

Școala Doctorală de Fizică

Dirrecția de studii

***Fizică atomică, Fizică nucleară,
Fizica particulelor elementare, Astrofizică***

Fizică teoretică și experimentală

- curs general -

Tema I

***Statistică pentru Fizică nucleară și Fizica particulelor
elementare***

ERORI EXPERIMENTALE.

METODE DE ÎNREGISTRARE A DATELOR EXPERIMENTALE

I.1. Definiții. Tipuri de erori. Metode de aproximare

Definiție: Studiul măsurătorilor fizice are ca obiect dezvoltarea posibilităților de a concepe experimente adecvate pentru înțelegerea fenomenelor fizice, furnizarea de tipuri speciale de "instrumente" mentale și dezvoltarea de tipuri speciale de atitudini mentale care rezultă din forma corespunzătoare și analiza diferitelor tipuri de măsurători cu privire, în principal, la precizia și acuratețea (corectitudinea) lor.

Orice **experiment științific** se bazează pe **măsurători**. Analiza experimentelor conduce la fapte științifice care pot sau nu să fie puse sub semnul întrebării. *Același fapt științific poate fi pus în evidență prin diferite forme de investigare și, de aceea, este necesar ca oamenii de știință să aibă un limbaj comun în prezentarea rezultatelor experimentale.* Pentru aceasta *este necesar să existe metode precise și repetabile de prelucrare a datelor experimentale* [1-5].

Fizicienii trebuie să aibă la îndemână metodele consacrate de investigare și prelucrarea a datelor experimentale și de prezentare a rezultatelor experimentale, trebuie să cunoască și să folosească astfel de metode, având în vedere faptul că, în prezent, *există metode și căi de obținere a unor rezultate experimentale sigure*.

Un prim pas pe calea stabilirii unor rezultate experimentale sigure îl reprezintă distingerea între **erorile care afectează o măsurătoare fizică** și **greșelile care se pot face la realizarea măsurătorilor**.

Greșelile sunt datorate *neatenției, neglijenței sau incompetenței experimentatorului*. **Erorile** sunt inerente oricărei metode sau tehnici de măsurare. Pentru reducerea sau eliminarea erorilor există o serie de metode speciale.

Erorile se pot clasifica în două categorii mari:

(a) erori sistematice;

(b) erori aleatoare (statistice).

Erorile sistematice se pot clasifica, la rândul lor, în următoarele tipuri: (i) *erori teoretice*; (ii) *erori instrumentale*; (iii) *erori personale*.

Erorile sistematice de pot fi reduse, corectate sau chiar înlăturate. Pentru toate acestea există metode speciale.

Erorile statistice sunt datorate fluctuațiilor. În cazul reducerii la minimul posibil sau eliminării erorilor sistematice se poate afirma că principala sursă de eroare și de imprecizie asupra unor măsurători fizice o reprezintă erorile statistice. Acest fapt impune acordarea unei atenții deosebite acestui tip de eroare și metodelor de calculare statistice asociate pentru obținerea de rezultate experimentale cât mai sigure și precise.

Pentru a avea posibilitatea analizării corecte a datelor experimentale trebuie să fie respectate o serie de aspecte de interes la colectarea acestora. Un prim aspect de interes este legat de *modul de înregistrare a datelor experimentale*. Aici trebuie avute în vedere eliminarea evenimentelor care sunt afectate de greșeli în timpul măsurătorilor fizice, precum și a celor afectate de erori prea mari. De asemenea, este necesară respectarea cu strictețe a procedurilor de măsurare, citire și înregistrare a datelor experimentale. Un alt aspect este determinat de *modul de scriere a datelor experimentale* și de legătura dintre forma de scriere și *eroarea de citire specifică aparaturii folosite în experiment*. Trebuie avut în vedere faptul că o practică comună este ca *eroarea de citire a unui instrument să fie considerată diviziunea cea mai mică posibilă și observabilă în experiment*.

Corectitudinea unei măsurători poate fi descrisă folosind 2 termeni: **(a) acuratețea (exactitatea)**; **(b) precizia**. În general, *noțiunea de acuratețe este legată de erorile sistematice, iar noțiunea de precizie de erorile statistice*.

În prezent nu există un experiment care să nu fie afectat de erori. De aceea, *nu se poate determina valoarea adevărată a unei mărimi și numai o valoare care se stabilește cu o anumită acuratețe sau precizie*. Acestea din urmă impun un anumit număr de cifre semnificative la scrierea valorii mărimii fizice determinate experimental. Dacă această valoare este folosită în diferite calcule este necesar ca numărul de cifre semnificative să se conserve. Acolo unde este cazul, după calcule, *se va proceda la rotunjiri pentru a păstra numărul de cifre semnificative*. Păstrarea numărului de cifre semnificative, precum și rotunjirea numerelor se face cu respectarea unor **reguli** care permit să nu se introducă erori suplimentare semnificative asupra rezultatelor finale.

Pentru aceasta este necesar să se ia în considerare următoarele relații de calcul pentru cazul în care se folosesc mărimi fizice determinate în experimente:

$$(1+x)^n = 1 + nx + n(n-1)x^2/2 + \dots$$

$$(1+a)^l(1+b)^m(1+c)^n = 1 + la + mb + nc, \quad a, b, c \ll 1, \quad l, m, n < 5$$

$$1/(1+a) = 1-a$$

$$(1+a)^{1/2} = 1+a/2$$

$$(1+a)/(1+b) = 1+a-b$$

$$(A^2+d)^{1/2} = A+d/2A$$

O altă serie de aproximații se bazează pe calculul diferențial. De obicei, se folosesc primii 2-3 termeni din dezvoltarea în serie Taylor.

Fie o dependență de tipul $y = f(x)$. Dacă se cunoaște o valoare y_1 a funcției y , pentru o valoare x_1 a lui x , atunci valoarea lui y la $x_1+\Delta x$ se poate scrie astfel:

$$y_2 \approx y_1 + (dy/dx)_1 \cdot \Delta x + (1/2) \cdot (d^2y/dx^2)_1 \cdot (\Delta x)^2 + \dots \quad (I.1)$$

În general, primii doi termeni sunt suficienți. În relația (III.1) termenul $(dy/dx)_1$ reprezintă rata de creștere a funcției $y = f(x)$ la o creștere a variabilei x , considerând o valoare particulară a variabilei x , anume x_1 .

Metoda de aproximare prezentată mai sus se poate aplica și în cazul funcțiilor de mai multe variabile, $f(x_1, x_2, \dots, x_n)$. În acest caz rata de creștere se poate scrie în modul următor:

$$\Delta y = (\partial f / \partial x_1)_{x'} \cdot \Delta x_1 + (\partial f / \partial x_2)_{x'} \cdot \Delta x_2 + \dots + (\partial f / \partial x_n)_{x'} \cdot \Delta x_n, \quad (I.2)$$

unde x' este setul de valori pentru care se calculează derivatele parțiale.

I.2. Analiza grafică

După culegerea/înregistrarea datelor experimentale un pas important în analiza măsurărilor fizice este **reprezentarea grafică**. De cele mai multe ori reprezentării grafice îi este asociată și *reprezentarea curbei care fit-ează cel mai bine punctele experimentale* incluse. În acest mod se obține o imagine mai clară a asupra experimentului și se oferă posibilitatea repetării lui. *Formele curbelor de fit obținute pot servi la verificarea legilor existente sau pot sugera legi noi.*

Curba de fit dă legătura dintre variabilele măsurate. De obicei, curba de fit se trasează printre punctele experimentale. La trasarea ei se respectă anumite reguli și se folosesc anumite metode specifice.

Cel mai important aspect este să se găsească o *ecuație matematică* care să fit-eze curba respectivă. În acest mod se poate obține mult mai multă informație. La obținerea ecuațiilor matematice se pleacă de la cea mai simplă formă - cea a *liniei drepte* - mergând spre forme din ce în ce mai complicate. Legea liniei drepte este destul de des întâlnită în Fizică și, în particular, în Fizica nucleară, deoarece numeroase date experimentale urmează în mod natural o astfel de dependență sau pot fi puse într-o formă care să urmeze o astfel de dependență. De exemplu, timpul de înjumătățire se poate obține din curbele de dezintegrare punând relația dintre vitezele de numărare adevărate - $R = R_0 e^{-\lambda t}$ - sub forma $\ln R = \ln R_0 - \lambda t$ [6].

Multe din legile Fizicii nu sunt însă liniare. În acest caz sunt două aspecte care trebuie avute în vedere, anume:

(a) *legea neliniară de variație este cunoscută din considerente teoretice*; în acest caz problema care se pune este aceea de a stabili *constantele ecuației matematice prin fit-area datelor experimentale*;

(b) *legea neliniară de variație nu este cunoscută*; de data aceasta se pune problema efectuării unei *aproximații empirice la datele experimentale*.

Remarcă. Introducerea de mai mulți termeni în funcția de fit mărește posibilitatea de a sesiza imprecizia aproximației utilizate în problemele de tip (b).

În analiza datelor experimentale un rol fundamental îl au metodele statistice. Ele dau o metodă clară de a construi o linie dreaptă ca rezultat al unui fit la datele experimentale (metoda celor mai mici pătrate, de exemplu) și pun la dispoziția fizicianului testele necesare pentru stabilirea unui fit corect în toate situațiile [1-6].

Pentru o mai corectă înțelegere a acestor aspecte în cele ce urmează vor fi prezentate unele aspecte legate de noțiuni de teoria probabilităților și statistică matematică.

Bibliografie

- [1]. H.G.Worthing, J.Geffner - Prelucrarea datelor experimentale, Editura Tehnică, București, 1959
- [2]. B.R.Martin - Statistics for Physicists - Academic Press, London and New York, 1971
- [3]. A.Solmitz - Annual Review of Nuclear Science (1963)
- [4]. W.T.Eadie et al - Statistical Methods in Experimental Physics, North-Holland Publishing Company, Amsterdam, 1971
- [5]. F.James - Proceedings of the 1970 CERN Computing and Data Processing School - Via Monastero, Varenna, Italy, 30 August-12 September 1970 - Preprint CERN 71-6 (1971)
- [6]. Colectiv de catedră - Fizică nucleară - îndrumător de laborator, Tipografia Universității București, 1987
- [7]. Louis Lyons – Statistics for nuclear and particle physicists – Cambridge University Press, 1992

NOȚIUNI DE TEORIA PROBABILITĂȚILOR

II.1. Noțiuni fundamentale

Pentru analiza datelor experimentale, prelucrarea lor și prezentarea rezultatelor experimentale este importantă definirea noțiunii de **probabilitate**. Trebuie avute în vedere două căi de definire a probabilității: *calea matematică, respectiv, calea fizică*. Pentru discutarea acestor probleme este necesară definirea unor noțiuni [1-6].

