GEORGE CIUCU ★ VIRGIL CRAIU ★ ION SĂCUIU

PROBLEME DE STATISTICĂ MATEMATICĂ

EDITURA TEHNICĂ
BUCUREŞTI — 1974

Lucrarea oferă un bogat material informativ general referitor la cunoaşterea şi aplicarea principiilor statisticii matematice. Prin conţinutul pe care-l are, ea răspunde unor necesităţi reale ale practicii construcţiei socialiste, avînd în vedere rolul major al ştiinţei matematice în producţia contemporană prin unirea organică a acestei ştiinţe cu producţia. Problemele sînt date gradat pornind de la cele mai simple pînă la probleme mai dificile, şi sînt rezolvate în întregime.

Cuprins: Probabilităţi. Repartiţii. Teoria selecţiei. Corelaţie şi regresie. Teoria estimaţiei. Verificarea ipotezelor statistice.

Lucrarea se adresează studenţilor, inginerilor, matematicienilor, economiştilor, medicilor, biologilor, fizicienilor, psihologilor, agronomilor, profesorilor de liceu şi cercetătorilor în domenii aplicative.

PREFAȚĂ

Știința și tehnica cunoaște astăzi un avînt nemaiîntîlnit. În această cursă a dezvoltării spectaculoase, ramurile științifice se întrepătrund și apar ramuri științifice noi, la frontierele unor ramuri bine delimitate pînă nu de mult. Nu întîmplător, tocmai în aceste ramuri de frontieră apar rezultate noi, dintre cele mai valoroase, care utilizînd metode de investigare proprii uneia sau mai multor ramuri adiacente măresc considerabil sfera cunoașterii în domeniul respectiv. Dintre disciplinele ale căror instrumente de cercetare au pătruns în majoritatea ramurilor științifice, matematica ocupă unul dintre primele locuri. Aproape că nu există domeniu de activitate științifică care să nu facă astăzi apel la metode și modele matematice, pentru prelucrarea, analiza și interpretarea rezultatelor specifice domeniului considerat. Multe dintre fenomenele studiate au o natură stochastică și ca atare, fac apel la teoria probabilităților și statistica matematică.

Necesității și dorinței diferiților specialiști de a utiliza teoria probabilităților și statistica matematică li se răspunde prin publicarea diferitelor monografii și culegeri de probleme. Cartea de față răspunde, credem noi, acestui deziderat. Ea urmează după o altă culegere de probleme de teoria probabilităților care a apărut deja în ediția a doua.

Specificul problemelor de statistică matematică, modul de abordare a acestora și posibilitățile ca unele dintre ele să fie aplicate direct în practică, subliniază necesitatea apariției separate a unei culegeri de probleme în acest domeniu. Problemele cuprinse în lucrare sînt complet rezolvate, fiind prezentate în ordinea dificultății lor. Am considerat că, pentru o mai bună înțelegere a problemelor, este indicat să cuprindem atît probleme teoretice cît și aplicații numerice, cu toate calculele și interpretările ce decurg din rezultatele obținute.

Cele cinci capitole pe care le cuprinde cartea acoperă partea clasică a statisticii matematice. O eventuală nouă ediție va trebui să cuprindă probabil și alte părți ale statisticii matematice, mai puțin sau chiar deloc reprezentate în această lucrare cum ar fi: analiza dispersională și factorială, seriile dinamice, controlul statistic al calității produselor, siguranța în funcționare, planificarea experimentelor etc. Asupra acestor aspecte vom insista în funcție de sugestiile și părerile cititorilor, cărora le vom acorda o deosebită atenție.

AUTORII

CUPRINS

Capitolul I

PROBABILITĂŢI. REPARTIŢII

I.1. *Să se arate că dacă* $P(S_2/S_1) < P(S_2)$, *atunci* $P(S_1/S_2) < P(S_1)$ *şi* $P(F_2/F_1) < P(F_2)$, *unde* $F_j = \bar{S}_j$, $j = 1,2$.

S o l u ţ i e. Din definiţia probabilităţii condiţionate

$$P(S_1/S_2) = \frac{P(S_1 \cap S_2)}{P(S_2)} = \frac{P(S_2/S_1)\,P(S_1)}{P(S_2)}$$

şi ţinînd seama de ipoteza făcută, obţinem

$$P(S_1/S_2) < P(S_1). \tag{I.1}$$

Putem scrie

$$P(F_2/F_1) = \frac{P(F_2 \cap F_1)}{P(F_1)} = \frac{P(\bar{S}_1 \cap \bar{S}_2)}{P(\bar{S}_1)} = \frac{P(\overline{S_1 \cup S_2})}{P(\bar{S}_1)} =$$

$$= \frac{1 - P(S_1 \cup S_2)}{1 - P(S_1)} = \frac{1 - P(S_1) - P(S_2) + P(S_1 \cap S_2)}{1 - P(S_1)} =$$

$$= \frac{1 - P(S_1) - P(S_2) + P(S_1/S_2)\,P(S_2)}{1 - P(S_1)}.$$

Din (I.1) obţinem

$$P(F_2/\mathrm{F}_1) = \frac{1 - P(S_1) - P(S_2) + P(S_1/S_2)\,P(S_2)}{1 - P(S_1)} <$$

$$< \frac{1 - P(S_1) - P(S_2) + P(S_1)\,P(S_2)}{1 - P(S_1)} = \frac{[1 - P(S_1)]\,[1 - P(S_2)]}{1 - P(S_1)} =$$

$$= 1 - P(S_2) = P(F_2).$$

I.2. *Definim variabilele aleatoare X_1 și X_2 în felul următor:*

$$X_1 = \begin{cases} 0 \text{ dacă subsistemul } S_1 \text{ se află în stare de funcțiune,} \\ 1 \text{ dacă subsistemul } S_1 \text{ nu se află în stare de funcțiune.} \end{cases}$$

Analog

$$X_2 = \begin{cases} 0 \text{ dacă } S_2 \text{ se află în stare de funcțiune,} \\ 1 \text{ dacă } S_2 \text{ nu se află în stare de funcțiune.} \end{cases}$$

Definim

$$P(X_1 = 1) = p_1 = 1 - R_1$$

cu

$$P(S_i) = R_i, \qquad i = 1, 2. \qquad P(X_2 = 1) = p_2 = 1 - R_2.$$

Fie

$$P(S_2/S_1) = \delta.$$

Să se calculeze coeficientul de corelație al variabilelor X_1 și X_2.

Soluție.

Avem

$$M(X_i) = 1 - R_i,$$

$$M(X_i^2) = 1 - R_i, \qquad i = 1, 2.$$

$$D^2(X_i) = R_i(1 - R_i),$$

$$M(X_1X_2) = P(\bar{S}_1 \cap \bar{S}_2) = 1 - P(S_1) - P(S_2) + P(S_1 \cap S_2) =$$

$$= 1 - R_1 - R_2 + P(S_2/S_1)\,P(S_1) = 1 - R_1 - R_2 + \delta R_1.$$

$$M(X_1X_2) - M(X_1)\,M(X_2) = R_1(\delta - R_2)$$

de unde

$$\rho(X_1, X_2) = \frac{R_1(\delta - R_2)}{\sqrt{R_1R_2(1 - R_1)(1 - R_2)}}.$$

I.3. *În condițiile exercițiului precedent, să presupunem că sistemul S alcătuit din subsistemele S_1 și S_2 este un sistem în serie. Se cere să se exprime siguranța sistemului, $R = P(X_1 = 0,\ X_2 = 0)$ cu ajutorul coeficientului de corelație.*

Soluție. Avem

$$R = P(X_1 = 0, X_2 = 0) = P(S_1 \cap S_2) = P(S_2/S_1)\, P(S_1) =$$
$$= \delta R_1 = R_1R_2 + \rho\sqrt{R_1R_2(1 - R_1)(1 - R_2)}.$$

I.4. *Să se arate că*

$$R_1R_2 - \frac{1}{4} \leqslant R \leqslant R_1R_2 + \frac{1}{4}$$

Pe baza acestui fapt să se arate că un sistem în serie format cu două subsisteme, nu poate avea siguranță mai mare decît $R_1R_2 + \frac{1}{4}$ și că pentru a atinge această limită este necesar ca $R_1 = R_2 = \frac{1}{2}$.

Soluție. Deoarece $|\rho| \leqslant 1$, obținem

$$-1 \leqslant \frac{R_1\delta - R_1R_2}{\sqrt{R_1R_2(1 - R_1)(1 - R_2)}} \leqslant 1.$$

Cum $\delta R_1 = R$, avem

$$R_1R_2 - \sqrt{R_1R_2(1 - R_1)(1 - R_2)} \leqslant R \leqslant R_1R_2 +$$
$$+ \sqrt{R_1R_2(1 - R_1)(1 - R_2)}.$$

Dar

$$\max\{R_1R_2(1 - R_1)(1 - R_2)\} = \frac{1}{16}$$

și deci

$$R_1R_2 - \frac{1}{4} \leqslant R \leqslant R_1R_2 + \frac{1}{4}.$$

Ca să arătăm că $R = R_1R_2 + \frac{1}{4}$ atrage după sine $R_1 = R_2 = \frac{1}{2}$ este suficient să observăm că

$$\max\left\{R_1R_2 + \frac{1}{4}\right\} = \max\{R_1R_2\} + \frac{1}{4}.$$

Însă $0 \leqslant R_1 \leqslant 1$ și $0 \leqslant R_2 \leqslant 1$ și atunci R_1R_2 atinge maximul cînd $R_1 = R_2 = \frac{1}{2}$, ceea ce demonstrează complet afirmația.

I.5. *Să se arate că siguranța unui sistem serie, alcătuit din două subsisteme nu depășește cea mai mică dintre siguranțele subsistemelor.*

Soluție. Am văzut că siguranța unui sistem în serie alcătuit din două subsisteme se poate scrie

$$R = \delta R_1$$

cu $\delta = P(S_2/S_1)$.

Maximul lui δR_1, se atinge cînd δ își atinge maximul care este 1. Însă din faptul că

$$P(S_1/S_2) = \frac{P(S_1 \cap S_2)}{P(S_2)} = \frac{\delta R_1}{R_2}$$

și $\delta \leqslant 1$ urmează că

$$R_1 \leqslant R_2$$

și deci că

$$R_1 = \min (R_1, R_2),$$

care ne conduce la

$$R \leqslant \min (R_1, R_2),$$

inegalitate care dovedește afirmația.

I.6. *Se consideră o metodă electronică pentru determinarea fisurilor la blocurile motor. Metoda dă următoarele rezultate:*

— dacă blocul motor prezintă o fisură, metoda o indică în 95% din cazuri, iar în 5% nu o indică;

— dacă blocul motor nu prezintă o fisură, metoda indică acest lucru în 98% din cazuri, însă de două ori dintr-o sută indică greșit existența unei fisuri.

Dacă într-un lot mare de blocuri motor se știe că 0,05% *din blocurile motor prezintă fisuri, care este probabilitatea ca un bloc motor luat la întîmplare să fie indicat de metodă ca prezentînd o fisură?*

Soluție. Considerăm evenimentele:

A_1 — blocul motor luat la întîmplare prezintă într-adevăr fisură,

A_2 — blocul motor luat la întîmplare nu prezintă fisură,

X — blocul motor luat la întîmplare este considerat ca prezentînd fisura dacă se aplică metoda arătată.

Avem

$$P(A_1) = 0{,}0^{3*}5; \qquad P(A_2) = 0{,}9995,$$

$$P(X/A_1) = 0{,}98; \qquad P(X/A_2) = 0{,}02,$$

de unde

$$P(A_1/X) = \frac{(0{,}0^35)\,(0{,}98)}{(0{,}0^35)\,(0{,}98) + (0{,}9^35)\,(0{,}02)} = 0{,}0239,$$

$$P(A_2/X) = \frac{(0{,}9^{3**}5)\,(0{,}02)}{(0{,}0^35)\,(0{,}98) + (0{,}9^35)\,(0{,}02)} = 0{,}976.$$

I.7. *Să se arate că*

$$\sum_{j=0}^{k} C_{b+k-j-1}^{k-j} C_{a+j-1}^{j} = C_{a+b+k-1}^{k}.$$

Soluție. Să considerăm două variabile aleatoare independente X_1 și X_2 fiecare avînd o repartiție binomială cu exponent negativ cu parametrul p și exponenții a și respectiv b.

Atunci

$$P(X_1 = j) = C_{a+j-1}^{j} p^a (1-p)^j, \qquad j = 0, 1, 2, \ldots$$

Variabila aleatoare $X = X_1 + X_2$ are de asemenea o repartiție binomială cu exponent negativ cu parametrul p și exponentul $a + b$. Dar de aici urmează că funcția de frecvență a variabilei X_1 condiționată de faptul că $X_1 + X_2 = k$ este dată de

$$P(X_1 = j / X_1 + X_2 = k) = \frac{P(X_1 = j;\ X_1 + X_2 = k)}{P(X_1 + X_2 = k)} =$$

$$= \frac{P(X_1 = j;\ X_2 = k - j)}{P(X_1 + X_2 = k)}; \qquad j = 0, 1, 2, \ldots, k.$$

Ținînd seama de ipotezele în care lucrăm avem

$$P(X_1 = j / X_1 + X_2 = k) = \frac{C_{a+j-1}^{j} p^a (1-p)^j\, C_{b+k-j-1}^{k-j} p^b (1-p)^{k-j}}{C_{a+b+k-1}^{k} p^{a+b} (1-p)^k} =$$

$$= \frac{C_{a+j-1}^{j} C_{b+k-j-1}^{k-j}}{C_{a+b+k-1}^{k}}, \qquad j = 0, 1, 2, \ldots, k$$

---

*) $0{,}0^35 = 0{,}0005$; ** $0{,}9^35 = 0{,}9995$

Cum

$$\sum_{j=0}^{k} P(X_1 = j / X_1 + X_2 = k) = 1,$$

urmează că

$$\sum_{j=0}^{k} C_{a+j-1}^{j} C_{b+k-j-1}^{k-j} = C_{a+b+k-1}^{k}.$$

Generalizare. Să se arate că

$$\sum_{j_i=0}^{k} \prod_{i=1}^{n} C_{a_i+j_i-1}^{j_i} = C^{k}_{\sum\limits_{i=1}^{n} a_i+k-1}.$$

Soluție. Fie $X_1, X_2, \ldots, X_n$, n variabile aleatoare independente fiecare avînd repartiția binomială negativă cu parametrul p și exponenții $a_1, a_2, \ldots, a_n$ respectiv.

Atunci

$$P\left(X_i = j_i,\ i = 1, 2, \ldots, n \,\middle|\, \sum_{i=1}^{n} X_i = k\right) = \prod_{i=1}^{n} C_{a_i+j_i-1}^{j_i} \Big| C^{k}_{\sum\limits_{i=1}^{n} a_i+k-1}$$

cu

$$j_i \geqslant 0, \sum_{i=1}^{n} j_i = k.$$

De aici urmează că

$$\sum_{j_i=0}^{k} \prod_{i=1}^{n} C_{a_i+j_i-1}^{j_i} = C^{k}_{\sum\limits_{i=1}^{n} a_i+k-1}.$$

I.8. *Într-o experiență poate să apară evenimentul A cu probabilitatea p_1, evenimentul B cu probabilitatea p_2, evenimentul AB cu probabilitatea p_{11} sau evenimentul $\bar{A}\bar{B}$ cu probabilitatea p_{00}. Se repetă experimentul de n ori și se cere să se determine probabilitatea ca în cele n experimente să apară de x ori evenimentul A și de y ori evenimentul B.*

Soluție. La fiecare experiment obținem următoarea clasificare a evenimentelor

	B	$\bar{B}$	
A	p_{11}	p_{10}	p_1
$\bar{A}$	p_{01}	p_{00}	q_1
	p_2	q_2	1

Atunci probabilitatea ca în n experiențe să apară de x ori A, de y ori B, de i ori AB și de $n - x - y + i$ ori $\bar{A}\bar{B}$ este

$$\frac{n!}{i!(x-i)!\,(y-i)!\,(n+i-x-y)!}\,p_{11}^{i}p_{10}^{x-i}p_{01}^{y-i}p_{00}^{n+i-x-y}$$

cu $i \leqslant \min(x, y)$.

De aici rezultă că probabilitatea ca în n experiențe să apară de x ori A și de y ori B este

$$b(x, y) = \sum_{i=0}^{\min(x,y)} \frac{n!}{i!(x-i)!\,(y-i)!\,(n+i-x-y)!}\,p_{11}^{i}p_{10}^{x-i}p_{01}^{y-i}p_{00}^{n+i-x-y} \tag{I.2}$$

Repartiția dată prin funcția de frecvență (I.2) poartă numele de repartiție binomială bidimensională.

I.9. *În condițiile exercițiului precedent să se stabilească repartiția limită pentru x și y fixați, cînd $n \to \infty$ și $p_{11}, p_{01}, p_{10} \to 0$, astfel încît $np_{11} = \lambda$, $np_1 = \lambda_1$, $np_2 = \lambda_2$*

S o l u ț i e Putem scrie

$$\lim_{n\to\infty} b(x, y) = \sum_{i=0}^{\min(x,y)} \lim_{n\to\infty}\left[\frac{n!}{i!(x-i)!\,(y-i)!\,(n+i-x-y)!}\times\right.$$

$$\left.\times\left(\frac{\lambda}{n}\right)^{i}\left(\frac{\lambda_1-\lambda}{n}\right)^{x-i}\left(\frac{\lambda_2-\lambda}{n}\right)^{y-i}\left(1-\frac{\lambda_1+\lambda_2-\lambda}{n}\right)^{n+i-x-y}\right]=$$

$$= \sum_{i=0}^{\min(x,y)} \lim_{n\to\infty}\left[\frac{n(n-1)\ldots(n-i)\,(n-i-1)\ldots(n-x)\ldots(n+i-x-y+1)}{i!(x-i)!\,(y-i)!}\times\right.$$

$$\left.\times\left(\frac{\lambda}{n}\right)^{i}\left(\frac{\lambda_1-\lambda}{n}\right)^{x-i}\left(\frac{\lambda_2-\lambda}{n}\right)^{y-i}\left(1-\frac{\lambda_1+\lambda_2-\lambda}{n}\right)^{n+i-x-y}\right]=$$

$$= \sum_{i=0}^{\min(x,y)} \lim_{n\to\infty}\left[\frac{1\left(1-\frac{1}{n}\right)\ldots\left(1-\frac{i}{n}\right)\ldots\left(1-\frac{x}{n}\right)\ldots\left(1-\frac{x+y-i-1)}{n}\right)}{i!(x-i)!\,(y-i)!}\times\right.$$

$$\left.\times\,\lambda^{i}(\lambda_1-\lambda)^{x-i}\,(\lambda_2-\lambda)^{y-i}\left(1-\frac{\lambda_1+\lambda_2-\lambda}{n}\right)^{n+i-x-y}\right].$$

Urmează că

$$\lim_{n\to\infty} b(x, y) = e^{-(\lambda_1+\lambda_2-\lambda)} \sum_{i=0}^{\min(x, y)} \frac{\lambda^i}{i!} \frac{(\lambda_1 - \lambda)^{x-i}}{(x-i)!} \frac{(\lambda_2 - \lambda)^{y-i}}{(y-i)!} =$$

$$= p(x, y; \lambda). \tag{I.3}$$

Repartiţia dată prin funcţia de frecvenţă (I.3) este numită repartiţia Poisson bidimensională.

I.10. *Fie X o variabilă aleatoare a cărei funcţie de repartiţie F este continuă şi strict crescătoare. Să se arate că variabila aleatoare $F(X)$ este repartizată uniform.*

S o l u ţ i e. Fie G funcţia de repartiţie a variabilei $F(X)$; avem

$$G(y) = P[F(X) < y].$$

Ipotezele făcute permit să se definească funcţia inversă F^{-1} (pentru $0 < y < 1$); astfel

$$F(X) < y \Leftrightarrow F^{-1}[F(X)] < F^{-1}(y).$$

Urmează că

$$G(y) = P[X < F^{-1}(y)] = F[F^{-1}(y)] = y,$$

ceea ce arată că $F(X)$ este repartizată uniform.

