1., -1., 1., 1., 1., 1.}

V.1.1. Transformata Fourier unidimensională (1D)

Fie funcția continuă și integrabilă $f(x)$ și funcția $F(u)$ integrabilă, atunci ecuațiile de transformare Fourier se pot defini astfel

$$\mathfrak{F}\{f(x)\} = F(u) = \int_{-\infty}^{+\infty} f(x) \cdot e^{-j2\pi xu} \cdot dx$$

$$\mathfrak{F}^{-1}\{F(u)\} = f(x) = \int_{-\infty}^{+\infty} F(u) \cdot e^{j2\pi xu} \cdot du$$
(5.1)

Dacă se preferă utilizarea frecvenței unghiulare $\omega = 2\pi u$, atunci formulele de mai sus devin

$$\mathfrak{F}\{f(x)\} = F(\omega) = \int_{-\infty}^{+\infty} f(x) \cdot e^{-j\omega x} \cdot dx$$

$$\mathfrak{F}^{-1}\{F(\omega)\} = f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) \cdot e^{j\omega x} \cdot d\omega$$
(5.2)

sau, având în vedere periodicitatea transformatei Fourier $F(u)$, domeniul de integrare în ultima relație se poate restrânge la o perioadă, astfel încât se poate scrie

$$\mathfrak{F}^{-1}\{F(\omega)\} = f(x) = \frac{1}{2\pi} \int_{\omega=-\pi}^{\pi} F(\omega) \cdot e^{j\omega x} \cdot d\omega$$

V.1.1.1. Proprietățile transformatelor Fourier

Proprietățile de simetrie ale funcției $f(x)$ în domeniul spațial (temporal) conduc în domeniul frecvenței la relații speciale între $F(u)$ și $F(-u)$ prezentate în tabelul de mai jos. Utilizarea acestor proprietăți crește semnificativ viteza de calcul și în consecință eficiența utilizării transformatelor Fourier [15-21].

Tabel 5.1 Proprietăți de simetrie a perechilor de transformate Fourier

Dacă ...	atunci ...
$f(x)$ este reală	$F(-u) = F^*(u)$
$f(x)$ este pur imaginară	$F(-u) = -F^*(u)$
$f(-x) = f(x)$ adică $f(x)$ este pară	$F(-u) = F(u)$
$f(-x) = -f(x)$ adică $f(x)$ este impară	$F(-u) = -F(u)$
$f(x)$ este reală și pară	$F(u)$ este reală și pară
$f(x)$ este reală și impară	$F(u)$ este imaginară și impară
$f(x)$ este imaginară și pară	$F(u)$ este imaginară și pară
$f(x)$ este imaginară și impară	$F(u)$ este reală și impară

Reprezentarea unei funcții simple și transformata Fourier corespunzătoare sunt ilustrate în figura 5.1.

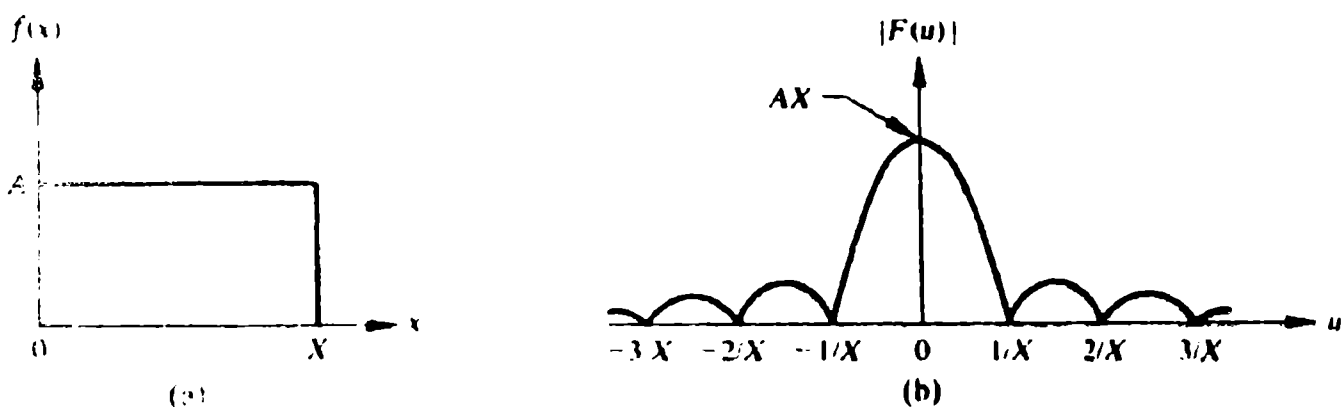


Figura 5.1 Reprezentarea unei funcții simple (a) și a transformatei sale Fourier (b).

În continuare sunt prezentate proprietățile elementare ale transformatelor Fourier. Vom utiliza în continuare pentru perechile de

transformate Fourier notația $f(x) \Leftrightarrow F(u)$ simplificând considerabil formalismul prezentării acestor proprietăți.

- **Liniaritate.** Proprietatea de liniaritate a transformatei Fourier decurge din liniaritatea integralei prin care se definește. Fie n semnale $f_1(x), f_2(x), \dots, f_n(x)$ care au transformatele Fourier respectiv $F_1(u), F_2(u), \dots, F_n(u)$, atunci orice combinație liniară a acestor

semnale $\sum_{i=1}^n a_i f_i(x)$, $a_i \in \mathcal{R}$, $i = 1, 2, \dots, n$, are transformata Fourier o

combinație liniară a transformatelor Fourier ale funcțiilor considerate,

$$\text{astfel încât } \sum_{i=1}^n a_i f_i(x) \Leftrightarrow \sum_{i=1}^n a_i F_i(u).$$

- **Corelație între lărgimea semnalului și lărgimea spectrului.** Cu cât semnalul va fi mai larg, cu atât spectrul său va fi mai îngust, și reciproc. Aceasta înseamnă că FT a unui semnal foarte scurt în domeniul temporal se va întinde considerabil în domeniul frecvenței conținând frecvențe de valoare ridicată. Spre exemplu, la semnale dreptunghiulare din ce în ce mai înguste având lățimile $a, \frac{a}{2}, \frac{a}{4}, \dots, \frac{a}{2k}$

și înălțimi din ce în ce mai mari $\frac{1}{a}, \frac{2}{a}, \frac{4}{a}, \dots, \frac{2k}{a}$ (astfel încât suprafața

lor să rămână unitară $\int_{-\infty}^{+\infty} f(x) \cdot dx = 1$), transformatele Fourier

corespunzătoare sunt sinusuri cardinale din ce în ce mai largi $\text{sinc } \pi u, \text{sinc } 2\pi u, \text{sinc } 4\pi u, \dots, \text{sinc } k\pi u$. La limită, când $k \rightarrow \infty$, semnalul tinde în domeniul temporal către o funcție de tip Dirac

plasat în origine $f(x) \rightarrow \delta(0)$ iar transformatei sale Fourier îi corespunde o dreaptă paralelă cu axa frecvenței în domeniul spectral $F(u) \rightarrow 1$. Reciproc, să considerăm cazul opus al unor semnale dreptunghiulare de amplitudine 1 și lățimi succesive $a, 2a, \dots, 2ka$. Spectrele corespunzătoare vor fi din ce în ce mai înguste astfel încât aria subgraficelor transformatelor Fourier corespunzătoare să rămână

constantă la valoarea $\int_{-\infty}^{+\infty} F(u) du = 1$. În cazul limită al unui semnal de amplitudine constantă și invariabil în timp (spațiu) $f(x) = 1$ transformata sa Fourier este funcția Dirac $\delta(u)$. În concluzie $\delta(x) \Leftrightarrow 1$ și $1 \Leftrightarrow \delta(u)$.

- *Interschimbabilitate între rolurile funcției directe și a transformatei Fourier corespunzătoare.* Cu alte cuvinte, dacă $f(x)$ are transformata

Fourier $F(u)$ avem $f(x) = \int_{-\infty}^{+\infty} F(u) \cdot e^{j2\pi xu} du$; schimbând rolurile

variabilelor x și u între ele rezultă $f(u) = \int_{-\infty}^{+\infty} F(x) \cdot e^{j2\pi ux} dx$. Dacă

notăm transformata Fourier a funcției $F(x)$ sub forma $\bar{F}(u)$, atunci

avem $\bar{F}(u) = \int_{-\infty}^{+\infty} F(x) \cdot e^{-j2\pi ux} dx = f(-u)$. În cazul particular al unui

semnal par $f(-u) = f(u)$ se obține imediat $\bar{F}(f) = f(u)$. Semnificația este următoarea: dacă un semnal par $f(x)$ are o transformată Fourier

$F(u)$ de asemenea pară, atunci și semnalul $F(x)$ are ca transformată Fourier funcția $f(u)$.

- **Scalarea în timp (spațiu).** Se poate arăta imediat că

$$f(ax) \Leftrightarrow \frac{1}{|a|} F\left(\frac{u}{a}\right) \quad (5.3)$$

Într-adevăr avem $\int_{-\infty}^{+\infty} f(ax) e^{-j2\pi xu} dx = \frac{1}{|a|} \int_{-\infty}^{+\infty} f(ax) e^{-j2\pi \frac{u}{a} ax} d(ax)$ și

dacă efectuăm schimbarea de variabilă $ax \rightarrow x'$, ultima integrală

devine $\frac{1}{|a|} \int_{-\infty}^{+\infty} f(x') e^{-j2\pi \frac{u}{a} x'} dx' = \frac{1}{|a|} F\left(\frac{u}{a}\right)$, ceea ce înseamnă că

transformata Fourier a semnalului $f(ax)$ este $\frac{1}{|a|} F\left(\frac{u}{a}\right)$.

- **Scalarea în frecvență.** În mod analog se arată că este adevărată și proprietatea

$$\frac{1}{|b|} f\left(\frac{x}{b}\right) \Leftrightarrow F(bu) \quad (5.4)$$

- **Deplasarea în domeniul temporal (spațial).** O deplasare τ în domeniul temporal are ca efect multiplicarea transformatei sale Fourier $F(u)$ cu un factor de fază $e^{-j2\pi u\tau}$, adică

$$f(x - \tau) \Leftrightarrow F(u) \cdot e^{-j2\pi u\tau} \quad (5.5)$$

De fapt, dacă considerăm $\int_{-\infty}^{+\infty} f(x - \tau) \cdot e^{-j2\pi ux} dx$ și notăm $x - \tau = x'$,

atunci prin substituție obținem

$$\int_{-\infty}^{+\infty} f(x') \cdot e^{-j2\pi u(x'+\tau)} dx' = e^{-j2\pi u\tau} \int_{-\infty}^{+\infty} f(x') \cdot e^{-j2\pi ux'} dx' = e^{-j2\pi u\tau} \cdot F(u)$$

- *Deplasarea în domeniul frecvenței.*

$$f(x) \cdot e^{j2\pi u_0 x} \Leftrightarrow F(u - u_0) \quad (5.6)$$

Această proprietate este un corolar al celei anterioare și se demonstrează rapid pornind de la

$$\int_{-\infty}^{+\infty} f(x) \cdot e^{j2\pi u_0 x} \cdot e^{-j2\pi ux} dx = \int_{-\infty}^{+\infty} f(x) \cdot e^{-j2\pi(u-u_0)x} dx = F(u - u_0).$$

Deoarece $e^{j2\pi u_0 x} = 1 \cdot e^{j2\pi u_0 x}$ iar 1 are ca transformată Fourier semnalul Dirac $\delta(u)$, rezultă că transformata Fourier a exponențialei $e^{j2\pi u_0 x}$ este $\delta(u - u_0)$ reprezentată printr-un segment de lungime

(amplitudine) 1 la frecvența u_0 . Dezvoltând raționamentul de mai

sus, transformata Fourier a unui semnal pur sinusoidal $\cos 2\pi u_0 t = \frac{1}{2}(e^{j2\pi u_0 t} + e^{-j2\pi u_0 t})$ este, pe baza liniarității

transformatei Fourier, o sumă de funcții Dirac

$\frac{1}{2}[\delta(u - u_0) + \delta(u + u_0)]$ care este reprezentată prin două segmente

de amplitudine $\frac{1}{2}$ localizate simetric la frecvențele u_0 și $-u_0$.

Rezultă astfel posibilitatea de a exprima transformata Fourier a unui semnal periodic utilizând funcția Dirac, deși am pornit de la semnale

neperiodice. Concluzia calitativă este că, aparent, transformata Fourier este mai generală decât seria Fourier.

Deși funcția original $f(x)$ poate fi reală, transformata sa Fourier $F(u)$ este în general o funcție complexă care se poate scrie evidențiind partea sa reală și imaginară $F(u) = R(u) + jI(u)$, sau sub formă exponențială $F(u) = |F(u)| \cdot e^{j\phi(u)}$ unde $|F(u)| = [R^2(u) + I^2(u)]^{1/2}$ și $\phi(u) = \tan^{-1}\left(\frac{I(u)}{R(u)}\right)$. Magnitudinea funcției $|F(u)|$ este denumită *spectrul Fourier* al funcției $f(x)$, iar pătratul spectrului $P(u) = |F(u)|^2 = R^2(u) + I^2(u)$ este denumit curent *puterea spectrală* sau *densitatea spectrală a puterii*. Denumirea de variabilă în frecvență a lui u este consecința expresiei termenului exponențial care poate fi scris conform formulei lui Euler

$$\exp[-j2\pi ux] = \cos 2\pi ux - j \sin 2\pi ux$$

Dacă considerăm integrala din definiția transformatei Fourier (5.1) ca limita sumării unor termeni discreți, atunci suntem conduși la interpretarea transformatei $F(u)$ ca o sumă infinită de termeni sinusoidali și cosinusoidali astfel încât fiecare valoare a lui u determină “frecvența” perechii corespunzătoare de termeni sinus și cosinus.

Puterea totală a unui semnal $f(x)$ are aceeași valoare calculată fie în domeniul spațial fie în domeniul frecvenței (teorema lui Parseval)

$$P_f \equiv \int_{-\infty}^{+\infty} |f(x)|^2 dx = \int_{-\infty}^{+\infty} |F(u)|^2 du \quad (5.7)$$

Pentru a preciza puterea semnalului conținută în domeniul spectral cuprins între valorile u și $u + du$ se preferă în general a nu se face distincție între frecvențe pozitive și negative utilizându-se plaja de variație de la 0 (DC) la $+\infty$ a frecvenței u . Se definește astfel *densitatea spectrală de putere PSD (one-sided power spectral density)* sub forma următoare

$$P_f(u) \equiv |F(u)|^2 + |F(-u)|^2, \quad 0 \leq u < +\infty \quad (5.8)$$

astfel încât *puterea totală* se obține prin integrarea frecvenței de la $u = 0$ la $u = +\infty$. În cazul particular în care f este reală, atunci am văzut că $F(-u) = F^*(u)$ și PSD se poate scrie sub o formă mai simplă $P_f(u) \equiv 2|F(u)|^2, 0 \leq u < +\infty$. Totuși, în anumite lucrări, se întâlnește această relație fără factorul 2; strict vorbind ea corespunde mărimii *two-sided PSD*. Diferența este ilustrată în figura 5.2 în care sunt prezentate cele trei puteri introduse. Se pot remarca ariile egale cuprinse sub cele 3 curbe care reprezintă pentru același semnal f cele 3 mărimi: (a) puterea totală, (b) *one-sided PSD*, și (c) *two-sided PSD*.

În general, dacă funcția $f(x)$ este nenulă pe toată axa reală, atunci puterea sa totală ca și densitatea spectrală a puterii sale PSD devin infinite. Se definește atunci *densitatea spectrală a puterii (one-sided sau two-sided) pe unitatea de timp* pentru un interval oricât de lung dar finit,

prin calculul PSD corespunzătoare unei funcții care are forma lui $f(x)$ pe intervalul considerat, dar care se anulează în exteriorul acestui interval, după care rezultatul se împarte la valoarea intervalului. Teorema lui Parseval în acest caz afirmă că integrala densității spectrale de putere pe unitatea de timp (*one-sided* PSD) pe domeniul frecvențelor pozitive este egală cu pătratul modulului amplitudinii semnalului $f(x)$. Un alt rezultat util care se referă la PSD (*one-sided*) pe unitatea de timp îl constituie convergența sa la valori finite pentru toate frecvențele cu excepția celor pentru care $f(x)$ are componente discrete sinusoidale (sau cosinusoidale) de amplitudine finită.