Fie un set de *condiții inițiale, reproductibile*, care definesc un *experiment*. Prin realizarea unei *observații* sau a unui *set de observații* se produce un *efect (rezultat) al experimentului*. Fie x_i , cu $i = 1, 2, \dots, n$, rezultatele experimentului. Trebuie menționat că mărimile x_i pot fi numere sau seturi de numere.

Definiție *Setul tuturor rezultatelor posibile $\{x_i\}$ ($i = 1, 2, \dots, n$) ale unui experiment se numește spațiul probelor sau populație, iar x_i este un punct din acest spațiu. Se notează în modul următor:*

$$S = \{x_i / i = 1, 2, \dots, n\}$$

Definiție *Un subset de puncte din populație $\{x_k\}$, cu $k = 1, 2, \dots, m$, unde $m < n$, se numește eveniment. Se notează astfel: $E = \{x_k / k = 1, 2, \dots, m\}$.*

Atunci când $m = n$ toată populația este inclusă în eveniment. **Realizarea unui eveniment** înseamnă că un punct din populație este inclus în subsetul de puncte din populație care definesc un anumit eveniment.

Calea matematică presupune definirea unei populații cu proprietăți specifice. În acest caz, teoria probabilităților se dezvoltă *axiomatic* și implică *stabilirea exactă a parametrilor și naturii populației* [7,8].

Calea fizică este strâns legată de situațiile reale, situații în care parametrii și natura populației sunt foarte rar cunoscute. De aceea, *scopul analizei statistice este tocmai acela de a stabili natura populației din care face parte eșantionul (proba, mostra, ...) de date experimentale, precum și valorile parametrilor populației*. În acest mod se încearcă găsirea acelei expresii matematice care descrie corect o anumită situație când se cunoaște o anumită parte a populației [1-4]. În acest caz se introduc probabilități operaționale, iar rezultatele obținute se prezintă în termenii acestor probabilități.

Definiție Se consideră o secvență de n încercări (*extrageri, probe*) în care evenimentul E se realizează de n_E ori. Raportul n_E/n se numește **frecvență relativă a unui eveniment E** , de clasă dată. Se notează cu $R[E]$.

Probabilitatea $P[E]$ a unui eveniment E este limita lui $R[E]$, când n crește nedefinit, presupunând că limita există.

Aceasta este definiția fizică a probabilității. Limitările sunt determinate de faptul că se poate realiza doar un număr finit de încercări (extrageri).

În terminologia curentă se mai întâlnește noțiunea de **probabilitate "a posteriori"**, respectiv, cea de **probabilitate "a priori"**. Prima este legată de *observațiile experimentale*, iar cea de a doua de *modelarea matematică* a unui eveniment.

Conform definițiilor și comentariilor de mai sus se poate defini **probabilitatea unui eveniment E** ca un număr cuprins în intervalul închis $[0,1]$ pentru care se realizează condiția:

$$0 \leq P[E] \leq 1.$$

Dacă $E \equiv S$, atunci $P[E] = 1$.

Complementul unui eveniment E se notează prin E^* .

Definirea evenimentelor s-a făcut folosind noțiuni specifice mulțimilor. De aceea se poate defini intersecția și reuniunea a două evenimente. Rezultatul intersecției a două evenimente A și B este un eveniment de tip "A sau B", iar reuniunea acestor evenimente dă un eveniment de tip "A și B". Ele au reflectări diferite în teoria probabilităților. *Două evenimente sunt distincte* dacă *intersecția* lor este mulțimea vidă.

Se poate defini o **probabilitate condițională**, anume: dacă un eveniment poate rezulta din n efecte reciproc exclusive - realizarea unui eveniment exclude realizarea celorlalte - și egale ca posibilități de realizare, din care n_B corespund la realizarea evenimentului B , iar n_{AB} corespund la realizarea evenimentului A , în condițiile în care evenimentul B s-a realizat, atunci probabilitatea unui eveniment A obținut după realizarea unui eveniment B este:

$$P[A / B] = \frac{n_{AB}}{n_A}$$

și se numește probabilitate condițională a evenimentului A .

Expresia probabilității condiționale a lui A se mai poate scrie astfel:

$$P[A / B] = \frac{P[A \cap B]}{P[B]}$$

Evenimentele se pot clasifica după diferite criterii. Fie A, B, C trei criterii de clasificare. În aceste condiții se poate defini **probabilitatea marginală**.

Dacă clasificările în criterii sunt $A_1, A_2, \dots, A_r, B_1, B_2, \dots, B_s$ și C_1, C_2, \dots, C_t , iar condiția:

$$\sum_{j=1}^r P[A_j] = \sum_{k=1}^s P[B_k] = \sum_{l=1}^t P[C_l] = 1$$

este îndeplinită, atunci probabilitatea marginală a lui A_j și C_l se definește astfel:

$$P[A_j \cap C_l] = \sum_{k=1}^s P[A_j \cap B_k \cap C_l].$$

Se poate defini și probabilitatea marginală a lui C_l prin relația următoare:

$$P[C_l] = \sum_{j=1}^r \sum_{k=1}^s P[A_j \cap B_k \cap C_l] = \sum_{j=1}^r P[A_j \cap C_l] = \sum_{k=1}^s P[B_k \cap C_l].$$

Pe baza noțiunilor definite până în prezent se poate defini **independența evenimentelor**, astfel:

Evenimentul A este independent de evenimentul B dacă $P[A \cap B] = P[A] \cdot P[B]$.

Folosind relația de definiție de mai sus se pot scrie următoarele relații:

$$P[A^*] = 1 - P[A],$$

$$P[A \cap B] = P[A] \cdot P[B/A] = P[A] \cdot P[B],$$

$$P[A \cup B] = P[A] + P[B] - P[A \cap B].$$

Dacă evenimentele A și B sunt independente se poate scrie următoarea relație:

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] = P[A] + P[B].$$

Pentru analiza și prelucrarea datelor experimentale *teorema lui Thomas Bayes* - care datează din anul 1763 - este destul de des folosită. Enunțul acestei teoreme este următorul: Dacă B_i ($i = 1, 2, \dots, n$) sunt evenimente *exclusive reciproc și exhaustive* - adică, toate evenimentele posibile sunt incluse în B_i - și dacă evenimentul A se poate realiza numai în combinație cu unul din cele n evenimente B_i , atunci:

$$P[B_i / A] = \frac{P[B_i] \cdot P[A / B_i]}{\sum_{j=1}^n (P[B_j] \cdot P[A / B_j])}$$

Teorema Bayes dă *probabilitatea "a posteriori"* de a avea evenimentul B_i când evenimentul A este cunoscut și realizat. Mărimea $P[B_i/A]$ se numește **verosimilitate**. Se alege, în general, acea *situație care are cea mai mare probabilitate "a posteriori"* și, de aceea,

metoda se mai numește *metoda verosimilității maxime*. Pentru folosirea metodei este necesar să se cunoască și probabilitățile "a priori" $P[B_i]$. Trebuie menționat faptul că aceste probabilități sunt - pentru cele mai multe situații de interes - necunoscute. Prin teorema Bayes toate probabilitățile "a priori" sunt luate egale.

Această teoremă, alături de aranjamente, permutări și combinații, este de mare utilitate în procesul complex și delicat al *deducerii statistice*.

II.2. Parametrii populației

Într-un experiment *nu* se dispune de o populație completă ci numai de diferite probe (eșantioane, mostre) care reprezintă submulțimi (subseturi) ale populației totale. *Problema fizică care se pune este cea a estimării proprietăților pornind de la natura probei prin deducție statistică.*

Printre cei mai folosiți parametri ai populației se numără: *media aritmetică, mediana (valoarea mediană), modul, abaterea medie, varianța, abaterea standard, momentele asociate de diferite ordine.*

Media aritmetică a unui set de N valori x_i ($i = 1, 2, \dots, N$) se definește prin relația de mai jos:

$$m_a = \frac{\sum_{j=1}^N x_j}{N}. \quad (II.1)$$

*Dacă mărimile x_1, x_2, \dots, x_N sunt aranjate în ordine crescătoare sau descrescătoare și sunt renumerotate ca $x_{(1)}, x_{(2)}, \dots, x_{(N)}$ se definește **mediana** ca valoarea de mijloc a noului set - pentru N număr impar - respectiv, ca valoarea de mijloc a perechii mijlocii - pentru N număr par.*

Un alt parametru de interes este **modul**. Acesta reprezintă acea valoare din setul de x_1, x_2, \dots, x_N care se realizează cu frecvență maximă.

Pentru a avea o măsură a dispersiei datelor și rezultatelor experimentale se pot folosi mai mulți parametri ai populației. Ca și cei definiți anterior ei dau o măsură a localizării.

*Media aritmetică a valorilor absolute ale abaterilor observațiilor de la mediană (m_m) se numește **abatere medie** și are următoarea expresie:*

$$\delta_m = \frac{\sum_{j=1}^N |x_j - m_m|}{N}. \quad (II.2)$$

Varianța unei populații - notată prin σ^2 - se definește ca media aritmetică a abaterilor mărimilor x_i , din setul dat, de la media aritmetică, m_a . Relația de definiție are următoarea formă:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - m_a)^2}{N}. \quad (II.3)$$

De interes în analiza datelor experimentale și în prezentarea rezultatelor experimentale este **abaterea standard**, σ . Ea se definește ca rădăcina pătrată a varianței.

O altă mărime de interes este **coeficientul de variație**, definit ca raportul dintre abaterea standard și media aritmetică, anume σ/m_a .

Alături de mărimile menționate mai sus, de mare interes în analiza datelor experimentale și în obținerea de informații dinamice în ciocniri nucleare la diferite energii, cu deosebire la energii relativiste, sunt momentele asociate unei distribuții de probabilitate specifice unei anumite populații. Se folosesc mai multe tipuri de momente. Dintre aceste de mare interes sunt **momentele simple (ordinare)** și **momentele factoriale**.