I.11. *Fie X o variabilă aleatoare cu funcţia de repartiţie*

$$F(x) = \begin{cases} 0 \quad , & x < 0, \\ \sum\limits_{j=0}^{[x]} C_N^j p^j (1-p)^{N-j}, & 0 \leqslant x < N, \\ 1 & x \geqslant N. \end{cases}$$

şi Y o variabilă aleatoare cu funcţia de repartiţie

$$G(z) = \begin{cases} 0, & z < 0, \\ z, & 0 \leqslant z < 1, \\ 1, & z \geqslant 1. \end{cases}$$

Să se afle funcţia de repartiţie a variabilei aleatoare $X + Y$

S o l u ţ i e. Folosind, convoluţia putem să scriem

$$H(x) = P(X + Y \leqslant x) = \int_{-\infty}^{\infty} F(x - z)\, \mathrm{d}G(z) =$$

$$= \int_0^1 \sum_{j=0}^{[x-z]} C_N^j p^j (1-p)^{N-j}\, \mathrm{d}z = \int_{x-1}^{x} \sum_{j=0}^{[u]} C_N^j p^j (1-p)^{N-j}\, \mathrm{d}u =$$

$$= \int_{x-1}^{[x]} \sum_{j=0}^{[u]} C_N^j p^j (1-p)^{N-j}\, \mathrm{d}u + \int_{[x]}^{x} \sum_{j=0}^{[u]} C_N^j p^j (1-p)^{N-j}\, \mathrm{d}u =$$

$$= \sum_{j=0}^{[x-1]} C_N^j p^j (1-p)^{N-j}([x] - x + 1) + \sum_{j=0}^{[x]} C_N^j p^j (1-p)^{N-j}(x - [x]) =$$

$$= \sum_{j=0}^{[x-1]} C_N^j p^j (1-p)^{N-j}([x] - x + 1 + x - [x]) +$$

$$+ C_N^{[x]} p^{[x]} (1-p)^{N-[x]} (x - [x]).$$

Urmează că

$$H(x) = P(X + Y \leqslant x) =$$

$$= \begin{cases} 0, & x < 0, \\ \sum_{j=0}^{[x-1]} C_N^j p^j (1-p)^{N-j} + C_N^{[x]} p^{[x]} (1-p)^{N-[x]} (x - [x]) & 0 \leqslant x \leqslant N+1, \\ 1 & x \geqslant N+1. \end{cases}$$

I.12. *Fie $X_1, X_2, \ldots, X_n$, n variabile aleatoare independente, identic repartizate cu funcţia de frecvenţă*

$$P(X = x) = (\mathrm{e}^{\lambda} - 1)^{-1} \frac{x}{x!}, \quad x = 1, 2, \ldots \tag{I.4}$$

Să se arate că funcţia de frecvenţă a variabilei aleatoare $Y_n = \sum_{i=1}^{n} X_i$ este

$$g_n(y_n) = (\mathrm{e}^{\lambda} - 1)^{-n} \frac{\lambda^{y_n}}{y_n!} \sum_{j=0}^{n-1} C_n^j (-1)^j (n - j)^{y_n};$$

$$y_n = n, \quad n + 1, \ldots, \tag{I.5}$$

D e m o n s t r a ț i e. Vom dovedi (I.5) prin inducție. Pentru $n = 1$, (I.5) se reduce la (I.4) cu $y_1 = x$. Presupunem că (I.5) este adevărată pentru $n = k$.

$$g(y_k) = (e^{\lambda} - 1)^{-k} \frac{\lambda^{y_k}}{y_k!} \sum_{j=0}^{k-1} C_k^j (-1)^j (k-j)^{y_k}, \tag{I.6}$$

Repartiția comună corespunzătoare variabilelor Y_k și X_{k+1} este

$$h(y_k, x_{k+1}) = (e^{\lambda} - 1)^{-(k+1)} \frac{\lambda^{y_k + x_{k+1}}}{y_k! x_{k+1}!} \sum_{j=0}^{k-1} C_k^j (-1)^j (k-j)^{y_k}. \tag{I.7}$$

Dacă în (I.7) punem $y_{k+1} = y_k + x_{k+1}$, atunci

$$g(y_{k+1}) = (e^{\lambda} - 1)^{-(k+1)} \frac{\lambda^{y_{k+1}}}{y_{k+1}!} \sum_{x_{k+1}=1}^{y_{k+1}-1} C_{y_{k+1}}^{x_{k+1}} \times$$

$$\times \left[\sum_{j=0}^{k-1} C_k^j (-1)^j (k-j)^{y_{k+1} - x_{k+1}} \right]. \tag{I.8}$$

Cum

$$\sum_{x_{k+1}=1}^{y_{k+1}-1} C_{y_{k+1}}^{x_{k+1}} \left[\sum_{j=0}^{k-1} C_k^j (-1)^j (k-j)^{y_{k+1} - x_{k+1}} \right] =$$

$$= \sum_{j=0}^{k-1} C_k^j (-1)^j [(k+1-j)^{y_{k+1}} - (k-j)^{k_{k+1}} - 1]$$

și

$$\sum_{j=0}^{k-1} C_k^j (-1)^j [(k+1-j)^{y_{k+1}} - (k-j)^{y_{k+1}} - 1] =$$

$$= \sum_{j=0}^{k} C_{k+1}^j (-1)^j (k+1-j)^{y_{k+1}} \tag{I.9}$$

Ținînd seama de (I.9) în (I.8), urmează că (I.5) este adevărată pentru $n = k + 1$; deci (I.5) este adevărată pentru orice n.

Observație. În cazul în care variabila X_i are funcția de frecvență

$$P(X_i = x_i) = (e^{\lambda_i} - 1)^{-1} \frac{\lambda_i^{x_i}}{x_i!}; \quad x_i = 1, 2, \ldots, i = 1, 2, \ldots, n$$

(I.5) devine

$$(\text{I.5}')\quad g(y_n)=\left[\prod_{i=1}^{n}(e^{\lambda_i}-1)\,y_n!\right]^{-1}\left[\left(\sum_{i=1}^{n}\lambda_i\right)^{y_n}-\right.$$

$$\left.-\sum{}'\Big(\sum_{i\neq j}\lambda_i\Big)^{y_n}+\sum{}''\Big(\sum_{i\neq j\neq k}\lambda_i\Big)+\ldots+(-1)^{n+1}\sum_{i=1}^{n}\lambda_i^{y_n}\right],$$

unde $\sum'$ reprezintă suma după cei C_n^2 termeni, astfel încît $i\neq j$; $\sum''$ reprezintă suma după cei C_n^3 termeni, astfel încît $i\neq j\neq k$.

I.13. *Se dă vectorul aleator (X, Y) a cărui densitate de repartiţie este dată de*

$$f(x,y)=\begin{cases}\dfrac{1}{2(e-1)}\left[\dfrac{1}{x}+\dfrac{1}{y}\right] & \text{dacă } 1\leqslant x\leqslant e;\ 1\leqslant y\leqslant e,\\ 0, & \text{în rest}\end{cases}$$

Să se afle densitatea de repartiţie marginală a variabilei aleatoare X.

S o l u ţ i e. Din definiţia densităţii de repartiţie marginale şi datele problemei

$$f(x)=\int_{-\infty}^{\infty}f(x,y)\,dy=\int_{1}^{e}\frac{1}{2(e-1)}\left[\frac{1}{x}+\frac{1}{y}\right]dy$$

de unde

$$f(x)=\frac{1}{2}\left[\frac{1}{x}+\frac{1}{e-1}\right].$$

I.14. *Repartiţiile marginale ale unei repartiţii normale bidimensionale sînt la rîndul lor repartiţii normale.*

Să se arate că există repartiţii bidimensionale care nu sînt normale şi ale căror repartiţii marginale unidimensionale sînt normale.

S o l u ţ i e. Fie $f_1(x)$ şi $f_2(y)$ densităţile de repartiţie ale variabilelor aleatoare X şi Y simetrice faţă de valorile θ_1 şi θ_2 respectiv şi fie $g(x, y)$ o funcţie astfel încît

$$g(\theta_1+x,y)=-g(\theta_1-x,y)\quad\text{şi}\quad g(x,\theta_2+y)=-g(x,\theta_2-y).$$

Atunci din densitatea de repartiţie a vectorului aleator bidimensional (X, Y):

$$f(x, y) = f_1(x) f_2(y) (1 - g(x, y)),$$

rezultă că $f_1(x)$ şi $f_2(x)$ sînt densităţile de repartiţie marginale ale variabilelor X şi respectiv Y.

În particular pentru $\theta_1 = \theta_2 = 0$ şi

$$f_1(x) = \frac{1}{2} e^{-\frac{x^2}{2}}; \quad f_2(y) = \frac{1}{2} e^{-\frac{y^2}{2}},$$

$$g(x, y) = \frac{xy}{(1 + x^2)(1 + y^2)},$$

obţinem

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \left(1 - \frac{xy}{(1 + x^2)(1 + y^2)}\right)$$

$$(x, y) \in (-\infty, \infty) \times (-\infty, +\infty).$$

De aici rezultă

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) \mathrm{d}x = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

adică repartiţiile marginale unidimensionale ale variabilelor X şi Y sînt normale unidimensionale.

I.15. *Să se arate că există un vector aleator (X, Y) cu X şi Y dependente astfel încît X^2 şi Y^2 să fie independente.*

Soluţie. Fie $f_1(x)$ $[f_2(y)]$ densitatea de repartiţie a variabilei aleatoare continue $X[Y]$ şi fie $g(x, y)$ o funcţie aleasă astfel încît să fie pară în x pentru orice y fixat sau pară în y pentru orice x fixat. Pentru a ne preciza ideile să presupunem că $g(x, y)$ este pară în x.

Atunci funcţia de repartiţie a vectorului aleator (X^2, Y^2), ştiind că densitatea de repartiţie a vectorului aleator (X, Y) este

$$f(x, y) = f_1(x) f_2(y) (1-g(x, y))$$

poate fi calculată după cum urmează:

$$F_{X^2, Y^2}(u, v) = P(X^2 < u,\ Y^2 < v) = P(-\sqrt{u} < X < \sqrt{u},\ -\sqrt{v} <$$

$$< Y < \sqrt{v}) = \int_{-\sqrt{v}}^{\sqrt{v}} \int_{-\sqrt{u}}^{\sqrt{u}} f_1(x)\, f_2(y)\, (1 - g(x, y))\, \mathrm{d}x\, \mathrm{d}y =$$

$$= \int_{-\sqrt{v}}^{\sqrt{v}} \int_{-\sqrt{u}}^{\sqrt{u}} f_1(x)\, f_2(y)\, \mathrm{d}x\, \mathrm{d}y - \int_{-\sqrt{v}}^{\sqrt{v}} f_2(y)\, \mathrm{d}y \int_{-\sqrt{u}}^{\sqrt{u}} f_1(x)\, g(x, y)\, \mathrm{d}x =$$

$$= \int_{-\sqrt{u}}^{\sqrt{u}} f_1(x)\, \mathrm{d}x \int_{-\sqrt{v}}^{\sqrt{v}} f_2(y)\, \mathrm{d}y = P(-\sqrt{u} < X < \sqrt{u}) P(-\sqrt{u} <$$

$$< Y < \sqrt{v}) = P(X^2 < u)\, P(Y^2 < v)$$

deoarece

$$\int_{-\sqrt{u}}^{\sqrt{u}} f_1(x)\, g(x, y)\, \mathrm{d}x = 0$$

întrucît

$$f_1(x)\, f(x, y)$$

este funcţie pară în x pentru orice y.

Acest rezultat ne spune că variabilele aleatoare X^2 şi Y^2 sînt independente.

I.16. *Fie* (X, Y) *un vector aleator caracterizat prin densitatea de repartiţie*

$$f(x, y) = \begin{cases} \frac{1}{4}(1 + xy), & \text{dacă } |x| < 1, \quad |y| < 1, \\ 0, & \text{în rest.} \end{cases}$$

a) *Să se calculeze densităţile de repartiţie marginale ale variabilelor aleatoare* X *şi* Y;

b) *Să se arate că variabilele aleatoare* X^2 *şi* Y^2 *sînt independente*;

c) *Să se calculeze densitatea de repartiţie a vectorului aleator* (X^2, Y^2), *precum şi densităţile de repartiţie marginale ale variabilelor* X^2, *respectiv* Y^2.

S o l u ţ i e. a) Densităţile de repartiţie marginale ale variabilelor aleatoare X şi Y sînt

$$f_1(x) = \frac{1}{4}\int_{-1}^{1}(1 + xy)\,dy = \frac{1}{2},$$

$$f_2(y) = \frac{1}{4}\int_{-1}^{1}(1 + xy)\,dx = \frac{1}{2},$$

ceea ce arată că variabilele X şi Y sînt repartizate uniform pe $(-1, 1)$.

b) Dacă $F_{X^2, Y^2}(u, v)$ reprezintă funcţia de repartiţie a vectorului aleator (X^2, Y^2), atunci

$$F_{X^2, Y^2}(u, v) = P(X^2 < u,\ Y^2 < v) = P\,(-\sqrt{u} < X < \sqrt{u};$$

$$-\sqrt{v} < Y < \sqrt{v}) = \int_{-\sqrt{u}}^{\sqrt{u}}\int_{-\sqrt{v}}^{\sqrt{v}} \frac{1}{4}\,(1 + xy)\,dx\,dy =$$

$$= \frac{1}{4}\,x\Big|_{-\sqrt{u}}^{\sqrt{u}}\,y\Big|_{-\sqrt{v}}^{\sqrt{v}} + \frac{1}{4}\,\frac{x^2}{2}\Big|_{-\sqrt{u}}^{\sqrt{u}}\,\frac{y^2}{2}\Big|_{-\sqrt{v}}^{\sqrt{v}} =$$

$$= \sqrt{u}\sqrt{v} = F_{X^2}(u)\,F_{Y^2}(v),$$

adică variabilele aleatoare X^2 şi Y^2 sînt independente.

c) Din

$$F_{X^2, Y^2}(u, v) = \begin{cases} 0 & u \leqslant 0 \quad \text{sau} \quad v \leqslant 0, \\ \sqrt{u}\sqrt{v} & 0 < u < 1; \quad 0 < v < 1, \\ 1 & u \geqslant 1, \quad v \geqslant 1, \end{cases}$$

rezultă

$$g(u, v) = \begin{cases} \dfrac{1}{4\sqrt{u}\sqrt{v}} & 0 < u < 1; \quad 0 < v < 1, \\ 0, & \text{în rest} \end{cases}$$

şi de aici

$$g_1(u) = \int_0^1 g(u, v)\,dv = \begin{cases} \dfrac{1}{2\sqrt{u}} & 0 < u < 1, \\ 0 & \text{în rest,} \end{cases}$$

$$g_2(v) = \int_0^1 g(u, v)\,du = \begin{cases} \dfrac{1}{2\sqrt{v}} & 0 < v < 1, \\ 0 & \text{în rest.} \end{cases}$$

I.17. *Se consideră $n \geqslant 3$ variabile aleatoare $X_1, X_2, \ldots, X_n$ a căror densitate comună de repartiţie este*

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2}\left[1 + \prod_{i=1}^{n}\left(x_i e^{-\frac{1}{2}x_i^2}\right)\right].$$

Atunci orice submulţime proprie de variabile aleatoare a mulţimii $\{X_1, X_2, \ldots, X_n\}$ este formată din variabile aleatoare independente cu repartiţia comună normală multidimensională în timp ce $X_1, X_2, \ldots, X_n$ sînt dependente iar repartiţia lor comună este diferită de cea normală.

S o l u ţ i e. Să considerăm vectorul aleator

$$X^{(j)} = (X_1, X_2, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n)$$

şi să notăm densitatea de repartiţie comună marginală corespunzătoare prin

$$g_j(x^{(j)}) = g_j(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n).$$

Atunci din datele problemei rezultă că

$$g_j(x^{(j)}) = \int_{-\infty}^{\infty} f(x)\, dx_j = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2} dx_j =$$

$$= \frac{1}{(2\pi)^{(n-1)/2}} e^{-\frac{1}{2}\sum_{i\neq j} x_i^2},$$

deoarece

$$\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2}\left[\prod_{i=1}^{n} x_i e^{-\frac{1}{2}x_i^2}\right] dx_j =$$

$$= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i\neq j} x_i^2}\left[\prod_{i\neq j} x_i e^{-\frac{1}{2}x_i^2}\right]\int_{-\infty}^{\infty} x_j e^{-x_j^2} dx_j = 0$$

(întrucît sub ultima integrală avem o funcţie impară)

De aici rezultă că

$$g_j(x^{(j)}) = \frac{1}{(2\pi)^{(n-1)/2}} e^{-\frac{1}{2}\sum_{i\neq j} x_i^2} = \prod_{i\neq j}\left(\frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}x_i^2}\right),$$

ceea ce dovedeşte că variabilele $X_1, X_2, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n$ constituie $n-1$ variabile aleatoare independente repartizate normal $N(0, 1)$, cu repartiţie comună $g_j(x^{(j)})$.

Tot de aici rezultă că orice submulţime a acestei submulţimi este formată din variabile aleatoare independente repartizate normal $N(0, 1)$, ceea ce dovedeşte afirmaţia deoarece din densitatea de repartiţie

$$f(x) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2}\left[1 + \prod_{i=1}^{n}\left(x_i e^{-\frac{1}{2}x_i^2}\right)\right],$$

rezultă că $(X_1, X_2, ..., X_n)$ sînt dependente iar repartiţia lor comună este diferită de cea normal n-dimensională.

I.18. *O variabilă aleatoare X cu valori în $(0, 1)$ are densitatea de repartiţie $k(\theta)\, e^{\theta x}$, unde θ este un număr real fixat. Fiind dat un şir de întregi $a_1, a_2, ..., a_n, ...$ superiori lui 1, fie $q_n = a_1 \cdot a_2 ... a_n$.*

Să se arate că există variabilele aleatoare independente $X_1, ..., X_n, ...$ cu valori întregi astfel încît $0 \leqslant X_n < a_n$, satisfăcînd relaţia

$$X = \sum_{n=1}^{\infty} \frac{X_n}{q_n}.$$

Soluţie. Toate numerele x aparţinînd lui $(0, 1)$ care nu sînt de forma p/q_n, unde p este întreg, se pot scrie în mod unic sub forma

$$x = \sum_{n=1}^{\infty} \frac{x_n}{p_n},$$

unde x_n sînt întregi, astfel încît $0 \leqslant x_n < a_n$.

Dacă $x = p/q_n$, x are două reprezentări posibile:
$x_k = a_k - 1$ sau $x_k = 0$ pornind de la un anumit rang.

Mulţimea p/q_n fiind nenumărabilă şi X avînd o repartiţie continuă, întregii X_n corespunzători lui X sînt aproape sigur definiţi.

Rămîne să arătăm că $X_1, ..., X_n$ sînt independente.

Dacă $x_1, ..., x_n$ sînt întregi astfel încît $0 \leqslant x_i < a_i$, $i = 1, ..., n$ atunci

$$P(X_1 = x_1, ..., X_n = x_n) =$$

$$= P\left(0 < X - \left(\frac{x_1}{q_1} + \frac{x_2}{q_2} + ... + \frac{x_n}{q_n}\right) < \frac{1}{q_n}\right) =$$

$$= A\, e^{\frac{\theta x_1}{q_1}} e^{\frac{\theta x_2}{q_2}} ... e^{\frac{\theta x_n}{q_n}},$$

unde A este o constantă independentă de $x_1, x_2, ..., x_n$.

Se observă că această probabilitate este un produs de funcţii de $x_1, x_2, ..., x_n$. Urmează că $X_1, X_2, ..., X_n$ sînt variabile independente.

Avem

$$P(X_n = x_n) = K_n\, e^{\frac{\theta x_n}{q_n}},$$

unde constanta K_n se determină din condiţia

$$P(0 \leqslant X_n < a_n) = 1,$$

ca fiind

$$K_n = \frac{1 - e^{\frac{\theta}{q_{n-1}}}}{1 - e^{\frac{\theta}{q_n}}}.$$

Observaţie. Fie

$M(a_1, a_2, \ldots, a_n, \ldots)$ clasa repartiţiilor variabilelor aleatoare de forma $\sum_{n=1}^{\infty} \frac{X_n}{q_n}$, unde variabilele X_n sînt independente şi unde X_n ia valorile $0, 1, \ldots, a_n - 1$.

M_0 clasa repartiţiilor pe $(0, 1)$ a căror densitate este $K(\theta)\, e^{\theta x}$.

Chatterji S. D. a arătat că pentru k fixat singurele repartiţii $M(k, k, \ldots, k, ..)$ avînd o densitate continuă sînt acelea din M_0.

I.19. *Să se afle modulul şi mediana variabilei aleatoare X avînd densitatea de repartiţie*

$$f(x) = \frac{ak(x - x_0)^{a-1}}{[1 + k(x - x_0)^a]^2}\; k > 0, \quad a > 1, \quad x_0 \leqslant x < \infty.$$

S o l u ţ i e. Modulul unei repartiţii este prin definiţie acea valoare a variabilei căreia îi corespunde densitatea maximă.