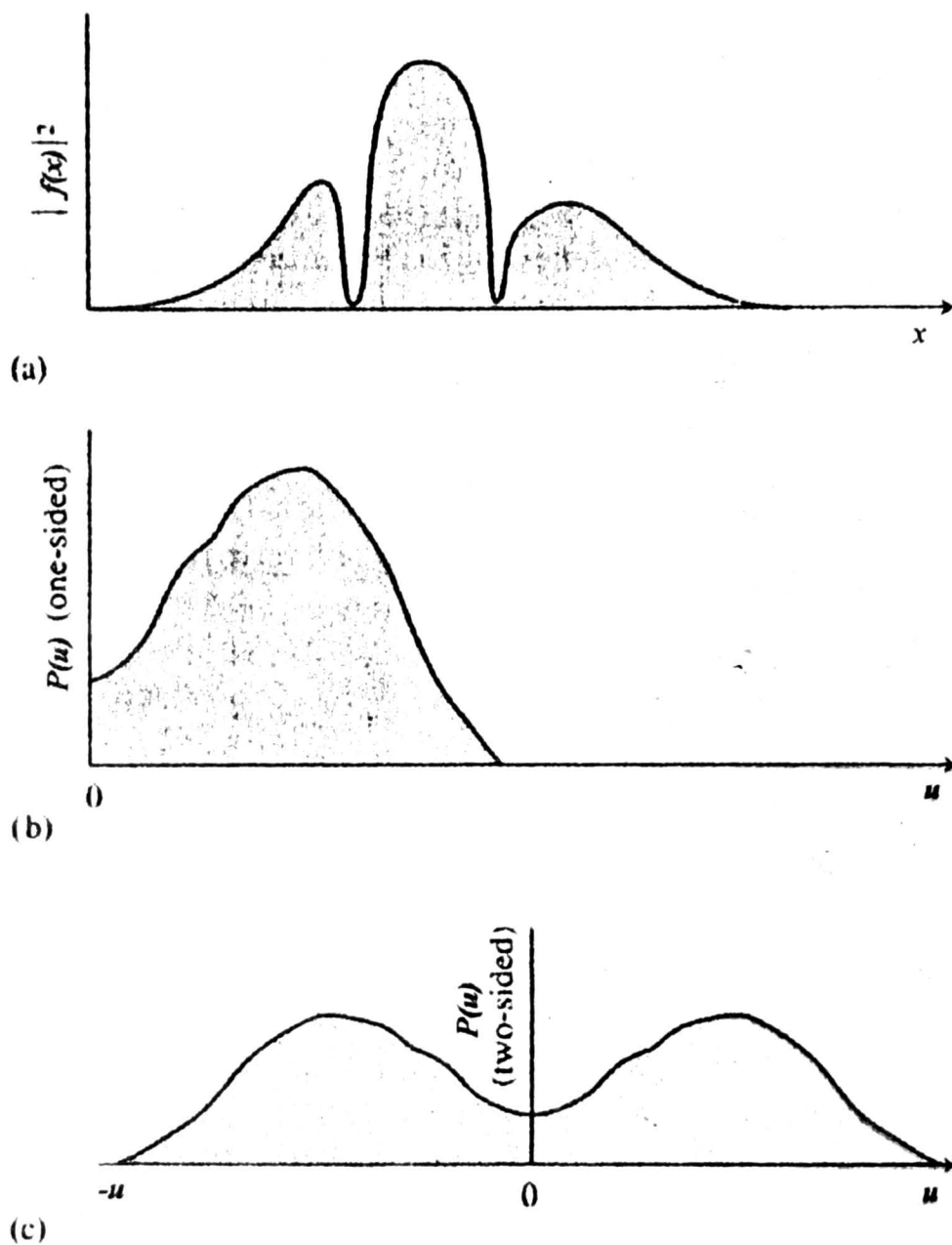


Figura 5.2

Normalizarea celor două puteri spectrale;

- (a) *Aria subgraficului modulului pătrat al*
- (b) *(b) Aria subgraficului funcției one-sided PSD la frecvențe pozitive,*
- (c) *(c) Aria subgraficului funcției two-sided PSD la frecvențe negative și pozitive. Toate cele trei arii au aceeași valoare. Corespunzător acestor frecvențe, PSD tinde către o funcție delta, adică un puls ascuțit a cărui lățime se îngustează progresiv, dar a cărui*

arie tinde către valoarea pătratică medie a amplitudinii componentei sinusoidale corespunzătoare acelei frecvențe.

Foarte multe din subiectele discutate în acest capitol sunt ilustrate de imagini ale funcțiilor și spectrele corespunzătoare Fourier afișate ca funcții de intensitate. Dinamica spectrelor Fourier este în mod uzual mult mai mare decât abilitatea monitoarelor tipice de a le reproduce cu finețe și precizie, fapt care se reflectă prin afișarea cu preponderență numai a regiunilor strălucitoare din imagini. Un mod de a compensa această incompatibilitate este de a afișa funcția

$$D(u, v) = c \cdot \log[1 + |F(u, v)|] \quad (5.9)$$

în locul spectrului $|F(u, v)|$, unde c este o constantă de scalare iar funcția logaritmică realizează compresia dorită.

5.1.1.2. Teorema de eșantionare

În cele mai multe dintre cazurile practice, în mod special la prelucrarea datelor experimentale, nu se operează direct cu funcții continue. De regulă $f(x)$ este eșantionată (adică sunt înregistrate valorile sale) la intervale (spațiale sau temporale) echidistante. Să notăm valoarea comună a intervalelor diviziunii cu Δx , atunci secvența valorilor eșantionate este $f_k = f(k \cdot \Delta x)$, $k = \dots - 3, -2, -1, 0, 1, 2, 3, \dots$. Inversul valorii intervalului de eșantionare $(\Delta x)^{-1}$ se numește *rata eșantionării*. Dacă Δx se măsoară în secunde (corespunzător domeniului temporal) atunci rata de eșantionare reprezintă numărul de eșantioane înregistrate într-o secundă și se măsoară în *herz*, dacă Δx se măsoară în metri

(corespunzător domeniului spațial) atunci rata de eșantionare reprezintă numărul de eșantioane înregistrate pe unitatea de lungime și se măsoară în $(metru)^{-1}$.

Pentru orice interval de eșantionare (diviziune) Δx există o frecvență specială denumită *frecvența critică Nyquist* dată de expresia

$$u_c \equiv \frac{1}{\Delta x} \quad (5.10)$$

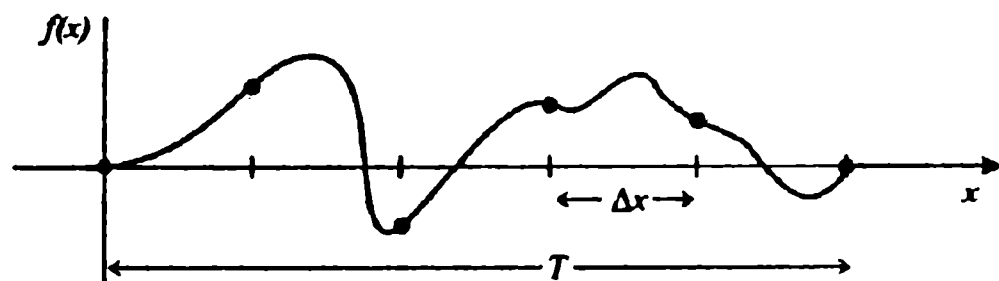
Această relație exprimă faptul că eșantionarea critică a unei unde sinusoidale corespunde la două puncte de eșantionare per perioadă. Adesea este preferat ca unitate de timp intervalul de eșantionare Δx . În acest caz frecvenței critice Nyquist îi corespunde constanta $1/2$.

Semnificația și importanța frecvenței Nyquist sunt relevante din două puncte de vedere distincte. Primul îl constituie *teorema de eșantionare*: dacă o funcție continuă $f(x)$ eșantionată cu diviziunea Δx este cu bandă limitată (*bandwidth limited*) la frecvențe mai mici decât u_c (în sensul că transformata sa Fourier $F(u) = 0$ pentru $|u| \geq u_c$), atunci funcția $f(x)$ este complet determinată de eșantioanele sale f_k și subzistă relația

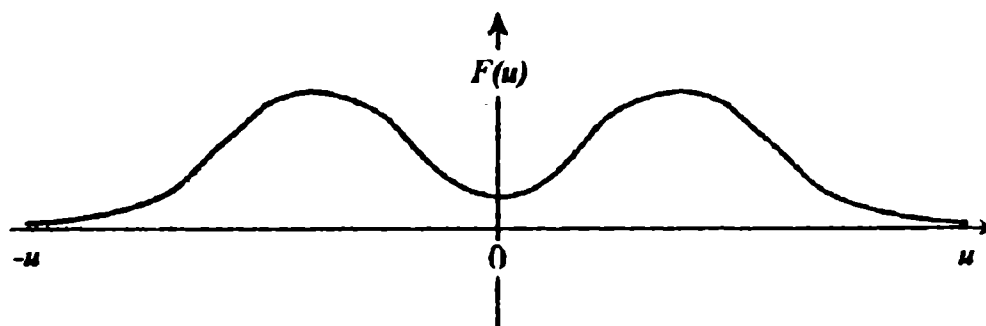
$$f(x) = \Delta x \sum_{k=-\infty}^{k=+\infty} f_k \cdot \frac{\sin [2\pi u_c (x - k\Delta x)]}{\pi(x - k\Delta x)} \quad (5.11)$$

Spre exemplu, dacă un semnal este prelucrat de un amplificator a cărui bandă de trecere în frecvență (finită) este cunoscută, atunci întreaga informație conținută de semnal poate fi înregistrată prin eșantionarea

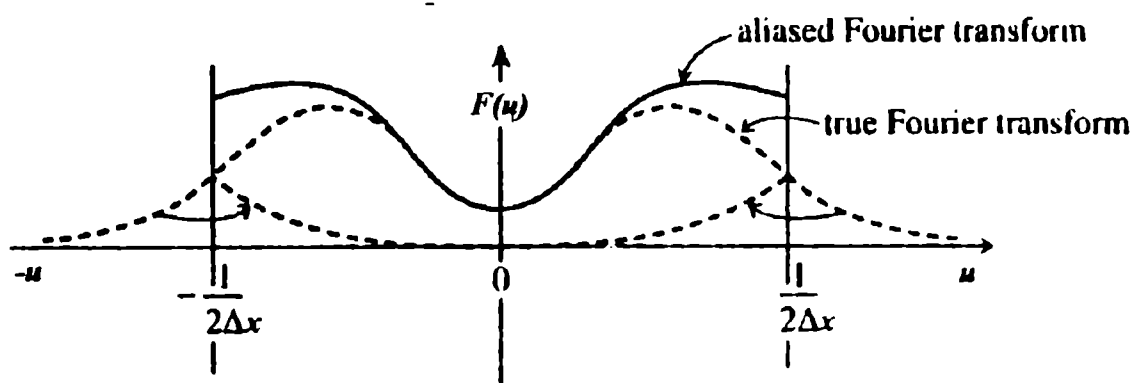
semnalului cu o rată egală cu dublul frecvenței maxime a amplificatorului.



(a)



(b)



(c)

Figura 5.3

Funcția continuă prezentată în (a) este diferită de zero pentru un interval finit T . Transformata sa Fourier, prezentată în (b), nu este limitată dar are amplitudinea finită pentru orice frecvență. Dacă funcția originală este eșantionată la intervale Δx , atunci transformata sa Fourier este definită pe intervalul mărginit $\left(-\frac{1}{2\Delta x}, \frac{1}{2\Delta x}\right)$. Puterea semnalului din exteriorul acestui interval este suprapusă peste cea corespunzătoare acestui interval. Efectul de aliasing poate fi eliminat prin filtrarea trece-jos a semnalului înainte de eșantionare.

Al doilea punct de vedere privește efectul eșantionării asupra unei funcții continue care nu este limitată în frecvență la domeniul $(-u_c, u_c)$. În acest caz densitatea spectrală a puterii care se află în exteriorul acestui interval de frecvență este deplasat în interiorul său, fenomen care poartă denumirea de *aliasing* (falsă translatare). Orice componentă în frecvență din exteriorul intervalului $(-u_c, u_c)$ este deplasată eronat în acest interval ca efect al procesului de eșantionare discretă. Două unde $\exp(j2\pi u_1 x)$ și $\exp(j2\pi u_2 x)$ vor produce aceleași eșantioane la un interval Δx dacă și numai dacă u_1 și u_2 diferă cu un multiplu de $1/\Delta x$, valoare care este tocmai lungimea intervalului de frecvență $(-\hat{u}_c, \hat{u}_c)$. Fenomenul de suprapunere poate fi evitat dacă (i) se cunosc limitele naturale în frecvență ale semnalului (sau se forțează limite cunoscute prin filtrare analogă a semnalului) și (ii) se eșantionează semnalul suficient de fin astfel încât să se înregistreze cel puțin două valori per perioada frecvenței celei mai ridicate prezente în semnal (figura 5.3).

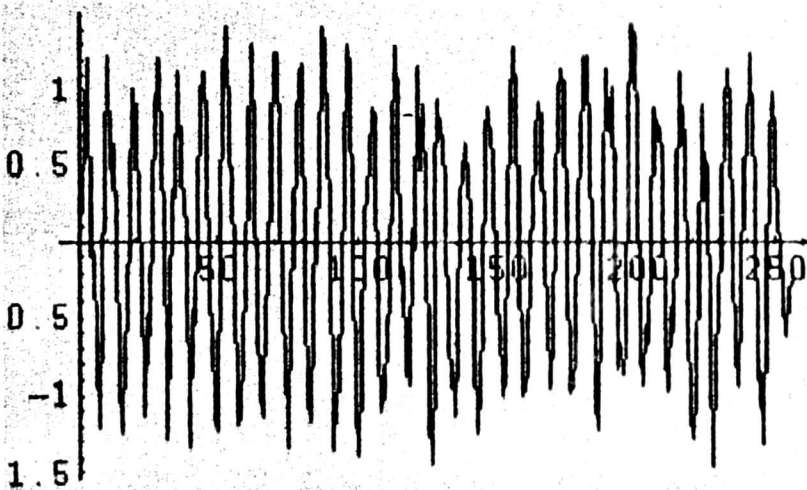
Se poate adopta un punct de vedere alternativ la cele de mai sus. Dacă o funcție continuă a fost eșantionată adecvat, atunci, la estimarea transformatei sale Fourier pe baza valorilor eșantioanelor, putem presupune că transformata Fourier se anulează în afara intervalului $(-u_c, u_c)$. Se poate admite că eșantionarea a fost adecvată dacă transformata Fourier a semnalului se apropie de zero atunci când frecvența tinde crescător către u_c și descrescător către $-u_c$. Dacă dimpotrivă, transformata sa Fourier tinde către valori finite, atunci este foarte probabil ca eșantionarea să fi fost făcută inadecvat iar

componentele din exteriorul intervalului $(-u_c, u_c)$ să fi fost deplasate eronat în interiorul acestui interval critic.

Exemplu

Codul următor scris în MATHEMATICA generează 256 de eșantioane ale unui semnal periodic peste care este suprapus aleatoriu zgomot. Reprezentarea grafică directă a semnalului întărește impresia de caracter aleatoriu al datelor

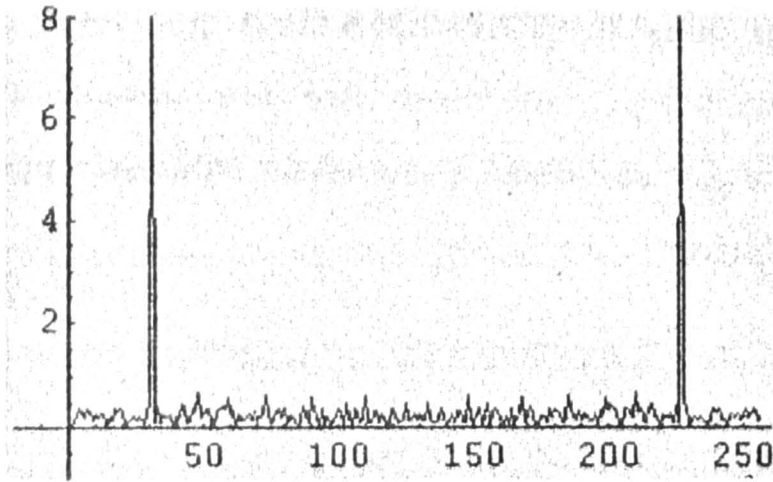
```
data=Table[N[Sin[30 2\[Pi] n/256]+(Random[]-1/2)], {n, 256}];  
ListPlot[data, PlotJoined->True]
```



Transformata sa Fourier arată însă un maxim puternic la 30, care este frecvența semnalului original.

```
ListPlot[Abs[Fourier[data]], PlotJoined->True,  
PlotRange->All]
```

Se poate clar remarca un al doilea vârf (*peak*) al reprezentării spectrale a semnalului, maxim care este situat la $256 - 30$ și care este consecința fenomenului de suprapunere (*aliasing*)



5.1.1.3. Transformata Fourier discretă

Să estimăm acum transformata Fourier a unei funcții $f(x)$ pornind de la un număr finit de valori eșantionate. Vom presupune că avem N valori eșantionate consecutiv la intervale echidistante Δx astfel încât se obține secvența de valori (figura 5.4)

$$\{f(x_0), f(x_0 + \Delta x), f(x_0 + 2\Delta x), \dots, f(x_0 + [N - 1]\Delta x)\} \quad (5.12)$$

Va fi convenabil în continuare să utilizăm variabila x atât ca variabilă continuă cât și ca variabilă discretă, depinzând de contextul în discuție. În acest sens vom utiliza notația

$$f(x) = f(x_0 + x \cdot \Delta x)$$

unde variabila x se presupune că ia valorile discrete $0, 1, 2, \dots, N - 1$. Cu alte cuvinte secvența

$$\{f(0), f(1), f(2), \dots, f(N - 1)\}$$

reprezintă *oricare* N eșantioane echidistante corespunzătoare funcției continue $f(x)$ (figura 5.4).

Din considerente de simplitate a expunerii vom presupune că N este par. Dacă funcția $f(x)$ ia valori diferite de zero numai pe un interval finit, atunci acest întreg interval îl vom presupune acoperit de cele N puncte ale diviziunii de eșantionare. Dacă $f(x)$ nu satisface această condiție, atunci vom presupune cel puțin că valorile eșantionate sunt într-un anume sens “tipice” pentru modul în care așază funcția $f(x)$ pe întreg domeniul său de definiție.

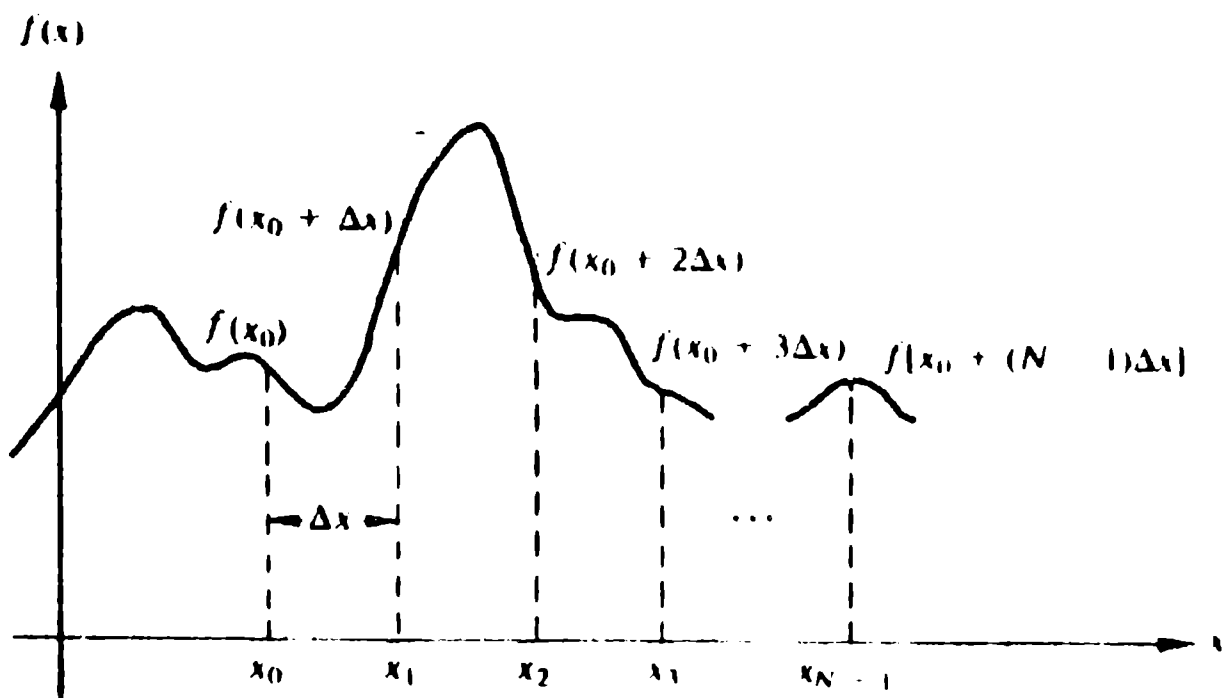


Figura 5.4

Eșantionarea unei funcții continue oarecare la intervale Δx echidistante.