Dacă momentele simple sunt calculate în raport cu un punct arbitrar m se obțin **momentele simple (ordinare) necentrate** definite astfel:

$$m'_k = \frac{\sum_{i=1}^N (x_i - m)^k}{N}. \quad (II.4)$$

Trebuie subliniat aici că momentul simplu necentrat de ordinul întâi este egal cu valoarea medie (media aritmetică).

Atunci când punctul ales este chiar valoarea medie m_a se obțin **momentele simple (ordinare) centrate**:

$$m_k = \frac{\sum_{i=1}^N (x_i - m_a)^k}{N}. \quad (II.5)$$

Trebuie menționat aici faptul că între cele două tipuri de momente simple există următoarele relații de recurență:

$$m_k = \sum_{j=0}^k C_k^j m_{k-j}' (-m_1')^j, \quad (II.6.1)$$

$$m_k' = \sum_{j=0}^k C_k^j m_{k-j} (-m_1)^j. \quad (II.6.2)$$

Momentele factoriale se definesc prin relația următoare:

$$\langle (n)_k \rangle = \sum_{n \geq k} (n)_k p_n, \quad (II.7)$$

unde $(n)_k = n(n-1)\dots(n-k+1)$.

Caracteristicile generale ale populației sunt reflectate și de câțiva parametri care pot fi definiți în funcție de valorile momentelor asociate [2-5,8,9]. Fiind determinați de forma distribuției de probabilitate care descrie populația ei pot fi legați de indicatorii de formă [2-5,8,9].

Parametrul de asimetrie se definește prin următorul raport:

$$\beta_1 = m_3^2 / m_2^3. \quad (II.8)$$

La definirea acestui parametru care indică abaterea de la forma simetrică a populației s-a avut în vedere faptul că pentru o populație distribuită simetric în jurul valorii medii momentul simplu centrat de ordinul al III-lea este nul ($m_3 = 0$).

Un alt parametru important este **parametrul de formare de maxime**. Ele se poate defini tot cu ajutorul momentelor simple centrate de ordin superior. Relația de definiție este următoarea:

$$\beta_2 = m_4 / m_2^2. \quad (II.9)$$

Acest parametru ia valori standard pentru populații diferite.

Toți parametrii menționați anterior sunt extrem de utili în analiza statistică a datelor experimentale, precum și în descrierea dinamicii diferitelor ciocniri hadronice [9-11]. Ei sunt strâns legați de noțiunea de distribuție, în general, și de distribuție de probabilitate, în particular. De aceea, în cele ce urmează vor fi abordate câteva aspecte legate de această noțiune.

II.3. Distribuții pentru populații. Legături cu momente și cumulanți

Noțiunea de distribuție este strâns legată de noțiunea de **variabilă aleatoare**. Se definește variabila aleatoare ca o funcție care poate lua o valoare definită în orice punct din

populație (spațiul probelor).

Fie o populație SP cu o funcție de probabilitate P și o variabilă aleatoare X care este definită în populația respectivă. În aceste condiții pentru fiecare punct din populație (spațiul probelor) - $x_i \in SP$ - se poate stabili o probabilitate $P[x_i]$ și o valoare numerică definită, $X(x_i)$, pentru o variabilă aleatoare. Variabila aleatoare poate fi *continuă* sau *discretă*.

Pentru o variabilă aleatoare continuă x se poate introduce o **funcție de densitate de probabilitate** (funcție de densitate), $f(x)$. Acest lucru este posibil numai dacă sunt satisfăcute următoarele condiții:

(i) $f(x)$ este un număr real, nenegativ, unic, pentru toate valorile reale ale lui x ;

(ii) $f(x)$ este normată la unitate, anume:

$$\int_{-\infty}^{+\infty} f(x)dx = 1, \quad (II.10);$$

(iii) probabilitatea cu care x cade între orice două valori reale a și b - pentru care $a < b$ - este dată de relația următoare:

$$P[a \leq x \leq b] = \int_a^b f(x)dx. \quad (II.11)$$

Se poate asocia și o **funcția de distribuție cumulativă** unei variabile aleatoare continue x . Ea se definește prin relația:

$$F(x) = \int_{-\infty}^x f(u)du. \quad (II.12)$$

Din relațiile de mai sus se poate deduce că probabilitatea ca un membru ales din întâmplare dintr-o distribuție să aibă valoarea x este chiar funcția de densitate $f(x)$. De asemenea, $F(x)$ este o funcție nedescrescătoare de x cu valori în intervalul $[0,1]$.

Folosind funcția de densitate definită mai sus se pot scrie expresiile unor parametri ai populației definiți în subcapitolul IV.2., anume:

(a) media în jurul unui punct arbitrar m :

$$m_m = \int_{-\infty}^{+\infty} f(x)(x - m)dx, \quad (II.13)$$

(b) varianța:

$$\sigma^2 = \int_{-\infty}^{+\infty} f(x)(x - m_a)^2 dx, \quad (II.14)$$

(c) momentele simple, centrate și necentrate, de ordin k:

$$m_k = \int_{-\infty}^{+\infty} f(x)(x - m_a)^k dx, \quad (II.15.1)$$

$$m'_k = \int_{-\infty}^{+\infty} f(x)(x - m)^k dx. \quad (II.15.2)$$

Remarcă. Pentru variabilele discrete se folosesc relații de definiție similare în care integralele trec în sume. De exemplu,

$$m_k = \sum_{j=1}^{\infty} f(x_j)(x_j - m_a)^k, \quad (II.15.1')$$

$$m'_k = \sum_{j=1}^{\infty} f(x_j)(x_j - m)^k. \quad (II.15.2')$$

În analiza statistică a datelor experimentale este de interes cunoașterea **valorii așteptate** pentru un anumit tip de populație. Pentru o variabilă aleatoare continuă x care are o funcție de densitate $f(x)$ **valoare așteptată a lui x , $A[x]$, se poate defini astfel:**

$$A[x] = \int_{-\infty}^x xf(x)dx, \quad (II.16)$$

Pentru o funcție $g(x)$ a lui x se poate scrie:

$$A[g(x)] = \int_{-\infty}^x g(x)f(x)dx, \quad (II.17)$$

Din relațiile de mai sus rezultă următoarele relații de legătură:

$$\begin{aligned} A[c] &= c, \\ A[cg(x)] &= cA[g(x)], \\ A[g_1(x) + g_2(x)] &= A[g_1(x)] + A[g_2(x)], \\ A[g_1(x) \cdot g_2(x)] &= A[g_1(x)] \cdot A[g_2(x)]. \end{aligned} \quad (II.18)$$

unde $c = \text{constantă}$.

Relații similare se pot scrie pentru momentele de diferite tipuri și diferite ordine.

Cunoașterea primelor câteva momente, în practică, determină caracteristicile esențiale ale distribuției. De aceea, este util să se stabilească o metodă generală de determinare a momentelor de orice ordin. Pentru aceasta este necesară introducerea unei funcții speciale, numită **funcție generatoare de momente (f.g.m.)**.

Funcția generatoare de momente simple necentrate se definește astfel, dacă variabila

aleatoare x are funcția de densitate $f(x)$:

$$M_x(z) = A[e^{xz}] = \int_{-\infty}^{+\infty} e^{xz} f(x) dx \quad . \quad (II.19)$$

Pentru momentele de diferite ordine se dezvoltă în serie e^{xz} și se obține, dacă $m = 0$:

$$M_x(z) = A[1 + xz + \frac{1}{2!}(xz)^2 + \dots] = \sum_{n=0}^{\infty} \frac{1}{n!} m'_n z^n \quad . \quad (II.20)$$

Dacă relația (IV.20) se diferențiază de n ori și se calculează pentru $z = 0$, atunci se obține următoarea relație generală pentru *momentele simple necentrate de ordin n* :

$$m'_n = \left. \frac{\partial^n M_x(z)}{\partial z^n} \right|_{z=0} \quad . \quad (II.21)$$

Funcția generatoare de momente simple, în jurul oricărui punct m , se poate scrie astfel:

$$M_x(z) = A[e^{(x-m)z}] \quad . \quad (II.22)$$

Pentru funcția generatoare de momente simple centrate se definește în modul următor:

$$M_m(z) = e^{-mz} M_x(z) \quad . \quad (II.23)$$

Logaritmiile funcțiilor generatoare de momente sunt folosiți pentru definirea *cumulanților de diferite ordine*. Fie dezvoltarea în serie Taylor a $\ln M_x(z)$:

$$\ln M_x(z) = k_1 z + k_2 \frac{z^2}{2!} + \dots \quad , \quad (II.24)$$

unde $k_i = \left. \frac{\partial^i M_x(z)}{\partial z^i} \right|_{z=0}$ reprezintă cumulanții de ordin i . Pentru fiecare tip de moment se pot defini cumulanți corespunzători.

Există distribuții pentru care nu se pot defini astfel de funcții simple. În aceste situații se introduce **funcția caracteristică**, $\Phi_x(t)$. Dacă variabila aleatoare x are funcția de densitate $f(x)$, atunci se poate defini următoarea funcție caracteristică:

$$\Phi_x(z) = A[e^{iz}] = \int_{-\infty}^{+\infty} f(x)e^{iz} dx = M_x(iz) \quad . \quad (II.25)$$

Legătura dintre funcția de densitate și funcția caracteristică este dată de teorema următoare, numită teorema de inversie: dacă $f(x)$ este o funcție de densitate cu o funcție de distribuție continuă peste tot și are o funcție caracteristică $\Phi_x(t)$, definită prin relația (II.25), atunci:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi_x(z)e^{-iz} dz \quad . \quad (IV.26)$$

Relația (IV.26) reprezintă transformata Fourier.

Observații

1. Toate mărimile și noțiunile introduse până în prezent se pot extinde și pentru distribuții de mai multe variabile. În acest caz variabila aleatoare x devine un vector de n componente, iar integralele, respectiv, sumările se vor face în spații cu n dimensiuni, respectiv, după n indici.