Pentru a afla modulul va trebui să rezolvăm ecuaţia:

$$\frac{df}{dx} = 0, \quad \left(\frac{d^2f}{dx^2} < 0\right)$$

care are forma

$$ak(x - x_0)^{a-2}[(a-1)(1 + k(x - x_0)^a) - 2ak(x - x_0)^a] = 0.$$

Ecuaţia anterioară se mai poate scrie

$$a - 1 - k(x - x_0)^a\,(a + 1) = 0$$

sau

$$(x - x_0)^a = \frac{a - 1}{k(a + 1)}.$$

Urmează că

$$M_0 = x_0 + \left(\frac{a-1}{k(a+1)}\right)^{1/a}.$$

Pentru obținerea medianei avem ecuațiile

$$ak\int_{x_0}^{M_e} \frac{(x-x_0)^{a-1}}{[1+k(x-x_0)^a]^2}\,dx = ak\int_{M_e}^{\infty} \frac{(x-x_0)^{a-1}}{[1+k(x-x_0)^a]^2}\,dx = \frac{1}{2}.$$

Din ecuația

$$ak\int_{x_0}^{M} \frac{(x-x_0)^{a-1}}{[1+k(x-x_0)^a]^2} = \frac{1}{2},$$

obținem

$$M_e = x_0 + \left(\frac{1}{k}\right)^{1/a}.$$

I.20. *Să se calculeze modulul și media variabilei aleatoare X care urmează o repartiție lognormală.*

S o l u ț i e. Variabila aleatoare X este caracterizată prin densitatea de repartiție

$$f(x) = \begin{cases} 0 & x \leqslant 0, \\ \dfrac{\lg e}{\sqrt{2\pi}\sigma x}\, e^{-\frac{(\lg x - \lg \mu)^2}{2\sigma^2}}, & x > 0, \end{cases}$$

Avem

$$f'(x) = -\frac{\lg e}{\sqrt{2\pi}\sigma x^2}\, e^{-\frac{(\lg x - \lg\mu)^2}{2\sigma^2}} - \frac{(\lg e)^2}{\sqrt{2\pi}\sigma x^2}\,\frac{\lg x - \lg\mu}{\sigma^2}\, e^{\frac{-(\lg x - \lg\mu)^2}{2\sigma^2}} =$$

$$= -\frac{\lg e}{\sqrt{2\pi}\sigma x^2}\, e^{\frac{-(\lg x - \lg\mu)^2}{2\sigma^2}}\left(1 + \frac{\lg e}{\sigma}\,\frac{\lg x - \lg\mu}{\sigma}\right)$$

și cum modulul este soluția ecuației $f'(x) = 0$, obținem

$$\lg x = -\frac{\sigma^2}{\lg e} + \lg\mu$$

sau

$$M_0 = \mu\, 10^{-\frac{\sigma^2}{\lg e}}.$$

Din definiţia valorii medii rezultă

$$M(X)=\frac{\lg e}{\sqrt{2\pi}\sigma}\int_0^{\infty} x\,\frac{e^{-\frac{(\lg x-\lg \mu)^2}{2\sigma^2}}}{x}\,dx.$$

Dacă facem schimbarea de variabilă

$$\frac{\lg x-\lg \mu}{\sigma}=u,$$

obţinem

$$M(X)=\frac{\lg e}{\sqrt{2\pi}\sigma}\,\frac{\sigma\mu}{\lg e}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}+\frac{\sigma u}{\lg e}}\,du=$$

$$=\mu\, e^{\frac{\sigma^2}{2(\lg e)^2}}\,\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(u-\frac{\sigma}{\lg e}\right)^2}\,du$$

sau

$$M(X)=\mu\, e^{\frac{\sigma^2}{2(\lg e)^2}}.$$

I.21. *Se consideră variabila aleatoare X a cărei densitate de repartiţie este*

$$f(x)=\begin{cases}(-\ln p)\,p^x & \text{dacă } x\in[0,\infty),\ (0<p<1),\\ 0, & \text{în rest.}\end{cases}$$

Să se afle:

a) *Mediana variabilei aleatoare X.*

b) *Modulul variabilei aleatoare X.*

Soluţie. a) Variabila aleatoare X este o variabilă continuă, aşa că mediana ei este acea valoare M_e care satisface relaţiile

$$\int_{-\infty}^{M_e} f(x)\,dx=\int_{M_e}^{\infty} f(x)dx=\frac{1}{2}\cdot$$

Aşadar determinăm M_e din

$$(-\ln p)\int_0^{M_e} p^x\,dx=\frac{1}{2}$$

sau

$$\frac{(-\ln p)}{\ln p}p^x\Big|_0^{M_e}=\frac{1}{2},$$

adică

$$p^{M_e} = \frac{1}{2},$$

de unde rezultă

$$M_e = -\frac{\ln 2}{\ln p}.$$

b) Se observă că pe intervalul $[0,\infty]$, $f(x)$ este strict descrescătoare şi deci cea mai mare valoare o ia $f(x)$ în punctul 0. Deci, $x = 0$ este modulul variabilei aleatoare X.

I.22. *Fie X o variabilă aleatoare avînd funcţia de frecvenţă*

$$P(X = x) = \frac{1}{N}, \quad x = 1, 2, \ldots, N.$$

Să se determine $D^2(X)$.

Soluţie. Fie Y o variabilă aleatoare independentă de X şi avînd densitatea de repartiţie

$$f(y) = \begin{cases} 1, & -\frac{1}{2} \leqslant y \leqslant \frac{1}{2}, \\ 0, & \text{în rest.} \end{cases}$$

Variabila aleatoare $Z = X + Y$ are densitatea de repartiţie

$$g(z) = \frac{1}{N}, \quad \frac{1}{2} \leqslant z \leqslant N + \frac{1}{2},$$

Pentru $z = x + y$, unde x este un întreg şi $|y| \leqslant 1/2$, obţinem

$$G(z) = P(Z \leqslant z) = P\left(Z < x - \frac{1}{2}\right) + P\left(x - \frac{1}{2} \leqslant Z \leqslant x + y\right) =$$

$$= P\left(X \leqslant x - 1 \cap Y < \frac{1}{2}\right) + P(X = x \cap Y \leqslant y) =$$

$$= \frac{x-1}{N} 1 + \frac{1}{N}\left(y - \left(-\frac{1}{2}\right)\right) = \frac{z - \frac{1}{2}}{N}$$

Urmează că

$$g(z) = \frac{dG(z)}{dz} = \frac{1}{N}.$$

Se ştie că variabila aleatoare U cu densitatea de repartiţie $h(u) = 1/(b-a)$, $a < u < b$ are dispersia $\frac{(b-a)^2}{12}$.

Deoarece

$$D^2(Z) = D^2(X) + D^2(Y),$$

obţinem

$$D^2(X) = D^2(Z) - D^2(Y) = \frac{N^2}{12} - \frac{1}{12} = \frac{N^2-1}{12}$$

I.23. *Ca măsură a asimetriei se consideră expresia*

$$S_K = \frac{\text{media} - \text{modulul}}{\text{abaterea medie pătratică}}$$

Considerîndu-se variabila aleatoare X, avînd densitatea

$$f(x) = Kx^r(1-x)^s, \qquad 0 \leqslant x \leqslant 1$$

să se compare valorile lui S_K şi cele ale lui $\sqrt{\beta_1}$ pentru un şir de valori ale lui r şi s.

S o l u ţ i e. Din egalitatea

$$K\int_0^1 x^r(1-x)^s\,\mathrm{d}x = KB(r+1, s+1) = 1,$$

rezultă

$$K = \frac{1}{B(r+1, s+1)} = \frac{\Gamma(r+s+2)}{\Gamma(r+1)\Gamma(s+1)}.$$

Deci variabila aleatoare X are densitatea de repartiţie

$$f(x) = \begin{cases} \dfrac{\Gamma(r+s+2)}{\Gamma(r+1)\Gamma(s+1)}\, x^r(1-x)^s & \text{dacă } 0 \leqslant x \leqslant 1, \\ 0, & \text{în rest.} \end{cases}$$

Avem

$$M(X) = \frac{\Gamma(r+s+2)}{\Gamma(r+1)\Gamma(s+1)}\int_0^1 x^{r+1}(1-x)^s\,\mathrm{d}x = \frac{r+1}{r+s+2}.$$

Modulul se obţine din ecuaţia

$$\frac{\mathrm{d}f}{\mathrm{d}x} = 0 \left(\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} < 0\right).$$

Derivînd, obținem

$$\frac{df}{dx} = Kx^{r-1}(1-x)^{s-1}[r(1-x) - sx] = 0$$

care are soluțiile

$$x = 0, \quad x = 1, \quad x = \frac{r}{r+s}.$$

Pe de altă parte

$$\frac{d^2f}{dx^2} = Kx^{r-2}(1-x)^{s-2}\{(r-1)(1-x)[r-(r+s)x] -$$

$$-(r-1)x[r-(r+s)x] - (r+s)x(1-x)\}$$

se constată că este negativă pentru $x = \frac{r}{r+s}$.

Urmează că modulul este egal cu

$$M_0 = \frac{r}{r+s}.$$

Avem

$$M_2(X) = \frac{\Gamma(r+s+2)}{\Gamma(r+1)\Gamma(s+1)}\int_0^1 x^{r+2}(1-x)^s\,dx = \frac{(r+1)(r+2)}{(r+s+2)(r+s+3)},$$

de unde

$$D^2(X) = \frac{(r+1)(r+2)}{(r+s+2)(r+s+3)} - \frac{(r+1)^2}{(r+s+2)^2} = \frac{(r+1)(s+1)}{(r+s+2)^2(r+s+3)}.$$

Abaterea medie pătratică este

$$D(X) = \frac{1}{(r+s+2)}\sqrt{\frac{(r+1)(s+1)}{r+s+3}}.$$

Urmează că

$$S_k = \frac{\dfrac{r+1}{r+s+2} - \dfrac{r}{r+s}}{\dfrac{1}{r+s+2}\sqrt{\dfrac{(r+1)(s+1)}{r+s+3}}} = \frac{\dfrac{s-r}{s+r}\,\dfrac{1}{r+s+2}}{\dfrac{1}{r+s+2}\sqrt{\dfrac{(r+1)(s+1)}{r+s+3}}}.$$

Prin urmare

$$S_k = \frac{s-r}{s+r}\sqrt{\frac{r+s+3}{(r+1)(s+1)}}.$$

Cantitățile necesare pentru calcularea coeficientului de asimetrie notat $\sqrt{\beta_1}$ și egal cu $\frac{\mu_3}{\mu_2^{3/2}}$ sînt date în cele ce urmează.

Avem

$$\mu_3 = M((X - M(X))^3) = M(X^3) - 3M(X)M(X^2) + 2M^3(X),$$

și cum

$$M(X^3) = \frac{\Gamma(r+s+2)}{\Gamma(r+1)\Gamma(s+1)}\int_0^1 x^{r+3}(1-x)^s\,\mathrm{d}x =$$

$$= \frac{(r+1)(r+2)(r+3)}{(r+s+2)(r+s+3)(r+s+4)},$$

urmează că

$$\mu_3 = \frac{(r+1)(r+2)(r+3)}{(r+s+2)(r+s+3)(r+s+4)} - 3\,\frac{r+1}{r+s+2}\,\frac{(r+1)(r+2)}{(r+s+2)(r+s+3)} +$$

$$+ 2\,\frac{(r+1)^3}{(r+s+2)^3} = \frac{2(s-r)(r+1)((s+1)}{(r+s+2)^3(r+s+3)(r+s+4)}.$$

Rezultă că

$$\sqrt{\beta_1} = \frac{\mu^3}{\mu_2^{3/2}} = \frac{\dfrac{2(s-r)(r+1)(s+1)}{(r+s+2)^3(r+s+3)(r+s+4)}}{\dfrac{1}{(r+s+2)^3}\sqrt{\dfrac{(r+1)^3(s+1)^3}{(r+s+3)^3}}} =$$

$$= \frac{2(s-r)}{r+s+4} = \sqrt{\frac{r+s+3}{(r+1)(s+1)}}.$$

Deoarece

$$\frac{S_k}{\sqrt{\beta_1}} = \frac{\dfrac{s-r}{s+r}\sqrt{\dfrac{r+s+3}{(r+1)(s+1)}}}{\dfrac{2(s-r)}{s+r+4}\sqrt{\dfrac{r+s+3}{(r+1)(s+1)}}} = \frac{1}{2} + \frac{2}{s+r}$$

urmează că

$$\frac{S_k}{\sqrt{\beta_1}} > 1 \qquad \text{dacă } s + r < 4,$$

$$\frac{S_k}{\sqrt{\beta_1}} = 1 \qquad \text{dacă } s + r = 4,$$

$$\frac{S_k}{\sqrt{\beta_1}} < 1 \qquad \text{dacă } s + r > 4.$$

I.24. *Să se găsească valoarea raportului lui Geary:*

$$G = \frac{M(|X - m|)}{D(X)}$$

pentru:

a) *Repartiția normală* $N(m, \sigma)$.
b) *Repartiția rectangulară pe* $(0, 1)$.
c) *Repartițiile caracterizate prin probabilitățile elementare*

$$f(x)\mathrm{d}x = \frac{2}{\pi}\frac{\mathrm{d}x}{(1 + x^2)^2}, \qquad -\infty < x < \infty.$$

d) $f(x)\mathrm{d}x = \frac{1}{2}\mathrm{e}^{-|x|}\,\mathrm{d}x, \qquad -\infty < x < \infty.$

Soluție

a) Avem

$$M(|X - m|) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}|x - m|\,\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}\,\mathrm{d}x =$$

$$= \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}|y|\,\mathrm{e}^{-\frac{y^2}{2}}\,\mathrm{d}y = \frac{2\sigma}{\sqrt{2\pi}}\int_{0}^{\infty}y\,\mathrm{e}^{-\frac{y^2}{2}}\,\mathrm{d}y = \sigma\sqrt{\frac{2}{\pi}},$$

și cum $D(X) = \sigma$ rezultă că

$$G = \frac{\sigma\sqrt{\frac{2}{\pi}}}{\sigma} = \sqrt{\frac{2}{\pi}}.$$

b) Deoarece această repartiție este caracterizată prin densitatea de repartiție

$$f(x) = \begin{cases} 0, & x \notin (0, 1), \\ 1, & x \in (0, 1), \end{cases}$$

rezultă

$$m = M(X) = \frac{1}{2},$$

$$M\left(\left|X - \frac{1}{2}\right|\right) = \int_0^1 \left|x - \frac{1}{2}\right| \mathrm{d}x = -\int_0^{1/2}\left(x - \frac{1}{2}\right)\mathrm{d}x + \\ + \int_{1/2}^1 \left(x - \frac{1}{2}\right)\mathrm{d}x = \frac{1}{2^2}$$

$$D^2(X) = \int_0^1 \left(x - \frac{1}{2}\right)^2 \mathrm{d}x = \frac{1}{2^2 \cdot 3},$$

de unde $G = \frac{\sqrt{3}}{2}$.

c) În acest caz

$$m = \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{x\,\mathrm{d}x}{(1+x^2)^2} = 0$$

întrucît avem de integrat o funcţie impară pe intervalul $(-\infty, +\infty)$. Deci

$$M(|X - m|) = M(|X|) = \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{|x|\,\mathrm{d}x}{(1+x^2)^2} = \frac{-2}{\pi}\int_{-\infty}^{0} \frac{x\,\mathrm{d}x}{(1+x^2)^2} + \\ + \frac{2}{\pi}\int_0^{\infty} \frac{x\,\mathrm{d}x}{(1+x^2)^2} = \frac{4}{\pi}\int_0^{\infty} \frac{x\,\mathrm{d}x}{(1+x^2)^2} = -\frac{2}{\pi}\left.\frac{1}{(1+x^2)}\right|_0^{\infty} = \frac{2}{\pi}.$$

Ţinînd cont de observaţia făcută anterior avem

$$D^2(X) = \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{x^2\,\mathrm{d}x}{(1+x^2)^2} = \frac{4}{\pi}\int_0^{\infty} \frac{x^2\,\mathrm{d}x}{(1+x^2)^2}.$$

Facem schimbarea de variabilă $x^2 = y$ şi obţinem

$$D^2(X) = \frac{2}{\pi}\int_0^{\infty} y^{1/2}(1+y)^{-2}\,\mathrm{d}y = \frac{2}{\pi} B\left(\frac{3}{2}, \frac{1}{2}\right) = 1.$$

Cu acest rezultat obţinem valoarea raportului G:

$$G = \frac{M(|X - m|)}{D(X)} = \frac{2}{\pi}.$$

d) Ca şi în celelalte cazuri calculăm întîi valoarea medie

$$m = \frac{1}{2}\int_{-\infty}^{\infty} x\,\mathrm{e}^{-|x|}\,\mathrm{d}x = \frac{1}{2}\int_{-\infty}^{0} x\,\mathrm{e}^{x} + \frac{1}{2}\int_0^{\infty} x\,\mathrm{e}^{-x}\,\mathrm{d}x = 0.$$

Urmează că

$$M(|X-m|) = M(|X|) = \frac{1}{2}\int_{-\infty}^{\infty} |x|\, e^{-|x|}\, dx = \int_0^{\infty} x\, e^{-x}\, dx = 1,$$

$$D^2(X) = \frac{1}{2}\int_{-\infty}^{\infty} x^2\, e^{-|x|}\, dx = \int_0^{\infty} x^2\, e^{-x}\, dx = 2,$$

de unde

$$G = \frac{M(|X-m|)}{D(X)} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}.$$

I.25. *Se consideră variabila aleatoare X astfel încît*

$$M(X^k) = \frac{1}{k+1}, \qquad k = 1, 2, 3.$$

Să se calculeze momentul centrat de ordinul 3.

Soluție. Va trebui să calculăm $M((X-M(X))^3)$.

Dar $M(X) = \frac{1}{2}$ și deci

$$M\left(\left(X-\frac{1}{2}\right)^3\right) = M\left(X^3 - \frac{3}{2}X^2 + \frac{3}{2^2}X - \frac{1}{2^3}\right) =$$

$$= \frac{1}{2^2} - \frac{3}{2}\cdot\frac{1}{3} + \frac{3}{2^3} - \frac{1}{2^3} = 0.$$

I.26. *Să se afle primele patru momente centrate în jurul valorii medii ale repartiției caracterizate prin densitatea*

$$f(x) = \begin{cases} a\, e^{-ax} & 0 \leqslant x < \infty, \quad a > 0, \\ 0 & \text{în rest.} \end{cases}$$

Soluție. Avem

$$m_1 = a\int_0^{\infty} x\, e^{-ax}\, dx = -\int_0^{\infty} x\, d(e^{-ax}) = \frac{1}{a},$$

$$\mu_1 = a\int_0^{\infty}\left(x - \frac{1}{a}\right) e^{-ax}\, dx = 0,$$

$$\mu_2 = a\int_0^{\infty}\left(x-\frac{1}{a}\right)^2 e^{-ax}\, dx = \frac{1}{a}\int_0^{\infty}(ax-1)^2\, e^{-ax}\, dx.$$

Dacă facem schimbarea de variabilă $ax = y$ obţinem

$$\mu_2 = \frac{1}{a^2}\int_0^\infty (y-1)^2 e^{-y}\,dy = \frac{1}{a^2}\int_0^\infty y^2 e^{-y}\,dy -$$

$$-\frac{2}{a^2}\int_0^\infty y\, e^{-y}\,dy + \frac{1}{a^2}\int_0^\infty e^{-y}\,dy,$$

Utilizînd funcţia $\Gamma(p)$ rezultă:

$$\mu_2 = \frac{\Gamma(3)}{a^2} - \frac{2}{a^2}\Gamma(2) + \frac{1}{a^2} = \frac{1}{a^2}.$$

Analog, obţinem

$$\mu_3 = a\int_0^\infty \left(x - \frac{1}{a}\right)^3 e^{-ax}\,dx = \frac{1}{a^3}\int_0^\infty (y-1)^3 e^{-y}\,dy = \frac{2}{a^3},$$

$$\mu_4 = a\int_0^\infty \left(x - \frac{1}{a}\right)^4 e^{-ax}\,dx = \frac{1}{a^4}\int_0^\infty (y-1)^4 e^{-y}\,dy = \frac{9}{a^4}.$$

I.27. *Repartiţia veniturilor este caracterizată aproximativ prin probabilitatea elementară*

$$a\left(\frac{x_0}{x}\right)^{1+a}\frac{dx}{x_0}; \qquad x_0 \leqslant x < \infty$$

unde x_0 şi a sînt constante.

Să se obţină în funcţie de x_0, a şi n expresiile:

1° *Valoarea medie a mediei aritmetice* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *a unor astfel de venituri;*

2° *Dispersia mediei aritmetice dată mai sus.*

Soluţie. Deoarece

$$M(X) = \int_{x_0}^\infty xa\left(\frac{x_0}{x}\right)^{1+a}\frac{dx}{x_0} = \frac{a}{x_0}x_0^{1+a}\int_{x_0}^\infty x^{-a}\,dx = \frac{ax_0}{a-1}, \quad (a>1),$$

urmează că

$$M(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} M(X_i) = \frac{nax_0}{n(a-1)} = \frac{ax_0}{a-1}, \qquad (a>1).$$

Prin definiție și din datele problemei

$$D^2(\overline{X}) = M\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{ax_0}{a-1}\right)^2\right] =$$

$$= M\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{ax_0}{a-1}\right)\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n} M\left[\left(X_i - \frac{ax_0}{a-1}\right)^2\right] +$$

$$+ \frac{2}{n^2}\sum_{i<j} M\left[\left(X_i - \frac{ax_0}{a-1}\right)\left(X_j - \frac{ax_0}{a-1}\right)\right] = \frac{1}{n^2}\sum_{i=1}^{n} M\left[\left(X_i - \frac{ax_0}{a-1}\right)^2\right].$$

Deoarece

$$M\left[\left(X_i - \frac{ax_0}{a-1}\right)^2\right] = M(X_i^2) - \frac{a^2x_0^2}{(a-1)^2},$$

iar

$$M(X_i^2) = \int_{x_0}^{\infty} x^2 a \left(\frac{x_0}{x}\right)^{1+a} \frac{\mathrm{d}x}{x_0} = a x_0^a \int_{x_0}^{\infty} x^{-(a-1)}\,\mathrm{d}x =$$

$$= \frac{ax_0^2}{a-2} \qquad (a > 2)\; i = 1, 2, \ldots, n$$

obținem

$$D^2(\overline{X}) = \frac{ax_0^2}{n(a-1)^2(a-2)}, \quad a > 2.$$

I.28. *Fie* (X_1, X_2, X_3) *un vector aleator normal tridimensional pentru care*

$$\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{și} \quad A = \begin{pmatrix} 7 & 3 & 2 \\ 3 & 4 & 1 \\ 2 & 1 & 2 \end{pmatrix}.$$

unde A *este matricea formei pătratice* $x'Ax$.

Să se determine densitatea de repartiție și matricea de covarianță corespunzătoare vectorului (X_1, X_2, X_3).