Deoarece avem un număr finit de eșantioane, în loc să estimăm transformata Fourier $F(u)$ pe intervalul $(-u_c, u_c)$, vom evalua valorile sale corespunzătoare valorilor discrete

$$u_n \equiv \frac{n}{N\Delta x}, \quad n = -\frac{N}{2}, \dots, \frac{N}{2} \quad (5.13)$$

Aceste puncte sunt în număr de $N + 1$ deoarece cele două extreme care corespund frecvenței critice Nyquist nu sunt independente (de fapt sunt egale), astfel încât rămân exact N valori independente.

Să aproximăm acum valoarea integralei care definește transformata Fourier

$$F(u_n) = \int_{-\infty}^{+\infty} f(x) \cdot e^{j2\pi u_n x} dx \approx \sum_{x=0}^{N-1} f(x) \cdot e^{j2\pi u_n x} \Delta x = \Delta x \sum_{x=0}^{N-1} f(x) \cdot e^{j2\pi x n/N} \quad (5.14)$$

Să notăm ultima sumă numită și transformata Fourier discretă a celor N puncte ale lui $f(x)$ prin

$$F_n \equiv \sum_{x=0}^{N-1} f(x) \cdot e^{j2\pi x n/N} \quad (5.15)$$

Transformata Fourier discretă pune în corespondență N valori complexe $f(x)$ din domeniul spațial cu N valori complexe din domeniul frecvenței F_n , și nu depinde de parametri dimensionali, cum ar fi intervalul de eșantionare. Relația (5.14) între transformata Fourier discretă a unui șir de valori numerice și transformata Fourier corespunzătoare interpretării acestor valori ca eșantioane ale unei funcții continue eșantionată cu intervalul Δx poate fi rescrisă sub forma

$$F(u_n) \approx \Delta x \cdot F'_n \quad (5.16)$$

unde u_n este dat de (5.13). Vom utiliza în continuare aceeași convenție de scriere notând cu u atât frecvența ca variabilă continuă cât și discretă,

$$\text{astfel încât } F(u) = F(u_n) = F\left(\frac{n}{N\Delta x}\right), \quad n = -\frac{N}{2}, \dots, \frac{N}{2}.$$

O funcție simplă și eșantioanele spațiale corespunzătoare sunt ilustrate demonstrativ în figura 5.5.

Spre deosebire de cazul continuu, transformata Fourier discretă există întotdeauna, fapt ce poate fi simplu demonstrat prin substituția expresiei funcției original $f(x)$ în definiția transformatei Fourier discrete

$F(u)$ și rezultă

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} \left[\sum_{r=0}^{N-1} F(r) \cdot e^{j2\pi r x/N} \right] \cdot e^{-j2\pi u x/N} =$$

$$\frac{1}{N} \sum_{r=0}^{N-1} F(r) \cdot \left[\sum_{x=0}^{N-1} e^{j2\pi r x/N} \cdot e^{-j2\pi u x/N} \right] = F(u)$$

Ultima egalitate este adevărată în virtutea condiției de ortogonalitate

$$\sum_{x=0}^{N-1} e^{j2\pi r x/N} \cdot e^{-j2\pi u x/N} = \begin{cases} N & \text{dacă } r = u \\ 0 & \text{dacă } r \neq u \end{cases}$$

Se poate ușor arăta că F'_n în (5.15) este periodică în n cu perioada N , astfel încât $F'_{-n} = F'_{N-n}$, $n = 1, 2, \dots$. Transformata discretă Fourier are proprietăți de simetrie similare cu transformata Fourier continuă. Formula transformatei Fourier inverse care dă setul funcțiilor f_k pornind de la valorile F'_n este

$$f_k = \frac{1}{N} \sum_{n=0}^{N-1} F_n \cdot e^{-j2\pi kn/N}, \quad k = 0, 1, \dots, N-1 \quad (5.17)$$

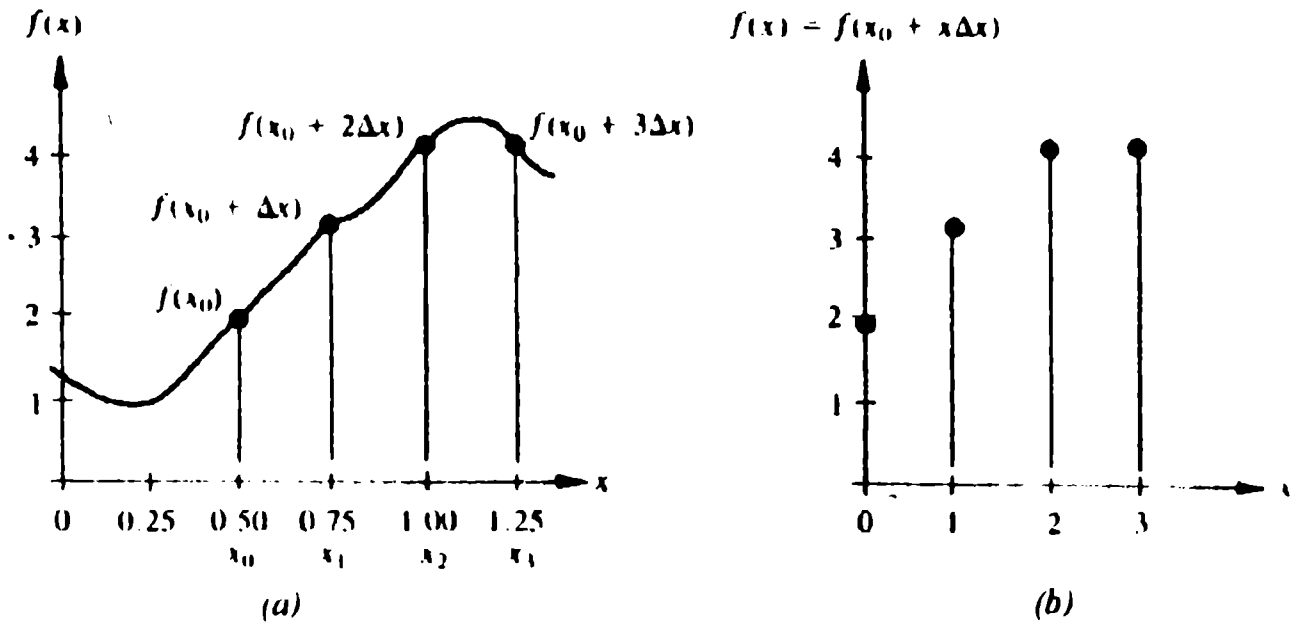


Figura 5.5

Eșantionarea unei funcții continue simple. În (a) variabila x este continuă, iar în (b) variabila x este discretă și $x_0 = 0$.

Diferențele minime (semnul exponențialei și factorul $1/N$ în transformata inversă) între expresiile care dau perechile de transformate (5.11) și (5.12), au drept consecință modificări minime ale rutinei care poate calcula atât transformata Fourier directă cât și cea inversă.

Forma discretă a teoremei lui Parseval în notațiile introduse mai sus este

$$\sum_{k=0}^{N-1} |f_k|^2 = \frac{1}{N} \sum_{n=0}^{N-1} |F_n|^2 \quad (5.18)$$

Exemplu

Transformata Fourier există și se poate determina și în cazul în care numărul datelor nu este o putere a lui 2. În orice caz însă, eficiența

compuțatională este indiscutabilă în cazul în care numărul datelor nu este o putere a lui 2. Codul următor scris în **MATHEMATICA** calculează transformata Fourier directă a trei valori numerice, după care sunt regăsite datele de intrare originale prin aplicarea transformatei Fourier inversă și trunchiere.

```
In[1]:=Fourier[{1, -1, 1}]
In[2]:=InverseFourier[%]
In[3]:=Chop[%]
```

```
Out[1]={0.57735 +0. I, 0.57735 - 1. I,
0.57735 +1. I}
Out[2]={1. +0. I, -1. -1.60247 10-16 I, 1. -
1.28198 10-16 I}
Out[3]={1., -1., 1.}
```

5.1.1.4. Convoluția

Produsul de convoluție a două funcții continue și integrabile $f(x)$ și $g(x)$ se definește pe domeniul spațial (temporal) și este o funcție de x

$$f(x) * g(x) = \int_{-\infty}^{+\infty} f(\alpha) \cdot g(x - \alpha) \cdot d\alpha \quad (5.19)$$

Se remarcă imediat că produsul de convoluție este simetric $f * g = g * f$, așadar funcția $f * g$ este un element al perechii de transformate

$$f(x) * g(x) \Leftrightarrow F(u) \cdot G(u) \quad (5.20')$$

Relația de mai sus stabilește că transformata Fourier a produsului de convoluție este dată de produsul transformatelor Fourier ale celor două

funcții. Un rezultat analog constă în faptul că și convoluția în domeniul frecvenței se reduce la multiplicare în domeniul spațial

$$f(x) \cdot g(x) \Leftrightarrow F(u) * G(u) \quad (5.20'')$$

Aceste două relații (5.20') și (5.20'') sunt cunoscute sub numele de *teorema de convoluție*.

Deoarece semnificația integralei de convoluție nu este transparentă direct, vom începe discuția cu două exemple care ilustrează grafic utilizarea ecuației (5.19).

Exemplu

Să considerăm funcțiile $f(x)$ și $g(x)$ reprezentate grafic în figurile 5.6a și respectiv 5.6b. Evaluarea integralei de definiție a convoluției (5.19) presupune anterior formarea funcției $g(x-\alpha)$. Aceasta înseamnă o operație de simetrie în raport cu originea pentru a obține $g(-\alpha)$ prezentată în figura 5.6c, urmată de o deplasare pe distanța x care conduce la evaluarea $g(x-\alpha)$ ilustrată în figura 5.6d. Apoi, pentru fiecare valoare a lui x se înmulțește $f(\alpha)$ cu $g(x-\alpha)$ și se integrează produsul după α de la $-\infty$ la $+\infty$. Produsul $f(\alpha) \cdot g(x-\alpha)$ corespunde regiunii hașurate din figura 5.6e, care constituie reprezentarea corespunzătoare intervalului $0 \leq x \leq 1$. Acest produs este nul pentru valori ale lui α în afara intervalului $[0, x]$, astfel încât $f * g = x/2$. Pentru x aparținând intervalului $[1, 2]$ rezultă $f * g = (1 - x/2)$ care este ilustrat în figura 5.6f. Pentru restul valorilor lui x produsul $f(\alpha) \cdot g(x-\alpha)$ se anulează, deci în afara intervalului $[0, 2]$ avem $f * g = 0$ și putem scrie în final expresia completă a produsului

$$f * g = \begin{cases} x/2 & 0 \leq x \leq 1 \\ 1 - x/2 & 1 \leq x \leq 2 \\ 0 & \text{in rest} \end{cases}$$

Figura 5.6g prezintă grafic acest rezultat

Un caz de interes special în procesarea semnalelor îl constituie convoluția unei funcții $f(x)$ cu funcția impuls $\delta(x - x_0)$ definită de relația

$$\int_{-\infty}^{\infty} f(x) \cdot \delta(x - x_0) \cdot dx = f(x_0) \quad (5.21)$$

Funcția $\delta(x - x_0)$ poate fi privită ca având aria unitară într-o vecinătate infinitezimală a lui x_0 și anulându-se în rest, adică

$$\int_{-\infty}^{\infty} \delta(x - x_0) \cdot dx = \int_{x_0}^{x_0^+} \delta(x - x_0) \cdot dx = 1 \quad (5.22)$$

În cele mai multe cazuri este rezonabil să considerăm că $\delta(x - x_0)$ este localizată în $x = x_0$ și că “intensitatea” impulsului este determinată de valoarea funcției $f(x)$ în punctul $x = x_0$. De exemplu, dacă $f(x) = A$, atunci $A \cdot \delta(x - x_0)$ este un impuls de intensitate A localizat la $x = x_0$. Convenția grafică uzuală este de a reprezenta impulsul de o săgeată situată la $x = x_0$, având înălțimea egală cu intensitatea impulsului. Figura 5.7 ilustrează această convenție pentru impulsul $A \cdot \delta(x - x_0)$.

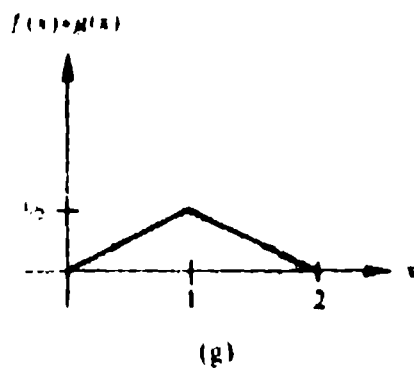
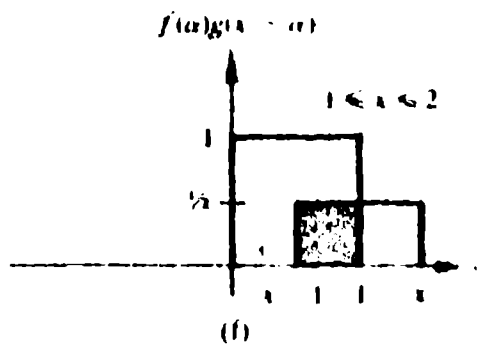
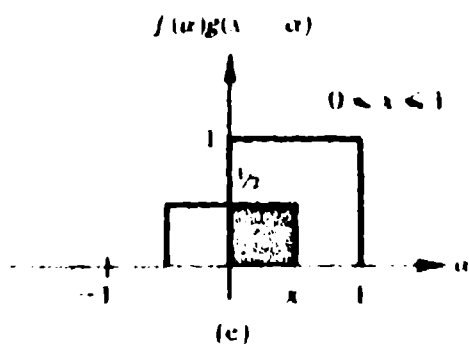
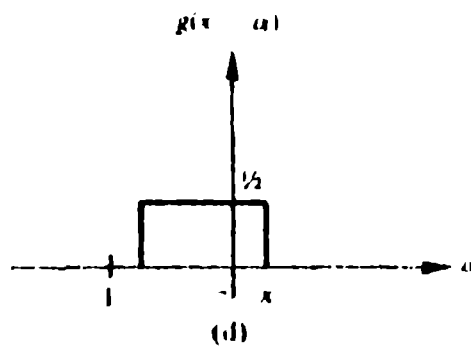
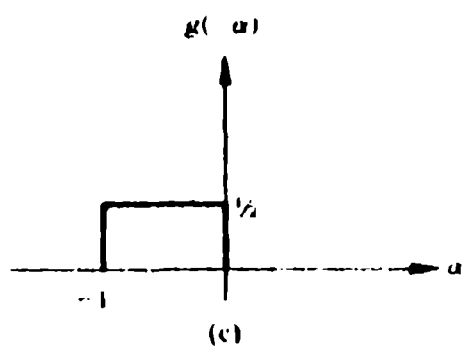
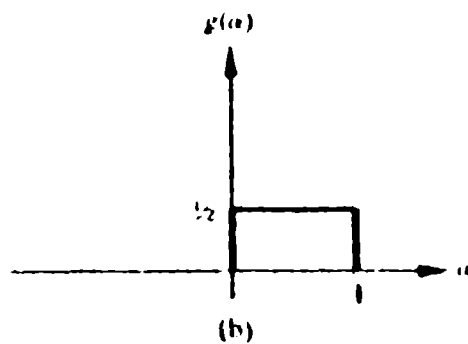
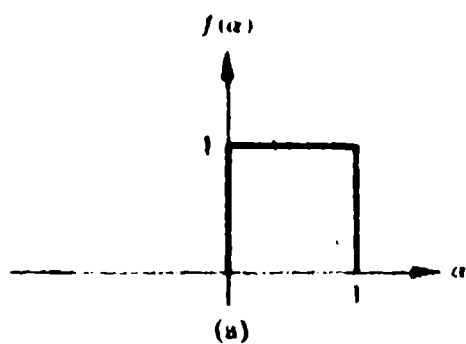


Figura 5.6

Ilustrarea grafică a convoluției. Ariile hașurate corespund regiunilor în care produsul $f(a) \cdot g(x-a)$ este nenul.