2. Dacă variabila aleatoare este o funcție de o variabilă x , $y(\{x\})$, iar $y(\{x\})$ este o funcție monotonă, atunci funcția de densitate se poate fi scrisă sub forma următoare:

$$f(y\{x\}) = f(x\{y\}) \left| \frac{dx}{dy} \right| \quad . \quad (II.27)$$

3. Există cazuri în care funcția de densitate se calculează numai dacă sunt satisfăcute condițiile:

(i) $dy/dx \neq 0$;

(ii) $y = y(\{x\})$ are o soluție reală, finită, iar expresia este de forma:

$$f(y\{x\}) = \prod_{\forall x} f(x\{y\}) \left| \frac{dy}{dx} \right|^{-1} \quad . \quad (IV.28)$$

În cazurile în care condițiile de mai sus nu sunt respectate $f(y\{x\}) = 0$.

Din multitudinea de distribuții folosite în Fizica nucleară, Fizica particulelor elementare și Fizica nucleară relativistă cele mai des folosite sunt: **distribuția Poisson, distribuția binomială, distribuția Gauss și distribuția binomială negativă [1-9]**. În multe situații de interes sunt utile combinații ale acestor distribuții [2,4,5,9,11,12]. Unele aspecte de interes legate de aceste distribuții vor fi considerate în curs și în diferitele lucrări de laborator incluse în acest manual.

Bibliografie

- [1]. H.G.Worthing, J.Geffner - Prelucrarea datelor experimentale, Editura Tehnică, București, 1959
- [2]. B.R.Martin - Statistics for Physicists, Academic Press, London and New York, 1971
- [3]. A.Solmitz - Annual Review of Nuclear Science (1963)
- [4]. W.T.Eadie et al - Statistical Methods in Experimental Physics, North-Holland Publishing Company, Amsterdam, 1971
- [5]. F.James - Proceedings of the 1970 CERN Computing and Data Processing School - Via Monastero, Varenna, Italy, 30 August-12 September 1970 - Preprint CERN 71-6 (1971)
- [6]. Colectiv de catedră - Fizică nucleară - îndrumător de laborator, Tipografia Universității București, 1987
- [7]. B.Gndenko - Theory of probability, MIR , Moscow,1982
- [8]. Gh.Mihoc, V.Craiu - Tratat de Statistică matematică, Editura Academiei RSR, București, 1981
- [9]. P.Carruthers, C.C.Shih - International Journal of Modern Physics A2(5)(1987)1447-1547
- [10].Isac Stern - Teză de doctorat, IFIN București-Măgurele, 1981
- [11].Al.Jipa - Teză de doctorat, Facultatea de Fizică, Universitatea București, 1989
- [12].Al.Jipa, C.Beșliu, R.Zaharia, A.M.David - Journal of Physics G: Nuclear and Particle Physics 22(2)(1996)221-230

PROBE EXPERIMENTALE DIN POPULAȚII

III.1. Noțiuni fundamentale

O problemă majoră în deducerea statistică este legată de faptul că **într-un experiment nu se poate avea acces la întreaga populație**. Într-un experiment se are acces la o parte din populație, parte care se numește **eșantion (probă sau mostră) din populație** [1-8]. Din această cauză este foarte important să se aleagă corect metoda de caracterizare a probei, astfel încât concluziile asupra populației să rămână relativ stabile de la o probă la alta. **Acest lucru înseamnă că parametrii variază puțin de la o probă la alta.**

Proprietățile de dorit pentru diferite probe din populații sunt legate de o serie de definiții și teoreme.

Se definește o probă de dimensiune n ca fiind setul de valori numerice x_1, x_2, \dots, x_n pentru cele n observații selectate dintr-un set mai mare.

Statistica reprezintă o valoare numerică determinată din probă. Tot prin **statistică** se înțelege totalitatea valorilor probei.

Media unei probe de dimensiune n, $\langle m_a^p \rangle$, se calculează astfel:

$$\langle m_a^p \rangle = \frac{1}{n} \sum_{i=1}^n x_i \quad . \quad (V.1)$$

Pentru proba de dimensiune n considerată se poate calcula **varianța** folosind următoarea relație:

$$\sigma_p^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - m_a^p)^2 \quad . \quad (V.2)$$

$\sigma_p = \sqrt{\sigma_p^2}$ reprezintă **abaterea standard** a probei experimentale.

Notă. Factorul $1/N$ folosit în definirea varianței populației se înlocuiește în cazul varianței probei experimentale prin factorul $1/(n-1)$ pentru a avea asigurări că valoarea așteptată a întregii statistici de un tip dat, calculată pentru o probă experimentală de dimensiune n, va fi egală cu parametrul corespunzător pentru populație.

Fie o probă aleatoare de dimensiune n - x_1, x_2, \dots, x_n - cu o funcție de densitate $f(x)$. În acest caz, funcția de distribuție a unei probe statistice $y(x_1, x_2, \dots, x_n)$, este dată de o relație de forma următoare:

$$F(y) = \int \dots \int \prod_{j=1}^n f(x_j) dx_j \quad . \quad (III.3)$$

Observații

- (a) Integrarea se face pentru regiunea în care $y \geq y(x_1, x_2, \dots, x_n)$.
- (b) Se poate considera că $y(x_1, x_2, \dots, x_n)$ este o nouă variabilă. În acest caz se aleg (n-1) variabile - funcții de x_j - astfel încât integrandul n-dimensional din relația (V.3) să ia o formă simplă.
- (c) În foarte multe lucrări de interes din domeniu se folosește convenția următoarea: parametrii populației sunt notați cu litere grecești, iar parametrii probei experimentale din populație sunt notați cu litere latine.

III.2. Distribuții asociate probelor experimentale

Fie o probă, PS, de n observații x_j ($j = 1, 2, \dots, n$) selectate la întâmplare. Proba experimentală PS se numește **probă aleatoare cu înlocuire** sau *probă aleatoare simplă* dacă - în general - observația x_{n-1} este înapoiată populației înainte ca observația x_n să fie selectată. Dacă observația x_{n-1} nu este înapoiată populației, atunci PS este numită probă aleatoare fără înlocuire. În cele mai multe situații de interes se întâlnește cea de a doua situație.

Legăturile dintre parametrii populației și parametrii probei experimentale sunt exprimate în câteva teoreme de interes.

Teorema I. Fie N dimensiunea unei populații finite și fie n dimensiunea unei probe experimentale fără înlocuire. În acest caz, pentru toate probele experimentale de dimensiune n, media mediilor este egală cu media populației, iar varianța mediilor este egală cu varianța populației înmulțită cu un factor (N-n)/[n(N-1)].

Conform teoremei de mai sus se poate scrie:

$$\begin{aligned} \langle m_a^p \rangle &= m_a \\ \sigma_p^2 &= \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \quad . \quad (III.4) \end{aligned}$$

Remarcă. Dacă proba este cu înlocuire, atunci relațiile de mai sus, (V.4), se modifică astfel:

$$\begin{aligned} \langle m_a^p \rangle &= m_a \\ \sigma_p^2 &= \frac{\sigma^2}{n} \quad . \quad (III.5) \end{aligned}$$

Observație. Pentru populații discrete infinite se realizează numai relațiile (III.5), indiferent de tipul probei.

Pentru unele populații continue infinite este utilă următoarea teoremă:

Teorema II. Fie x o variabilă aleatoare continuă distribuită cu media $\langle m_a \rangle$, varianța σ^2 și funcția de densitate $f(x)$. Fie niște probe aleatoare de dimensiune n scoase din această distribuție. Atunci distribuția asociată mediilor are media $\langle m_a^p \rangle$ egală cu media populației, m_a , și varianța, σ_p^2 , egală cu varianța populației, σ^2 , înmulțită cu un factor $1/n$.

Conform teoremei de mai sus sunt îndeplinite relațiile:

$$\begin{aligned} \langle m_a^p \rangle &= m_a \\ \sigma_p^2 &= \frac{\sigma^2}{n}. \end{aligned} \quad . \quad (III.6)$$

Cele două teoreme conduc la următoarea concluzie: pe măsură ce dimensiune probei crește varianța mediei probei descrește, astfel încât probabilitatea ca media probei să fie o estimare bună a mediei populației crește. Această concluzie este strâns legată de **legea slabă a numerelor mari**. Enunțul acestei legi este următorul:

Fie x_i o populație de variabile aleatoare independente cu media m_a și varianță finită. Fie $\langle m_a^p \rangle$ media unei probe de dimensiune n , definită prin relația:

$$\langle m_a^p \rangle = \frac{1}{n} \sum_{j=1}^n x_j \quad . \quad (III.7)$$

Atunci, pentru orice valori date $\varepsilon > 0$ și $0 < \delta < 1$, există un număr întreg n , astfel încât, pentru toate numerele $m \geq n$, este satisfăcută relația:

$$P[|m_a^p(m) - m_a| < \varepsilon] \geq 1 - \delta \quad . \quad (III.8)$$

Legea numerelor mari este o consecință (un caz special) de inegalitate Cebîșev. Teorema asociată acestei inegalități se enunță astfel:

Fie $f(x)$ o funcție de densitate pentru o populație cu media m_a și varianță finită σ^2 . Fie p orice număr pozitiv și fie $\langle m_a^p \rangle$ media unei probe aleatoare de dimensiune n obținută din $f(x)$. În acest caz este satisfăcută relația:

$$P[|m_a^p(m) - m_a| \leq \frac{p\sigma}{\sqrt{n}}] \geq 1 - \frac{1}{p^2} \quad . \quad (III.9)$$

Teoremele enunțate anterior permit să se introducă una din cele mai importante

teoreme pentru analiza statistică, anume: **teorema limitei centrale**. Teorema se aplică atât pentru distribuții discrete cât și pentru distribuții continue și se enunță în modul următor:

Fie variabilele aleatoare independente x_i , de funcție de densitate necunoscută, identic distribuite, cu media m_a și varianța σ^2 , ambele finite. Atunci, distribuția având media probei $\langle m_a^p \rangle$ tinde la distribuția normală cu media m_a și varianța σ^2/n , când n devine mare. Dacă $u(t)$ este forma standard a distribuției normale, atunci, pentru t_1 și t_2 arbitrari, se realizează următoarea relație de legătură:

$$\lim_{n \rightarrow \infty} P\left\{t_1 \leq \frac{\langle m_a^p \rangle - m_a}{\frac{\sigma}{\sqrt{n}}} \leq t_2\right\} = \int_{t_1}^{t_2} u(t) dt \quad . \quad (III.10)$$

O altă teoremă de interes este următoarea:

Fie $l = \sum_{j=1}^n a_j x_j$, unde a_j sunt constante reale și x_j sunt variabile aleatoare cu media m_a , varianța σ^2 și covarianțe σ_{ij} ($i, j = 1, 2, \dots, n$ și $i \neq j$). Atunci

$$m_l = \sum_{j=1}^n a_j m_j \quad , \quad (III.11)$$

$$\sigma_l^2 = \sum_{j=1}^n a_j^2 \sigma_j^2 + 2 \sum_{j < k} a_j a_k \sigma_{jk} \quad . \quad (III.12)$$

Dacă variabilele aleatoare x_j sunt independente, atunci:

$$\sigma_l^2 = \sum_{j=1}^n a_j^2 \sigma_j^2 \quad . \quad (III.13)$$

Odată stabilite aceste reguli teoretice importante pentru analiza statistică este necesară găsirea unei "punți" cu diferite situații experimentale concrete.