Soluție. Avem

$$n(x; 0, A) = \frac{\sqrt{\det A}}{(2\pi)^{3/2}} \exp\left[-\frac{1}{2}\, x'Ax\right]$$

şi cum det $A = 27$,

$$x'Ax = (x_1, x_2, x_3)\begin{pmatrix}7 & 3 & 2\\ 3 & 4 & 1\\ 2 & 1 & 2\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix} =$$

$$= 7x_1^2 + 4x_2^2 + 2x_3^2 + 6x_1x_2 + 4x_1x_3 + 2x_2x_3 =$$

$$= (\sqrt{3}x_1 + \sqrt{3}x_2)^2 + (2x_1 + x_3)^2 + (x_2 + x_3)^2 > 0,$$

de unde

$$n(x; 0, A) = \frac{3\sqrt{3}}{(2\pi)^{3/2}} \exp\left\{-\frac{1}{2}[(\sqrt{3}x_1 + \sqrt{3}x_2)^2 + \right.$$

$$\left. + (2x_1 + x_3)^2 + (x_2 + x_3)^2]\right\}.$$

Avem

$$\Sigma = A^{-1} = \frac{1}{27}\begin{pmatrix}7 & -4 & -5\\ -4 & 10 & -1\\ -5 & -1 & 19\end{pmatrix}.$$

Elementele matricei de covarianţă se mai pot calcula şi direct din relaţia

$$\sigma_{ij} = \int_{R_3} x_i x_j n(x; 0; A)\, dx_1\, dx_2\, dx_3; \quad 1 \leqslant i, j \leqslant 3$$

prin transformarea

$$y_1 = x_2 + x_3, \quad y_2 = 2x_1 + x_3, \quad y_3 = \sqrt{3}x_1 + \sqrt{3}x_2.$$

I.29. *Să se determine a şi A astfel ca densitatea de repartiţie*

$$\frac{1}{2\pi}\left\{-\frac{1}{2}[(x_1 - 1)^2 + (x_2 - 2)^2]\right\}$$

să se pună sub forma

$$n(x; a, A) = \frac{\sqrt{\det A}}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(x - a)'\, A(x - a)\right].$$

Soluţie. Avem

$$\mu_1 = \frac{1}{2\pi}\int_{R^2} x_1 \exp\left\{-\frac{1}{2}[(x_1-1)^2+(x_2-2)^2]\right\} dx_1dx_2,$$

sau făcînd transformarea

$$x_1 - 1 = u$$
$$x_2 - 2 = v$$

pentru care

$$\frac{D(x_1, x_2)}{D(u, v)} = 1,$$

obţinem

$$\mu_1 = \frac{1}{2\pi}\int_{R^2} (u+1)\exp\left[-\frac{1}{2}(u^2+v^2)\right] dudv = 1.$$

Prin aceeaşi transformare, obţinem

$$\mu_2 = \frac{1}{2\pi}\int_{R^2} (v+2)\exp\left[-\frac{1}{2}(u^2+v^2)\right] dudv = 2.$$

Deci $a' = (1, 2)$.

Avem

$$\sigma_{11} = \frac{1}{2\pi}\int_{R^2} x_1^2 \exp\left\{-\frac{1}{2}[x_1-1)^2+(x_2-2)^2] dx_1dx_2 =\right.$$

$$= \frac{1}{2\pi}\int_{R^2} (u^2+2u+1)\exp\left[-\frac{1}{2}(u^2+v^2)\right] dudv = 2.$$

$$\sigma_{12} = \sigma_{21} = \frac{1}{2\pi}\int_{R^2} x_1x_2 \exp\left\{-\frac{1}{2}[(x_1-1)^2+(x_2-2)^2]\right\} dx_1dx_2 =$$

$$= \frac{1}{2\pi}\int_{R^2} (u+1)(v+2)\exp\left[-\frac{1}{2}(u^2+v^2)\right] dudv = 2$$

şi

$$\sigma_{22} = \frac{1}{2\pi}\int_{R^2} x_2^2 \exp\left\{-\frac{1}{2}[(x_1-1)^2+(x_2-2)^2]\right\} dx_1dx_2 = 5,$$

ceea ce ne permite să scriem

$$\Sigma = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}, \qquad \rho = \frac{\operatorname{cov}(X_1, X_2)}{\sigma_1\sigma_2} = 0.$$

Deoarece $\det \Sigma = 6$ urmează că

$$A = \Sigma^{-1} = \begin{pmatrix} \frac{5}{6} & -\frac{2}{6} \\ -\frac{2}{6} & \frac{2}{6} \end{pmatrix}, \qquad \det A = 1/6.$$

Deci

$$f(x) = \frac{1}{2\pi\sqrt{6}} \exp\left\{ -\frac{1}{2}(x_1 - 1, x_2 - 2) \begin{pmatrix} \frac{5}{6} & -\frac{2}{6} \\ -\frac{2}{6} & \frac{2}{6} \end{pmatrix} \begin{pmatrix} x_1 - 1 \\ x_2 - 2 \end{pmatrix} \right\}.$$

I.30. *Să se arate că dacă $X_1, X_2, \ldots, X_n$ sînt variabile aleatoare nenegative, simetrice şi astfel încît evenimentul că orice mulţime de variabile X egale cu zero este de probabilitate zero, atunci*

$$M\left(\frac{X_1 + X_2 + \ldots + X_k}{X_1 + X_2 + \ldots + X_n}\right) = \frac{k}{n}, \qquad 1 \leqslant k \leqslant n.$$

Soluţie. Din proprietatea de aditivitate a mediei rezultă că

$$M\left(\frac{X_1 + X_2 + \ldots + X_n}{X_1 + X_2 + \ldots + X_n}\right) =$$

$$= M\left(\frac{X_1}{X_1 + X_2 + \ldots + X_n}\right) + \ldots + M\left(\frac{X_n}{X_1 + X_2 + \ldots + X_n}\right) = 1.$$

Întrucît variabilele aleatoare $X_1, X_2, \ldots, X_n$ sînt simetrice şi esenţial pozitive putem scrie

$$M\left(\frac{X_i}{X_1 + \ldots + X_n}\right) = M\left(\frac{X_j}{X_1 + \ldots + X_n}\right), \quad i, j = 1, 2, \ldots, n.$$

Deci

$$M\left(\frac{X_i}{X_1 + X_2 + \ldots + X_n}\right) = \frac{1}{n}$$

şi de aici urmează că

$$M\left(\frac{X_1 + X_2 + \ldots + X_k}{X_1 + X_2 + \ldots + X_n}\right) = \sum_{j=1}^{k} M\left(\frac{X_j}{X_1 + \ldots + X_n}\right) = \frac{k}{n}.$$

I.31. *Fie X şi Y două variabile aleatoare astfel încît vectorul aleator (X, Y) să aibe ca densitate funcţia*

$$g(x, y) = \frac{B^2}{2\pi} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}} - \lambda(x)\lambda(y).$$

Să se dea condiţii suficiente asupra funcţiei $\lambda(x)$ şi asupra constantei B pentru ca g să fie o densitate, X şi Y fiind repartizate $N(0, 1)$.
În ce condiţii suplimentare corelaţia variabilelor X şi Y este nulă.

S o l u ţ i e. Din condiţia $g(x, y) > 0$ obţinem

$$\lambda(x)\,\lambda(y) < \frac{B^2}{2\pi} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}}$$

sau acelaşi lucru

$$\lambda(x) \leqslant \frac{|B|}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

g satisfăcînd condiţiile $g \geqslant 0$, şi g sumabilă este densitate.
Trebuie să determinăm λ şi B astfel încît

$$\int_{R^2} g(x, y)\, dx\, dy = 1.$$

Densitatea de repartiţie a variabilei aleatoare X este

$$\int_R g(x, y)\, dy.$$

Pentru ca

$$\int_R g(x, y)\, \mathrm{d}y = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}$$

este suficient ca

$$\int_R \lambda(y)\, \mathrm{d}y = 0 \quad \text{şi} \quad B = 1.$$

Pentru ca corelaţia dintre variabilele X şi Y să fie nulă trebuie ca $M(XY) = 0$.

Avem

$$M(X, Y) = \frac{1}{2\pi} \iint_{R^2} xy\, \mathrm{e}^{-\frac{x^2+y^2}{2}} \mathrm{d}x\, \mathrm{d}y + \frac{1}{2\pi} \iint_{R^2} xy\lambda(x)\, \lambda(y)\, \mathrm{d}x\, \mathrm{d}y =$$

$$= \left(\int_R x\lambda(x)\, \mathrm{d}x\right)^2.$$

Deci $M(X\,Y) = 0$ dacă

$$\int_R x\lambda(x)\, \mathrm{d}x = 0.$$

Fie

$$\lambda(x) = \begin{cases} \frac{1}{k}(3x^2 - 1) & \text{pentru } |x| < 1, \\ 0 & \text{în rest.} \end{cases}$$

Avem

$$\int_{-1}^{1} \lambda(x)\, \mathrm{d}x = 0$$

deoarece este pară

$$\int x\lambda(x)\, \mathrm{d}x = 0.$$

Pentru ca

$$|\lambda(x)| < \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\frac{x^2}{2}}$$

este suficient ca

$$\sup_{|x|<1} |\lambda(x)| < \inf_{|x|<1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{2\pi e}}$$

sau

$$\sup_{|x|<1} \lambda(x) = \frac{2}{k},$$

de unde $k > 2\sqrt{2\pi}$ e este o condiţie suficientă pentru ca vectorul (X, Y) de densitate $g(x, y)$ să aibă componentele repartizate $N(0, 1)$ şi de corelaţie nulă şi astfel ca vectorul (X, Y) să nu fie normal.

I.32. *Să se arate că există o variabilă aleatoare pentru care valoarea medie este de aceeaşi formă cu densitatea.*

Soluţie. Fie X o variabilă aleatoare repartizată $N(0, 1)$ şi

$$Y = \begin{cases} X, & \text{dacă} \quad X(\omega) > x_0, \\ 0, & \text{dacă} \quad X(\omega) < x_0. \end{cases}$$

Avem

$$M(Y) = \int_{-\infty}^{x_0} 0 \cdot f(x)\,\mathrm{d}x + \int_{x_0}^{\infty} x f(x)\,\mathrm{d}x =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{x_0}^{\infty} x\,e^{-\frac{x^2}{2}}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_0^2}{2}}.$$

Găsim $g(x)$ astfel încît

$$\int_x^{\infty} u g(u)\,\mathrm{d}u = g(x).$$

Derivînd ambii membri obţinem ecuaţia

$$-x g(x) = \frac{\mathrm{d}g(x)}{\mathrm{d}x}$$

a cărei soluţie generală este

$$g(x) = k e^{-\frac{x^2}{2}}.$$

Constanta de integrare k se determină astfel ca g să fie densitate; se obţine

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

I.33 *Considerăm un vector aleator* $X_{(n)} = (X_1, X_2, \ldots, X_n)'$ *ale cărui componente sînt variabile aleatoare discrete sau continue (nu neapărat independente) cu* $M(X_i) = \mu_i$, $D^2(X_i) = \sigma_i^2$, $1 \leqslant i \leqslant n$.

Dacă pentru orice $t \geqslant 1$ *fixat definim următoarele elemente*

$$A_j = \{X_{(n)} : | X_j - \mu_j | \leqslant n^{1/2} t\sigma_j\},$$

$$B_n = \bigcap_{j=1}^{n} A_j,$$

atunci

$$P(B_n) \geqslant 1 - \frac{1}{t^2}.$$

Soluţie. Avem

$$P(\bar{A}_j) = P(\{X_{(n)} : | X_j - \mu_j | > n^{1/2} t\sigma_j\} \leqslant \frac{1}{nt^2}$$

de unde

$$\sum_{j=1}^{n} P(\bar{A}_j) \leqslant \sum_{j=1}^{n} \frac{1}{nt^2} = \frac{1}{t^2}.$$

Dar

$$\bar{B}_n = \overline{\bigcap_{j=1}^{n} A_j} = \bigcup_{j=1}^{n} \bar{A}_j,$$

de unde urmează că

$$P(\bar{B}_n) \leqslant \sum_{j=1}^{n} P(\bar{A}_j) \leqslant \frac{1}{t^2},$$

adică

$$P(B_n) = 1 - P(\bar{B}_n) \geqslant 1 - \frac{1}{t^2}.$$

I.34. *Fie* (A) *o arie plană şi* $M(x, y)$ *un punct din* (A). *Să se arate că se poate presupune că probabilitatea elementară este proporţională cu aria elementară* $\mathrm{d}x\mathrm{d}y$, *în aşa fel încît probabilitatea ca* M *să fie într-o* *arie* (R) *interioară ariei* (A) *este egală cu* $\frac{R}{A}$, *unde am notat* $A = $ *măs* (A), $R = $ *măs* (R).

S o l u ţ i e. Fiind dată o a doua arie (A') şi un punct interior ei $M'(x', y')$, se poate presupune că probabilitatea ca M' să se afle într-o arie $(R') \subset (A')$ este reprezentată prin raportul $\frac{R'}{A'}$ (păstrînd notaţiile din enunţ).

Fie $\varphi(x, y; x', y')$ o funcţie de coordonatele punctelor M şi M' pe care o presupunem continuă. Cînd M şi M' parcurg ariile A şi A', valorile funcţiei φ variază într-un anumit interval (α, β).

Ne propunem să determinăm probabilitatea ca valoarea lui φ să aparţină intervalului $(\gamma, \gamma + \mathrm{d}\gamma) \subset (\alpha, \beta)$.

Probabilitatea căutată va fi $p(\gamma)\,\mathrm{d}\gamma$ care va trebui să verifice egalitatea $\int_\alpha^\beta p(\gamma)\,\mathrm{d}\gamma = 1$.

Dacă fixăm poziţia lui M, punctele M' pentru care $\gamma \leqslant \varphi \leqslant \gamma + \mathrm{d}\gamma$ sînt cuprinse între două arce de curbă infinit de apropiate şi măsura ariei limitate de aceste două arce este de forma $h(\gamma; x, y)\,\mathrm{d}\gamma$. Determinarea este o problemă de analiză.

Probabilitatea ca M' să fie interior acestei arii este $\frac{1}{A'} h(\gamma; x, y)\,\mathrm{d}\gamma$.

Dacă în loc ca M să fie fixat este numai supus la condiţia de a fi interior unui element de arie $\mathrm{d}x\,\mathrm{d}y$ ce cuprind punctul (x, y), iar $\mathrm{d}x$ şi $\mathrm{d}y$ sînt suficient de mici în raport cu $d\gamma$, modificarea adusă de $h(\gamma; x, y)\,\mathrm{d}\gamma$ este neglijabilă.

Acest raţionament este folosit frecvent şi avem evident dreptul ca să-l facem pentru că $\mathrm{d}\gamma$ este dat.

Însă probabilitatea ca M să se găsească în acest element de arie este $\frac{\mathrm{d}x\mathrm{d}y}{A}$ şi în virtutea probabilităţilor compuse, probabilitatea ca M fiind în acest element de arie, φ să fie cuprins între γ şi $\gamma + \mathrm{d}\gamma$ este egal cu produsul

$$h(\gamma; x, y)\,\frac{\mathrm{d}\gamma}{A'}\,\frac{\mathrm{d}x\,\mathrm{d}y}{A}.$$

Făcînd acum să varieze punctul $M(x, y)$ în aria (A) obţinem probabilitatea căutată

$$p(\gamma)\,\mathrm{d}\gamma = \frac{\mathrm{d}\gamma}{AA'} \iint_{(A)} h(\gamma; x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

I.35. *Fie s un întreg pozitiv. Să se arate că probabilitatea ca alegînd la întîmplare două numere întregi pozitive, ele să aibe cel mai mare divizor comun pe s este $\frac{1}{s^2\zeta(2)}$ unde $\zeta(t)$ este funcția zeta a lui Riemann.*

Soluție. Fie $\varphi_k(n)$ numărul întregilor a, astfel încît $0 < a \leqslant n$, $(a, n) = k$ unde prin (a, n) înțelegem cel mai mare divizor comun al numerelor a și n. Atunci

$$\varphi_k(kn) = \varphi(n),$$

unde $\varphi(n)$ este funcția φ a lui Euler.

Se constată că există s^2n^2 perechi de numere întregi (a, b) astfel încît $0 < a \leqslant sn$ și $0 < b \leqslant sn$.

Fie N_n numărul perechilor de numere întregi (a, b), astfel încît $(a, b) = s$.

Atunci calculînd numărul acestor perechi obținem

$$N_n = \varphi_s(s) + 2\varphi_s(2s) + \ldots + 2\varphi_s(ns).$$

care mai poate fi scris:

$$N_n = 2[\varphi(1) + \varphi(2) + \ldots + \varphi(n)] - 1.$$

Însă

$$\varphi(1) + \varphi(2) + \ldots + \varphi(n) = \frac{3}{\pi^2} n^2 + 0(n \lg n)$$

și de aici urmează că

$$\lim_{n\to\infty} \frac{N_n}{s^2n^2} = \frac{6}{s^2\pi^2} = \frac{1}{s^2\zeta(2)},$$

care este tocmai probabilitatea căutată.

Observație. Pentru calcularea acestei probabilități am folosit următoarele: Fie A mulțimea k — uplurilor de întregi pozitivi

$$A = \{(a_1, a_2, \ldots, a_k) : a_i \in N^*, \quad i = 1, 2, \ldots, k\}$$

și fie

$$A(n) = \{(a_1, a_2, \ldots, a_n) : a_i \in N^* ; a_i < n ; \quad 1 < i < k\}$$

atunci

$$P(A) = \lim_{n\to\infty} \frac{A(n)}{n^k},$$

care este densitatea asimptotică a mulțimii A.

I.36. *Fie X o variabilă aleatoare cu valori în N, şi Y o variabilă aleatoare astfel încît*

$$P(Y = y / X = x) = C_x^y p^y q^{x-y}, \qquad y = 0, 1, \ldots, x,$$

unde p şi q sînt fixe, astfel încît $p, q > 0$ şi $p + q = 1$.

Fie $Z = X - Y$. În ipoteza că Y şi Z sînt variabile aleatoare independente să se determine repartiţia lui X.

Soluţie. Fie $p_x = P(X = x)$. Pentru $0 \leqslant s, t \leqslant 1$,

$$M(s^X t^Z) = M\left(t^X\left(\frac{s}{t}\right)^Y\right) = \sum_{x=0}^{\infty} p_x t^x \sum_{y=0}^{x} C_y^x p^y q^{x-y}\left(\frac{s}{t}\right)^y =$$

$$= \sum_{x=0}^{\infty} p_x (ps + qt)^x = f(ps + qt).$$

Funcţia f este analitică pe $0 \leqslant s < 1$, deci indefinit derivabilă şi condiţia de independenţă revine la

$$\frac{\partial^2}{\partial s \partial t} \operatorname{Log} f(ps + qt) = 0.$$

Punînd $g(s) = \operatorname{Log} f(s)$, se obţine

$$pqg''(ps + qt) = 0, \qquad 0 \leqslant s, t < 1.$$

Deci $g(s) = \lambda(s - 1) + c$ unde λ şi c sînt constante.

$f(1) = 1$ implică $c = 0$ şi $f(s) \leqslant 1$ dacă $0 \leqslant s < 1$ implică $\lambda > 0$.

Se observă că este necesar ca X să urmeze o repartiţie Poisson de parametru λ.

Verificăm că această condiţie este suficientă, în particular că λ nu depinde de p.

Dacă $P(X = x) = e^{-\lambda} \dfrac{\lambda^x}{x!}$ şi dacă Y/X urmează o repartiţie binomială de parametru p şi X, repartiţia variabilei Y este dată prin

$$P(Y = y) = \sum_{z=y}^{\infty} C_n^y p^y q^{x-y} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda p} \frac{(\lambda p)^y}{y!}.$$

Deci Y urmează o repartiţie Poisson de parametru λp.

Dacă Y şi Z sînt două variabile aleatoare independente, urmînd repartiţii Poisson de parametri λp respectiv λq, atunci $Y + Z$ urmează o repartiţie Poisson de parametru λ.

Urmează că

$$P(Y = y \mid Y + Z = x) = C_x^y p^y q^{y-x}, \qquad y = 0, 1, \ldots, x.$$

I.37. *Fie $X_1, \dots, X_n$ variabile aleatoare independente, fiecare urmînd aceeaşi repartiţie definită prin $P(X_n > x) = e^{-x}$ dacă $x \geqslant 0$.*

a) *Să se calculeze, în R^{n-1} densitatea vectorului*

$$\left(\frac{S_1}{S_n}, \frac{S_2}{S_n}, \dots, \frac{S_{n-1}}{S_n}\right),$$

unde

$$S_k = X_1 + X_2 + \dots + X_k, \qquad k = 1, 2, \dots, n.$$

b) *Fie $X'_1, \dots, X'_n$ şirul format cu variabilele X_i aranjate în ordine crescătoare şi $T_1 = X'_1$ şi $T_k = X'_k - X'_{k-1}$, $k = 2, 3, \dots, n$.*

Să se arate că $(n - 1 + k)\, T_k$ sînt variabile aleatoare independente fiecare urmînd aceeaşi repartiţie şi să se calculeze $M(T_k \mid S_n)$.

c) *Se rupe un segment de lungime l în $(n - 1)$ puncte alese uniform pe segment şi se ordonează segmentele astfel formate în ordinea lungimii. Să se determine lungimea medie a segmentului de ordin k.*

S o l u ţ i e. a) În R^n densitatea vectorului $(X_1, \dots, X_n)$ este pe $(x_i \geqslant 0,\ i = 1, 2, \dots, n)$, $\exp\,(-x_1 - x_2 - \dots - x_n)$.

Densitatea vectorului $(S_1, S_2, \dots, S_n)$ este e^{-s_n} pe $0 \leqslant s_1 \leqslant s_2 \leqslant \dots \leqslant s_n$, iar a vectorului

$$\left(\frac{S_1}{S_n}, \frac{S_2}{S_n}, \dots, \frac{S_{n-1}}{S_n}, S_n\right)$$

este $s_n^{n-1}\, e^{-s_n}$ pe $0 \leqslant s_1 \leqslant \dots \leqslant s_{n-1} \leqslant 1;\ s_n \geqslant 0$.

Integrînd această densitate în raport cu S_n, găsim că densitatea vectorului $\left(\frac{S_1}{S_n}, \frac{S_2}{S_n}, \dots, \frac{S_{n-1}}{S_n}\right)$ este $(n-1)!$ pe $0 \leqslant s_1 \leqslant s_2 \dots \leqslant s_{n-1} \leqslant 1$.

b) În R^n densitatea vectorului $(X_1, \dots, X_n)$ este pe $(x_1 \geqslant 0,\ i = 1, 2, \dots, n)$ $\exp\,(-x_1 - x_2 - \dots - x_n)$.