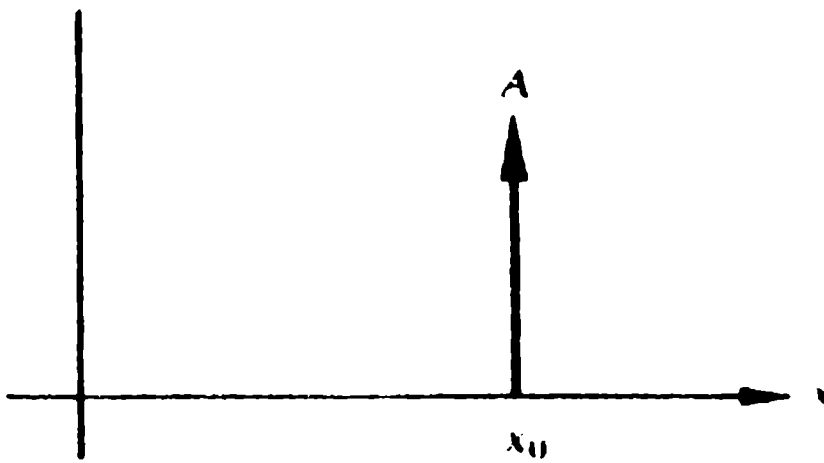


Figura 5.7
 Reprezentarea grafică a
 semnalului $A \cdot \delta(x - x_0)$

Exemplu

În continuarea ilustrării semnificației integralei de convoluție (5.19), să considerăm produsul de convoluție al funcției $f(x)$ definită ca și mai înainte (figura 5.8a) cu funcția impuls $g(x) = \delta(x + T) + \delta(x) + \delta(x - T)$ (figura 5.8b). Procedând în mod similar exemplului anterior se obține rezultatul din figura 5.8c, care în esență prezintă convoluția ca o “copiere” a funcției $f(x)$ în pozițiile corespunzătoare fiecărui impuls din $g(x)$.

Să presupunem acum că funcțiile $f(x)$ și $g(x)$, în loc să fie continue, reprezintă eșantioanele pe ariile de dimensiune A , respectiv B , adică $\{f(0), f(1), \dots, f(A-1)\}$ și respectiv $\{g(0), g(1), \dots, g(B-1)\}$. Atunci, așa după cum s-a specificat anterior, transformata Fourier discretă și inversa sa sunt funcții periodice. În consecință, formularea unei teoreme de convoluție în cazul discret consistentă cu periodicitatea implică prezumția că funcțiile discrete $f(x)$ și $g(x)$ sunt periodice cu o perioadă oarecare M . Atunci produsul de convoluție este de asemenea periodic cu

aceeași perioadă M . Se poate arăta [Brigham 1974] că atâta timp cât se alege M astfel încât

$$M \leq A + B - 1 \quad (5.23)$$

perioadele individuale ale convoluției se vor suprapune, fenomen cunoscut sub numele de *eroare de suprapunere (wraparound error)*.

Dacă $M = A + B - 1$ perioadele sunt adiacente, iar dacă $M > A + B - 1$ perioadele vor fi separate de un interval egal cu diferența dintre M și $A + B - 1$. Deoarece perioada presupusă este mai mare atât decât A cât și B , lungimea secvențelor de eșantionare trebuie extinsă până la lungimea M . Secvențele extinse se obțin prin bordare cu 0 astfel încât

$$f_e(x) = \begin{cases} f(x) & 0 \leq x \leq A - 1 \\ 0 & A \leq x \leq M - 1 \end{cases} \quad (5.24')$$

și în mod similar

$$g_e(x) = \begin{cases} g(x) & 0 \leq x \leq B - 1 \\ 0 & B \leq x \leq M - 1 \end{cases} \quad (5.24'')$$

Convoluția discretă a funcțiilor $f(x)$ și $g(x)$ bazată pe extensiile $f_e(x)$ și $g_e(x)$ este atunci definită de expresia

$$f_e(x) * g_e(x) = \frac{1}{M} \sum_{m=0}^{M-1} f_e(m) \cdot g_e(x - m), \quad x = 0, 1, \dots, M - 1 \quad (5.25)$$

Funcția de convoluție este discretă, periodică cu perioada M , astfel încât valorile $x = 0, 1, \dots, M - 1$ descriu o perioadă completă a funcției $f_e(x) * g_e(x)$.

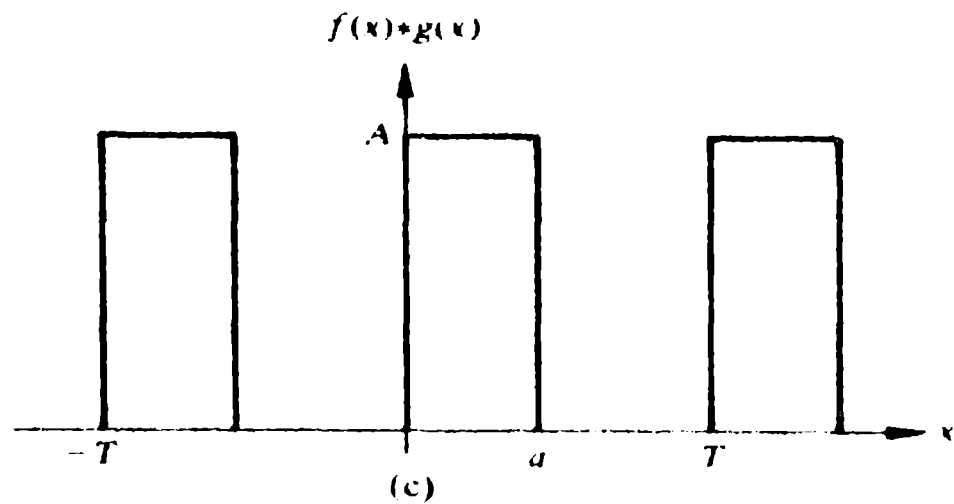
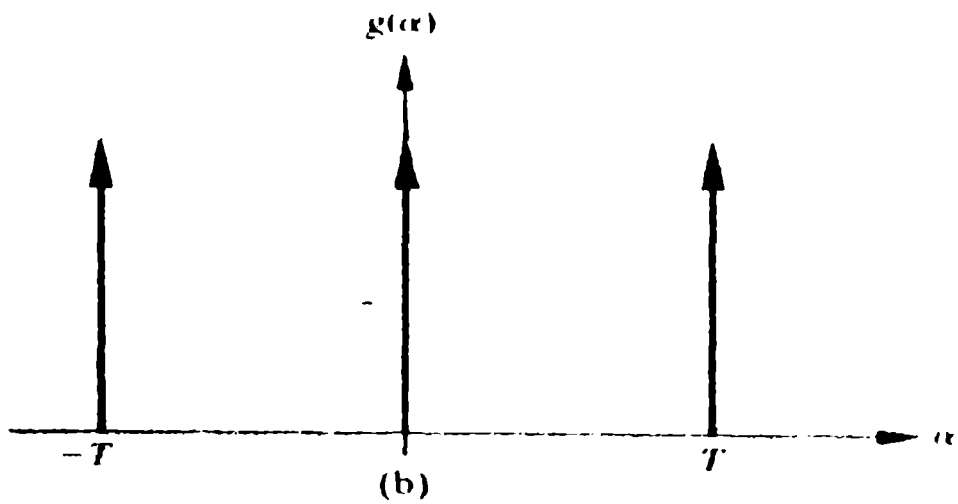
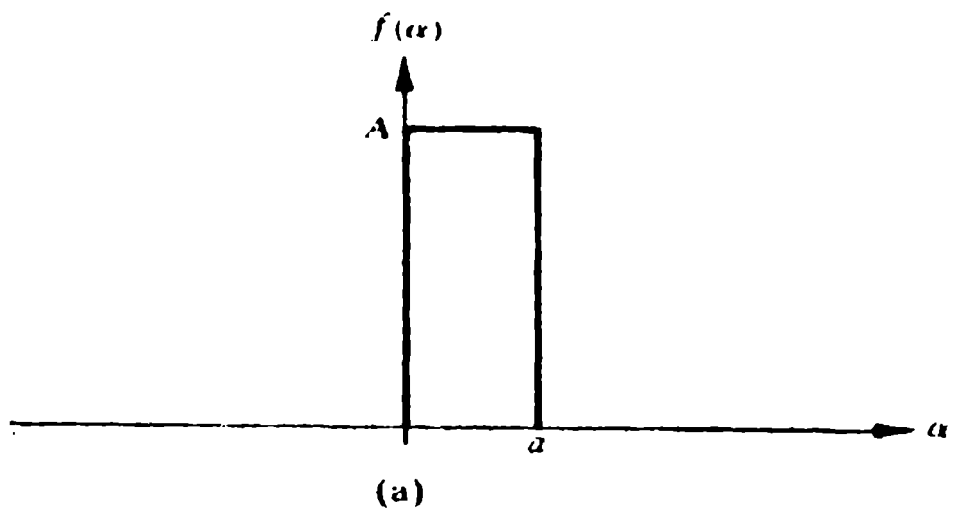


Figura 5.8
Convoluția funcției $f(x)$ definită în (a) cu funcția $g(x)$ de tip impuls (b) și rezultatul grafic al convoluției (c).

5.1.1.5. Corelația

Produsul de corelație a două funcții continue și integrabile $f(x)$ și $g(x)$ se definește de asemenea pe domeniul spațial (temporal) și este o funcție de x

$$f(x) \circ g(x) = \int_{-\infty}^{+\infty} f^*(\alpha) \cdot g(x + \alpha) \cdot d\alpha \quad (5.26)$$

unde simbolul $*$ desemnează complexul conjugat. Convoluția definită de ecuația (5.19) și corelația dată de ecuația (5.26) sunt relativ similare, cu deosebirea că în cazul corelației funcția $g(x)$ nu se înfășoară în jurul originii. Astfel, pentru a realiza produsul de corelație a două funcții $f(x)$ și $g(x)$ pur și simplu se deplasează $f(x)$ peste $g(x)$ și se integrează de la $-\infty$ la $+\infty$ pentru toate valorile deplasării x . Modul de evaluare a corelației este prezentat în figura 5.9.

Sub formă discretă, expresia echivalentă ecuației (5.26) se definește astfel

$$f_e(x) \circ g_e(x) = \frac{1}{M} \sum_{m=0}^{M-1} f_e^*(m) \cdot g_e(x + m), \quad x = 0, 1, \dots, M-1 \quad (5.27)$$

Considerațiile dezvoltate anterior privind $f_e(x)$ și $g_e(x)$, presupunerea privind periodicitatea lor ca și alegerea lui M rămân adevărate și în cazul corelației discrete (5.27).

Ca și la convoluție este adevărată o teoremă similară și la corelație, atât în cazul continuu cât și pentru cazul discret, care atestă apartenența produsului $f \circ g$ la perechea de transformate

$$f(x) \circ g(x) \Leftrightarrow F(-u) \cdot G(u) \quad (5.28)$$

Relația de mai sus este cunoscută sub numele de *teorema de corelație*.

Una din aplicațiile principale ale corelației în procesarea de semnale este în domeniul identificării de prototipuri, care constă în determinarea celei mai asemănătoare configurații (formă) dintr-un set de configurații cunoscute atunci când este prezentată o configurație necunoscută. Un mod de abordare al acestei probleme constă în calculul integralelor de corelație dintre configurația necunoscută și toate configurațiile din baza de date (cunoscute). Cea mai asemănătoare configurație este selectată corespunzător valorii maxime a integralei de corelație, așadar corelației de amplitudine maximă.

Ca și în cazul convoluției discrete, de multe ori calculul produsului de corelație discretă $f_e(x) \circ g_e(x)$ este mai comod de efectuat în domeniul frecvenței utilizând transformatele Fourier corespunzătoare calculate pe baza unor algoritmi rapizi de tipul FFT (*Fast Fourier Transform*).

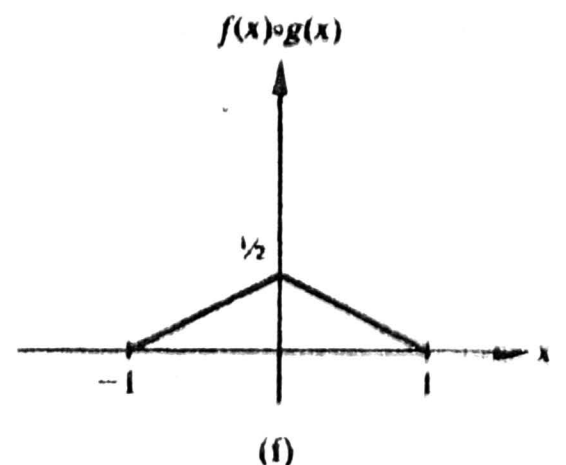
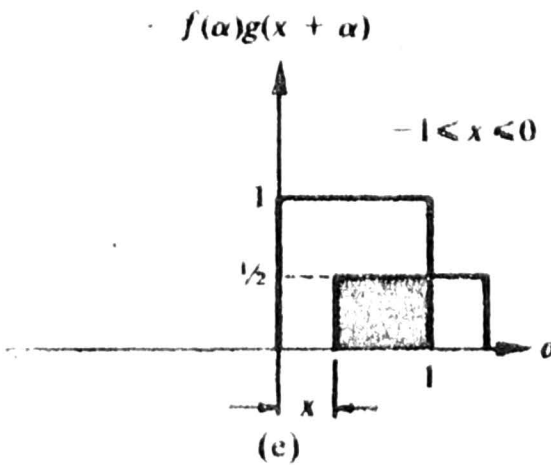
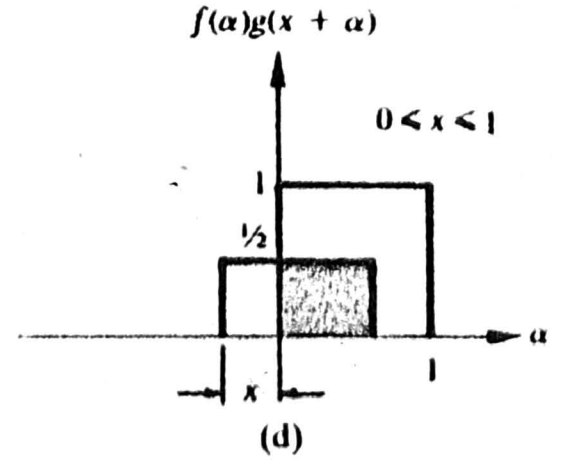
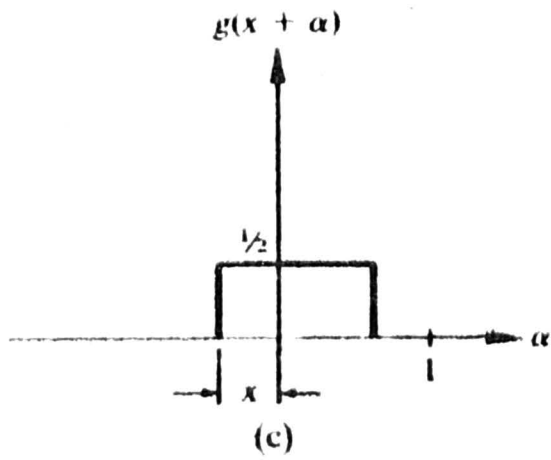
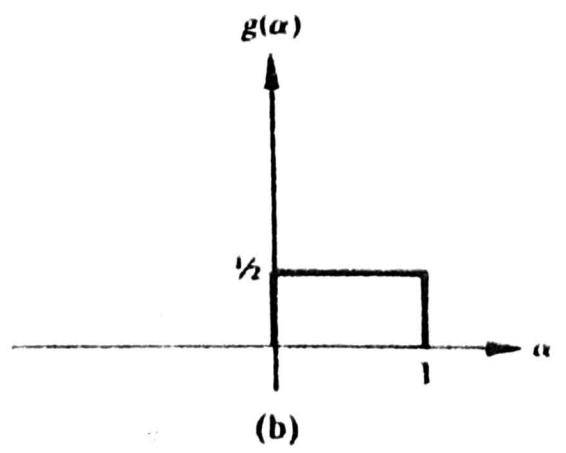
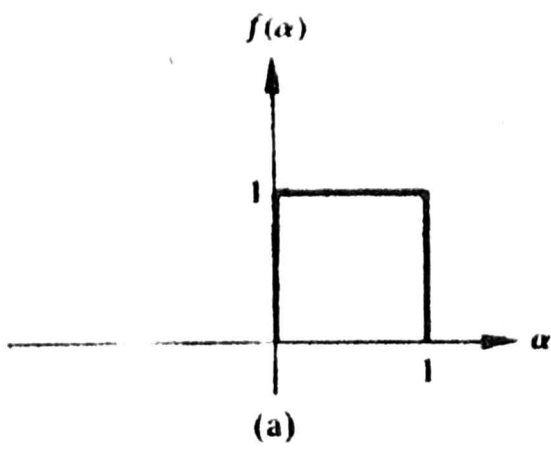


Figura 5.9

Ilustrarea grafică a corelației. Suprafețele hașurate corespund regiunilor în care produsul de corelație este nemul.

5.1.1.6. Eșantionarea funcțiilor continue unidimensionale

Transformatele Fourier și produsul de convoluție ne permit o abordare analitică mai profundă. În particular, ne interesează condițiile de eșantionare în care o imagine continuă poate fi complet refăcută dintr-un set de eșantioane discrete.

Să considerăm o funcție $f(x)$ continuă oarecare (figura 5.10a) care se extinde de la $-\infty$ la $+\infty$. Să presupunem că transformata sa Fourier $F(u)$ este nenulă numai în interiorul intervalului $[-W, W]$, atunci $F(u)$ ar putea avea forma din figura 5.10b. O funcție care are această proprietate pentru orice valoare finită a lui W se numește *funcție cu bandă limitată (band-limited)*.

Versiunea eșantionată a funcției $f(x)$ implică multiplicarea acestei funcții cu o funcție de eșantionare $s(x)$, care constă dintr-un tren de impulsuri echidistante la intervale Δx (figura 5.10c). Teorema de convoluție stabilește echivalența dintre multiplicarea în domeniul x (spațial) și convoluția în domeniul u (al frecvenței). Transformata Fourier $F(u) * S(u)$ a produsului $s(x) \cdot f(x)$ este prezentată în figura 5.10f.