III.3. Erori experimentale.

Formula de propagare a erorilor

După cum s-a arătat anterior **într-un experiment nu se poate determina valoarea unei mărimi cu o precizie absolută**. Cu alte cuvinte nu se pot reduce erorile făcute în măsurători la zero. În acest context este important să se găsească "*punți de legătură*" între statistica teoretică și diferitele situații experimentale și, mai ales, modalități de aplicare în situații concrete.

Înainte de a trece la aceste trebuie reamintit faptul că prin **precizie** - în statistica datelor și rezultatelor experimentale - se are în vedere micimea erorilor, iar prin **exactitatea (acuratețe, corectitudine)** se definește devierea (abaterea) observației de la valoarea "adevărată" - în ipoteza că are sens acest concept.

În mod convențional, ca măsură a erorilor aleatoare (întâmplătoare) se folosește *abaterea standard*, σ . De multe ori, în practică, ea mai este denumită și **eroare standard**. Trebuie menționat aici că în anumite situații se mai folosește și conceptul de **eroare probabilă**, definită prin următoarea relație:

$$\int_{m_a-p}^{m_a+p} f(x)dx = \frac{1}{2} \quad . \quad (III.14)$$

Determinarea valorii unei mărimi din date și rezultate experimentale afectate de diferite erori impune stabilirea unei metode sigure și repetabile de calculare sau estimare a erorii de care este afectată mărime respectivă. Această metodă poartă numele de **legea propagării erorilor**.

Fie $y=y(p)=y(p_1, p_2, \dots, p_m)$ o funcție de m parametri p_j ($j=1, 2, \dots, m$). Dacă se dorește cunoașterea erorii experimentale asupra lui y , atunci când se cunosc erorile experimentale asupra lui p_j , este necesar să se ia în considerare valorile "adevărate" pentru parametrii p_j . Fie p_j^* aceste valori "adevărate". În acest caz, dacă mărimile $(p_j - p_j^*)$ sunt mici, atunci funcția $y=y(p)=y(p_1, p_2, \dots, p_m)$ se poate dezvolta în serie Taylor în jurul punctului $p=p^*$. Se obține următoarea expresie:

$$y(p) = y(p^*) + \sum_{j=1}^m (p_j - p_j^*) \left. \frac{\partial y(p)}{\partial p_j} \right|_{p=p^*} + \dots \quad . \quad (III.15)$$

Observație. Pentru valori mici ale diferenței $(p_j - p_j^*)$ se poate considera numai aproximația de

ordinul întâi în dezvoltarea Taylor.

Varianța mărimii $y(p)$ se poate scrie sub forma următoare:

$$\text{var}[y(p)] = A[[y(p) - A[y(p)]]^2] \cong A[[y(p) - y(p^*)]^2] \quad . \quad (III.16)$$

Elementele matricei de varianță au expresii de forma:

$$V_{ij} = A[(p_i - p_i^*)(p_j - p_j^*)] \quad . \quad (III.17)$$

Fie $(\Delta y)^2 = \text{var}[y(p)]$. Atunci se poate scrie următoarea relație, luând în considerare mărimile calculate anterior:

$$(\Delta y)^2 = \sum_{i=1}^m \sum_{j=1}^m \left\{ \left. \frac{\partial y(p)}{\partial p_i} \right|_{p=p^*} V_{ij} \left. \frac{\partial y(p)}{\partial p_j} \right|_{p=p^*} \right\} \quad . \quad (III.18)$$

Relația (V.18) este cunoscută sub numele de **formula de propagare a erorilor**.

În cazul unor erori necorelate este îndeplinită următoarea relație pentru covarianță:

$$\text{cov}(p_i, p_j) = 0 \quad . \quad (III.19)$$

De aceea, $V_{ij} = 0$ - pentru $i \neq j$, respectiv, $V_{ij} = (\Delta p_i)^2$ - pentru $i=j$. În acest caz, formula de propagare a erorilor, pentru erori necorelate se poate scrie astfel:

$$(\Delta y)^2 = \sum_{i=1}^m \left[\left. \frac{\partial y(p)}{\partial p_i} \right|_{p=p^*} \Delta p_i \right]^2 \quad . \quad (III.20)$$

Remarcă. La utilizarea formulei de propagare a erorilor, indiferent de formă - (III.18) sau (III.20) - trebuie ca să se analizeze dacă mărimile Δp_i sunt suficient de mici pentru a se putea aplica formula lui Taylor, de dezvoltare în serie.

III.4. Metode de fit pentru distribuțiile experimentale

III.4.1. Considerații generale

În mod obișnuit distribuțiile experimentale sunt comparate cu diferite distribuții teoretice. Alegerea distribuției teoretice depinde de ipotezele făcute pentru descrierea unui anumit set de date experimentale [9-12].

Stabilirea acordului dintre rezultatele experimentale și diferitele modelări propuse se poate face cu ajutorul unor tipuri specifice de teste. Multe din aceste teste sunt legate de distribuția normală (Gauss), distribuție care se bucură de un număr de proprietăți speciale [1-9].

Măsurătorile fizice implică, în multe situații de interes, distribuții care au abateri standard relativ mici, atât de la o probă la alta, cât și de la valoarea adevărată (așteptată). Din acest motiv se poate considera că chiar cu un număr relativ mic de observații experimentale se poate defini o distribuție caracterizată de o valoare medie și o varianță suficient de bune pentru scopuri practice, în raport cu o populație de același tip de populație.

Aceste metode - *numite metode de fit (potrivire)* - trebuie să îndeplinească anumite condiții și să satisfacă anumite necesități practice. *Una din condițiile de bază este ca ele să fie aplicabile indiferent de numărul de "citiri" implicate.* De aceea, este necesară raportarea fiecărei "citiri".

Trebuie menționat aici faptul că intră în sarcina celui care face un experiment și prelucrează datele experimentale obținute să folosească o estimare descriptibilă și repetabilă în mod exact pentru erori experimentale și distribuțiile asociate acestora.

III.4.2. Metoda celor mai mici pătrate

III.4.2.1. Principiul metodei

Printr-un număr finit de citiri nu se poate determina exact distribuția erorilor. Din acest motiv nu se poate determina valoarea adevărată a oricărei mărimi măsurate. Printr-un experiment se poate obține valoarea cea mai probabilă.

Fie x_i ($i = 1, \dots, n$) valoarea unei citiri și fie x_o valoarea cea mai probabilă. Pentru valoarea cea mai probabilă trebuie avută în vedere următoarea definiție: *cea mai probabilă valoare care poate fi obținută dintr-un set dat de observații experimentale este cea care face ca setul de observații respectiv să fie cel mai probabil.*

Se poate constata că *metoda verosimilității maxime* este cea mai utilă în acest caz, pentru stabilirea setului de observații care dă probabilitatea maximă. De aceea, trebuie să se considere că x_o este o variabilă aleatoare, deoarece - în acest caz - mărimile x_1, x_2, \dots, x_n sunt cunoscute. Două direcții de studiu sunt importante: stabilirea probabilității de a găsi setul de observații experimentale care dă probabilitatea maximă și găsirea valorii maxime a expresiei considerate.

Probabilitatea de a găsi setul de citiri setul de "citiri" cu probabilitatea maximă se obține prin înmulțirea probabilităților individuale pentru toate "citirile". Se poate scrie o relație de forma:

$$P = \Delta P_1 \Delta P_2 \dots \Delta P_n \quad , \quad (III.21)$$

unde

$$\Delta P_i = Q(x_i - m_a \leq x \leq x_i - m_a + \Delta x) \quad . \quad (III.22)$$

Se face ipoteza că mărimile ΔP_i sunt distribuite conform distribuției normale (Gauss), anume:

$$\Delta P_i = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - m_a)^2}{2\sigma^2}} \cdot \Delta x \quad . \quad (III.23)$$

În acest mod se poate determina probabilitatea de a găsi grupul de "citiri" căutat. Expresia acestei probabilități este următoarea:

$$P = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n (\Delta x)^n e^{-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - m_a)^2} \quad . \quad (III.24)$$

Metoda celor mai mici pătrate este o metodă de fit care permite estimarea valorilor așteptate pentru distribuția considerată folosind valori $\delta_j = x_j - x_o$ diferite și modificând valoarea x_o până când probabilitatea P atinge valoarea maximă. Acest mod de lucru nu afectează mărimile n , Δx și $\frac{1}{\sqrt{2}\sigma}$.

Argumentul funcției exponențiale conține numai termeni pătratici. De aceea, valoarea maximă a probabilității P se obține atunci când valoarea sumei $\sum_{j=1}^n \delta_j^2$ este cea mai mică posibilă în condițiile date. *Acesta este principiul metodei celor mai mici pătrate.*

III.4.2.2. Aplicarea metodei celor mai mici pătrate

Ideea fundamentală pentru găsirea valorii celei mai probabile a mărimii măsurate este aceea de a lua din "citirile" existente pe cele mai probabile.

Înainte de a se discuta cazuri concrete de aplicare a metodei celor mai mici pătrate trebuie făcute două observații importante, anume:

A. Metoda se poate aplica atât pentru cazul în care se consideră o singură necunoscută, cât și pentru cazul în care se consideră mai multe necunoscute.