Densitatea vectorului $(X'_1, X'_2, \dots, X'_n)$ este

$$n!\ \exp\,(-x'_1 - x'_2 - \dots - x'_n)$$

pe $(0 \leqslant x'_1 \leqslant x'_2 \leqslant \dots \leqslant x'_n)$, iar a vectorului $(T_1, T_2, \dots, T_n)$ este

$$n! \exp\,(-nt_1 - (n-1)\,t_2 - \dots - t_n) \text{ pe } (t'_1 \geqslant 0, \qquad i = 1, \dots, n).$$

Se observă că această densitate este un produs de densităţi, ceea ce arată că variabilele T_i sînt independente şi că

$$P((n - k + 1)\, T_k > x) = e^{-x} \text{ dacă } x \geqslant 0.$$

Deci

$$M((n-k+1)\, T_k/(nT_1+(n-1)T_2+\ldots+T_n))=\frac{1}{2}$$

şi deci

$$M(T_k/S_n)=1/n(n-k+1).$$

c) Fie I_k lungimea segmentului de ordin k, aceasta fiind dată. Din a), repartiţia lui I_k este aceeaşi cu a lui X'_k/S_n. Cum $X'_k=T_1+\ldots+T_k$, după b)

$$M(I_k)=\frac{1}{n}\left(\frac{1}{n}+\frac{1}{n-1}+\ldots+\frac{1}{n-k+1}\right).$$

I.38. *Fie* $X_1, \ldots, X_n, \ldots$ *variabile aleatoare independente, fiecare urmînd aceeaşi repartiţie definită prin* $P(X_n=\pm 1)=1/2$ *şi*

$$S_n=X_1+X_2+\ldots+X_n.$$

Să se arate că sigur

$$\lim_{n\to\infty}\,(2n\,\mathrm{Log}\,n)^{-1/2}\,|\,S_n\,|\leqslant 1.$$

Soluţie. Avem

$$e^{xS_n}\geqslant e^{xS_n}I_{\{S_n\geqslant a\}}\geqslant e^{ax}I_{\{S_n\geqslant a\}}$$

şi cum

$$M(e^{xS_n})=(M(e^{xX_1})^n=(\mathrm{ch}\, x)^n.$$

Aplicînd media membrilor extremi ai acestei inegalităţi obţinem:

$$P(S_n\geqslant a)\leqslant e^{-ax}(\mathrm{ch}\, x)^n.$$

Considerînd $f(x)=\frac{x^2}{2}-\mathrm{Log}\;\mathrm{ch}\, x$, se observă că a doua derivată a acestei funcţii este întotdeauna pozitivă, şi deci că

$$f(x)\geqslant f(0)=0$$

Deci

$$P(S_n\geqslant a)\leqslant \exp\left(-ax+\frac{nx^2}{2}\right).$$

Minimul membrului doi al acestei ultime inegalităţi fiind atins în $x=a/n$ urmează că

$$P(S_n\geqslant a)\leqslant \exp\left(-\frac{a^2}{2n}\right).$$

Făcînd $a = c(2n \operatorname{Log} n)^{1/2}$, se obține ținînd seama de simetria lui S_n.

$$P(|S_n|) \geqslant c(2n \operatorname{Log} n)^{1/2} \leqslant 2n^{-c^2}.$$

Al doilea membru al acestei inegalități converge dacă $c > 1$. Deci, după lema Borel-Cantelli, evenimentul

$$\{|S_n| > c(2n \operatorname{Log} n)^{1/2}\}$$

nu are loc aproape sigur decît de un număr finit de ori, adică, aproape sigur

$$\overline{\lim}\,(2n \operatorname{Log} n)^{-1/2} |S_n| \leqslant c \quad \text{oricare ar fi} \quad c > 1.$$

I.39. *Fie* $X_1, \ldots, X_n, \ldots,$ *variabile aleatoare independente, fiecare urmînd aceeași repartiție definită prin* $P(X_n = \pm 1) = \frac{1}{2}$ *și șirul de numere* $c_1, \ldots, c_n, \ldots,$ *astfel încît* $\gamma^2 = \sum_1^\infty c_n^2 < \infty$.

Să se arate că $\sum_{n=1}^{\infty} c_n X_n$ *converge în probabilitate.*

S o l u ț i e. Fie

$$R = \sum_{n<k\leqslant N} c_k X_k \quad \text{și} \quad \gamma_n^2 = \sum_{k>n} c_k^2.$$

Avem

$$M(e^{xR}) = \prod_{n<k\leqslant N} \operatorname{ch}(xc_n)$$

și cum

$$\operatorname{ch} x \leqslant \exp \frac{x^2}{2},$$

obținem

$$M(e^{xR}) \leqslant \exp\left(\frac{x^2\gamma_n^2}{2}\right) \text{ oricare ar fi } x.$$

Ca și în problema I.38

$$P(|R| > \varepsilon) < 2P(R > \varepsilon) \leqslant 2 \inf_x e^{-x\varepsilon} M(e^{xR}) \leqslant$$

$$\leqslant 2 \inf_x \exp\left\{-x\varepsilon + \frac{x^2\gamma_n^2}{2}\right\} = 2\exp\left(-\frac{2}{2\gamma_n^2}\right).$$

Deci $P(|R| > \varepsilon) \xrightarrow[n\to\infty]{} 0$ oricare ar fi $\varepsilon > 0$.

Criteriul lui Cauchy fiind verificat, seria converge în probabilitate.

I.40. *Se consideră şirul de numere reale* $(a_n)_{n\in N^*}$, $a_n > a_{n+1} > 0$, $n = 1, 2, \ldots$, *şi se definesc variabilele aleatoare* $(X_n)_{x\in N^*}$ *astfel încît*

$$P(X_n = a_n) = P(X_n = -a_n) = \frac{1}{2}.$$

Să se arate că dacă variabilele aleatoare X_n *sînt independente şi* $\sum\limits_{n=1}^{\infty} a_n^2 < \infty$, *atunci seria* $\sum\limits_{n=1}^{\infty} X_n$ *converge aproape sigur către o variabilă aleatoare* X *a cărei funcţie de repartiţie este continuă.*

Soluţie. Din enunţ, rezultă

$$M(X_n) = 0, \qquad D^2(X_n) = a_n^2,$$

şi

$$\sum_{n=1}^{\infty} D^2(X_n) = \sum_{n=1}^{\infty} a_n^2 < \infty.$$

Aceasta ne spune că şirul $S_n = \sum\limits_{k=1}^{n} X_k$ converge aproape sigur către o variabilă aleatoare X.

Să arătăm că funcţia de repartiţie a variabilei aleatoare X este continuă. Într-adevăr, pentru orice $l \geqslant 0$ avem

$$Q_X(l) = \sup_x [F_X(x+l) - F_X(x-0)].$$

Pe de altă parte, dacă X şi Y sînt două variabile aleatoare independente, atunci

$$Q_X(l) \geqslant Q_{X+Y}(l) \quad (\forall)\, l \geqslant 0.$$

Dacă punem

$$Q_n(l) = \sup_x [F_n(x+l) - F_n(x-\text{o})],$$

unde $F_n(x)$ este funcţia de repartiţie a variabilei aleatoare

$$S_n = \sum_{k=1}^{x} X_k,$$

atunci putem scrie

$$Q_{n-1}(l) \geqslant Q_n(l) \geqslant Q(l).$$

În particular

$$Q_{n-1}(0) \geqslant Q_n(0) \geqslant Q(0).$$

Însă datorită faptului că $a_n > a_{n+1} > 0$ rezultă că

$$Q_n(0) = \sup\,[F_n(x) - F_n(x-0)] \leqslant \frac{n}{2^n}$$

adică

$$\frac{n}{2^n} \geqslant Q_n(0) \geqslant Q(0) = \sup_x\,[F(x) - F(x-0)].$$

Cum

$$\lim_{n\to\infty} Q_n(0) = 0,$$

rezultă că $F(x) - F(x-0) = Q(0) = 0$ pentru orice x real, de unde tragem concluzia că $F(x)$ este o funcţie de repartiţie continuă.

I.41. *Fie* $X_1, X_2, \ldots, X_n, \ldots,$ *variabile aleatoare independente şi uniform repartizate pe* $[0, 1]$. *Fie*

$$Y_n = \min_{1\leqslant k\leqslant n} X_k \text{ şi } S_n = Y_1 + Y_2 + \ldots + Y_n,$$

Să se arate că $\frac{S_n}{\log n}$ *converge în medie pătratică către* 1 *dacă* n *tinde către* ∞.

Soluţie. Deoarece

$$P(Y_i > x) = (1-x)^i \quad \text{dacă} \quad x \in [0, 1],$$

$$M(Y_i) = \int_0^1 ix(1-x)^{i-1}\,\mathrm{d}x = \frac{1}{i+1}.$$

Deci

$$s_n = M(S_n) = \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n+1}.$$

Cum

$$\int_2^{n+1} \frac{\mathrm{d}x}{x} < s_n < \int_1^{n+1} \frac{\mathrm{d}x}{x}, \quad s_n \sim \log n \text{ dacă } n \to \infty.$$

Deci după egalitatea

$$M\left[\left(\frac{S_n}{\log n}-1\right)^2\right]\leqslant\frac{\sigma^2(S_n)}{(\mathrm{Log}\,n)^n}+\left(\frac{s_n}{\mathrm{Log}\,n}-1\right)^2,$$

este suficient să arătăm că

$$\frac{\sigma^2(S_n)}{(\mathrm{Log}\,n)^2}\to 0 \quad \text{dacă} \quad n\to\infty.$$

Avem

$$M(S_n^2)=\sum_1^n M(Y_i^2)+2\sum_{i=1}^{n-1}\sum_{k=1}^{n-i}M(Y_iY_{i+k}),$$

$$M(Y_i^2)=\int_0^1 ix^2(1-x)^{i-1}\,\mathrm{d}x=\frac{2}{(i+1)(i+2)}.$$

Repartiţia lui (Y_n, Y_{n+k}) este dată de

$$P(y\leqslant Y_{i+k}\leqslant Y_i\leqslant x)=$$

$$=P(y\leqslant X'_{i+1},\ldots,X'_{i+k})\,P(Y_i\leqslant x)=$$

$$=(1-(1-x)^i)\,(1-y)^k \quad \text{dacă} \quad y\leqslant x.$$

Această repartiţie are o densitate în domeniul $0\leqslant y<x<1$ dată de derivata parţială $-\dfrac{\partial^2}{\partial x\partial y}$ a expresiei precedente

$$ki(1-x)^{i-1}(1-y)^{k-1}.$$

Însă probabilitatea $P(Y_{i+k}=Y_i\leqslant x)$ nu este nulă şi este egală cu

$$P(X_{i+1},\ldots,X_{i+k}\geqslant Y_i;\quad Y_i\leqslant x)=\int_0^x n(1-u)^{n-1}(1-u)^k\,du.$$

Deci

$$M(Y_iY_{i+k})=kn\iint\limits_{0\leqslant y<x\leqslant 1}xy(1-x)^{n-1}(1-y)^{k-1}\,\mathrm{d}x\mathrm{d}y+$$

$$+n\int_0^1 x^2(1-x)^{n-k+1}\,\mathrm{d}x.$$

Avem

$$M(Y_iY_{i+k})=\frac{1}{i+k+2}\left(\frac{1}{i+1}+\frac{1}{i+k+1}\right)$$

şi

$$M(S_n^2) = 2\sum_{j=1}^{n} \frac{1}{j+2}\left(\frac{j-1}{j+1} + s_j\right).$$

Cum

$$\frac{j-1}{(j+2)(j+2)} = \frac{3}{j+2} - \frac{2}{j+1},$$

urmează că

$$2\sum_{j=1}^{n} \frac{j-1}{(j+1)(j+2)} \sim 2s_n.$$

Cum

$$2\sum_{j=1}^{n} \frac{s_j}{j+2} = 2\sum_{j=2}^{n+1}\sum_{i<j} \frac{1}{(j+1)(i+1)} = s_n^2 + 0(1),$$

$$\sigma^2(S_n) = M(S_n^2) - s_n^2 \sim 2s_n$$

şi deci

$$\lim_{n\to\infty} \frac{\sigma^2(S_n)}{(\mathrm{Log}\, n)^n} = \lim_{n\to\infty} \frac{2s_n}{(\mathrm{Log}\, n)^2} = 0.$$

I.42. *Fie $X_1, \ldots, X_n, \ldots,$ variabile aleatoare independente, fiecare urmînd aceeaşi repartiţie definită prin $P(X_n = \pm 1) = 1/2$. Să se arate că variabila aleatoare $Y = \sum_1^{\infty} \frac{X_n}{n}$ există şi posedă o densitate.*

Soluţie. 1°. Din problema I.39, deoarece $\sum_1^{\infty} \frac{1}{n^2}$ este convergentă, variabila aleatoare Y există.

Dacă punem

$$X = \sum_1^{\infty} \frac{X_{2^n}}{2^n} \quad \text{şi} \quad Z = Y - X,$$

X este repartizată uniform pe $[-1, +1]$.

Într-adevăr, după problema I.18

$$\frac{X+1}{2} = \sum_1^{\infty} \frac{Y_n}{2^n} \quad \text{sau} \quad Y_n = \frac{X_{2^n}+1}{2}$$

are o densitate constantă pe $[0,1]$.

2°. A doua soluţie se obţine cu ajutorul funcţiei caracteristice.

Funcţia caracteristică a variabilei $\sum_{1}^{N} \frac{X_{2n}}{2^n}$,

$$\cos\frac{t}{2}\cos\frac{t}{2^n}\dots\cos\frac{t}{2^N}=\frac{\sin t}{2^N\sin\frac{t}{2^N}},$$

pentru $N\to\infty$ este $\frac{\sin t}{t}$, care este funcţia caracteristică corespunzătoare unei repartiţii uniforme.

Y este deci suma a două variabile aleatoare independente X şi Z, dintre care una posedă o densitate. Deci Y posedă de asemenea o densitate.

I.43. *Fie variabila aleatoare care dă numărul necesar de aruncări ale unui zar pentru a obţine prima oară faţa 1. Să se determine funcţia generatoare de momente factoriale $\theta(t)$ a variabilei X, media şi dispersia variabilei X.*

Soluţie. Variabila aleatoare X avînd repartiţia

$$X:\begin{pmatrix} 1 & 2 & 3 & \dots & n & \dots \\ \frac{1}{6} & \left(\frac{1}{6}\right)^2 & \left(\frac{1}{6}\right)^2 & \dots & \left(\frac{1}{6}\right)^n & \dots \end{pmatrix}$$

urmează că

$$\theta(t)=M(t^x)=\sum_{1}^{\infty}\frac{t^n}{6^n}=\frac{\frac{t}{6}}{1-\frac{t}{6}}$$

sau

$$\theta(t)=\frac{t}{6-t}\cdot$$

Ţinînd seama că

$$M(X)=\theta'(t)\,|_{t=1};\quad M(X)=\frac{6}{25}$$

şi

$$M(X^2)-M(X)=\theta''(t)\,|_{t=1}$$

obţinem

$$M(X^2)=\frac{6}{5^2}+\frac{12}{5^3}=\frac{42}{5^3}\cdot$$

Deci

$$D^2(X) = \frac{42}{5^3} - \frac{36}{5^4} = \frac{174}{54}.$$

I.44. *Fie*

$$f(x, y) = \begin{cases} 0, & \textit{dacă min } (x, y) < 0, \\ xye^{-(x+y)}, & \textit{dacă } \min (x, y) \geqslant 0. \end{cases}$$

Se cere:

a) $F(x, y)$, $F_X(x)$, $F_Y(y)$, $f_X(x)$, $f_Y(y)$.
b) *Funcția caracteristică a vectorului* (X, Y).
c) *Momentele de diferite ordine.*

Soluție. a). Prin definiție

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)\,\mathrm{d}u\mathrm{d}v =$$

$$= \begin{cases} 0, & \text{dacă } \min (x, y) < 0. \\ \displaystyle\int_0^x\int_0^y uv\, \mathrm{e}^{-(u+v)}\,\mathrm{d}u\mathrm{d}v, & \text{dacă } \min (x, y) \geqslant 0. \end{cases}$$

Urmează deci, că

$$F(x, y) = \begin{cases} 0 & \text{dacă } \min (x, y) < 0, \\ [1 - (1 + x)\,\mathrm{e}^{-x}]\,[1 - (1 + y)\,\mathrm{e}^{-y}] & \text{dacă } \min (x, y) \geqslant 0. \end{cases}$$

$$F_X(x) = \lim_{y\to\infty} F(x, y) = \begin{cases} 0 & \text{dacă } x < 0, \\ 1 - (1 + x)\,\mathrm{e}^{-x}, & \text{dacă } x \geqslant 0. \end{cases}$$

Analog

$$F_Y(y) = \lim_{x\to\infty} F(x, y) = \begin{cases} 0, & \text{dacă } y < 0, \\ 1 - (1 + y)\,\mathrm{e}^{-y}, & \text{dacă } y \geqslant 0. \end{cases}$$

De aici rezultă imediat că

$$f_X(x) = \begin{cases} 0 & \text{dacă } x < 0, \\ x\,\mathrm{e}^{-x} & \text{dacă } x \geqslant 0. \end{cases}$$

$$f_Y(y) = \begin{cases} 0 & \text{dacă } y < 0, \\ y\,\mathrm{e}^{-y} & \text{dacă } y \geqslant 0. \end{cases}$$

b) Din definiţia funcţiei caracteristice, rezultă

$$\varphi(s, t) = \int_0^\infty \int_0^\infty e^{i(st+ty)}\, xy\, e^{-(x+y)}\, dx\, dy =$$

$$= \int_0^\infty x\, e^{-(1-is)x}\, dx \int_0^\infty y\, e^{-(1-it)y}\, dy$$

de unde

$$\varphi(s, t) = (1 - is)^{-2} \cdot (1 - it)^{-2}.$$

c) Folosind fie definiţia momentelor de diferite ordine, exprimate cu ajutorul densităţii de repartiţie, fie cu ajutorul funcţiei caracteristice, găsim

$$M(X^k \cdot Y^l) = (k + 1)!\, (l + 1)!$$

Observaţie. Se vede că variabilele X şi Y sînt independente.

I.45. *Dacă o variabilă aleatoare X are o densitate de repartiţie f simetrică faţă de origine, să se arate că φ ia numai valori reale.*

Soluţie. Densitatea de repartiţie fiind simetrică satisface egalitatea $f(x) = f(-x)$. Urmează că

$$\varphi(t) = M(e^{itX}) = \int_{-\infty}^\infty \cos tx\, f(x)\, dx + i \int_{-\infty}^\infty \sin tx\, f(x)\, dx$$

şi cum

$$\int_{-\infty}^\infty \sin txf(x)\, dx = 0$$

obţinem

$$\varphi(t) = \int_{-\infty}^\infty \cos txf(x)\, dx$$

I.46. *Fie X o variabilă aleatoare simetrică astfel încît $|X| \geqslant 1$. Să se arate că funcţia sa caracteristică $\varphi(t)$ posedă întotdeauna un zero în $(0, \pi)$.*

Soluţie. X fiind simetrică, $\varphi(t)$ este reală. Dacă $F(x) = P(|X| < x)$, atunci

$$\varphi(t) = \int_1^\infty \cos tx\, dF(x).$$

Obţinem

$$\int_0^\pi \varphi(t) \sin t\, dt = \int_1^\infty dF(x) \int_0^\pi \sin t \cos tx\, dt =$$

$$= \int_1^\infty \frac{1 + \cos \pi x}{1 - x^2}\, dF(x) \leqslant 0.$$

Deci $\varphi(t)$ nu poate fi tot timpul pozitivă în $(0, \pi)$.

I.47. *Fie n variabile aleatoare independente $X_1, X_2, \ldots, X_n$ fiecare repartizată normal $N(\mu, \sigma_j)$, $1 \leqslant j \leqslant n$.*

Să se arate că suma ponderată

$$X = \sum_{j=1}^{n} W_j X_j$$

este de asemenea normal repartizată şi să se afle dispersia variabilei aleatoare X în funcţie de dispersiile σ_j^2, $1 \leqslant j \leqslant n$.

Ştiind că $\sum_{j=1}^{n} W_j = 1$ să se deducă sistemul de ponderi $W_1, W_2, \ldots, W_n$ astfel încît variabila X să aibă dispersia minimă.

Soluţie. Deoarece funcţia caracteristică a variabilei aleatoare X_j, $1 \leqslant j \leqslant n$ este

$$\varphi_j(t) = e^{i\mu t - \frac{t^2\sigma_j^2}{2}}, \quad 1 \leqslant j \leqslant n,$$

urmează că

$$\psi_X(t) = M(e^{itX}) = M\left(e^{it\sum_{j=1}^{n} W_j X_j}\right) = M\left(\prod_{j=1}^{n} e^{itW_jX_j}\right) =$$

$$= \prod_{j=1}^{n} M(e^{itW_jX_j}) = \prod_{j=1}^{n} e^{itW_j\mu - \frac{t^2W_j^2\sigma_j^2}{2}},$$

ceea ce dovedeşte că variabila aleatoare X urmează o repartiţie normală

$$N\left(\mu; \sqrt{\sum_{j=1}^{n} W_j^2\sigma_j^2}\right) \text{ de medie } \mu \text{ şi dispersie } \sum_{j=1}^{n} W_j^2\sigma_j^2.$$

Pentru a afla sistemul de ponderi W_j, $1 \leqslant j \leqslant n$, va trebui să rezolvăm o problemă de minimum cînd variabilele W_j, $1 \leqslant j \leqslant n$ sînt legate prin relaţia:

$$\sum_{j=1}^{n} W_j = 1.$$

Aplicînd metoda multiplicatorilor lui Lagrange putem scrie

$$D^2(X) = \sum_{j=1}^{n} W_j^2\sigma_j^2 - \lambda\left(\sum_{j=1}^{n} W_j - 1\right),$$

care trebuie să fie minimă.