Transformata este periodică cu perioada $1/\Delta x$ și repetarea transformatei se poate face cu suprapunere. Spre exemplu, în prima perioadă centrul

regiunii suprapuse se află la $u = \frac{1}{2\Delta x}$ dacă $\frac{1}{2\Delta x} < W$. Pentru a evita

această problemă, se alege intervalul de eșantionare Δx astfel încât să

avem satisfăcută condiția $\frac{1}{2\Delta x} \geq W$, ceea ce revine la

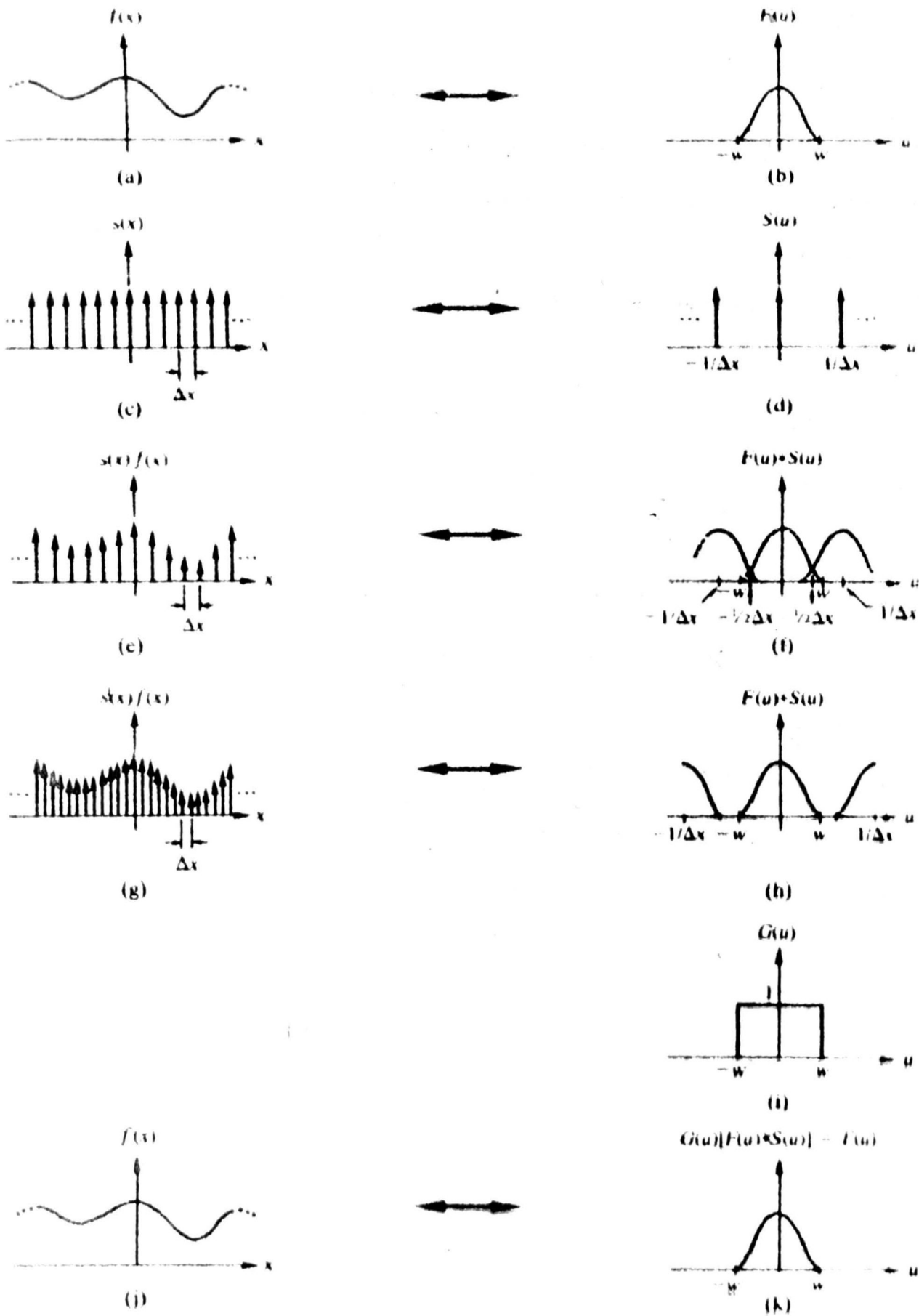


Figura 5.10 Dezvoltarea grafică a conceptului de eşantionare.

$$\Delta x \leq \frac{1}{2W} \quad (5.29)$$

Rezultatul descreșterii intervalului de eșantionare este ilustrat în figurile 5.10g și 5.10h și constă în separarea perioadelor transformatei $F(u)*S(u)$ care acum nu se mai suprapun. Importanța acestei alegeri rezidă în faptul că multiplicarea transformatei $F(u)*S(u)$ din figura 5.10h cu funcția (figura 5.10i)

$$G(u) = \begin{cases} 1 & -W \leq u \leq W \\ 0 & \text{in rest} \end{cases} \quad (5.30)$$

face posibilă izolarea completă a funcției $F(u)$ așa după cum se poate observa în figura 5.10k. Atunci transformata Fourier inversă a acestui produs $G(u) \cdot [F(u)*S(u)]$ conduce la funcția *continuă* original $f(x)$. Refacerea completă a unei funcții cu bandă limitată eșantionată conform expresiei (5.29) este cunoscută sub numele de *teorema de eșantionare Whittaker-Shannon* [13,14]. Toată informația din domeniul frecvenței a unei funcții cu bandă limitată este complet conținută în intervalul $[-W, W]$. Dacă relația (5.29) nu este satisfăcută, atunci transformata produsului $F(u)*S(u)$ în acest interval este afectată de contribuțiile perioadelor adiacente.

5.1.2. Transformata Fourier bidimensională (2D)

Dacă $f(x, y)$ este continuă și integrabilă iar $F(u, v)$ este integrabilă, atunci se definește transformata Fourier directă $\mathfrak{F}\{\}$ și

respectiv transformata Fourier inversă $\mathfrak{F}^{-1}\{\}$ prin expresiile integralele următoare:

$$\begin{aligned}\mathfrak{F}\{f(x, y)\} = F(u, v) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \cdot e^{-j2\pi(ux+vy)} \cdot dx \cdot dy \\ \mathfrak{F}^{-1}\{F(u, v)\} = f(x, y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(u, v) \cdot e^{j2\pi(ux+vy)} \cdot du \cdot dv\end{aligned}, \quad j = \sqrt{-1}$$

(5.31)

Evident că, așa cum s-a arătat în cazul $1D$, și transformata Fourier $2D$ este un *operator liniar* datorită integralei din definiție. Pentru situațiile în care se preferă *frecvențe unghiulare* în locul frecvențelor u și v , se introduc $\omega_1 = 2\pi u$ (*frecvență orizontală*) și $\omega_2 = 2\pi v$ (*frecvență verticală*) și rezultă perechea de relații

$$\begin{aligned}\mathfrak{F}\{f(x, y)\} = F(\omega_1, \omega_2) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \cdot e^{-j(\omega_1 x + \omega_2 y)} \cdot dx \cdot dy \\ \mathfrak{F}^{-1}\{F(\omega_1, \omega_2)\} = f(x, y) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(\omega_1, \omega_2) \cdot e^{j(\omega_1 x + \omega_2 y)} \cdot d\omega_1 \cdot d\omega_2\end{aligned}$$

(5.32)

Funcția $F(\omega_1, \omega_2)$ este periodică atât în ω_1 cât și ω_2 cu perioada 2π , astfel încât putem rescrie ultima ecuație sub forma

$$\mathfrak{F}^{-1}\{F(\omega_1, \omega_2)\} = f(x, y) = \frac{1}{2\pi} \int_{\omega_1=-\pi}^{\pi} \int_{\omega_2=-\pi}^{\pi} F(\omega_1, \omega_2) \cdot e^{j(\omega_1 x + \omega_2 y)} \cdot d\omega_1 \cdot d\omega_2$$

(5.33)

BIBLIOGRAFIE

1. R. Mutihac, *Introducere în procesarea de imagini*, Editura Universității din București, București, 1999.
2. T.W. Parks și C.S. Burrus. *Digital Filter Design*. New York: John Wiley & Sons, 206-209, 1987.
3. IEEE. *Programs for Digital Signal Processing*. IEEE Press. New York: John Wiley & Sons, 1979. Algorithm 8.1.
4. E.C. Levi, "Complex-Curve Fitting." *IRE Trans. on Automatic Control*. Vol. AC-4, 37-44, 1959.
5. J.E. Denis, Jr., și R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ: Prentice Hall, 1983.
6. A.K. Jain, *Fundamentals of Digital Image Processing*. EnglewoodCliffs,NJ: Prentice-Hall, 1989.
7. A.V. Oppenheim, și R.W. Schafer. *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1975.
8. A.V. Oppenheim și R.W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
9. I.W. Selesnick, M. Lang și C.S. Burrus. "Constrained Least Square Design of FIR Filters without Specified Transition Bands." *IEEE Transactions on Signal Processing*, 44, No. 8, 1996.
10. D.B. Percival și A.T. Walden. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge: Cambridge University Press, 1993.
11. W.B. Pennebaker și J.L. Mitchell. *JPEG Still Image DataCompression Standard*. New York, NY: Van Nostrand Reinhold, Chapter 4, 1993.

12. L.R. Rabiner, și B. Gold. *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 393-399, **1975**.
13. P.D. Welch, The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Trans. Audio Electroacoust.* Vol. AU-15 70-73, **1967**.
14. L.J. Karam, și J.H. McClellan. Complex Chebyshev Approximation for FIR Filter Design. *IEEE Trans. on Circuits and Systems II*. 207-216, **1995**.
15. V.F. Demjanjov, și V.N. Malozemov. *Introduction to Minimax*. New York: John Wiley & Sons, **1974**.
16. S.M. Kay, "*Modern Spectral Estimation*." Englewood Cliffs, NJ: Prentice-Hall, 454, **1988**.
17. L.B. Jackson, "*Digital Filters and Signal Processing*." 3rd ed Boston: Kluwer Academic Publishers, Chapter 11, **1996**.
18. S.K. Mitra, "*Digital Signal Processing: A Computer-Based Approach*." New York: McGraw-Hill, Chapter 9, **1998**.
19. P.P. Vaidyanathan, "Robust Digital Filter Structures" in *Handbook for Digital Signal Processing*. S.K. Mitra and J.F. Kaiser, ed. Chapter 7. New York: John Wiley & Sons, **1993**.
20. J.S. Bendat și A.G. Piersol, "*Random Data: Analysis and Measurement Procedures*." New York: John Wiley & Sons, 332, **1971**.
21. R.E. Crochiere, și L.R. Rabiner. *Multi-Rate Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 88-91, **1983**.

VI. APLICAȚII ALE REȚELELOR NEURONALE ÎN CHIMIE

VI.1. Introducere

Rețelele neuronale artificiale au fost concepute pentru rezolvarea unor probleme specifice inteligenței artificiale, dintre care cele mai semnificative sunt: sistemele expert, recunoașterea de forme (configurații), procesarea imaginii, controlul automat al sistemelor, modelarea datelor analitice, etc. Inteligența artificială a fost definită ca fiind studiul proceselor inteligenței umane prin intermediul metodelor computaționale [1]. Importanța cercetărilor asupra sistemului nervos al diferitelor organisme aflate pe variate trepte de evoluție a materiei vii este atât teoretică, prin studii fundamentale asupra proceselor gândirii, cât și practică, prin transferul de tehnologie de la biosisteme spre tehnică și tehnologie [2,3]. Relațiile complexe dintre organisme și mediul ambiant, ca și necesitatea funcționării coordonate a subsistemelor unui organism, impun circulația permanentă de informație referitoare atât la parametrii mediului extern cât și la funcționarea subsistemelor constituente. Sistemul nervos constituie o amplă rețea de comunicație în cadrul organismului, rețea în care semnalele utilizate constau din fluxuri nervoase (impulsuri). Deși sistemul nervos central (SNC) procesează continuu informație relevantă situațiilor naturale mai rapid decât orice sistem cibernetic, experiențele de psihofiziologie umană arată că intervalul de timp scurs din momentul recepționării unui stimul și până în momentul declanșării răspunsului nu poate coborî sub 100 ms, reacție care poate fi suplimentar

întârziată de factori specifici ființelor vii. Sistemele vii sunt superioare sistemelor cibernetice, cu timpi specifici de răspuns sub 1 ms, în cel puțin următoarele privințe [4,5]:

- Capacitate de analiză și sinteză,
- Flexibilitate,
- Fiabilitate,
- Redundanță,
- Miniaturizare,
- Grad înalt de conectivitate.

VI.2. Noțiuni de teoria modelării

Modelarea este o metodă de cercetare a unor sisteme sau procese numite originale, prin studiul asupra altor sisteme mai simple ca structură și proprietăți care se află în anumite raporturi cu originalele. Modelele constituie imagini homomorfe ale sistemelor originale, care păstrează proprietățile esențiale ale acestora în sensul unor aplicații specifice. Parametrii unui model dat pot fi modificați pe parcursul modelării, crescând astfel eficiența sa. Modificarea modelelor se poate face extern adoptând o structură axiomatică diferită, sau intern, prin evidențierea unei relații noi în interiorul modelului existent. Neuronul constituie elementul de bază morfologic și fiziologic al sistemului nervos, fiind o celulă excitabilă cu un înalt grad de specializare, care asigură recepția, codificarea, transmisia și prelucrarea informației conținută de stimuli

Un neuron biologic este alcătuit dintr-un corp neuronal (*soma*) și din două tipuri de prelungiri: *axoni* și *dendrite* (Figura 6.1). Organizarea internă a neuronului este similară cu a oricărei alte celule.

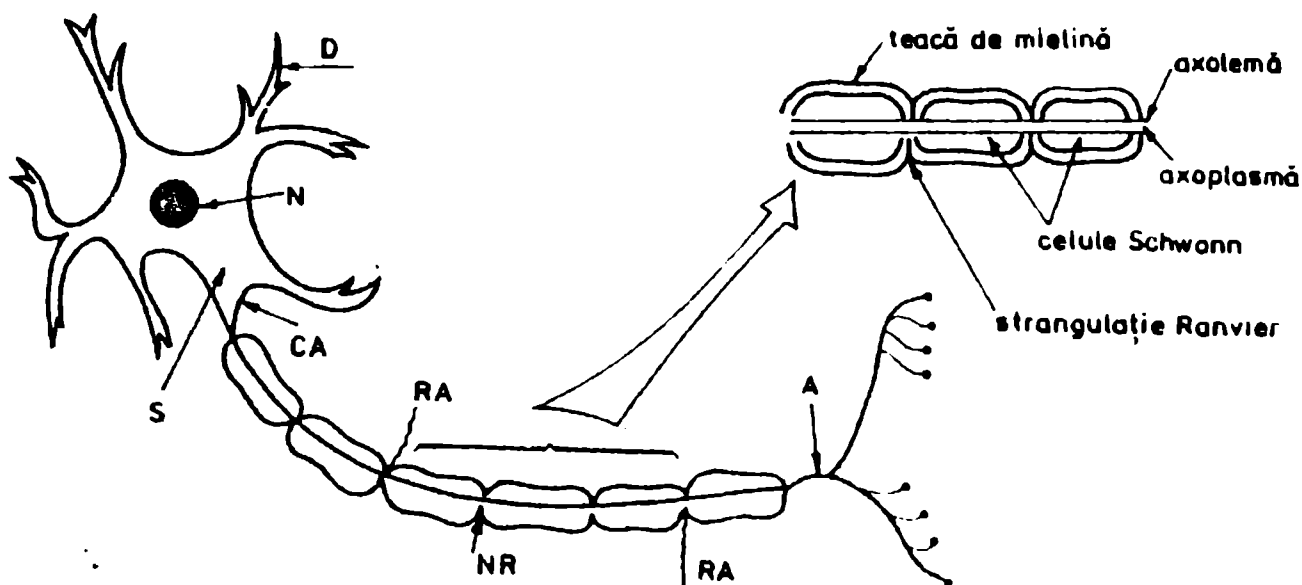


Figura 6.1

Structura schematică a unui neuron biologic;
A axon, FA fibre axonale, RA ramuri axonale inserate în dreptul
nodurilor Ranvier, CA canal axonal.

Corpul neuronului (pericarion) poate avea dimensiuni foarte diferite, de la 5 ÷ 10 μm (neuronii mici din scoarța cerebelului) până la 120 ÷ 125 μm (neuronii mari din măduva spinării). El posedă un nucleu în jurul căruia este dispusă citoplasma neuronală (neuropasma) care conține organite celulare comune, neurofibrile și incluziuni citoplasmatică. Dintre organitele celulare comune sunt de menționat: (i) corpusculii Nissl, (ii) ribozomii formați din acizi nucleici (ADN și ARN) cu rol în sinteza proteinelor, (iii) lizozomii, care conțin enzime (hidrolaze) implicate în procesul de digestie celulară, și (iv) mitocondrii responsabile de respirația

celulară. Citoplasma neuronală conține și organite specifice, în special *neurofibrile* formate din proteine dispuse în spirală.

Citoplasma neuronală se află într-o strânsă dependență de *nucleul* celulei nervoase, care este constituit dintr-o veziculă cu aspect clar ce conține unul sau mai mulți *nucleoli*. Acizii nucleici constituie componenții specifici ai nucleului (ADN este localizat în nucleoli). Învelișul nucleului este format dintr-o membrană dublă, cu pori, care face posibil transportul de la nucleu la citoplasmă a acizilor ribonucleici mesageri (purtători ai mesajelor privitoare la sinteza proteinelor).

Dendritele sunt prelungiri citoplasmatică în general bogat ramificate; ele sunt separate de mediu printr-o membrană prelungire a membranei plasmatică a corpului neuronal. Citoplasma dendritelor conține corpusculi Nissl, mitocondrii și neurofibrile. Dendritele primesc excitații și conduc impulsurile nervoase spre pericarion, deci în sens aferent.

Axonul, prelungire citoplasmatică unică pentru fiecare celulă de lungime mult mai mare decât dendritele, conduce impulsurile nervoase în sens eferent. Diametrul axonilor variază de la 1 μm la circa 20 μm , iar lungimea sa este cuprinsă între 10 μm (axonii neuronilor din SNC) la mai mult de 1 m (axonii neuronilor motori din coarnele anterioare ale măduvei spinării). Axonul se termină cu o multitudine de ramificații (arborele axonal). Elementele constitutive principale ale axonului sunt redată schematic în figura 6.1.