B. În general, x_o este media aritmetică a observațiilor.

Fie cazul unei singure necunoscute. Pentru a respecta condiția ca suma $\sum_{j=1}^n \delta_j^2$ să fie minimă este necesar să fie satisfăcută următoarea relație:

$$\frac{\partial \sum_{j=1}^n \delta_j^2}{\partial x_o} = -2[(x_1 - x_o) + (x_2 - x_o) + \dots + (x_n - x_o)] = 0 \quad . \quad (III.25)$$

Din relația (V.25) se obține:

$$x_o = \frac{\sum_{j=1}^n x_j}{n} \quad . \quad (III.26)$$

Se confirmă astfel observația de la punctul A.

Fie cazul general în care se consideră m necunoscute care satisfac n ecuații, unde $m < n$. Fie A_1, A_2, \dots, A_m mărimile necunoscute, iar x_1, x_2, \dots, x_n observațiile experimentale. Dacă

variabilele experimentale cunoscute sunt $a_1, b_1, \dots, a_n, b_n, \dots$, iar între ele există o relație liniară, atunci se pot scrie următoarele ecuații:

$$A_j a_k + A_j b_k + \dots + A_j q_k = x_k \quad , \quad k = 1, n, \quad j = 1, m \quad . \quad (III.27)$$

Pentru a păstra liniaritatea ecuațiilor trebuie ca: $a_k = b_k^2$.

Este necesară calcularea sumei pătratelor mărimilor δ_i . În acest mod se obține un sistem de m ecuații cu m necunoscute. Suma pătratelor se poate scrie în modul următor:

$$\sum_{j=1}^n \delta_j^2 = \sum_{j=1}^n \sum_{k=1}^m [x_j - (\sum_{l=1}^n A_k a_l + A_k b_l + \dots + A_k q_l)]^2 \quad . \quad (III.28)$$

Folosind o relație de tipul relației (V.25) se obțin ecuațiile normale pentru coeficienții A_k :

$$\frac{\partial \sum_{j=1}^n \delta_j^2}{\partial A_k} = -2\{a_1[x_1 - (A_1 a_1 + A_2 b_1 + \dots + A_m q_1)] + a_2[x_2 - (A_1 a_2 + A_2 b_2 + \dots + A_m q_2)] + \dots + a_n[x_n - (A_1 a_n + A_2 b_n + \dots + A_m q_n)]\} = 0. \quad (III.29)$$

Pentru rezolvarea sistemului se introduc următoarele notații:

$$[aa] = \sum_{j=1}^n a_j^2; \quad [bb] = \sum_{j=1}^n b_j^2; \dots; [ab] = \sum_{j=1}^n a_j b_j; \dots; [aq] = \sum_{j=1}^n a_j q_j; \quad [ax] = \sum_{j=1}^n a_j x_j \quad . \quad (III.30)$$

Cu ajutorul notațiilor de mai sus sistemul de m ecuații normale (III.29), cu m necunoscute, se poate scrie în modul următor:

$$\begin{aligned} [aa]A_1 + [ab]A_2 + \dots + [aq]A_m &= [ax] \\ [ab]A_1 + [bb]A_2 + \dots + [bq]A_m &= [bx] \\ \cdot & \\ \cdot & \\ \cdot & \\ [aq]A_1 + [bq]A_2 + \dots + [qq]A_m &= [qx] \end{aligned} \quad . \quad (V.31)$$

Pentru rezolvarea sistemului de ecuații (V.31) se pot folosi diferite metode. O cale folosită relativ frecvent este cea care implică introducerea determinanților. Calculul unui coeficient A_1 se poate face folosind proprietățile determinanților, anume:

$$A_i = \frac{\begin{vmatrix} [aa] & [ab] \dots [ax] \dots [aq] \\ [ab] & [bb] \dots [bx] \dots [bq] \\ \cdot & \cdot \cdot \cdot \\ [aq] & [bq] \dots [qx] \dots [qq] \end{vmatrix}}{\begin{vmatrix} [aa] & [ab] \dots [al] \dots [aq] \\ [ab] & [bb] \dots [bl] \dots [bq] \\ \cdot & \cdot \cdot \cdot \\ [aq] & [bq] \dots [ql] \dots [qq] \end{vmatrix}} \quad . \quad (V.32)$$

Există și alte căi de rezolvarea a sistemului de ecuații (V.31).

Metoda celor mai mici pătrate se poate folosi pentru fit-area cu o dreaptă a unui set de date experimentale, situație des întâlnită în experimentele de Fizică nucleară [6].

Fie cazul în care nu există nici un motiv "a priori" de a presupune că datele experimentale nu sunt de încredere. Fie $y = A + Bx$ ecuația dreptei cu care se face fit-area și fie y_i valoarea observată a mărimii y atunci când mărimea x are valoarea x_i . Aplicând principiul fundamental al metodei se obține:

$$\sum_i \delta_i^2 = \sum_i (A + Bx_i - y_i)^2 \quad , \quad (V.33)$$

de unde se ajunge, prin derivare, la următoarele ecuații:

$$\frac{\partial \sum_i \delta_i^2}{\partial A} = 2 \sum_i (A + Bx_i - y_i) = 0 \quad , \quad (V.34)$$

$$\frac{\partial \sum_i \delta_i^2}{\partial B} = 2 \sum_i (A + Bx_i - y_i)x_i = 0 \quad . \quad (V.35)$$

Cele două ecuații ale sistemului se mai pot scrie astfel:

$$nA + B \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad , \quad (V.36)$$

$$A \sum_{i=1}^n x_i + B \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \quad . \quad (V.37)$$

Aici n este numărul de date experimentale considerate în eșantionul respectiv.

Soluțiile pentru parametrii A și B sunt următoarele:

$$A = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}, \quad (V.38)$$

$$B = \frac{n \left(\sum_{i=1}^n x_i y_i\right) - \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}. \quad (V.39)$$

Prin efectuarea calculelor se obțin cele mai bune valori ale parametrilor A și B pentru situația considerată. Cu ajutorul lor se poate trasa - printre punctele experimentale - dreapta care descrie cel mai bine eșantionul respectiv.

În multe experimente fiecare observație experimentală este caracterizată prin precizie specifică, diferită de a celorlalte. De aceea, este necesară introducerea unei distribuții "părinte" a erorilor de dimensiune infinită. Eroarea fiecărei observații (date) experimentale poate fi caracterizată prin valori diferite ale mărimii $h = \frac{1}{\sqrt{2}\sigma}$ care intră în expresia ecuației pentru funcția de distribuție normală [1-6].

Fie un set de n date experimentale având erori ζ_i , caracterizate de un indice de precizie $h_i = \frac{1}{\sqrt{2}\sigma_i}$, dar având toate aceeași valoare așteptată [1-8]. Probabilitatea de a obține un astfel de set este următoarea:

$$P = \left(\prod_{i=1}^n \frac{h_i}{\sqrt{\pi}} \right) (\Delta\zeta)^n e^{-\sum_{i=1}^n h_i^2 \zeta_i^2}. \quad (V.40)$$

Valoarea așteptată, comună pentru cele n date, se estimează pentru valoarea maximă a probabilității P . Această valoare maximă se obține atunci când $\sum_{i=1}^n h_i^2 \delta_i^2$ are cea mai mică valoare posibilă. Cele mai probabile valori ale necunoscutelor se obțin atunci când $\sum_{i=1}^n h_i^2 \delta_i^2$ este minimă.

Fie $h_i^2 = w_i h^2$, unde h este o constantă. Pe baza relațiilor anterioare se poate scrie:

$$h^2 \left(\sum_{i=1}^n w_i \delta_i^2 \right) = \min \quad . \quad (V.41)$$

Se consideră δ_i de forma următoare:

$$\delta_i = x_i - x \quad ,$$

unde x este valoarea așteptată a mărimii necunoscute, putând fi considerată ca o variabilă. Prin înlocuirea în ecuația (V.41) și diferențierea în raport cu variabila aleatoare x se obține, din

condiția de minim $\frac{\partial \left(\sum_{i=1}^n w_i \delta_i^2 \right)}{\partial x} = 0$, următoarea expresie a valorii așteptate:

$$x = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad . \quad (V.42)$$

Relația de mai sus este utilă în obținerea mediilor ponderate. De aceea, se consideră că w_i reprezintă ponderile observațiilor, iar h reprezintă precizia măsurării pentru cazul în care ponderea este egală cu unitatea.

Din relațiile anterioare - (V.40)÷(V.42) - se obține următoarea relație de legătură:

$$\frac{h_1}{h_2} = \sqrt{\frac{w_1}{w_2}} = \frac{\sigma_2}{\sigma_1} = \frac{p_2}{p_1} \quad , \quad (V.43)$$

Dacă h_o , respectiv, σ_o , sunt precizia măsurării, respectiv, abaterea standard pentru distribuția mediilor aleatoare pentru n date experimentale, atunci ecuația (V.43) conduce la următoarea relație de legătură:

$$\sigma_o^2 = \frac{1}{n} \sigma^2 \quad . \quad (V.44)$$

Pentru această situație se obține un nou sistem de ecuații, asemănător cu cel din ecuația (V.31), anume:

$$\begin{aligned} [waa]A_1 + [wab]A_2 + \dots + [waq]A_m &= [wax] \\ [wab]A_1 + [wbb]A_2 + \dots + [wbq]A_m &= [wbx] \\ \cdot & \\ \cdot & \\ [waq]A_1 + [wbq]A_2 + \dots + [wqq]A_m &= [wqx] \end{aligned} \quad . \quad (V.45)$$

Aici $[waa] = \sum_{i=1}^n w_i a_i^2$ ș.a.m.d. Rezolvarea sistemului se face ca și în cazul sistemului de ecuații (V.31).

Există situații când un set de observații/date experimentale poate să conțină câteva mărimi (necunoscute) care satisfac exact una sau mai multe condiții teoretice care se stabilesc între necunoscute. În aceste situații se reduce numărul necunoscutelor care trebuie să fie calculate. Numărul necunoscutelor care nu mai trebuie să fie calculate este dat de numărul de condiții care sunt satisfăcute de mărimile considerate.

Trebuie menționat aici faptul că există experimente în care ecuațiile care definesc condițiile conțin necunoscute neliniare. În aceste cazuri metoda celor mai mici pătrate se poate aplica numai dacă se cunosc valorile aproximative pentru necunoscute. Aceste valori pot fi obținute prin metode care necesită calcule mai puțin complicate și laborioase.