De aici rezultă

$$\frac{\partial D^2(X)}{\partial W_j} = 0, \quad 1 \leqslant j \leqslant n$$

sau

$$2W_j\sigma_j^2 - \lambda = 0, \quad 1 \leqslant j \leqslant n,$$

$$\lambda = \frac{1}{2W_j\sigma_j^2}, \quad 1 \leqslant j \leqslant n,$$

relaţie care mai poate fi scrisă

$$W_j = \frac{1}{2\lambda\sigma_j^2}, \quad 1 \leqslant j \leqslant n.$$

Sumînd după $j = 1, \ldots, n$ şi ţinînd seama de faptul că

$$\sum_{j=1}^{n} W_j = 1$$

obţinem

$$\lambda = \frac{1}{2}\sum_{j=1}^{n}\frac{1}{\sigma_j^2},$$

ceea ce ne conduce în final la rezultatul

$$W_j = \frac{1}{\sigma_j^2 \sum\limits_{j=1}^{n} \frac{1}{\sigma_j^2}} \; 1 \leqslant j \leqslant n,$$

I.48. *Să se determine funcţia de frecvenţă a variabilei aleatoare X care are funcţia caracteristică*

$$\varphi(t) = \frac{e^{it}(1 - e^{int})}{n(1 - e^{it})}.$$

S o l u ţ i e. Putem scrie

$$\varphi(t) = \frac{e^{it}[1 - (e^{it})^n]}{n(1 - e^{it})} = \frac{e^{it}(1 - e^{it})}{n(1 - e^{it})}(1 + e^{it} + e^{2it} + \ldots + e^{(n-1)it}) =$$

$$= \frac{e^{it} + e^{i2t} + \ldots + e^{int}}{n}$$

şi cum $\varphi(t) = M(e^{itX})$, urmează că variabila aleatoare X are repartiţia

$$X: \begin{pmatrix} 1 & 2 & \dots & n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

I.49. *Să se determine densitatea de repartiţie a variabilei aleatoare X care are funcţia caracteristică*

$$\varphi(t) = \frac{1}{1+t^2}.$$

Soluţie. Aplicînd formula de inversiune putem scrie

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-itx}}{1+t^2}\,dt.$$

Vom calcula ultima integrală cu ajutorul reziduurilor. În acest scop vom considera în planul complex conturul (fig. I.1)

Atunci

$$\int_C \frac{e^{-iaz}}{1+z^2}\,dz = \int_\Gamma \frac{e^{-iaz}}{1+z^2}\,dz + \int_{-R}^{-r} \frac{e^{-ita}}{1+t^2}\,dt + \int_\gamma \frac{e^{-iza}}{1+z^2}\,dz +$$

$$+ \int_r^R \frac{e^{-ita}}{1+t^2}\,dt = 2\pi i \operatorname{rez}\left(\frac{e^{-iaz}}{1+z^2}\right)_{z=+i}$$

Avem

$$\operatorname{rez}\left(\frac{e^{-iaz}}{1+z^2}\right)_{z=i} = \frac{e^{+a}}{2i}$$

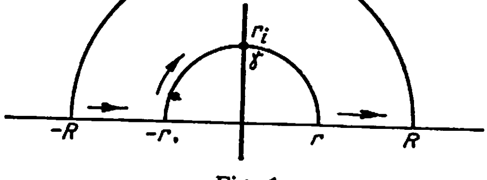

Fig. 1

şi cum

$$\lim_{R\to\infty} \frac{|z|\,|e^{-iaz}|}{|1+z^2|} = \lim_{R\to\infty} \frac{Re^{ay}}{|1+z^2|} = 0 \text{ pentru } a < 0,$$

$$\lim_{r\to 0} \frac{|z|\,|e^{-iaz}|}{|1+z^2|} \leqslant \lim_{r\to 0} \frac{r\,e^{ay}}{1-r^2} = 0$$

la limită rezultă

$$\int_{-\infty}^{\infty} \frac{e^{-itx}}{1+t^2}\, dt = \pi e^x \text{ pentru } x < 0$$

şi dacă înlocuim în expresia lui $f(x)$ obţinem

$$f(x) = \frac{1}{2}\, e^x, \text{ pentru } x < 0.$$

Pentru $x > 0$ se consideră conturul simetric acestuia faţă de axa Ox.

Se obţine în final

$$f(x) = \frac{1}{2}\, e^{-|x|}, \quad x \in R_1.$$

I.50. *Să se determine funcţia caracteristică corespunzătoare variabilei aleatoare X caracterizate prin densitatea de repartiţie*

$$f(x) = \frac{k}{\pi[k^2 + (x-\mu)^2]}, \quad -\infty < x < \infty,$$

unde k şi μ sînt constante reale $(k > 0)$.

S o l u ţ i e. Prin definiţie şi din datele problemei

$$\varphi(t) = \frac{k}{\pi} \int_{-\infty}^{\infty} \frac{e^{itx}}{k^2 + (x-\mu)^2}\, dx.$$

Considerăm integrala

$$I(z) = \int_C \frac{e^{itz}}{k^2 + (z-\mu)^2}\, dz$$

şi conturul C din figura I.2.

Aplicînd teorema reziduurilor acestui contur închis, obţinem

$$\int_r^R \frac{e^{itx}}{(x-\mu^2)+k^2}\,dx + \int_\Gamma \frac{e^{itz}}{k^2+(z-\mu)^2}\,dz + \int_{-R}^{-r} \frac{e^{itx}}{(x-\mu)^2+k^2}\,dx +$$

$$+\int_\gamma \frac{e^{itz}}{(z-\mu)^2+k^2}\,dz = 2\pi i \operatorname{rez}\left(\frac{e^{itz}}{(z-\mu)^2+k^2}\right)_{z=\mu+ik}$$

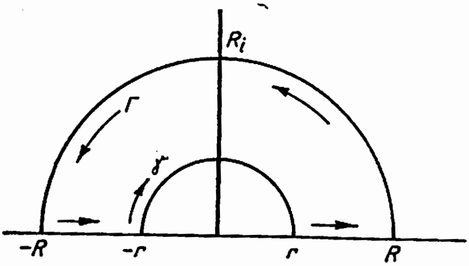

Fig. 2

Avem

$$\operatorname{rez}\left(\frac{e^{itz}}{(z-\mu)^2+k^2}\right)_{z=\mu+ik} = \frac{e^{it\mu-tk}}{2ik}$$

şi cum

$$\lim_{R\to\infty} \frac{|z|\,e^{-ty}}{|(z-\mu)^2+k^2|} = 0, \quad t>0;$$

$$\lim_{r\to 0} \frac{|z|\,e^{-ty}}{|(z-\mu)^2+k^2|} = 0,$$

rezultă că pentru $r\to 0$ şi $R\to\infty$

$$\int_{-\infty}^{\infty} \frac{e^{itx}}{(x-\mu)^2+k^2}\,dx = \frac{\pi}{k}\, e^{t(i\mu-k)},$$

de unde

$$\varphi(t) = e^{t(i\mu-k)} \text{ pentru } t>0.$$

Pentru $t<0$ se consideră conturul simetric acestuia faţă de axa Ox şi se obţine în final

$$\varphi(t) = e^{it\mu-k|t|}.$$

I.51. *Fie X_1 şi X_2 două variabile aleatoare independente repartizate χ^2 cu $2n$ şi respectiv $2m$ grade de libertate. Să se arate că funcţia caracteristică a variabilei aleatoare*

$$Y = aX_1 - bX_2, \quad a>0,\ b>0$$

este dată de

$$\varphi_Y(t) = \sum_{j=1}^{n} C_{n+m-j-1}^{m-1} \left(\frac{a}{a+b}\right)^m \left(\frac{b}{a+b}\right)^{n-j} (1-2\mathrm{i}\,at)^{-j} +$$

$$+ \sum_{k=1}^{m} C_{n+m-k-1}^{n-1} \left(\frac{b}{a+b}\right)^n \left(\frac{a}{a+b}\right)^{m-k} (1+2\mathrm{i}\,bt)^{-k}.$$

S o l u ţ i e. Din faptul că variabilele aleatoare X_1 şi X_2 sînt independente şi repartizate χ^2 cu $2n$ şi respectiv $2m$ grade de libertate, rezultă că

$$\varphi_Y(t) = M(\mathrm{e}^{\mathrm{i}tY}) = M(\mathrm{e}^{\mathrm{i}t(aX_1 - bX_2)}) = M(\mathrm{e}^{\mathrm{i}atX_1})\, M(\mathrm{e}^{\mathrm{i}(-b)tX_2})$$

care ne conduce la

$$\varphi_Y(t) = (1 - 2\mathrm{i}at)^{-n} (1 + 2\mathrm{i}\,bt)^{-m}.$$

Să introducem notaţiile $x = 2\mathrm{i}t$; $p = 1 - q = \frac{a}{a+b}(1 + bx)$; $u = n - j$; $v = m - k$.

Atunci problema revine la a arăta că

$$p^m \sum_{u=0}^{n-1} C_{m-u-1}^{u} q^u + q^n \sum_{v=0}^{m-1} C_{n+v-1}^{v} p^v = 1,$$

adică membrul al doilea al expresiilor sub care apare $\varphi_Y(t)$ reprezintă acelaşi lucru.

Pentru a dovedi egalitatea de mai sus procedăm prin inducţie după n şi m. Pentru $n = m = 1$ egalitatea se reduce la $p + q = 1$ care este adevărată.

O vom presupune adevărată pentru m fixat şi $n = k$ şi vom arăta că se menţine adevărată pentru $n = k + 1$ apoi vom schimba rolurile lui m cu n.

Deci pentru m fixat şi $n = k + 1$ obţinem:

$$p^m \sum_{u=0}^{k} C_{m+u-1}^{u} q^u = 1 - q^{k+1} \sum_{v=0}^{m-1} C_{k+v}^{v} p^v.$$

Cum relaţia am presupus-o adevărată pentru $n = k$, obţinem

$$\sum_{v=1}^{m-1} [C_{k+v}^{v} - C_{k+v-1}^{v}]\, p^v = \sum_{v=1}^{m-1} C_{k+v-1}^{v-1} p^v.$$

Dar cum

$$C_{k+v}^{v} - C_{k+v-1}^{v} = C_{k+v-1}^{v-1},$$

rezultă că pentru m fixat şi $n = k + 1$ relaţia se menţine adevărată. Schimbăm acum rolului n cu m, obţinem justificarea relaţiei care probează afirmaţia făcută.

I.52. *Se consideră vectorul aleator* (X, Y) *a cărui densitate de repartiţie este* $f(x, y)$, $-\infty < x < \infty$, $-\infty < y < \infty$ *şi cu funcţia caracteristică* $\varphi(t_1, t_2)$.

Să se calculeze $\varphi(t_1 / Y = y)$ *în funcţie de* $\varphi(t_1, t_2)$.

S o l u ţ i e. Putem scrie

$$f(x, y) = f(x/y)\, f_Y(y),$$

unde $f_Y(y)$ este densitatea de repartiţie marginală a variabilei aleatoare Y. Atunci

$$\varphi(t_1/y) = \int_{-\infty}^{\infty} e^{it_1x} f(x/y)\, dx.$$

De aici rezultă că

$$\varphi(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it_1x}\, e^{it_2y} f(x, y)\, dxdy =$$

$$= \int_{-\infty}^{\infty} e^{it_2y} \varphi(t_1/y) f_Y(y)\, dy.$$

Aplicînd teorema de inversiune obţinem

$$\varphi(t_1/y) f_Y(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it_2y}\varphi(t_1, t_2)\, dt_2$$

şi deci

$$\varphi(t_1/y) = \frac{\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it_2y}\varphi(t_1, t_2)\, dt_2}{f_Y(y)}.$$

Cum

$$f_Y(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it_2y}\varphi(0, t_2)\, dt_2$$

rezultă că

$$\varphi(t_1/y) = \frac{\int_{-\infty}^{\infty} e^{-it_2y}\varphi(t_1, t_2)\, dt_2}{\int_{-\infty}^{\infty} e^{-it_2y}\varphi(0, t_2)\, dt_2}.$$

1.53. *Fie* (X_1, X_2, X_3) *un vector aleator normal tridimensional pentru care*

$$M(X_1) = M(X_2) = M(X_3) = 0,$$

$$M(X_1^2) = M(X_2^2) = M(X_3^2) = 1.$$

$$M(X_1X_2) = M(X_1X_3) = M(X_2X_3) = \frac{1}{2}.$$

Să se determine funcția caracteristică și densitatea de repartiție a vectorului (X_1, X_2, X_3).

Soluție. Cum matricea de covarianțe asociată vectorului $X = (X_1, X_2, X_3)$ este

$$\Sigma = \begin{pmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{pmatrix},$$

urmează că

$$\varphi_X(t) = M(e^{it'X}) = e^{it'\mu - \frac{1}{2}t'\Sigma t} = e^{-\frac{1}{2}t'\Sigma t},$$

întrucît

$$\mu' = (M(X_1), \quad M(X_2), \quad M(X_3)) = (\mu_1, \mu_2, \mu_3) = 0.$$

Cum

$$t'\Sigma t = (t_1, t_2, t_3) \begin{pmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} =$$

$$= t_1^2 + t_2^2 + t_3^2 + t_1t_2 + t_1t_3 + t_2t_3,$$

obținem

$$\varphi_X(t) = e^{-\frac{1}{2}(t_1^2+t_2^2+t_3^2+t_1t_2+t_1t_3+t_2t_3)}.$$

Avem

$$n(x: 0, \Sigma) = \frac{1}{(2\pi)^{3/2}\sqrt{\det \Sigma}} \exp\left[-\frac{1}{2}\, x'\Sigma^{-1}x\right]$$

şi cum

$$\det \Sigma = \frac{1}{2},$$

$$\Sigma^{-1} = 2\begin{pmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

$$x'\Sigma^{-1}x = (x_1, x_2, x_3)\begin{pmatrix} \frac{3}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{3}{2} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} =$$

$$= \frac{3}{2}(x_1^2 + x_2^2 + x_3^2) - (x_1x_2 + x_1x_3 + x_2x_3),$$

obţinem

$$n(x;\, 0,\, \Sigma) = \frac{2}{(2\pi)^{3/2}} \exp\left\{-\frac{1}{2}\left[\frac{3}{2}(x_1^2 + x_2^2 + x_3^2) - \right.\right.$$

$$\left.\left. - (x_1x_2 + x_1x_3 + x_2x_3)\right]\right\}$$

I.54. *Fie X o variabilă aleatoare repartizată uniform pe $[-1,1]$ şi Y o variabilă aleatoare avînd densitatea de repartiţie*

$$g(x) = \frac{k \sin^2 x}{x^2}.$$

Să se calculeze funcţia caracteristică corespunzătoare variabilei aleatoare Y.

Soluţie. X fiind repartizată uniform pe segmentul $[-1,1]$ are densitatea de repartiţie

$$f(x) = \begin{cases} \frac{1}{2} & \text{dacă} \quad x \in [-1, 1], \\ 0 & \text{dacă} \quad x \notin [-1, 1] \end{cases}$$

şi funcţia caracteristică

$$\varphi_X^\circ(t) = \frac{1}{2}\int_{-1}^{1} e^{itX}\, dx = \frac{\sin t}{t}.$$

Dacă X_1 şi X_2 sînt două variabile aleatoare repartizate uniform pe $[-1, 1]$, atunci

$$\varphi_{X_1+X_2}(t) = \int_{-\infty}^{\infty} h(x)\, e^{itX}\, dx = \frac{\sin^2 t}{t^2},$$

unde am notat prin $h(x)$ densitatea de repartiţie a variabilei aleatoare $X_1 + X_2$.

$\frac{\sin^2 t}{t^2}$ fiind o funcţie integrabilă, formula de inversiune ne dă

$$h(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itX}\frac{\sin^2 t}{t^2} dt$$

Pe de altă parte, funcţia caracteristică a variabilei aleatoare Y este

$$\varphi_Y(t) = k\int_{-\infty}^{\infty} e^{itX}\frac{\sin^2 x}{x^2}\, dx.$$

Fie

$$\varphi_Y(-t) = 2\pi k h(t).$$

Vom evalua funcţia h. Din compunerea funcţiilor de repartiţie, rezultă

$$h(y) = \int_{-1}^{1} f(y)\, f(y-x) dx = \frac{1}{2}\int_{-1}^{1} f(y-x)\, dx,$$

de unde

$$h(y) = \begin{cases} \frac{2-|y|}{4} & \text{dacă} \quad -2 < y < 2, \\ 0 & \text{dacă} \quad y < -2,\ y > 2. \end{cases}$$

Putem de asemenea determina densitatea de repartiţie a variabilei $X_1 + X_2$ în felul următor:

Probabilitatea ca $X_1 + X_2 < y$, pentru $-2 < y < 0$, este egală cu probabilitatea ca un punct să se găsească în triunghiul dublu haşurat (fig. I.3).

Avem

$$H(y) = \frac{1}{2} \cdot \frac{(2 - |y|)^2}{2^2}$$

de unde

$$h(y) = \frac{2-|y|}{4} = \frac{2+y}{4} \quad \text{dacă} \quad y < 0.$$

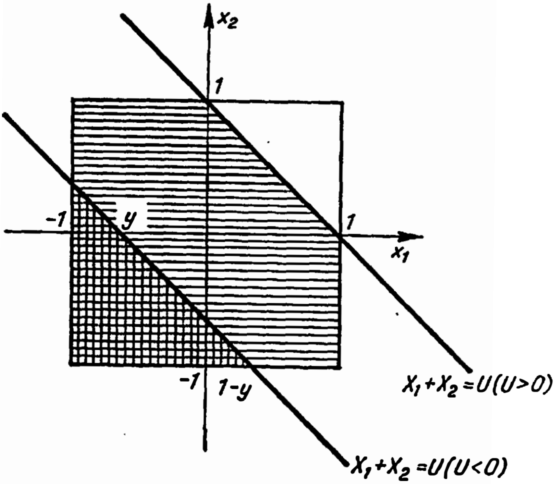

Fig. I.3.

Dacă $0 > y > 2$

$$H(y) = 1 - \frac{(2-y)^2}{2 \times 4} = 1 - \frac{(2-y)^2}{8},$$

$$h(y) = \frac{2-y}{4}.$$

Rezultă că

$$h(y) = \frac{2-|y|}{4} \quad \text{pentru} \quad -2 < y < 2.$$

Avem

$$\varphi_Y(-t) = 2\pi k \frac{2-|t|}{4}.$$

Ţinînd seama că $\varphi_Y(0) = 1$, rezultă că $k = \dfrac{1}{\pi}$.

Deci densitatea variabilei Y este

$$g(x) = \frac{1}{\pi} \frac{\sin^2 x}{x^2},$$

iar funcţia caracteristică

$$\varphi_Y(t) = \frac{2 - |t|}{2}.$$

I.55. *Să se arate că dacă X_1 şi X_2 sînt două variabile aleatoare independente iar $X = X_1 + X_2$ este o variabilă aleatoare Poisson ($\varphi_X(t) = e^{\lambda(e^{it}-1)}$) atunci X_1 şi X_2 sînt variabile aleatoare Poisson.*

Soluţie. Din faptul că X este o variabilă aleatoare Poisson urmează că spectrul ei este mărginit inferior şi deci spectrele variabilelor X_1 şi X_2 sînt mărginite inferior. (Spectrul unei variabile aleatoare Poisson este mulţimea numerelor întregi pozitive sau zero).

Dacă N_1 reprezintă marginea inferioară a spectrului variabilei X_1 şi N_2 cea a spectrului variabilei X_2, atunci $N_1 + N_2 = 0$.

Să presupunem că în spectrul valorilor lui X_1 ar exista şi alte valori decît $N_1, N_1 + 1, N_2 + 2, \ldots, N_1 + n, \ldots$, ($n$ fiind un număr între pozitiv sau nul) adică o valoare $N_1 + n + h$ (h nefiind număr întreg).

Atunci

$$N_1 + n + h + N_2 = n + h$$

ar conduce la faptul că spectrul variabilei X nu ar fi format din numere întregi pozitive sau zero. Acelaşi lucru fiind valabil pentru variabila X_2, rezultă că spectrul variabilei X_1 este conţinut în mulţimea

$$N_1, N_1 + 1, \ldots, N_1 + n, \ldots$$

iar spectrul variabilei X_2 în mulţimea

$$N_2, N_2 + 1, \ldots, N_2 + n, \ldots$$

cu $N_1 + N_2 = 0$.

Scăzînd din variabila aleatoare X_1 cantitatea N_1 şi din X_2 cantitatea N_2, nu modificăm cu nimic pe $X = X_1 + X_2$ căci $N_1 + N_2 = 0$ şi deci X este descompusă într-o sumă de două variabile aleatoare independente fiecare avînd spectrul

$$\{0, 1, 2, \ldots\}$$

Funcţia caracteristică a variabilei $X_j (j = 1, 2)$ va fi de forma

$$\varphi_{X_j}(t) = \sum_{n=0}^{\infty} p_n^{(j)} e^{itn}, \quad j = 1, 2$$

cu

$$p_n^{(j)} = P(X_j = n), \quad j = 1, 2,$$

X_1 şi X_2 fiind independente putem scrie

$$e^{\lambda(e^{it}-1)} = \left(\sum_{n=0}^{\infty} p_n^{(1)} e^{itn}\right)\left(\sum_{n=0}^{\infty} p_n^{(2)} e^{itn}\right)$$

sau cu notaţia $e^{it} = Z$ avem

$$e^{\lambda(Z-1)} = \left(\sum_{n=0}^{\infty} p_n^{(1)} Z^n\right)\left(\sum_{n=0}^{\infty} p_n^{(2)} Z^n\right) = g_1(Z)\, g_2(Z).$$

Întrucît $p_n^{(1)}$ şi $p_n^{(2)}$ sînt probabilităţi, cele două serii scrise în membrul doi sînt convergente pentru $|Z| < 1$ şi deci definesc două funcţii $g_1(Z)$ şi $g_2(Z)$ (funcţiile generatoare ale variabilelor X_1 şi X_2 respectiv).