Se observă din figura 6.2 că expansiunea axonică este formată din *axoplasmă* conținând neurofibrile, lizozomi și mitocndrii, precum și dintr-o serie de învelișuri. Primul dintre acestea este *axolema* (membrana

axonului) care este implicată major în procesele de excitație și în cele de transmitere a impulsurilor nervoase. Al doilea înveliș este *teaca de mielină*, întreruptă la intervale neregulate de 0,1÷2 mm de nodurile (strangulațiile) Ranvier. Deoarece nu conduce electric, mielina joacă un rol important în propagarea impulsului nervos de-a lungul fibrei. Astfel, la fibrele mielinizate în transmiterea semnalului sunt implicate numai porțiunile de membrană liberă (din nodurile Ranvier), regiuni care permit schimburi ionice cu mediul (Figura 6.1). Ca urmare, transmisia se face cu pierderi reduse de energie și, ceea ce este mai important, cu perturbații mai mici. În general, din axon pleacă o ramură laterală ce se inserează la nivelul unui nod Ranvier și care se îndreaptă spre alți neuroni, fie direct, fie prin intermediul altor ramuri axonale. Aceste fibre colaterale au rol în mecanismele de reverberație sau de inhibiție prin retroacțiune (*feedback*) sau prin anticipație (*feedforward*).

Deși impulsul nervos se poate propaga în ambele sensuri de-a lungul unui axon, totuși *in vivo* propagarea se face unidirecțional și anume *ortodromic*: de la soma neuronală spre terminațiile axonale. Propagarea unidirecțională a impulsului nervos se datorează unor contacte anatomice interneuronale numite *sinapse*, care se comportă ca niște diode lăsând să treacă potențialele de acțiune numai într-un sens, iar în sens opus prezentând o impedanță foarte mare.

Chiar dacă din punct de vedere morfologic există o întreagă varietate de sinapse, totuși din punct de vedere funcțional se pot diferenția în esență două tipuri:

1. Sinapse cu transmisie chimică (Figura 6.2a);
2. Sinapse cu transmisie electrică prin curenți locali (Figura 6.2b).

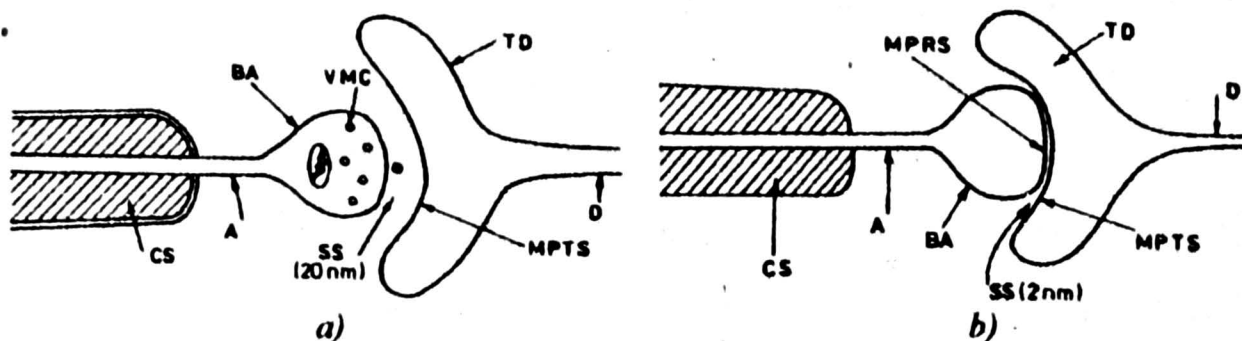
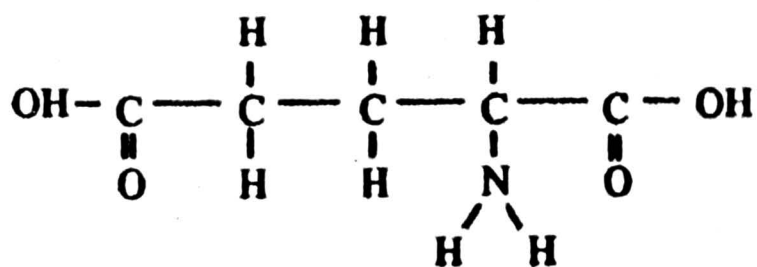


Figura 6.2

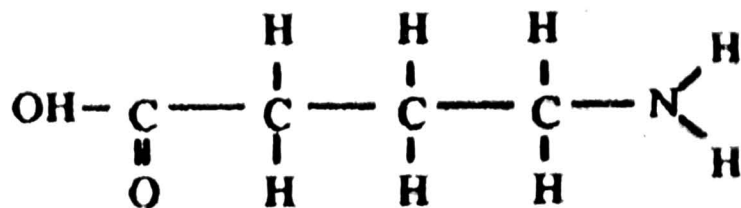
*Reprezentarea schematică a sinapselor neuro-dendritice;
a) cu transmisie chimică, b) cu transmisie electrică.*

În cazul contactelor directe dintre celulele piramidale ale cortexului, mediatorul chimic în transmisia sinaptică excitatorie este glutamatul



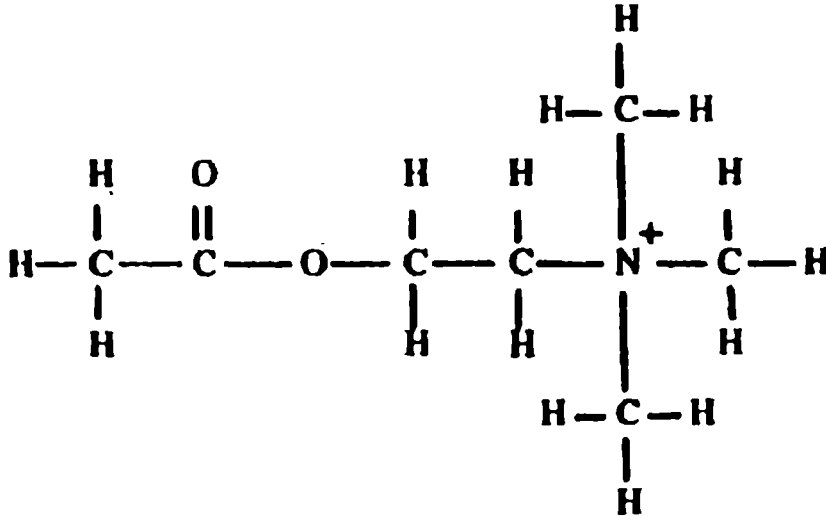
Glutamat

Neurotransmițătorul pentru sinapsele inhibitorii este acidul gama-aminobutiric (GABA)



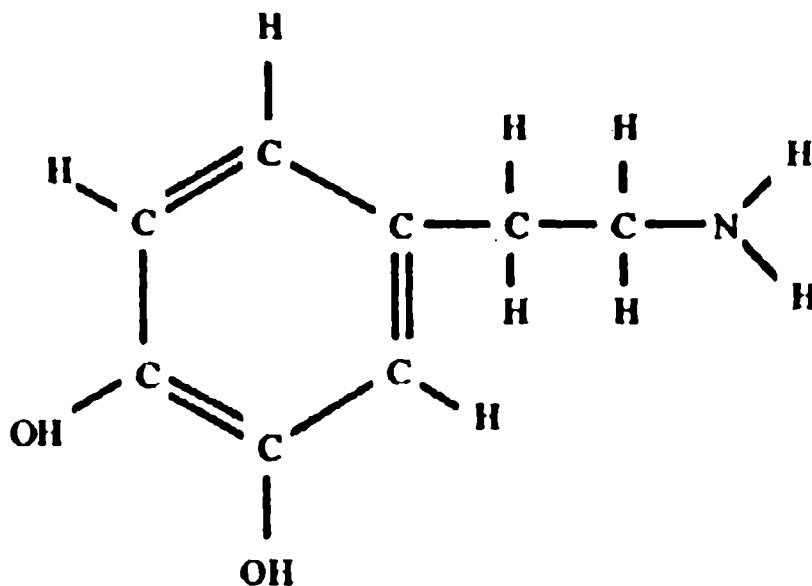
GABA

Hipocampusul este controlat de *septum* și implicit de programele fundamentale (configurații de activitate) ale hipotalamusului. Neurotransmițătorul pentru septum este acetilcolina (același mediator ca și în cazul joncțiunilor neuromusculare)



Acetilcolina

Activități intense și de durată ale formațiunilor nigrostriate stimulează secreția unui neurotransmițător specific, dopamina, care blochează efectele produse de *globus pallidus* permițând excitarea simultană a numeroase regiuni restrânse ale cortexului motor.



Dopamina

Conexiunile interneuronale se realizează prin contactele sinaptice, specificitatea rețelelor neuronale fiind determinate în esență, nu atât de numărul și natura neuronilor constituenți, cât de numărul extrem de mare al interconexiunilor și caracterul tridimensional al rețelei.

Modelarea și simularea rețelelor neuronale sunt importante atât pentru elaborarea de modele de procesare a informației în sistemul nervos animal ("*computational neuroscience*"), cât și pentru conceptele de origine neuronală operante în sistemele inteligente artificiale ("*neural computing*"). Problema centrală în modelare este alegerea unui *model optimal*, care implică existența unor criterii de comparație.

Referitor la rețelele neuronale, se utilizează frecvent termenul de *paradigmă*. În acest context, paradigma cuprinde tipic descrierea formei și funcției unității elementare de procesare (celulă, nod, neuron,...), topologia rețelei, care specifică configurația interconexiunilor dintre unități, precum și o regulă de adaptare (învățare, instruire) a ponderilor interconexiunilor. Trei clase de sisteme pot constitui paradigme ale

sistemelor inteligente: *sistemele continui*, *sistemele logice* și *sistemele simbolice*. Acestea le corespund în calculatoarele convenționale respectiv nivelele de *circuitistică*, de *logică* și de *programare*. Paradigmele rețelelor neuronale, deși distincte în detaliu, selectează o submulțime comună de atribute tipice rețelelor neuronale biologice [6]:

- Elemente simple de procesare,
- Nivel ridicat de conectivitate,
- Procesare distribuită paralel a datelor,
- Funcție de transfer neliniară pentru neuroni,
- Retroacțiune ("feed-back"),
- Propagare neamortizată și cu viteză constantă energiei,
- Procesare nonalgoritmă,
- Adaptabilitate,
- Autoorganizare
- Toleranță la defecte,
- Răspunsuri relevante la date de intrare vagi ("fuzzy"),
- Putere de generalizare.

Abilitatea procesării simbolice către care tind modelele performante de rețele neuronale actuale este o consecință tehnică neașteptată a calculatoarelor digitale, care nu au fost inițial concepute în acest scop. În mod similar poate fi privită facilitatea de a raționa logic și rezolva probleme matematice, care apare ca o "deficiență tehnică" inițială neașteptată a creierului uman, în mod indiscutabil inutilă supraviețuirii în zorii apariției speciei umane [7,8].

Problema centrală în modelarea rețelelor neuronale este alegerea unui *model optimal*, care implică existența unor criterii de comparație. Aceste criterii sunt mai mult sau mai puțin semnificative în contextul unor aplicații particulare. Alegerea unui model existent sau construcția unui model nou, depind de mai mulți factori cu ponderi diferite:

- Disponibilitatea și tipul facilităților computaționale de prelucrare a datelor și simulare asistate de calculator;
- Suport tehnic și financiar,
- Experiența și nivelul științific al grupului de cercetare,
- Antecedente istorice, etc

Sursele de incertitudine (nedeterminare), care afectează modelarea sunt clasificate în trei categorii distincte, având la origine următoarele [9]

1. Măsurători imprecise sau nesigure (cu o anumită lipsă de acuratețe);
2. Abateri aleatoare;
3. Descreri sau prezentări vagi (incerte).

Se pot considera că aceste trei clase calitativ diferite de la care poate proveni incertitudinea în procesul modelării, conduc la trei tipuri de modele distincte [2, 10, 11] :

1. *Modelul determinist;*
2. *Modelul probabilist (stochastic);*
3. *Modelul vag (fuzzy).*

VI.3. Modele de rețele neuronale

VI.3.1 Rețele neuronale formale cu instruire supervizată (memorii asociative)

Conceptul de *memorie asociativă* caracterizează abilitatea unui sistem dinamic de a parcurge o secvență de reprezentări interne sau de reconstituire a unei reprezentări complexe pornind de la o versiune distorsionată sau incompletă a sa. Memoria asociativă are la bază conceptul de reprezentare distribuită a informației prin stări colective ale rețelei neuronale. Dinamica interacțiilor poate fi făcută să corespundă structurii informaționale, fără un echivalent fizic direct în sistem, realizându-se prin efecte globale și reflectându-se în procesele de redare a informației memorate. Problemele fundamentale ridicate de memoria asociativă sunt: *diafonia* ("cross-talk"), *invarianța* și *inabilitatea* de a opera cu structuri.

Ca exemple de modele de rețele neuronale formale cu instruire supervizată pot fi menționate: modelul *Sutton-Barto*, modelul *heterostat*, modelul *Hopfield-Tank*, modelul *Hoppenstead*, modelul *Hamming*, modelul *Hopfield stochastic* și *perceptronul*.

Perceptronul constituie o clasă reprezentativă de rețele neuronale introdusă de Rosenblat [12], care este capabilă de instruire supervizată, având astfel caracterul unei *memorii asociative*. Caracterul paralel al procesării informației este finalizat în modelele matematice ale *suptafeșelor decizionale* (modelul *liniar* și modelul *descreșterii*

gradientului), care susțin premiza fundamentală a lui Rosenblatt relativă la caracterul probabilistic al perceptronului. Dacă teorema de convergență a perceptronului oferă a priori succesul algoritmului recursiv de instruire, în cazul perceptronilor multistratificați pentru care nu există asemenea teoreme de convergență, funcționează criterii particulare de suficiență ("*credit assignment problem*").

Modelele matematice fundamentale ale rețelelor neuronale includ inevitabil, modelul neuronului binar clasic, de inspirație biologică introdus de McCulloch și Pitts [13]. Totuși, trebuie remarcat caracterul analogic (liniar) al comportamentului unui neuron biologic: funcționarea sa se bazează pe curenți ionici care variază continuu și mai puțin pe valori discrete cărora să li se poată asocia nivelele de 0 și 1.

VI.3.2. Rețele neuronale formale cu instruire nesupervizată (clasificatori vectoriali)

Rețelele neuronale cu instruire nesupervizată și autosupervizată sunt denumite generic *clasificatori vectoriali*. În cazul rețelelor neuronale cu instruire nesupervizată numai o singură ieșire este activă la un moment de timp. Aceste rețele organizează datele de intrare sub formă de clase sau de categorii. Rețelele neuronale cu instruire nesupervizată au nevoie de un număr de celule de ieșire egal cu numărul de categorii care se dorește memorat. De exemplu N unități de ieșire pot clasifica numai N categorii și nu 2^N ca în cazul memoriilor asociative. Prezentăm mai jos câteva exemple de rețele neuronale formale cu instruire nesupervizată.

Rețeaua Carpenter-Grosberg este reprezentativă prin comparație cu rețeaua Hamming, în sensul că asigură clasificarea exemplarelor pe măsura prezentării lor la intrare prin identificarea unor clase de configurații asemănătoare specificate prin câte un exemplar reprezentativ.

Modelul instruirii competitive, model de învățare nesupervizată corespunzător detectorului de regularitate, este în esență un algoritm de instruire statistică neasociativă și este analizat în legătură cu abilitatea rețelelor neuronale adaptive de a descoperi caracteristici relevante ale stimulilor de intrare. Cele mai reprezentative arhitecturi structurale și mecanisme de instruire competitivă sunt: *modelul Linsker* și *modelele Fukushima*.

Modelul activării interactive este calitativ un algoritm capabil să explice elementele fundamentale ale percepției vizuale și auditive.

Teoria rezonanței adaptive continuă seria modelelor de instruire competitivă, care încearcă să modeleze stabilitatea algoritmilor competitivi în cazul rețelelor plurineuronale.

Un alt grup important de modele caracterizate prin instruire autosupervizată cuprinde *algoritmii dezvoltați pe baza principiului darwinist de evoluție a speciilor prin selecție*, reflectat în rețelele neuronale prin competiția dintre neuroni în procesele de învățare. Astfel următoarele modele pot fi considerate cele reprezentative [2]:

modelul Malsburg, *modelul Kohonen*, care propune un mecanism de autoorganizare a rețelelor neuronale prin care numărul neuronilor alocați unei regiuni din spațiul stimulilor de intrare este proporțional cu nivelul ei de activitate; *modele "genetice" de instruire*, care se bazează pe un

ansamblu de opțiuni inițiale (populația de cromozomi) în procesele iterative de instruire și *modelul Hecht-Nielsen*, care combină regula lui Hebb cu conceptele instruirii competitive, în scopul identificării aspectelor structurale caracteristice câmpului datelor de intrare.

VI.4. Aplicații ale rețelelor neuronale

VI.4.1. Noțiuni fundamentale

Rețelele neuronale constituie, în esență, un mediu specific de procesare a informației în sisteme care operează cu date de intrare multivariate și răspunsuri multiple. *Datele multivariate* sunt asociate cu *un obiect* caracterizat de mai multe variabile; ca exemplu, analiza unor probe de apă potabilă pentru determinarea poluanților Pb, Cl₂ și CO constituie o problemă cu date de intrare multivariate. *Răspunsurile multiple* sunt asociate cu date multivariate studiate în funcție de mai mulți *factori* care pot influența răspunsul sistemului; în exemplul de mai sus factorii ar putea fi temperatura, presiunea, etc.

Relația generală dintre intrarea m -variata $X_s = (x_{s1}, x_{s2}, \dots, x_{sm})$ și răspunsul n -variata $Y_s = (y_{s1}, y_{s2}, \dots, y_{sn})$ al sistemului, unde $s = 1, 2, \dots, r$ indică eșantioane (probe) diferite, poate fi explicitată compact sub forma:

$$Y_s = A X_s \quad (6.1)$$

unde A este în general un operator complex. În particular, A poate fi o matrice de dimensiune $(m \times n)$ care transformă liniar vectorul X_s în

vectorul Y_s , astfel încât ecuația (6.1) constituie *modelul matematic al problemei multivariate multirăspuns*.

În principiu, problema este următoarea: se dă un set de date m -variate $\{X_s\}_{s=1,2,\dots,r}$ care conduce la setul răspunsurilor n -variate $\{Y_s\}_{s=1,2,\dots,r}$, în timp ce operatorul A este de cele mai multe ori necunoscut. Mai mult, în majoritatea aplicațiilor se solicită estimarea chiar a răspunsurilor posibile $\{Y_s\}_{s=1,2,\dots,r}$. Sarcina fundamentală este de a trage concluzii relevante din probe necunoscute pe baza răspunsurilor obținute pornind de la un set de date multivariate selecționate, $\{Y_s\}_{s=1,2,\dots,p}$, unde $s \in \wp$, $\text{card } \wp = p < r$, \wp fiind denumit *setul de antrenament, învățare sau instruire*. Nu există un algoritm universal care să selecționeze din mulțimea perechilor de vectori intrare-ieșire $\{(X_s, Y_s)\}$, numite pe scurt configurații, acele perechi care sunt cele mai relevante prin conținutul informațional în procesul de instruire a rețelei.