Fie Z_1 și Z_2 astfel de necunoscute. Se consideră relații de legătură de forma $f_i(Z_1, Z_2) = X_i$, $i=1, n$ și ponderile w_i corespunzătoare. Se presupune că funcțiile $f_i(Z_1, Z_2)$ sunt neliniare în cele două variabile, Z_1 și Z_2 .

Dacă se presupune că valorile aproximative ale lui Z_1 și Z_2 au fost obținute prin alte metode, atunci se poate considera că $Z_1 = A + z_1$ și $Z_2 = B + z_2$, unde z_1 și z_2 sunt noile necunoscute de determinat. Prin dezvoltarea în serie Taylor a funcțiilor $f_i(Z_1, Z_2) = X_i$, pentru z_1 și z_2 foarte mici, se obține:

$$f_i(Z_1, Z_2) = f_i(A, B) + \left. \frac{\partial f_i}{\partial Z_1} \right|_{A, B} z_1 + \left. \frac{\partial f_i}{\partial Z_2} \right|_{A, B} z_2 + \dots, \quad i=1, n. \quad (V.46)$$

Se introduc notațiile: $X_i - f_i(A, B) = m_i$, $i=1, n$. Mărimile m_i sunt foarte mici și pot fi considerate ca noi variabile. În acest caz se obține un nou set de ecuații, și anume:

$$X_i - f_i(A, B) = \left. \frac{\partial f_i}{\partial Z_1} \right|_{A, B} z_1 + \left. \frac{\partial f_i}{\partial Z_2} \right|_{A, B} z_2 = m_i, \quad i=1, n. \quad (V.47)$$

Ecuațiile (V.47) sunt liniare în mărimile z_1 și z_2 . Ele se rezolvă prin metoda celor mai mici pătrate, în modul arătat anterior.

Metoda celor mai mici pătrate se poate utiliza în estimarea împrăștierii unor mărimi determinate din date experimentale. Trebuie avut în vedere faptul că între abaterea standard a populației, σ , și abaterea standard a eșantionului (probei), σ_p , există unele diferențe (subcapitolul V.1). De obicei, interesează cea mai bună estimare a abaterii standard a populației.

Fie o distribuție "mamă" centrată pe valoarea așteptată, m_a . Fie o distribuție specifică unei probe centrată pe valoarea cea mai probabilă, m_n , rezultată din n date experimentale sau observații. Cunoscând cea mai bună estimare a centrului de simetrie pentru distribuția "mamă" este necesară stabilirea lărgimii sale, în ipoteza că sub curba specifică probei este inclusă 0.6827 din aria unitate a curbei "mamă".

Fie $d = m_n - m_a$. Pentru o valoare experimentală e_i se pot scrie următoarele relații:

$$\begin{aligned}\xi_i &= e_i - m_a \\ \delta_i &= e_i - m_n \quad (V.48) \\ \xi_i &= \delta_i + d\end{aligned}$$

Deoarece valoarea lui m_n se găsește punând condiția $\sum_{i=1}^n \delta_i = 0$ din relațiile (V.48) se obține:

$$\sum_{i=1}^n \xi_i^2 = \sum_{i=1}^n \delta_i^2 + nd^2 \quad (V.49)$$

Valoarea lui d trebuie să fie, cel mult, de același ordin de mărime cu una din măsurile împrăștierii pentru o distribuție a mediilor, în cazul unor probe de dimensiune n . Toate măsurile împrăștierilor au aceeași formă, și sunt legate unele de altele prin constante.

În ipoteza că:

$$d^2 = c\sigma_p^2 = c \frac{\sigma^2}{n} = c \frac{\sum_{i=1}^n \xi_i^2}{n^2}, \quad (V.50)$$

relația (V.49) se poate scrie astfel:

$$\sum_{i=1}^n \xi_i^2 = \sum_{i=1}^n \delta_i^2 + c \frac{\sum_{i=1}^n \xi_i^2}{n} \quad (V.51)$$

Soluția este de forma:

$$\frac{\sum_{i=1}^n \xi_i^2}{n} = \frac{\sum_{i=1}^n \delta_i^2}{n - c}. \quad (V.52)$$

Observații

1. Pentru valori foarte mari ale lui n corecția nu este importantă.
2. Corecția este importantă numai pentru valori mici ale lui n .

3. În cazul $n=1$ se ajunge la o valoare nedeterminată a raportului $\frac{\sum_{i=1}^n \xi_i^2}{n}$. De aceea, pentru $n=1$ se alege $c=1$.

Având în vedere rezultatele anterioare, se consideră că pentru o singură variabilă cea mai bună estimare a abaterii standard este de forma următoare:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n \delta_i^2}{n-1}} \quad (V.53)$$

Cea mai bună estimare a abaterii standard pentru medie se poate scrie astfel:

$$\sigma_p = \sqrt{\frac{\sum_{i=1}^n \delta_i^2}{n(n-1)}} \quad (V.54)$$

În relația (V.54) mărimea $(n-1)$ reprezintă numărul gradelor de libertate ale sistemului.

Remarcă. În acest caz prin grad de libertate se înțelege numărul de date experimentale sau observații în exces în raport cu numărul minim teoretic necesar pentru a obține mărime necunoscută. În general, pentru n date experimentale sau observații asupra a q necunoscute este $(n-q)$.

Pentru q necunoscute abaterea standard - în cazul unor ponderi egale cu unitatea ale datelor experimentale sau observațiilor - are următoarea expresie:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n \delta_i^2}{n-q}} \quad (V.55)$$

Dacă ponderile sunt diferite de unitate și au valori specifice w_i , atunci abaterea standard are o formă nouă, anume:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n w_i \delta_i^2}{n-q}} \quad (V.56)$$

Abaterea standard pentru o dată experimentală sau observație de pondere w_k se poate scrie în modul următor:

$$\sigma_{w_k} = \frac{\sigma}{\sqrt{w_k}} = \sqrt{\frac{\sum_{i=1}^n w_i \delta_i^2}{w_k (n-q)}} \quad (V.57)$$

Pentru cazul unei singure necunoscute abaterea standard a mediei (eroarea) se poate scrie astfel:

$$\sigma_p = \frac{\sigma}{\sqrt{\sum_{k=1}^n w_k}} = \sqrt{\frac{\sum_{i=1}^n w_i \delta_i^2}{\sum_{k=1}^n w_k (n-1)}} \quad (V.58)$$

În cazul mai multor necunoscute soluția se complică și se calculează diferit, de la caz la caz.

Există situații în care printre datele experimentale se pot afla și unele afectate de greșeli. Pentru a le elimina se consideră că sunt afectate de greșeli cele pentru care valoarea este

mai mare decât $3.29 \sqrt{\frac{\sum_{i=1}^n \delta_i^2}{n-1}}$. Valoarea respectivă este stabilită cu ajutorul unei relații de tipul relației (V.23), integrând de la 0 la valoarea considerată ca admisibilă.

O altă metodă de eliminare a datelor experimentale (observațiilor) afectate de greșeli este cea cunoscută sub numele de *criteriul lui Chauvenet*. Această metodă dă probabilitatea limită de realizare pentru date experimentale (observații) "acceptabile", în funcție de numărul acestora. Această probabilitate este dată de următoarea relație:

$$P_{\text{lim}} = \frac{2n-1}{4n} \quad (V.59)$$

Estimarea gradului de **precizie** se poate face prin observarea diferenței dintre valorile cele mai mari și cele mai mici ale datelor experimentale (observațiilor). Diferența poartă numele de *domeniu* și nu are aceeași distribuție de probabilitate ca acestea.

Fie R dimensiunea unui domeniu. Fie $\delta = R/2$ diferența dintre cea mai mare și cea mai mică valoare din cele n date experimentale (observații). În aceste condiții este satisfăcută următoarea relație:

$$n-2 = 2nP\left(|\delta| < \frac{R}{2}\right), \quad (V.60)$$

unde $P\left(|\delta| < \frac{R}{2}\right)$ este probabilitatea de a observa o diferență δ mai mică decât $R/2$.

În general, sunt îndeplinite următoarele condiții:

$$2nP\left(|\delta| > \frac{R}{2}\right) = \frac{1}{2} \quad (V.61)$$

$$P\left(|\delta| < \frac{R}{2}\right) = 1 - P\left(|\delta| > \frac{R}{2}\right) \quad (V.62)$$

V.4.3. Distribuția χ^2

Pentru descrierea dispersiei unei populații folosind varianța probei a fost introdusă distribuția χ^2 . Definiția ei se poate face în cadrul următoarei **teoreme**:

Dacă x_i ($i=1,2,\dots,n$) sunt probe de variabile aleatoare distribuite normal și independente, de medii m_i și varianțe σ_i , atunci statistica

$$\chi^2 = \sum_{i=1}^n \left(\frac{x_i - m_i}{\sigma_i} \right)^2 \quad (V.63)$$

este distribuită cu funcția de densitate de probabilitate:

$$f(\chi^2, \mu) = \frac{1}{2^{\frac{\mu}{2}} \Gamma\left(\frac{\mu}{2}\right)} \chi^{2\left(\frac{\mu}{2}-1\right)} e^{-\frac{\chi^2}{2}}, \quad \chi^2 > 0, \quad (V.64)$$

de medie μ și varianță 2μ .

Statistica (V.63) se numește distribuție χ^2 cu n grade de libertate.