Dezvoltînd în serie în jurul originii funcţia $e^{\lambda(Z-1)}$ şi egalînd coeficienţii, avem pentru orice n:

$$e^{-\lambda} \frac{\lambda^n}{n!} = p_n^{(1)} p_0^{(2)} + p_{n-1}^{(1)} p_1^{(2)} + \dots + p_0^{(1)} p_n^{(2)} \geqslant p_n^{(1)} p_0^{(2)},$$

($p_n^{(j)}$ fiind probabilităţi) şi prin urmare

$$p_n^{(1)} \leqslant \frac{e^{-\lambda}\lambda^n}{p_0^{(2)} n!} \quad \text{şi} \quad p_n^{(2)} \leqslant \frac{e^{-\lambda}\lambda^n}{p_0^{(1)} n!},$$

$p_0^{(1)}$ şi $p_0^{(2)}$ sînt diferite de zero căci dacă cel puţin una din ele ar fi zero, atunci $g_1(Z)\, g_2(Z)$ ar fi zero pentru $Z = 0$, ceea ce nu este adevărat.

Aceasta ne arată că $p_n^{(1)}$ şi $p_n^{(2)}$ sînt inferiori coeficienţilor lui Z^n din dezvoltarea în serie a funcţiilor $\frac{e^{-\lambda}}{p_0^{(2)}} e^{\lambda Z}$, respectiv $\frac{e^{-\lambda}}{p_0^{(1)}} e^{\lambda Z}$. Deci $g_1(Z)$ şi $g_2(Z)$ sînt funcţii întregi şi cum $e^{\lambda(Z-1)}$ nu are zerouri rezultă că nici una din ele nu are zerouri.

Atunci funcţiile $g_1(Z)$ şi $g_2(Z)$ întregi şi fără zerouri sînt de forma $e^{h_1(Z)}$ respectiv $e^{h_2(Z)}$ cu h_1 şi h_2 întregi.

Pe de altă parte g_1 şi g_2 au toţi coeficienţii din dezvoltarea lor în serie pozitivi ceea ce implică

$$|g_1(Z)| \leqslant p_0^{(1)} + p_0^{(1)} R + \dots + p_n^{(1)} R^n + \dots \leqslant g_1(R)$$

pentru orice Z situat pe cercul de rază R.

Însă $e^{\lambda(R-1)} \geqslant g_1(R)\, p_0^{(2)}$ şi prin urmare pe un cerc de rază R, $g_1(R)$ care este maximul modulului $|g_1(Z)|$ este inferior lui $\frac{e^{-\lambda}}{p_0^{(2)}} e^{\lambda R}$, adică $g_1(Z)$ (deci şi $g_2(Z)$) este de ordinul 1 cel mult. Teorema lui Lionville-Hadamard ne conduce în acest caz la faptul că $g_1(Z) = e^{h_1(Z)}$(h_1 fiind o funcţie întreagă) este egală cu $e^{\mu_1 Z+\nu_1}$. Din faptul că toţi coeficienţii lui $g_1(Z)$ sînt pozitivi, rezultă că $\mu_1 > 0$ iar din $g_1(1) = \varphi_{X_1}(0) = 1$, rezultă că $\nu_1 = -\mu_1$, ceea ce ne conduce la $g_1(Z) = e^{\mu_1(Z-1)}$ sau încă $\varphi_{X_1}(t) = e^{\mu_1(e^{it}-1)}$ şi analog $\varphi_{X_2}(t) = e^{\mu_2(e^{it}-1)}$.

Deci X_1 urmează o repartiţie Poisson cu parametrul μ_1, X_2 urmează o repartiţie Poisson cu parametrul μ_2, cu $\mu_1 + \mu_2 = \lambda$.

I.56. *Să se arate că variabila aleatoare*

$$Y = \chi'^2(n;\ \theta)/(\lambda_1\chi_1^2(n_1) + \lambda_2\chi_2^2(n_2))$$

unde

— $\chi'^2(n;\theta)$ *este o variabilă aleatoare repartizată* χ^2 *necentrat cu parametrul de necentricitate* θ *şi* n *grade de libertate;*

— χ_1^2 *şi* χ_2^2 *sînt variabile aleatoare independente repartizate* χ^2 *cu* n_1 *respectiv* n_2 *grade de libertate;*

— λ_1 şi λ_2 *sînt constante arbitrare, nenule;*

— $n_i = 2g_i$ *sînt întregi pari, are funcţia de repartiţie*

$$F_Y(y) = \sum_{j=1}^{2}\sum_{s=1}^{g_j} \alpha_{js} F^{\theta}_{n,\,2s}(\lambda_j y),$$

unde $F^{\theta}_{n,k}$ *este funcţia de repartiţie a raportului* $\chi'^2(n,\theta)/\chi_k^2$, *iar constantele* α_{js} *sînt date de*

$$\alpha_{1s} = \frac{(-1)^{g_1-s}\,\Gamma(g_1+g_2-s)}{\Gamma(g_2)\,\Gamma(g_1-s+1)} \cdot \frac{\lambda_1^{g_2}\lambda_2^{g_1-s}}{(\lambda_1-\lambda_2)^{g_1+g_2-s}},$$

$$\alpha_{2s} = \frac{(-1)^{g_2-s}\,\Gamma(g_1+g_2-s)}{\Gamma(g_1)\,\Gamma(g_2-s+1)} \cdot \frac{\lambda_1^{g_2-s}\,\lambda_1^{g_1}}{(\lambda_2-\lambda_1)^{g_1+g_2-s}}.$$

I n d i c a ţ i e. Se ţine seama că funcţia caracteristică a numărătorului este

$$(1-2it)^{-\frac{n}{2}} \exp\left(\frac{\theta i t}{1-2it}\right)$$

iar a numitorului

$$(1-2it\lambda_1)^{-n_1}(1-2it\lambda_2)^{-n_2}=\sum_{j=1}^{2}\sum_{s=1}^{g_j}\alpha_{js}(1-2it\lambda_j)^{-s}$$

şi se aplică formula de inversiune pentru repartiţia raportului a două variabile aleatoare. Se obţine

$$F_Y(y)=\frac{1}{2}-\frac{1}{2\pi i}\sum_{j=1}^{2}\sum_{s=1}^{g_j}\alpha_{js}\int_{*}\frac{\left[\exp\left(\frac{\theta it}{1-2it}\right)\right](1-2it)^{-\frac{n}{2}}(1+2it\lambda_j y)^{-s}}{t}\,dt$$

unde

$$\int_{*}=\lim_{\substack{\varepsilon\to 0\\ T\to\infty}}\left(\int_{-T}^{-\varepsilon}+\int_{\varepsilon}^{T}\right).$$

I.57. *Fie X şi Y două variabile aleatoare care au funcţiile generatoare de momente $G_1(t)$ respectiv $G_2(t)$. Atunci dacă X şi Y sînt variabile aleatoare independente, rezultă că funcţia generatoare de momente $G(t)$ a variabilei $Z=X+Y$ verifică relaţia*

$$G(t)=G_1(t)\,G_2(t).$$

Să se arate că din relaţia $G(t)=G_1(t)G_2(t)$ nu rezultă că X şi Y sînt independente.

Soluţie. Fie $f_1(x)$ şi $f_2(y)$ densităţile de repartiţie ale variabilelor X şi Y, continue, simetrice faţă de zero şi cu funcţiile generatoare de momente $G_1(t), G_2(t)$ respectiv. Fie $g(.,.)$ o funcţie cu valori reale de două variabile, integrabilă şi astfel încît $|g(x, y)|\leqslant 1$ pentru orice x şi y.

Fie acum vectorul aleator (X, Y) a cărei densitate de repartiţie este

$$f(x, y)=f_1(x)\,f_2(y)\,(1-g(x, y)).$$

Să alegem acum $f_1=f_2$ şi g astfel încît $g(x, y)=(x-y)\,h(x, y)$ unde $h(x, y)$ este simetrică în x şi y. Atunci

$$\iint f_1(x)\,f_2(y)\,(x-y)\,h(x, y)\,e^{t(x+y)}\,dx\,dy=0,$$

aşa că

$$G(t) = M(e^{it(X+Y)}) = \iint f_1(x) f_2(y) [1 - (x - y)\ h(x, y)]\, e^{it(x+y)}\, dx\, dy =$$

$$\left(\int f_1(x)\ e^{itx}\, dx\right)^2 = G_1^2(t),$$

deşi variabilele X şi Y sînt dependente.

I.58. *Se consideră variabilele aleatoare independente* $X_1, X_2, \ldots, X_n$ *astfel încît*

$$X_j : \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}; \quad j = 1, 2, \ldots, n.$$

a) *Să se calculeze funcţia caracteristică a variabilei aleatoare*

$$Y_n = \sum_{j=1}^{n} a_j X_j,$$

unde $a_1, a_2, \ldots, a_n$ *sînt constante reale.*

b) *Să se arate că dacă* $a_j = \frac{1}{2^j}$, $j = 1, 2, \ldots, n$ *atunci variabila aleatoare* Y_n *are la limită o repartiţie uniformă pe intervalul* $[-1, 1]$.

S o l u ţ i e. a) Din enunţ rezultă că variabila aleatoare $a_j X_j$ are repartiţia:

$$a_j X_j : \begin{pmatrix} -a_j & a_j \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

şi deci funcţia caracteristică:

$$\varphi_j(t) = \frac{e^{it\,aj} + e^{-ita_j}}{2} = \cos a_j t.$$

Cum $X_1, X_2, \ldots, X_n$ sînt independente rezultă că funcţia caracteristică a variabilei aleatoare Y_n este dată de expresia

$$\varphi_{Y_n}(t) = M(e^{itY_n}) = M\left(\prod_{j=1}^{n} e^{ita_jX_j}\right) = \prod_{j=1}^{n} M(e^{ita_jX_j})$$

deci

$$\varphi_{Y_n}(t) = \prod_{j=1}^{n} \varphi_j(t) = \prod_{j=1}^{n} \cos ta_j.$$

b) Funcţia caracteristică ce corespunde unei variabile aleatoare repartizată uniform pe intervalul [—1, 1)] este dată de

$$\varphi(t) = M(e^{itX}) = \frac{1}{2}\int_{-1}^{1} e^{itX}\, dx = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t}.$$

Dacă acum scriem succesiunea de egalităţi

$$\sin t = \frac{1}{2}\sin\frac{t}{2}\cos\frac{t}{2} = \frac{1}{2^2}\sin\frac{t}{2^2}\cos\frac{t}{2^2}\cos\frac{t}{2} = \dots$$

$$\dots = \frac{1}{2^n}\sin\frac{t}{2^n}\cos\frac{t}{2^n}\cos\frac{t}{2^{n-1}}\dots\cos\frac{t}{2^2}\cos\frac{t}{2}$$

atunci

$$\prod_{j=1}^{n}\cos\frac{t}{2^j} = \frac{\sin t}{\frac{1}{2^n}\sin\frac{t}{2^n}} = \frac{\sin t}{\dfrac{t\sin\frac{t}{2^n}}{\frac{t}{2^n}}}.$$

De aici rezultă că

$$\lim_{n\to\infty}\varphi_{Y_n}(t) = \lim_{n\to\infty}\frac{\sin t}{t\dfrac{\sin\frac{t}{2^n}}{\frac{t}{2^n}}} = \frac{\sin t}{t} = \varphi(t),$$

ceea ce dovedeşte complet afirmaţia făcută.

Capitolul II

TEORIA SELECŢIEI

II.1. *Fie $X_1, X_2, \dots, X_n$, n variabile aleatoare independente repartizate:*

a) *Normal de medii $\mu_1, \mu_2, \dots, \mu_n$ şi dispersii $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$.*
b) *Gama cu parametrii $\alpha_1, \alpha_2, \dots, \alpha_n$* şi *$\beta_1, \beta_2, \dots, \beta_n$.*
c) *Poisson cu parametrii $\lambda_1, \lambda_2, \dots, \lambda_n$.*
d) *Binomial cu parametrii $k_1, k_2, \dots, k_n$* şi *$p_1, p_2, \dots, p_n$.*
e) *Binomial negativ cu parametrii $k_1, k_2, \dots, k_n$* şi *$p_1, p_2, \dots, p_n$.*

Să se determine repartiţia variabilei aleatoare $\overline{X}_n = \frac{1}{n}\sum_1^n X_t$.

S o l u ţ i e. Pentru a determina repartiţia variabilei aleatoare $\overline{X}_n$, vom folosi metoda funcţiei generatoare de momente.

Funcţia generatoare de momente pentru media a n variabile aleatoare este

$$Me^{t\overline{X}_n} = M\left[\exp \frac{t}{n}\sum_1^n X_t\right] = M\left[\prod_1^n Me^{\frac{t}{n}X_t}\right].$$

Cum variabilele $X_1, \dots, X_n$ sînt independente

$$M\, e^{t\overline{X}_n} = \prod_1^n Me^{\frac{t}{n}X_t},$$

iar dacă $(X_1, X_2, \dots, X_n)$ constituie o selecţie dintr-o populaţie în care caracteristica X este o variabilă aleatoare, atunci

$$Me^{t\overline{X}_n} = \left(M\, e^{\frac{t}{n}X}\right)^n.$$

a) Cum funcţia generatoare de momente a variabilei X_i este

$$M\, e^{tX_i} = \exp\left(\mu_i t + \frac{\sigma_i^2 t}{2}\right)$$

urmează că

$$M\, e^{t\bar{X}_n} = \exp\left(\frac{1}{n}\sum_1^n \mu_i t + \frac{1}{2n^2}\sum_1^n \sigma_i^2 t^2\right),$$

care este funcţia generatoare a unei variabile aleatoare

$$N\left(\frac{1}{n}\sum_1^n \mu_i, \quad \frac{1}{n^2}\sum_1^n \sigma_i^2\right).$$

Observaţie. Media de selecţie, în acest caz, are densitatea de repartiţie

$$f(\bar{x}_n) = \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left[-(\bar{x}_n - \mu)^2/(2\sigma^2/n)\right], \quad -\infty < \bar{x}_n < \infty$$

b) Cum funcţia generatoare de momente a variabilei X_i de densitate de repartiţie:

$$\frac{1}{\Gamma(\alpha_i)\,\beta_i^{\alpha_i}}\, x^{\alpha_i - 1}\, e^{-x/\beta_i}, \quad 0 \leqslant x < \infty, \quad \alpha_i, \beta_i > 0$$

este

$$M\, e^{tX_i} = (1 - \beta_i t)^{-\alpha_i}$$

urmează că

$$M\, e^{t\bar{X}_n} = \prod_1^n \left(1 - \frac{\beta_i t}{n}\right)^{-\alpha_i}.$$

Observaţie. Media de selecţie are densitatea de repartiţie

$$f(\bar{x}_n) = \frac{1}{\Gamma(n\alpha)(\beta/n)^{n\alpha}}\, \bar{x}_n^{n\alpha - 1}\, e^{-n\bar{x}_n/\beta}, \quad 0 < \bar{x}_n < \infty,$$

deoarece în acest caz

$$\alpha_1 = \alpha_2 = \ldots = \alpha_n = \alpha, \quad \beta_1 = \beta_2 = \ldots = \beta_n = \beta.$$

Aceasta arată că media de selecţie este de asemenea o variabilă aleatoare repartizată gama.

c) Cum funcţia generatoare de momente a variabilei X_i este

$$M\, e^{tX_i} = e^{-\lambda_i(1 - e^t)}$$

urmează că

$$M\, e^{t\bar{X}_n} = e^{-\sum\limits_1^n \lambda_i\left(1 - e^{\frac{t}{n}}\right)}.$$

Funcţia generatoare de momente a variabilei $Y = n\bar{X}_n$ se obţine înlocuind pe t prin nt în funcţia generatoare a variabilei $\bar{X}_n$ obţinem:

$$M\,e^{tY} = M\,e^{tn\bar{X}_n} = M\,e^{(tn)\bar{X}_n} = e^{-\sum\limits_1^n \lambda_i((1-e^t)},$$

care nu este altceva decît funcţia generatoare de momente a unei repartiţii Poisson de parametrii $\sum\limits_1^n \lambda_i = \lambda$.

Urmează că funcţia de frecvenţă a variabilei y este

$$P(y) = \frac{e^{-\Sigma\lambda_i}(\Sigma\lambda_i)^y}{y!}; \quad y = 0, 1, 2, \ldots$$

Făcînd transformarea $\bar{X}_n = Y/n$, obţinem

$$P(\bar{x}_n) = \frac{\exp\left(-\sum\limits_1^n \lambda_i\right)\left(\sum\limits_1^n \lambda_i\right)^{n\bar{x}_n}}{(n\bar{x}_n)!}; \quad \bar{x}_n = 0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots$$

Urmează că o sumă de variabile aleatoare independente repartizată Poisson este o variabilă repartizată Poisson însă media unor astfel de variabile aleatoare nu este o variabilă Poisson.

Observaţie. Media de selecţie are funcţia de frecvenţă

$$P(\bar{x}_n) = \frac{e^{-n\lambda}(n\lambda)^{n\bar{x}_n}}{(n\bar{x}_n)!}.$$

d) Funcţia generatoare de momente a variabilei X_i este

$$M\,e^{tX_i} = [p_i\,e^t + (1 - p_i)]^{k_i},$$

ceea ce permite să scriem

$$M\,e^{t\bar{X}_n} = \prod_1^n [p_i\,e^{\frac{t}{n}} + (1 - p_i)]^{k_i}.$$

În cazul în care $p_1 = p_2 = \ldots = p_n = p$, atunci

$$M\,e^{t\bar{X}_n} = [p\,e^{\frac{t}{n}} + (1 - p)]^{\sum\limits_1^n k_i},$$

care nu este funcţia generatoare corespunzătoare unei variabile binomiale.

Variabila aleatoare $Y = n\bar{X}_n = \sum\limits_1^n X_i$ are funcţia generatoare de momente

$$M\,e^{tY} = M\,e^{(tn)\bar{X}_n} = [p\,e^t + (1 - p)]^{\sum\limits_1^n k_i},$$

ceea ce arată că Y urmează o repartiție binomială de parametrii $\sum\limits_{1}^{n} k_i$ și p; deci

$$P(y) = C^{y}_{\sum\limits_{1}^{n} k_i} p^{y}(1-p)^{\sum\limits_{1}^{n} k_i - y}; \quad y = 0, 1, 2, \ldots, \sum_{1}^{n} k_i.$$

Făcînd transformarea $\bar{X}_n = Y/n$ obținem

$$P(\bar{x}_n) = C^{n\bar{x}_n}_{\sum\limits_{1}^{n} k_i} p^{n\bar{x}_n} (1-p)^{\sum\limits_{1}^{n} k_i - n\bar{x}_n}; \quad \bar{x}_n = 0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{1}{n}\sum_{1}^{n} k_i$$

care nu este funcția de frecvență corespunzătoare unei variabile repartizate Poisson.

Observație. Media de selecție are funcția de frecvență

$$P(\bar{x}_n) = C^{n\bar{x}_n}_{nk} p^{n\bar{x}_n}(1-p)^{nk-n\bar{x}_n}, \quad \bar{x}_n = 0, \frac{1}{n}, \frac{2}{n}, \ldots, k$$

e) Funcția generatoare de momente corespunzătoare variabilei X_i de funcție de frecvență

$$P(x) = C^{k_i-1}_{x+k_i-1} p_i^{k_i}(1-p_i)^{x}; \qquad x = 0, 1, 2, \ldots$$

este

$$M\, e^{tX_i} = p_i^{k_i}[1-(1-p_i)\, e^{t}]^{-k_i},$$

ceea ce permite să scriem

$$M\, e^{t\bar{X}_n} = \prod_{1}^{n} p_i^{k_i}[1-(1-p_i)\, e^{\frac{t}{n}}]^{-k_i}.$$

În cazul în care $p_1 = p_2 = \ldots = p_n = p$, avem

$$M\, e^{t\bar{X}_n} = p^{\sum\limits_{1}^{n} k_i}[1-(1-p)\, e^{\frac{t}{n}}]^{-\sum\limits_{1}^{n} k_i}.$$

Dacă punem $Y = n\bar{X}_n$, obținem

$$M\, e^{tY} = M\, e^{(tn)\bar{X}_n} = p^{\sum\limits_{1}^{n} k_i}[1-(1-p)\, e^{t}]^{-\sum\limits_{1}^{n} k_i},$$

expresie ce reprezintă funcţia generatoare de momente a unei variabile repartizate binomial negativ de parametrii $\sum_1^n k_i$ şi p; deci

$$P(y) = C^{\sum_1^n k_i - 1}_{y + \sum_1^n k_i - 1} p^{\sum_1^n k_i} (1-p)^y; \quad y = 0, 1, 2, \ldots$$

Folosind transformarea $\bar{X}_n = Y/n$ obţinem

$$P(\bar{x}_n) = C^{\sum_1^n k_i - 1}_{n\bar{x}_n + \sum_1^n k_i - 1} p^{\sum_1^n k_i} (1-p)^{n\bar{x}_n}, \quad \bar{x}_n = 0, \frac{1}{n}, \frac{2}{n}, \ldots$$

Observaţie. Media de selecţie are funcţia de frecvenţă

$$P(\bar{x}_n) = C^{nk-1}_{n\bar{x}_n + nk - 1} p^{nk} (1-p)^{n\bar{x}_n}, \quad \bar{x}_n = 0, \frac{1}{n}, \frac{2}{n}, \ldots$$

II.2. *Dintr-o populaţie normală cu media 12 probabilitatea cu media de selecţie corespunzătoare unei selecţii de volum 16, să depăşască 14 este 0,24. Care este probabilitatea ca o singură observaţie din această populaţie să aibă o valoare mai mare decît 14?*

S o l u ţ i e. Media de selecţie $\bar{x}_n$ este repartizată de asemenea normal de medie μ şi dispersie σ^2/n. Avem

$$P(\bar{x}_n > 14) = 0{,}24 \quad \text{sau} \quad 1 - P(\bar{x}_n < 14) = 0{,}24.$$

Urmează că

$$0{,}24 = 1 - P\left(\frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} < \frac{14 - 12}{\sigma/4}\right)$$

de unde

$$0{,}24 = 1 - \Phi\left(\frac{8}{\sigma}\right); \qquad \Phi\left(\frac{8}{\sigma}\right) = 0{,}76.$$

Din tabele găsim

$$\frac{8}{\sigma} = 0{,}71 \quad \text{sau} \quad \sigma = 11{,}26.$$

Deci

$$P(x > 14) = 1 - P(x < 14) = 1 - P\left(U < \frac{14-12}{11,26}\right) =$$

$$= 1 - \Phi(0,1776).$$

Din tabele găsim Φ (0,1776) = 0,9616, de unde P ($x > 14$) = 0,0384.