Alegerea unei strategii de instruire *supervizată* sau *nesupervizată* a rețelei depinde de specificul problemei și datele disponibile. Deoarece instruirea nesupervizată este mai flexibilă prin multitudinea răspunsurilor posibile, este recomandată ca etapă inițială de acumulare a informației. Ulterior, se trece la instruirea supervizată, proces în care datele multivariate cunoscute se împart în trei clase:

- *Setul de instruire (învățare)* \wp , care conține mulțimea configurațiilor de referință;
- *Setul de control*, care determină încheierea procesului de instruire;

- *Setul de testare*, care permite verificarea predictibilității rețelei.

Deși alegerea aleatorie a configurațiilor pentru cele trei clase considerate nu este cea mai indicată strategie, totuși este singura opțiune rezonabilă în absența unor date sau informații prealabile.

În caracterizarea performanțelor rețelelor neuronale doi factori sunt considerați esențiali și anume:

- *Reproducerea din memorie ("recall")*, care se exprimă prin procentajul configurațiilor recunoscute corect din *setul de instruire*, după ce procesul de învățare s-a încheiat,
- *Predicția ("prediction")*, care se exprimă prin procentajul configurațiilor recunoscute corect din *setul de testare*, configurații neutilizate în procesul de instruire a rețelei.

VI.4.2. Clase de aplicații specifice ale rețelelor neuronale

Clasificarea datelor multivariate sortează configurațiile în clase simple, sau ierarhizează în interiorul claselor. *Ierarhizarea* implică o secvență a nivelelor decizionale care să clasifice adecvat o configurație într-un grup sau subgrup, categorie sau subcategorie, clasă sau subclasă.

Specific aplicațiilor în chimie, un obiect poate aparține simultan mai multor clase distincte, astfel încât rezultatul este produsul logic al mai multor decizii (Figura 6.3).

Modelarea urmărește elaborarea de sisteme formale capabile să genereze răspunsuri continue pentru datele de intrare, în sensul că la modificări mici la intrare corespund modificări mici ale răspunsului sistemului. Prin esența sa, modelarea necesită întotdeauna instruire

supervizată și, în consecință, implică algoritmi de propagare inversă sau contra-propagare a erorii ("*back-propagation*" și "*counter-propagation*").

În general, criteriul alegerii parametrilor de modelare este minimizarea diferenței dintre setul răspunsurilor determinate experimental,

$\{Y_s^{\text{experimental}}\}_{s \in \varphi}$, și răspunsurile generate de model, $\{Y_s^{\text{model}}\}_{s \in \varphi}$,

corespunzătoare aceluiași set de configurații de intrare $\{X_s\}_{s \in \varphi}$, adică:

$$\min (Y_s^{\text{experimental}} - Y_s^{\text{model}})^2 \quad (6.2)$$

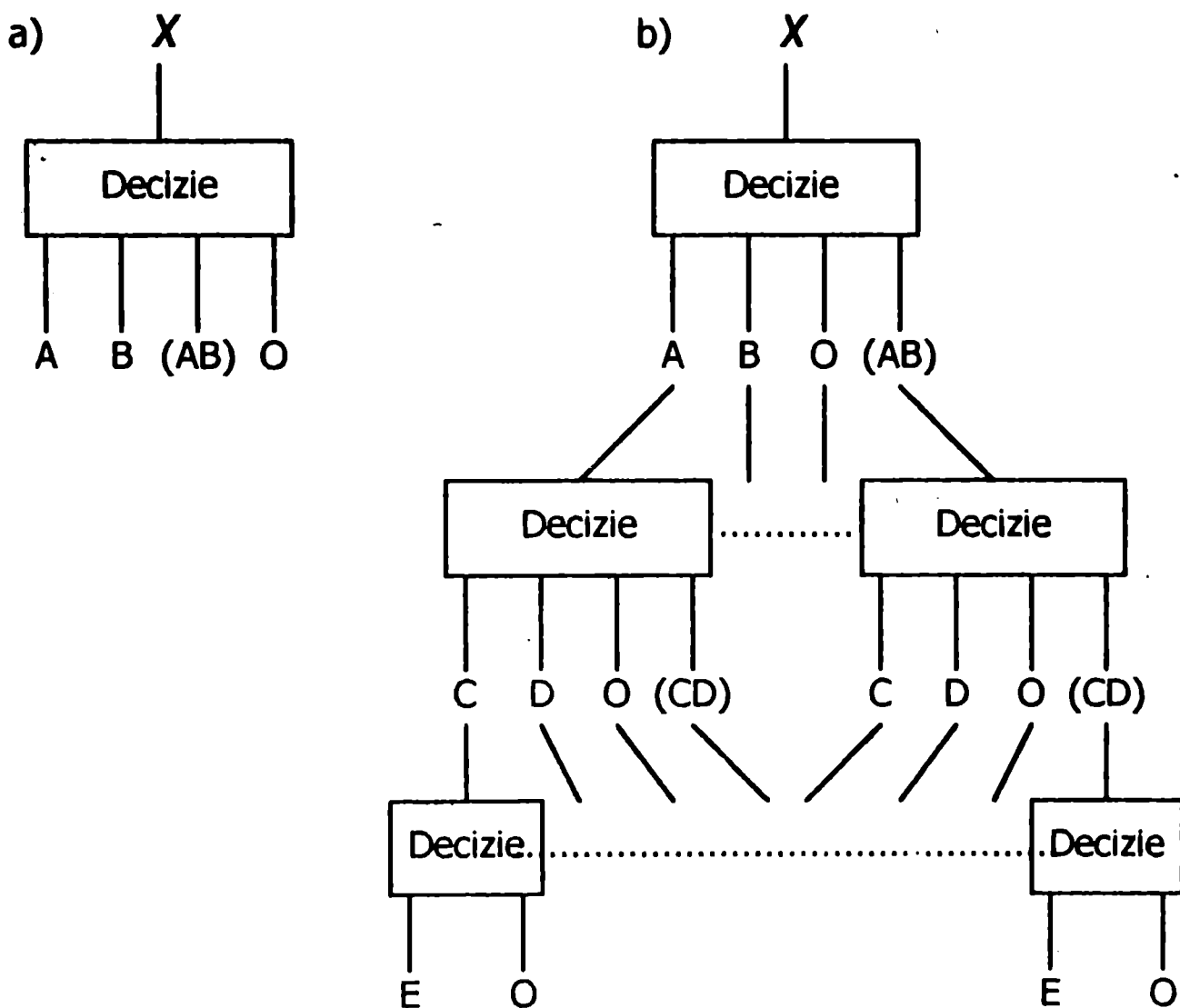


Figura 6.3

Clasificare pe un nivel decizional (a) și clasificare ierarhică (b).

Avantajul major al rețelelor neuronale în problemele multirăspuns asupra modelării clasice cu funcții analitice constă în faptul că "țintele", $\{Y_s^{\text{model}}\}_{s \in \varphi}$, pot fi multidimensionale. Cu alte cuvinte, modelele clasice necesită predefinirea unei funcții pentru fiecare răspuns, în timp ce rețelele neuronale pot modela un vector multirăspuns fără informații apriorice.

Modelarea este eficientă pentru configurații cu un număr relativ redus de variabile de intrare (în general mai mic de 10), care conduc la arhitecturi de rețele relativ limitate și putere de calcul solicitată inferioară celor implicate de tehnicile clasice de clasificare (Figura 6.4).

Alegerea algoritmului de instruire este consecința directă a distribuției obiectelor de intrare în spațiul variabilelor. Metoda propagării inverse a erorii ("*error back-propagation algorithm*") este indicată în cazul aplicațiilor cu un număr relativ redus de obiecte și repartizate relativ egal. Metoda contra-propagării erorii ("*error counter-propagation algorithm*") este adecvată în cazul unui număr mare de obiecte de intrare repartizate inegal în spațiul variabilelor, după ce au fost în prealabil reduse la o acoperire regulată a spațiului prin intermediul rețelelor Kohonen sau evaluări statistice ale intervalelor variabilelor.

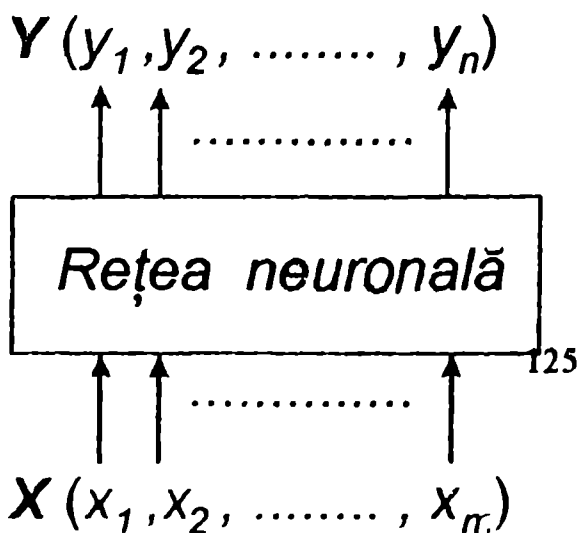


Figura 6.4

Modelarea de principiu cu rețele neuronale.

VI.4.3. Abordarea modelării neuronale

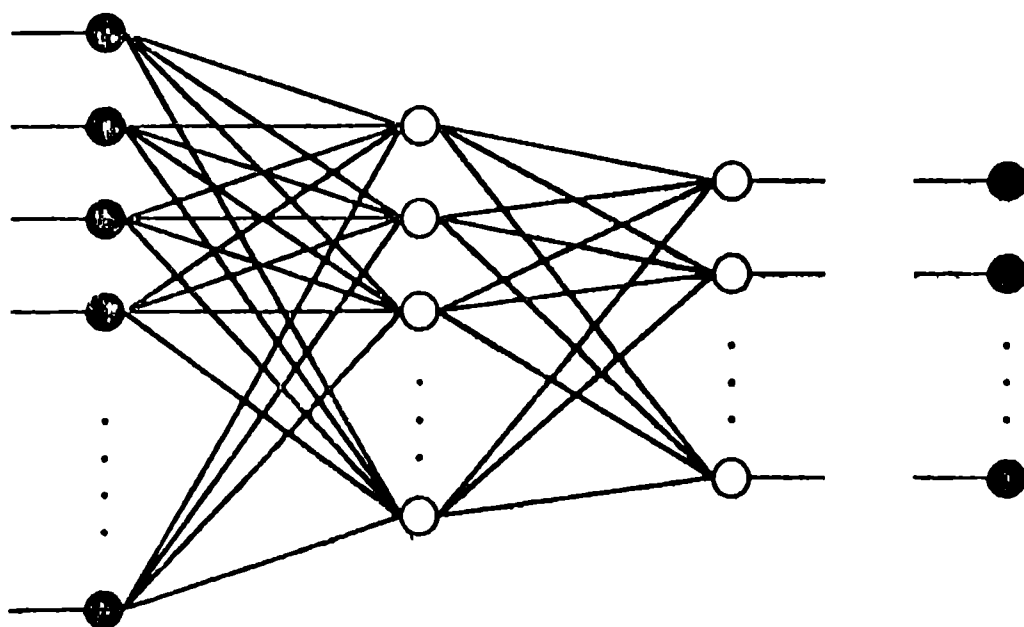
Utilizarea rețelelor neuronale este deschisă unei largi varietăți de aplicații. În general, tipuri diferite de probleme necesită diferite arhitecturi neuronale și strategii de instruire adaptate caracterului datelor de intrare [14].

Ca modalitate principală de abordare a unei probleme utilizând rețele neuronale, este în general indicat a se clasifica mai întâi problema de interes într-una din clasele de aplicații prezentate mai sus. În al doilea rând se alege și se adaptează topologia rețelei care corespunde aplicației identificate. În al treilea rând se optează pentru un algoritm de instruire adecvat capabil să rezolve rețeaua. Indicații generale privind clasificarea tipurilor de probleme care se pot aborda cu tehnici de inspirație neuronală sunt prezentate pe larg în literatura de specialitate [15-17].

Problemele de chimie conțin adesea elemente care aparțin la tipuri diferite de probleme, astfel încât rezolvarea globală se desfășoară utilizând mai multe tipuri de rețele simultan. Deoarece rețelele neuronale sunt metode generale de rezolvare, un accent deosebit trebuie pus pe reprezentarea informației chimice în context. Cu toate acestea, cele mai multe aplicații practice în chimie publicate în literatura de specialitate [18,19] dovedesc succesul rețelei directe, complet conectate, de tip perceptron, cu un singur strat intermediar (Figura 6.5) și care se instruieste utilizând algoritmul de propagare inversă a erorii.

Atunci când rețelele directe clasice sunt instruite pentru mai multe configurații cu reguli de propagare derivate din algoritmul de propagare inversă a erorii, pot apare interferențe puternice la stabilirea ponderilor sinaptice. Aceste interferențe au ca efect descreșterea convergenței procesului de instruire și scăderea puterii de generalizare a rețelei. O cale de a evita interferențele pentru anumite subclase de configurații constă în utilizarea simultană a mai multor *rețele locale diferite* și o *rețea-poartă* care să decidă rețeaua locală activă pentru fiecare subclasă în parte [20].

Ultimile realizări în această direcție se referă la elaborarea unor noi *funcții obiectiv* specializate pentru asociații adaptive de rețele neuronale locale. Mai clar, în locul unei combinații liniare a ieșirilor furnizate separat de rețelele locale, este sugerată abordarea unei rețele-poartă care decide statistic rețeaua locală responsabilă pentru clasificarea fiecărei dintre configurațiile de intrare [21].



Unitati de intrare Unitati intermediare Unitati de iesire Unitati tinta

Figura 6.5.

Rețea neuronală de tip perceptron cu un singur strat intermediar și unități țintă care permit instruirea supervizată.

VI.4.4. Aplicații ale rețelelor neuronale în chimie

Abilitatea de a procesa date indiferent de caracterul lor (discret, continuu sau vag), fără specificarea unor corelații prealabile, recomandă rețelele neuronale în aplicații complexe de recunoaștere de configurații. Mai mult, acolo unde nu există modele analitice liniare, rețelele neuronale se dovedesc eficiente datorită caracterului neliniar al răspunsurilor furnizate.

Aplicațiile rețelelor neuronale în științele naturale, și în mod special în chimie, oferă noi metode de cercetare care pot îmbunătăți sau releva noi perspective în modelarea, prelucrarea și interpretarea datelor experimentale. Creșterea continuă a publicațiilor anuale dedicate cercetărilor în domeniul rețelelor neuronale, ca și a diverselor aplicații raportate, dovedeste interesul crescând al comunității științifice pentru sistemele artificiale de inspirație neuronală

VI.4.4.1. Reactivitatea legăturilor chimice

Reacțiile chimice sunt inițiate prin desfacerea unor legături dintre molecule de aceea cunoașterea reactivității legăturilor chimice este o

problemă fundamentală în chimie. Parametrii care influențează desfacerea legăturilor chimice sunt limitați în modelare la următorul set reprezentativ:

- Δq_{tot} diferența de sarcină totală,
- Δq_{π} : diferența în sarcină π ,
- $\Delta \chi_{\sigma}$: diferența în electronegativitate σ ,
- Q_{σ} : măsura polarității legăturii,
- α_b : polarizabilitatea legăturii,
- BDE : energia de disociere a legăturii,
- R^{\pm} : măsura stabilității rezonanței disponibilă pentru particulele generate prin heteroliză.

Valorile acestor parametri sunt calculate utilizând un pachet de programe numit *PETRA* (*Parameter Estimation for the Treatment of Reactivity Applications*) [17]. Modelul trebuie să coreleze valorile parametrilor cu reactivitatea legăturilor de interes. O abordare eficientă a acestei probleme constă în combinația unei rețele Kohonen pentru selectarea setului de instruire, cu o rețea multistratificată echipată cu instruire supervizată și antrenată pentru clasificare. Ansamblul celor două rețele este capabil să prezică în moleculele alifatiche care dintre legăturile simple se vor desface preferențial polar și care nu, ba mai mult, indică și direcția în care se vor deplasa sarcinile după heteroliză. Deși rețeaua poate fi antrenată numai pe legături simple, ea dispune de caracteristica de generalizare, în sensul că asociază reactivitatea corectă inclusiv la unele legături duble (de exemplu C = O). Răspunsurile rețelei pot fi interpretate

ca *probabilități* de desfacere a legăturilor, ca și o *măsură cantitativă* a reactivității.

VI.4.4.2. *Optimizarea proceselor HPLC*

Metodele cromatografice joacă un rol important în *tehnicile de separare*. Cromatografia de lichide de înaltă performanță (HPLC) constituie o metodă răspândită de analiză în care un amestec este distribuit între o fază staționară, lichidă sau solidă, pe un suport solid, și o fază lichidă mobilă supusă la presiune ridicată. Diversele componente se distribuie diferit între cele două faze și astfel sunt separate. Principalul factor în acest caz îl constituie *factorul de performanță (FP)* al analizei, care include două aspecte: puterea de separare a coloanei și timpul necesar pentru desfășurarea integrală a procesului de separare.

Orice analiză de proces HPLC implică modelarea fazei mobile, în sensul specificării factorului de performanță ca funcție de parametrii săi. Apoi se alege cea mai potrivită fază mobilă pentru o analiză particulară.

În majoritatea aplicațiilor se consideră numai doi parametri de care depinde faza mobilă: *concentrația de etanol (C_e)* și *aciditatea fazei (pH)*. Orice modelare, clasică sau cu rețele neuronale, necesită date experimentale. În modelarea clasică, pentru calculul valorii factorului de performanță se utilizează în general o formă biliniară:

$$FP = ax^2 + by^2 + cxy + dx + ey = f \quad (6.3)$$

unde x semnifică valorile concentrației C_e , iar y valorile pH -ului. Parametrii a , b , c , d și f sunt obținuți prin orice tehnică de optimizare, cum ar fi *regresia liniară*, metoda *Simplex*, etc.