Funcția gamma este definită prin integrala următoare:

$$\Gamma(x) = \int_0^{\infty} du e^{-u} u^{x-1}, \quad 0 < x < \infty \quad (V.65)$$

Funcția caracteristică a distribuției χ^2 are următoarea expresie:

$$\Phi(t) = (1 - 2it)^{-\frac{\mu}{2}} \quad (V.66)$$

Ținând seama de expresia funcției generatoare, anume:

$$M_x(t) \equiv E[e^{xt}] = \int_{-\infty}^{\infty} dx e^{tx} f(x) \quad (V.67)$$

se poate scrie următoarea *funcție generatoare de momente pentru distribuția χ^2* :

$$M(t) = (1 - 2t)^{-\frac{\mu}{2}} \quad (V.68)$$

Proprietățile funcției generatoare de momente permit obținerea expresiilor pentru momentele simple necentrate, anume:

$$\begin{aligned}\frac{\partial M(t)}{\partial t} \Big|_{t=0} &= \mu = m_1' \\ \frac{\partial^2 M(t)}{\partial t^2} \Big|_{t=0} &= 2\mu = m_2' \\ \frac{\partial^3 M(t)}{\partial t^3} \Big|_{t=0} &= 8\mu = m_3' \\ \frac{\partial^4 M(t)}{\partial t^4} \Big|_{t=0} &= 12\mu(\mu + 4) = m_4'\end{aligned}$$

Cu ajutorul acestor momente se pot defini parametrii de asimetrie și de formare de maxime - β_1 , respectiv, β_2 . Se obțin următoarele expresii:

$$\begin{aligned}\beta_1 &= \frac{8}{\mu} \\ \beta_2 &= 3\left(1 + \frac{4}{\mu}\right)\end{aligned}$$

Se observă că, pentru $\mu \rightarrow \infty$, valorile parametrilor de mai sus tind spre cele caracteristice distribuției normale, și anume:

$$\begin{aligned}\beta_1^n &= 0 \\ \beta_2^n &= 3\end{aligned}$$

În condiția menționată - $\mu \rightarrow \infty$ - distribuția χ^2 însăși tinde lent spre distribuția normală.

Distribuția χ^2 este o distribuție uniparametrică. De aceea, în anumite situații, se folosește statistica $\sqrt{2\chi^2}$ care tinde rapid spre distribuția normală, atunci când $\mu \rightarrow \infty$, având media $\mu' = \sqrt{2\mu - 1}$ și varianța 1 .

O proprietate importantă a statisticii χ^2 este proprietatea de aditivitate. Această proprietate arată că *suma a n variabile independente χ_j^2 , $j=1,2,\dots,n$, fiecare având distribuții χ^2 cu ν_j , $j=1,2,\dots,n$, grade de libertate, este ea însăși distribuită ca χ^2 cu*

$$\nu = \sum_{j=1}^n \nu_j \text{ grade de libertate.}$$

În folosirea statisticii și distribuției χ^2 sunt utile și următoarele două *teoreme*.

Teorema I. Fie x_1, x_2, \dots, x_n o probă de dimensiune n extrasă dintr-o populație normală cu medie 0 și varianță unitate. Atunci, statistica $u = \sum_{i=1}^n (x_i - \langle x \rangle)^2$ este distribuită ca χ^2 cu $n-1$

grade de libertate, iar varianța probei este $\sigma_p = \frac{\sigma^2 \chi^2}{n-1}$, fiind distribuită, de asemenea, ca χ^2 cu $n-1$ grade de libertate și independentă de media probei, $\langle x \rangle$.

Teorema II. Media și varianța probei sunt variabile aleatoare independente atunci când proba este extrasă la întâmplare dintr-o populație normală.

Este utilă calcularea proporției α a ariei de sub curbele distribuției χ^2 a diferitelor puncte χ^2_α pentru care este satisfăcută următoarea condiție:

$$P(\chi^2 \geq \chi^2_\alpha) = \alpha = \int_{\chi^2_\alpha}^{\infty} f(\chi^2, \mu) d\chi^2 \quad (V.69)$$

Punctele definite de relația (V.69) sunt numite și *puncte de procentaj*.

Determinarea parametrilor - în cazul folosirii testului χ^2 - se face prin impunerea condiției de minimizare pentru distribuția χ^2 . Este cea mai utilizată metodă de analiză a datelor experimentale obținute prin măsurători fizice.

V.4.4. Distribuția t

În marea majoritate a situațiilor de interes nu sunt cunoscute media și varianța populației. De aceea, acestea se înlocuiesc cu estimări calculate din proba respectivă. Distribuția unei probe de medie $\langle x \rangle$ este aproximativ normală, cu o medie a populației μ și o varianță $\frac{\sigma^2}{n}$, unde σ^2 este varianța populației, iar n este dimensiunea probei.

Statistica $u = \frac{\langle x \rangle - \mu}{\frac{\sigma}{\sqrt{n}}}$ este distribuită aproximativ normal cu medie 0 și varianță 1 ,

pentru n mare.

În aceste condiții este important să se stabilească care este *distribuția care permite să se folosească varianța probei pentru a putea face afirmații cu privire la media populației*. Acest tip de distribuție se numește **distribuție t sau distribuție Student**. Ea se poate introduce pe baza următoarei teoreme:

Fie $u = \frac{\langle x \rangle - \mu}{\frac{\sigma}{\sqrt{n}}}$ cu o distribuție normală de medie 0 și varianță 1 . Fie w cu o

distribuție χ^2 cu n grade de libertate. Mărimile \sqrt{w} și u sunt independente statistic. Atunci, variabila aleatoare

$$t = \frac{u}{\sqrt{\frac{w}{n}}} \quad (\text{V.70})$$

are o funcție de densitate de probabilitate

$$f(t, n) = \frac{\Gamma\left[\frac{n+1}{2}\right]}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < t < +\infty \quad (\text{V.71})$$

de medie 0 și varianță $\frac{n}{n-2}$, pentru $n > 2$.

Statistica t se spune că are o distribuție t (Student) cu n grade de libertate.

Principalele proprietăți ale distribuției t (Student) sunt cuprinse în următoarele 3 teoreme:

Teorema I. Fie $x_i, i=1,2,\dots,n$, o probă de dimensiune n extrasă dintr-o populație normală de medie μ și varianță σ^2 . Atunci statistica

$$t = \frac{\sqrt{n}}{\sigma_p} (\langle x \rangle - \mu), \quad (\text{V.72})$$

unde $\sigma_p = \frac{1}{n-1} \sum_{i=1}^n (x_i - \langle x \rangle)^2$ și $\langle x \rangle = \frac{1}{n} \sum_{i=1}^n x_i$, este distribuită ca distribuția t (**Student**) cu $n-1$ grade de libertate.

Teorema II. Cu cât numărul gradelor de libertate ale distribuției t (**Student**) se apropie de infinit cu atât distribuția tinde la distribuția normală, în forma standard.

Teorema III. Fie probele aleatoare $x_{11}, x_{12}, \dots, x_{1n}$ și $x_{21}, x_{22}, \dots, x_{2n}$ de dimensiuni n_1 , respectiv, n_2 , independente, reprezentate prin populațiile normale 1, respectiv, 2, având medii μ_1 , respectiv, μ_2 , și aceeași varianță, σ^2 . Atunci, definind $\langle x_i \rangle = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$, $i = 1, 2$, statistica

$$t = \frac{(\langle x_1 \rangle - \langle x_2 \rangle) - (\mu_1 - \mu_2)}{\sqrt{\sigma_{12}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (\text{V.73})$$

are o distribuție t (**Student**) cu $n=n_1+n_2$ grade de libertate.

În relația (V.73) σ_{12}^2 este varianța sumei probelor 1 și 2 și este definită prin următoarea expresie:

$$\sigma_{12}^2 = \frac{\sum_{i=1}^2 \sum_{j=1}^{n_i} (x_{ij} - \langle x_i \rangle)^2}{n_1 + n_2 - 2} \quad (\text{V.74})$$

Ca și distribuția χ^2 și distribuția t (**Student**) este o familie uniparametrică de curbe. Valorile - în procente - ale punctelor din familia de curbe se obține ținându-se cont de faptul că distribuția este simetrică în jurul valorii $t=0$. Se obține următorul rezultat:

$$P[t < -t_\alpha(n)] = P[t > t_\alpha(n)] = \alpha \quad (\text{V.75})$$

V.4.5. Distribuția F

În cazul în care este necesar să se compare două varianțe sau mai mult de două valori medii se folosește un alt tip de distribuție, anume *distribuția F*.

Teorema care permite introducerea acestui tip de distribuție se enunță astfel:

Fie două variabile aleatoare χ^2_i ($i=1,2$), distribuite ca χ^2 cu n_i grade de libertate. Statistica

$$F = F(n_1, n_2) = \frac{\frac{\chi_1^2}{n_1}}{\frac{\chi_2^2}{n_2}} \quad (V.76)$$

este distribuită, în acest caz, cu funcția de densitate de probabilitate

$$f(F; n_1, n_2) = \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \frac{F^{\frac{n_1-2}{2}}}{\left(1 + \frac{n_1}{n_2} F\right)^{\frac{n_1+n_2}{2}}} \quad (V.77)$$

cu $F \geq 0$, medie $\mu = \frac{n_2}{n_2 - 2}$, $n_2 > 2$ și varianță $\sigma^2 = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_1 - 2)^2(n_2 - 4)}$, $n_2 > 4$.

Punctele din familia de curbe - în procente - se definesc prin relația de mai jos:

$$P[F \geq F_\alpha] = \alpha = \int_{F_\alpha}^{\infty} dF f(F; n_1, n_2). \quad (V.78)$$

Trebuie menționat, în încheierea acestui capitol, că există și alte metode de analiză a datelor experimentale [1-12]

Bibliografie

- [1]. H.G.Worthing, J.Geffner - Prelucrarea datelor experimentale, Editura Tehnică, 1959
- [2]. B.R.Martin - Statistics for Physicists, Plenum Press, 1971
- [3]. A.Solmitz - Annual Review of Nuclear Science (1963)
- [4]. W.T.Eadie et al - Statistical Methods in Experimental Physics, North-Holland Publishing Company, Amsterdam, 1971
- [5]. F.James - Proceedings of the 1970 CERN Computing and Data Processing School - Via Monastero, Varenna, Italy, 30 August-12 September 1970 - Preprint CERN 71-6 (1971)
- [6]. Colectiv de catedră - Fizică nucleară - îndrumător de laborator, Tipografia Universității București, 1987
- [7]. B.Gndenko - Theory of probability, MIR , Moscow,1982
- [8]. Gh.Mihoc, V.Craiu - Tratat de Statistică matematică, Editura Academiei RSR, București, 1981
- [9]. P.Carruthers, C.C.Shih - International Journal of Modern Physics A2(5)(1987)1447-1547
- [10].Isac Stern - Teză de doctorat, IFIN București-Măgurele, 1981
- [11].Al.Jipa - Teză de doctorat, Facultatea de Fizică, Universitatea București, 1989
- [12].Al.Jipa et al - Journal of Physics G: Nuclear and Particle Physics 22(2)(1996)221-230