Aceleași date pot fi utilizate pentru proiectarea unei rețele neuronale formale cu un singur strat intermediar, echipată cu o regulă de instruire de tipul propagării inverse a erorii. Alegerea acestui algoritm de instruire este consecința arhitecturii multistratificate a rețelei, precum și consecința necesității instruirii supervizate reclamată de modelare. Corespunzător proprietăților avute în vedere la modelare, rețeaua trebuie să conțină doi neuroni de intrare, un neuron de ieșire și un număr de neuroni intermediari.

Experiențe de optimizare a proceselor HPLC în analiza vinului au relevat ca cele mai performante rețelele de tipul $(2 \times 6 \times 1)$ care conțin $(2 + 1) \times 6 + (6 + 1) \times 1$ coeficienți sinaptici. Rezultatele obținute s-au dovedit în bun acord cu cele furnizate prin metode clasice.

VI.4.4.3. Relația structură-activitate

Unul dintre domeniile cele mai promițătoare de aplicații ale tehnicilor de modelare îl constituie determinarea relațiilor cantitative dintre structura chimică a unui compus și activitatea sa fizică, chimică sau biologică. În mod deosebit, corelarea structurii chimice a medicamentelor cu efectele farmacologice este de mare interes în industria farmaceutică. Datorită complexității care poate fi atinsă în modelarea rețelelor neuronale, acestea joacă un rol din ce în ce mai important în studiul

relației cantitative structură-activitate, problemă denumită în literatura de specialitate QSAR ("*Quantitative Structure-Activity Relationships*").

Un exemplu remarcabil în acest sens îl constituie lucrările lui Aoyama, Suzuki și Ichikawa [17] care urmăresc prognozarea dozei medicamentoase minime necesară pentru prelungirea cu 40% a vieții cobailor test, cărora li s-au inoculat celule cancerigene L-1210. *Doza efectivă minimă* depinde de *concentrația C* a substanței care produce efectul dorit și se exprimă în termeni logaritmici de forma $\log(1/C)$, consecință a plajei de mai multe ordine de mărime a concentrațiilor solicitate.

Setul de date cuprinde modificările caracteristicilor structurale ale carbochinonelor, o clasă de compuși care manifestă activitate *anticancerigenă* cu diverse intensități. Această activitate este în primul rând influențată de identitatea substituienților R^1 și R^2 . În analiza multiliniară standard, influența combinată a substituienților R^1 și R^2 este descrisă de următoarele variabile fizico-chimice:

- $MR_{1,2}$: refractivitatea molară,
- $\pi_{1,2}$: contribuția substituienților la hidrofobicitate,
- F' : suma constantelor substituienților la efectul de câmp,
- R : suma constantelor substituienților la efectul de rezonanță;

și două variabile locale responsabile de influența exclusivă a lui R^1 :

- MR_1 : refractivitatea molară a lui R^1 ,

- π_1 . contribuția lui R^1 la hidrofobicitate.

Au fost selecționate 35 de carbochinone diferite pentru care, fie au fost măsurate valorile celor 6 variabile ($MR_{1,2}$, $\pi_{1,2}$, MR_1 , π_1 , F și R), fie au fost preluate date din literatura de specialitate (π_1 , F și R), apoi doza efectivă minimă a fost măsurată și reprezentată ca $\log(1/C)$. În consecință, a fost utilizată o rețea neuronală cu 6 unități de intrare corespunzătoare celor 6 variabile ($MR_{1,2}$, $\pi_{1,2}$, MR_1 , π_1 , F și R), o unitate de ieșire și o unitate țintă, care specifică doza minimă efectivă $\log(1/C)$. În majoritatea aplicațiilor s-a dovedit suficient un singur strat intermediar al cărui număr de unități s-a stabilit experimental. În aceste condiții, a fost utilizată o variantă a regulii de instruire cu propagare inversă a erorii. Rezultatele obținute cu rețele neuronale sunt în general superioare, deși nu într-o măsură covârșitoare, celor furnizate de tehnica clasică MLRA ("*Multi-Linear Regression Analysis*").

VI.4.4.4. Caracteristicile generale ale structurii secundare a proteinelor

Determinarea structurii macromoleculelor biologice este o problemă fundamentală în biologia moleculară și constituie un prim pas în elucidarea proprietăților funcționale complexe care le caracterizează. Forma elementară a problemei privește prognoza structurii secundare a proteinelor, adică conformația spațială la nivel inferior (de exemplu α -elicii, β -plane și spirale aleatoare) pe baza structurii lor primare [21]. În această problemă, legea care determină structura secundară a lanțului

proteic în funcție de proprietățile colective ale aminoacizilor individuali este principala necunoscută. Modurile de abordare a problemei se încadrează principial în două categorii: *tehnici clasice și metode specifice inteligenței artificiale*.

Abordarea neuronală a modelării structurii secundare din perspectiva tehnicilor aparținând domeniului inteligenței artificiale este superioară deoarece rețelele neuronale sunt caracterizate de o procesare holistică a datelor de intrare și, mai mult, nu necesită reguli de instruire *ad hoc* sau informații extrem de *specifice*, ca o consecință directă a procesului de instruire prin exemple și a puterii de generalizare a rețelei.

Rețelele neuronale recomandate în acest gen de aplicații sunt de tip perceptron multistratificat, iar algoritmul de instruire de tipul cu propagare inversă a erorii. Configurația unei secvențe primare de reziduuri de lungime l este aplicată la intrarea perceptronului, care urmează să asocieze aminoacidului din mijlocul configurației (l este presupus impar) structura sa secundară. Dimensiunea l a ferestrei este astfel aleasă încât reziduurile configurațiilor de intrare să cuprindă cele mai multe din interacțiile care presumptiv ar stabiliza structura. Experimental, s-a determinat că dimensiuni ale ferestrei mai mari de 5 reziduuri nu aduc în mod necesar o îmbunătățire sensibilă, dar s-a preferat în aplicații valoarea 7. Studiul statistic al proteinelor a relevat că lungimea unei configurații afectează moderat performanța metodei. A fost utilizat un cod expandat, în care grupuri de câte 20 neuroni au fost folosiți la codificarea fiecărui aminoacid din cei 20. Dintre aceștia, 19 aminoacizi sunt inactivi, iar poziția celui de-al 20-lea neuron activ în interiorul grupului identifică

reziduurile considerate. Astfel se evită corelațiile nedorite între diverse reziduuri care sunt simultan prezente la intrarea perceptronului. În consecință, numărul neuronilor de intrare este 7-1. Fiecare neuron de ieșire reprezintă una din posibilele structuri a fi prognozate și ieșirea rețelei este făcută să corespundă cu neuronul având activitate maximă. La instruire au fost necesare cel puțin 100 de cicluri complete de parcurgere a setului de configurații de referință [22].

În concluzie rețelele neuronale testate pe clase *omogene* de proteine (setul de instruire și cel de testare sunt alese din aceeași clasă), fie globulare, fie membranare, manifestă o putere de generalizare superioară metodelor clasice de prognoză. Mai mult, rețelele de tip perceptron sunt capabile de a identifica și alte structuri secundare decât elicii transmembranare (de exemplu elicii extramembranare sau β -spirale). În al doilea rând, experiențele efectuate pe clase de proteine presupuse *eterogene* (setul de instruire și cel de testare sunt alese din clase diferite) au evidențiat abilitatea perceptronilor de a învăța și recunoaște configurații și proprietăți statistice cu o eficiență comparabilă cu cea obținută pe clase omogene. În al treilea rând, superioritatea sensibilă a performanței sistemelor neuronale de tip perceptron în raport cu metodele clasice aplicate proteinelor membranare, sugerează că metodele de abordare cu rețele neuronale sunt mai eficiente în identificarea regularităților statistice conținute în setul de instruire, precum și mai eficiente decât metodele statistice în extrapolarea datelor structurale de la proteinele globulare la proteinele membranare [3, 23].

VI.4.4.5. Corelația spectru-structură

Elucidarea structurii compușilor organici se bazează preponderent pe metode spectroscopice, deși relațiile dintre structură și datele spectrale sunt în general prea complexe pentru a fi exprimate sub forma unor ecuații explicite. Se cunosc numeroase tehnici de abordare a interpretării automate a spectrelor infraroșii printre care: căutarea și fitarea cu spectrele aparținând unei baze de date, utilizarea tabelelor de corelații, diferite metode statistice, teoria mulțimilor, precum și diverse metode de recunoștere de forme ("*pattern recognition*") și sisteme expert ("*expert systems*") [24]. Corelațiile precise dintre spectrele complete de înaltă rezoluție și structurile chimice asociate, precum și disponibilitatea unor largi baze de date computerizate, fac extrem de atractiv acest domeniu pentru aplicații de interes a rețelelor neuronale.

Sistemele neuronale formale constituie o clasă de metode elegante de clasificare a configurațiilor în care un vector de intrare, ale cărui valori reprezintă obiectul de intrare, este transformat într-un vector de ieșire, ale cărui valori reprezintă clasificarea dorită a obiectului de intrare. Transformarea este realizată prin intermediul unei formule de propagare, care combină vectorul de intrare cu coeficienții interni ai rețelei. Obiectele de intrare, care constituie subiectul transformării, sunt spectrele infraroșii ale unor compuși diverși, în timp ce ieșirea furnizează serii de substructuri care sunt conținute (sau nu) de compusul reprezentat la intrare. Procesul include *interpretarea spectrală*, adică exprimarea proprietăților spectroscopice ca fragmente substructurale, și *generarea structurii*, adică

reuniunea fragmentelor substructurale alcătuind toate structurile moleculare complete compatibile cu ele.

O tehnică convenabilă de a opera cu date complexe multivariate este *clasificarea*, care sortează obiecte în clase de echivalență sau ierarhizează subclase în clase. Uzual, stabilirea unei clasificări ierarhice în decizii complexe începe cu un studiu statistic prealabil al relațiilor dintre reprezentarea obiectelor și clasele lor de apartenență. Clasificările ierarhice implică mai multe nivele decizionale și, mai mult, sunt aplicatii în care obiectul de intrare este caracterizat imprecis sau vag ("*fuzzy*"). Modelele clasice necesită funcții analitice predefinite pentru fiecare răspuns, în timp ce rețelele neuronale prezintă avantajul că pot modela vectori multirăspuns fără cunoștințe apriorice și procesa atât funcții de apartenență boolene ("*crisp membership functions*"), cât și funcții de apartenență vagi ("*fuzzy membership functions*"). Analiza corelației spectru-structură constituie o problemă tipică de clasificare multidecizională: obiectul de intrare este configurația unui spectru, iar răspunsul este o listă de clase prezente (sau absente) în compusul care a produs spectrul de la intrare. Atât instruirea supervizată, cât și cea nesupervizată, pot fi utilizate în proiectarea rețelelor neuronale destinate clasificării; în primul caz se recomandă o arhitectură cu un singur strat intermediar instruită cu regula de propagare inversă a erorii, în cel de-al doilea caz se face invariabil apel la rețele de tip Kohonen [25].

Munk, Madison și Rob [26] sugerează diviziunea spectrului infraroșu de la 3960 la 400 cm^{-1} în 640 intervale de $5,6\text{ cm}^{-1}$ lățime. Intensitatea x , a radiației transmise de intervalul i este scalată la:

$$x_i = 1.00 - (\%T)/100.0 \quad (6.4)$$

unde T este factorul de transmisie. Numărul mare de unități de intrare a fost redus experimental la 256, pentru a elimina răspunsurile neconcludente. O altă îmbunătățire a metodei constă în îngustarea lățimii intervalelor la joasă frecvență și lărgirea lor la înaltă frecvență, cu scopul de a le adapta la puterea de discriminare variabilă de la un capăt al altuia al spectrului. În consecință, dependența de frecvență a lungimii intervalului i este dată de următoarea expresie:

$$i = E(6.0\sqrt{\nu} - 120.0) \quad (6.5)$$

unde $E(\cdot)$ este funcția parte întreagă, iar ν este frecvența. Domeniul de frecvență asignat intrării i este apoi explorat și sunt înregistrate vârfurile în amplitudine. Dacă un maxim este prezent într-un interval, atunci amplitudinea sa este scalată între 0.000 și 1.000, în caz contrar valoarea intrării este setată la *zero*.

Experimental, compușii au fost descriși în termenii a 36 grupări funcționale reprezentative (alcooli primari, fenoli, amine, aminoacizi, esteri,...), care au fost alese din 140 grupări funcționale studiate anterior [26]. Fiecărei grupări i s-a repartizat o unitate de ieșire arbitrară, astfel încât răspunsul țintă devine un vector 36-variant cu componente binare în care valoarea 1 indică prezența grupării funcționale asociate iar 0 indică absența sa. Pe scurt, esența acestei probleme 256-variante cu 36-răspunsuri este de a stabili prezența sau absența a oricărei grupări funcționale din cele 36 considerate relevante pe baza unor spectre diferite alcătuite din 256 puncte. Majoritatea compușilor analizați conțin mai multe grupări

funcționale, astfel încât multe unități de ieșire sunt simultan active. Prin încercări s-a ajuns la arhitectura optimă a rețelei: $(256 \times 34 \times 36)$. Biblioteca de spectre a cuprins 2915 compuși, dintre care au fost aleși aleatoriu 2499 pentru setul de instruire și restul de 416 pentru setul de test. A fost respectată astfel o regulă empirică generală referitoare la instruirea rețelelor mari utilizând algoritmul cu propagarea inversă a erorii, și anume ca numărul datelor alese pentru instruire să depășească numărul ponderilor sinaptice modificabile ale rețelei. În cazul studiat, setul de instruire conține aproximativ 650.000 valori ($2.499 \text{ spectre} \times 256 \text{ intensități}$), în timp ce numărul coeficienților modificabili este puțin sub 350.000 ($256 \times 34 \times 36$).

Rețelele neuronale s-au dovedit, deși pe un număr restrâns de experimente publicate până în prezent, a fi extrem de promițătoare în competiția cu alte sisteme expert dedicate elucidării structurale a compuşilor chimici prin prelucrare automată a datelor. Mai mult, rețelele neuronale pot fi combinate cu sisteme expert și implementa o combinație de metode care, adesea, se dovedește mai tare decât fiecare în parte. Domeniul de explorare a sistemului expert se limitează la compuși pentru care răspunsurile rețelei nu sunt concludente [27, 28].

BIBLIOGRAFIE

1. E. Cherniak și D. McDermott, *Introduction to Artificial Intelligence*, Addison-Wesley, Reading, MA, 1985.
2. R. Mutihac și L. Mutihac, *Roum. Chem. Quart. Rev.*, **3**, 329, 1995.

3. R. Mutihac și L. Mutihac, *Roum. Chem. Quart. Rev.*, **1**, 67, 1997.
4. P. Smolensky, în *Parallel Distributed Processing, Exploration in the Microstructure of Cognition* (D. Rumelhart și J McClelland, Eds.), The MIT Press, Cambridge, MA, 1986.
5. M. Otto, *Chemometrics, Statistics and Computer Application in Analytical Chemistry*, Wiley-VCH, Weinheim, 1999.
6. R. Mutihac and V. Gheorghe, *Rev. Roum. Phys.*, **37**, No. 5, 535, 1992.
7. R. Mutihac, *Romanian Reports in Physics*, **47**, No. 1, 76, 1995.
8. L. Mutihac și V. David, *Chemometria*, Ed. Universității din București, București, 1997.
9. J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New-York, 1982.
10. R. Mutihac, *St. Cerc. Fiz.*, **43**, Nos. 9-10, 603, 1991.
11. R. Mutihac, *Proc. 13th World Congress on Computation and Applied Mathematics*, Dublin, Ireland, p.1009, 1991.
12. F. Rosenblatt, *Principles of Neurodynamics*, Spartan Books, 1962.
13. W.S. McCulloch and W. Pitts, *Bull. Math. Biophys.*, **5**, 115, 1943.
14. R. Mutihac, *Modelarea și simularea rețelelor neuronale*, Teza de doctorat, Universitatea București, 1994.
15. S.M. Deming și S.L. Morgan, *Experimental Design: A Chemometric Approach*, Elsevier, Amsterdam, NL, 1987.
16. K. Varmuza, *Pattern Recognition in Chemistry*, Springer-Verlag, Berlin, 1980.

17. J. Zupan și J. Gasteiger, *Neural Networks for Chemists*, VCH, Weinheim, 1991.
18. D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, și L. Kaufmann, *Chemometrics: A Textbook*, Elsevier, Amsterdam, NL, 1988.
19. V. Kvasnička, *Neur. Net. World*, **2**, 171, 1993.
20. R.A. Jacobs, M.I. Jordan, S.J. Nowlan și G.E. Hinton, *Neur. Comp.*, **3**, 79, 1991.
21. W. Kabsch și C. Sander, *Dictionary of Protein Secondary Structure: Pattern of Hydrogen-Bonded and Geometrical Features*, *Biopolymers*, **22**, 2577, 1983.
22. M. Compiani, P. Fariselli, și R. Casadio, in *Third Italian Workshop on Parallel Architectures and Neural Networks* (E.R. Caianiello, Ed.), World Scientific Publishing Co., 1990.
23. J. Garnier, D.J. Osguthorpe, și B. Robson, *J. Mol. Biol.*, **120**, 97, 1978.
24. E.W. Robb și M.E. Munk, *Mikrochim Acta*, **1**, 131, 1990.
25. R. Mutihac, *Modelarea și Simularea Neuronală*, Ed. Universității din București, București, 2000.
26. M.E. Munk, M.S. Madison și E.W. Robb, *Mikrochim Acta* **11**, 505, 1991.
27. W.J. Melssen, J.R.M. Smits, G.H. Rolf, și G. Kateman, *Chemom. Intell. Lab. Syst.*, 1993.
28. M. Novic și J. Zupan, *Vestn. Slov. Kem. Drus.*, **39**, 195, 1992.

Tiparul s-a executat sub c-da nr. 661/2000, la
Tipografia Editurii Universității din București

VERIFICAT

VERIFICAT
2017

ISBN: 973-575-437-1

<https://biblioteca-digitala.ro> / <https://unibuc.ro>

Lei 25.500