II.3. *Dacă presupunem că greutatea unor oameni este repartizată normal cu abaterea medie pătratică 2, care trebuie să fie volumul selecției, pentru a fi siguri (probabilitatea 0,99) că media de selecție nu diferă de media populației în valoare absolută cu mai mult de o unitate?*

Soluție. Prin ipoteză

$$P(|\bar{X}_n - \mu| < 1) = 0,99.$$

Putem scrie

$$P(|\bar{X}_n - \mu| < 1) = P(-1 < \bar{X}_n - \mu < 1) = P\left(-\frac{\sqrt{n}}{2} < U < \frac{\sqrt{n}}{2}\right) =$$

$$= 2\Phi\left(\frac{\sqrt{n}}{2}\right) - 1,$$

și cum

$$2\Phi\left(\frac{\sqrt{n}}{2}\right) - 1 = 0,99,$$

$$\Phi\left(\frac{\sqrt{n}}{2}\right) = 0,995.$$

Din tabele găsim că $\frac{\sqrt{n}}{2} = 2,58$ de unde $n \simeq 26$.

II.4. *Se consideră două populații caracterizate de variabile aleatoare independente avînd repartiții normale de aceeași medie și abateri medii pătratice* 5,10 *respectiv* 8,30. *Presupunînd că din fiecare populație se extrage cîte o selecție de volum* 81, *să se găsească probabilitatea ca diferența dintre mediile celor două selecții, în valoare absolută, să depășească* 0,80.

Soluție. $\bar{X}_{n_1} - \bar{X}_{n_2}$ este o variabilă aleatoare repartizată

$$N\left(0, \frac{(5,1)^2 + (8,3)^2)}{81}\right) \quad \text{adică} \quad N(0; 1,16).$$

Urmează că

$$P(|\overline{X}_{n_1} - \overline{X}_{n_2}| > 0{,}80) = 1 - P(-0{,}80 < \overline{X}_{n_1} - \overline{X}_{n_2} < 0{,}80) =$$

$$= 1 - \left[\Phi\left(\frac{0{,}80}{1{,}16}\right) - \Phi\left(-\frac{0{,}80}{1{,}16}\right)\right] = 2\left[\Phi\left(\frac{0{,}80}{1{,}16}\right) - 1\right]$$

şi din tabelele $\Phi\left(\frac{0{,}80}{1{,}16}\right) = 0{,}7549$, ceea ce ne permite să scriem

$$P(|\overline{X}_{n_1} - \overline{X}_{n_2}| > 0{,}80) = 0{,}4902.$$

II.5. *Dintr-o populaţie normală de valoare medie* 80 *şi dispersie* 40 *se extrage o selecţie de volum 400 şi dintr-o populaţie normală de valoare medie* 76 *şi dispersie* 180 *se extrage o selecţie de volum* 200. *Să se găsească probabilitatea ca:*

a) *Media aritmetică a primei selecţii să fie mai mare decît media aritmetică a celeilalte selecţii cu cel puţin 5 unităţi.*

b) *Diferenţa mediilor celor două selecţii în valoare absolută să fie mai mare decît 6.*

S o l u ţ i e. a) Trebuie să calculăm $P(\overline{X}_{n_1} > \overline{X}_{n_2} + 5)$ sau acelaşi lucru $P(\overline{X}_{n_1} - \overline{X}_{n_2} > 5)$,

$\overline{X}_{n_1} - \overline{X}_{n_2}$ este repartizată $N\left(80 - 76, \frac{40}{400} + \frac{80}{200}\right)$ sau N (4,1). Deci

$$P(\overline{X}_{n_1} - \overline{X}_{n_2} > 5) = 1 - P(\overline{X}_{n_1} - \overline{X}_{n_2} < 5) =$$

$$= 1 - \Phi\left(\frac{5-4}{1}\right) = 1 - \Phi(1) = 0{,}1587.$$

b) Putem scrie

$$P(|\overline{X}_{n_1} - \overline{X}_{n_2}| > 6) = 1 - P(-6 < \overline{X}_{n_1} - \overline{X}_{n_2} < 6)$$

şi cum

$$P(-6 < \overline{X}_{n_1} - \overline{X}_{n_2} < 6) = \Phi\left(\frac{6-4}{1}\right) - \Phi\left(\frac{-6-4}{-1}\right) = 0{,}9772,$$

obţinem

$$P(|\overline{X}_{n_1} - \overline{X}_{n_2}| > 6) = 0{,}0228.$$

II. 6. *Se consideră o populaţie în care caracteristica subcercetare X are densitatea de repartiţie:*

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{dacă } x > 0, \\ 0 & \text{dacă } x \leqslant 0. \end{cases}$$

Pe baza unei selecţii de volum n să se afle funcţia de repartiţie a variabilei

$$Y = \frac{2n\bar{x}_n}{\theta}$$

unde

$$\bar{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j.$$

S o l u ţ i e. Prin definiţie avem

$$G(y) = P(Y < y) = P\left(\frac{2n\bar{x}_n}{\theta} < y\right) = P\left(\bar{x}_n < \frac{\theta y}{2n}\right).$$

Pentru a afla funcţia de repartiţie a variabilei $\bar{x}_n$, vom determina mai întîi funcţia caracteristică

$$\varphi_{\bar{x}_n}(u) = M(e^{i\,\bar{x}_n u}) = M\left(e^{i\frac{u}{n}\sum\limits_{j=1}^{n} x_j}\right) = \prod_{j=1}^{n} M(e^{i\frac{u}{n}x_j}) =$$

$$= \prod_{j=1}^{n} \varphi_j\left(\frac{u}{n}\right) = \left(\varphi\left(\frac{u}{n}\right)\right)^n.$$

Cum

$$\varphi\left(\frac{u}{n}\right) = \left(1 - \frac{i\theta u}{n}\right)^{-1},$$

urmează că

$$\varphi_{\bar{x}_n}(u) = \left(1 - \frac{i\theta u}{n}\right)^{-n},$$

funcţie caracteristică care corespunde densităţii:

$$g(x;\theta) = \begin{cases} \dfrac{\left(\frac{n}{\theta}\right)^n x^{n-1} e^{-\frac{nx}{\theta}}}{\Gamma(n)}, & x > 0. \\ 0 & x \leqslant 0. \end{cases}$$

Aşadar

$$G(y) = \frac{n^n}{\theta^n\Gamma(n)}\int_0^{\frac{\theta y}{2n}} x^{n-1}\,\mathrm{e}^{-\frac{nx}{\theta}}\,\mathrm{d}x,$$

sau punînd $(2nx/\theta) = z$

$$G(y) = \frac{1}{2^n\Gamma(n)}\int_0^y z^{n-1}\mathrm{e}^{-\frac{z}{2}}\,\mathrm{d}z.$$

Observăm că $G(y)$ este funcţia de repartiţie a unei variabile aleatoare $\chi^2(2n)$ şi cu parametrul 1 şi prin urmare $\frac{2n\bar{x}_n}{\theta}$ reprezintă o variabilă $\chi^2(2n)$ şi parametrul 1.

II. 7. *Se consideră variabilele aleatoare independentele* $X_1, X_2, \dots, X_n$ *cu repartiţiile caracterizate de funcţiile de frecvenţă*

$$P(X_j = x_j) = \frac{\alpha_j\theta_j^{x_j}}{x_j}, \quad j = 1, 2, \dots, n$$

$$x_j = 1, 2, 3, \dots; \quad 0 < \theta_j < 1, \quad \alpha_j^{-1} = -\lg(1-\theta_j).$$

Să se determine repartiţia sumei

$$Z = \sum_{j=1}^n X_j.$$

Soluţie. Calculăm mai întîi funcţia caracteristică a variabilei aleatoare X_j:

$$\varphi_{X_j}(t) = M(\mathrm{e}^{\mathrm{i}tX_j}) = \sum_{k=1}^{\infty}\mathrm{e}^{\mathrm{i}tk}\frac{\alpha_j\theta_j^k}{k} = \alpha_j\sum_{k=1}^{\infty}\frac{(\theta_j\mathrm{e}^{\mathrm{i}\,t})^k}{k} =$$

$$= \alpha_j(-\lg(1-\theta_j\,\mathrm{e}^{\mathrm{i}t})).$$

Cum $X_1, X_2, \dots, X_n$ sînt independente, rezultă

$$\varphi_z(t) = \prod_{j=1}^n \alpha_j[-\lg(1-\theta_j\,\mathrm{e}^{\mathrm{i}t})].$$

Aplicînd formula de inversiune putem scrie

$$p(z;\theta_1,\dots,\theta_n) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\varphi_z(t)\,\mathrm{e}^{-\mathrm{i}\,zt}\,\mathrm{d}t = \frac{B(z;\theta_1,\dots,\theta_n)}{z!\,g(\theta_1,\dots,\theta_n)}$$

pentru $z = n,\ n+1,\ \ldots;\ n = 1, 2, 3, \ldots$

şi unde

$$g(\theta_1, \ldots, \theta_n) = \prod_{j=1}^{n} [-\lg(1-\theta_j)],$$

$$B(z; \theta_1, \ldots, \theta_n) = \sum_{\substack{r_j > 0 \\ r_1 + \ldots + r_n = z}} \frac{z!\theta_1^{r_1}, \ldots, \theta_n^{r_n}}{r_1, \ldots, r_n}$$

II. 8. *Fie* $(x_1, x_2, \ldots, x_n)$ *o selecţie de volum n dintr-o populaţie caracterizată de variabila aleatoare*

$$X: \begin{pmatrix} 0 & 1 \\ p & 1-p \end{pmatrix}.$$

Să se afle repartiţia variabilei aleatoare

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

$$\left(\text{unde } \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i\right).$$

S o l u ţ i e. În condiţiile problemei putem scrie

$$P(X = x) = p^x(1-p)^{1-x}, \quad x = 0,1; \quad q = 1 - p.$$

Dacă

$$Y = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i^2,$$

atunci

$$P(Y = y) = C_n^y p^y q^{n-y}; \quad y = 0, 1, \ldots, n.$$

Cum

$$(n-1)\, s^2 = \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 / n = Y - \frac{Y^2}{n} = Y(n-Y)\frac{1}{n},$$

rezultă că s^2 poate lua numai valorile

$$c = i(n-i)\frac{1}{n(n-1)}; \qquad i = 0, 1, ..., [n/2],$$

unde $[n/2]$ este cel mai mare întreg, mai mic sau egal cu $\frac{n}{2}$.

De aici urmează că

$$P(s^2 = c) = P(nY - Y^2 = i(n-i)) = P\left\{(Y - n/2)^2 = \right.$$

$$\left. = \left(i - \frac{n}{2}\right)^2\right\} = P\{(Y = i) \cup (Y = n-i)\} =$$

$$= C_n^i p^i q^i (p^{n-2i} + q^{n-2i}),$$

pentru $i < n/2$.

Dacă n este par şi $i = n/2$,

$$P\left(s^2 = \frac{n}{4(n-1)}\right) = C_n^{\frac{n}{2}} p^{\frac{n}{2}} q^{\frac{n}{2}}.$$

II. 9. *Să se arate că dacă $x_1, x_2, ..., x_n$ este o selecţie dintr-o populaţie în care caractersitica sub cercetare este o variabilă aleatoare X repartizată N $(0, \sigma^2)$ atunci media de selecţie este independentă de dispersia de selecţie.*

Soluţie. Pentru $n = 2$, avem

$$(x_1 - \bar{x}_2)^2 + (x_2 - \bar{x}_2)^2 = \left(\frac{x_1 - x_2}{\sqrt{2}}\right)^2,$$

ceea ce arată că suma pătratelor abaterilor de la media aritmetică poate fi exprimată ca pătratul combinaţiei liniare

$$\frac{x_1}{\sqrt{2}} - \frac{x_2}{\sqrt{2}}.$$

Cum

$$\sum a_i = \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} = 0, \quad \sum a_i^2 = \frac{1}{2} + \frac{1}{2} = 1,$$

urmează că această combinaţie liniară are media 0 şi dispersia σ^2.

Similar, pentru $n = 3$

$$(x_1 - \bar{x}_3)^2 + (x_2 - \bar{x}_3)^2 + (x_3 - \bar{x}_3)^2 = \left(\frac{x_1 - x_2}{\sqrt{2}}\right)^2 + \left(\frac{x_1 + x_2 - 2x_3}{\sqrt{6}}\right)^2.$$

Pentru al doilea termen al membrului drept, avem

$$\sum a_i = \frac{1}{\sqrt{6}} + \frac{1}{\sqrt{6}} - \frac{2}{\sqrt{6}} = 0, \qquad \sum a_i^2 = 1,$$

ceea ce arată că combinația liniară

$$\frac{x_1}{\sqrt{6}} + \frac{x_2}{\sqrt{6}} - \frac{2x_3}{\sqrt{6}}$$

are media 0 și dispersia σ^2.

Folosind faptul că date variabilele $\sum a_i x_i$ și $\sum b_j x_j$, $\sum a_i b_i = 0$ este o condiție suficientă ca variabilele să fie independente, găsim

$$\sum a_i b_i = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{6}} - \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{6}} + 0 \cdot \frac{2}{\sqrt{6}} = 0,$$

ceea ce arată că

$$\frac{x_1 - x_2}{\sqrt{2}} \quad \text{și} \quad \frac{x_1 - x_2 - 2x_3}{\sqrt{6}}$$

sînt combinații independente.

În cazul unei selecții de volum n

$$\sum_{i=1}^{n} (x_i - \bar{x}_n)^2 = \sum_{i=1}^{n-1} z_i^2,$$

unde

$$\left\{\begin{aligned}
z_1 &= \frac{1}{\sqrt{2}} x_1 - \frac{1}{\sqrt{2}} x_2, \\
z_2 &= \frac{1}{\sqrt{6}} x_1 + \frac{1}{\sqrt{6}} x_2 - \frac{2}{\sqrt{6}} x_3, \\
&\dots\dots\dots\dots\dots\dots\dots\dots \\
z_r &= \frac{1}{\sqrt{r(r+1)}} x_1 + \frac{1}{\sqrt{r(r+1)}} x_2 + \dots + \\
&\qquad + \frac{1}{\sqrt{r(r+1)}} x_r - \frac{r}{\sqrt{r(r+1)}} x_{r+1}, \\
&\dots\dots\dots\dots\dots\dots\dots\dots \\
z_{n-1} &= \frac{1}{\sqrt{n(n-1)}} x_1 + \frac{1}{\sqrt{n(n-1)}} x_2 + \dots + \\
&\qquad + \frac{1}{\sqrt{n(n-1)}} x_{n-1} - \frac{n-1}{\sqrt{n(n-1)}} x_n.
\end{aligned}\right. \tag{II.1}$$

Variabilele $z_1, z_2, .., z_{n-1}$, sînt independente, toate avînd media 0 şi dispersia σ^2. Deci $\frac{z_i}{\sigma}$ este repartizată $N(0,1)$, şi

$$\frac{1}{\sigma^2}\sum_{i=1}^{n-1} z_i^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2}$$

este repartizată $\chi^2(n-1)$.

Fiecare combinaţie liniară din (II. 1) este independentă de $\bar{x} = \frac{\Sigma x_i}{n}$. Urmează că $\bar{x}$ este independentă de toţi z_i şi deci de dispersia de selecţie.

II. 10. *Să se arate că media $\bar{x}$ şi dispersia s^2 a unei selecţii dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X avînd o repartiţie simetrică sînt necorelate.*

II. 11. *Fie $x_1, ..., x_n$ o selecţie dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X avînd densitatea de repartiţie*

$$f^*(x) = \begin{cases} \frac{k}{\theta} e^{-x/\theta} & \textit{pentru} \quad 0 \leqslant x \leqslant x_0, \\ 0 & \textit{în rest,} \end{cases}$$

unde $k^{-1} = (1 - e^{-x_0/\theta})$.

Să se determine densitatea de repartiţie a mediei de selecţie $\bar{x}_n$.

Soluţie. Funcţia caracteristică a variabilei aleatoare X avînd densitatea $f^*(x)$ este dată de

$$\varphi^*(t; x) = k\left\{1 - \exp\left[-\frac{x_0}{\theta}(1 - i\theta t)\right]\right\} \Big/ (1 - i\,\theta t).$$

Prin urmare, funcţia caracteristică a statisticii $y_n = \sum_1^n x_i$ este.

$$\varphi^*(t; y_n) = \sum_{r=0}^{n} b_r e^{it + x_0}(1 - i\,\theta t)^{-n},$$

unde

$$b_r = (-1)^r C_n^r k^n e^{-rx_0/\theta}.$$

Din formula de inversiune găsim funcţia de repartiţie a statisticii y_n.

$$S^*(y_n) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1 - e^{-i t y_n}}{i t} \varphi^*(t, y_n)\, dt =$$

$$= \sum_{r=0}^{n} b_r \int_{-\infty}^{+\infty} \frac{1 - e^{-i\tau z_n}}{i\tau} e^{i\tau(rx_0/\theta)}(1 - i\tau)^{-n}\, d\tau \qquad \text{(II.2)}$$

unde

$$\tau = t\theta \quad \text{şi} \quad z_n = y_n/\theta.$$

Funcţia caracteristică a repartiţiei Gama, avînd funcţia de repartiţie

$$G_n(x) = \int_0^x \frac{1}{\Gamma(n)} e^{-y} y^{(n-1)}\, dy$$

este dată de

$$\psi_n(\tau) = (1 - i r)^{-n}.$$

Integrantul termenului de ordin r din (II. 2) este

$$[(1 - e^{-i\tau z_n})/i\tau]\, \psi_n(\tau)\, e^{i\tau(rx_0/\theta)} = [(1 - e^{-i\tau z_n})/i\tau]\, \lambda_r(\tau),$$

unde

$$\lambda_r(\tau) = e^{i\tau(rx_0/\theta)}\psi_n(\tau)$$

este funcţia caracteristică corespunzătoare funcţiei de repartiţie

$$G_n(z_n - rx_0/\theta).$$

Deci funcţia de repartiţie a statisticii Y_n este dată de

$$S^*(y_n) = \sum_{r=0}^{n} b_r G_n\left(\frac{y_n - rx_0}{\theta}\right), \qquad \text{(II.3)}$$

unde $G_n(y_n - rx_0) = 0$ pentru $y_n < rx_0$.

Din (II. 3) obţinem funcţia de repartiţie $F_n^*(\bar{x}_n/\theta)$ a statisticii $\bar{x}_n$.

Observaţie. Cînd X are densitatea de repartiţie

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & \text{pentru } 0 \leqslant x < \infty, \\ 0 & \text{în rest,} \end{cases}$$

$n\bar{x}_n/\theta$ urmează o repartiţie gama cu parametru n şi

$$dF_n^*(\bar{x}_n \mid \theta) = f_n(\bar{x}_n \mid \theta)\, d\bar{x}_n = \frac{1}{\Gamma(n)} e^{-n\bar{x}_n/\theta} (n\bar{x}_n/\theta)^{n-1}\, dF(n\bar{x}_n/\theta).$$

II. 12. *Să presupunem că arderea unui anumit tip de bec de luminat este astfel încît probabilitatea ca un bec de acest tip luat la întîmplare să se ardă în intervalul de timp* $(t, t + \mathrm{d}t)$ este dată de

$$f(t)\,\mathrm{d}t = c\,\mathrm{e}^{-ct}\,\mathrm{d}t, \quad t > 0,$$

unde c *este o constantă pozitivă. Fie* $t_1, t_2, \ldots, t_n$ *duratele de funcționare corespunzătoare a* n *becuri luate la întîmplare din aceste tipuri de becuri și* $t_{(1)}, t_{(2)}, \ldots, t_{(n)}$, *statisticile de ordine corespunzătoare. Să se determine probabilitatea ca becul cu cea mai mare durată de funcționare* $(t_{(n)})$ *să se ardă în intervalul de timp* $(u, u + \mathrm{d}u)$.

S o l u ț i e. Funcția de repartiție a duratei de funcționare, este

$$F(t) = \int_{-\infty}^{t} f(t)\,\mathrm{d}t = \int_0^t c\,\mathrm{e}^{-ct}\,\mathrm{d}t = 1 - \mathrm{e}^{-ct}$$

și cum

$$G(u) = P(t_{(n)} < u) = F^n(u) = (1 - \mathrm{e}^{-cu})^n,$$

urmează că

$$g(u) = \frac{\mathrm{d}G(u)}{\mathrm{d}u} = nc(1 - \mathrm{e}^{-cu})^{n-1}\,\mathrm{e}^{-cu},$$

ceea ce arată că probabilitatea ca becul cu cea mai mare durată de funcționare să se ardă în intervalul $(u, u + \mathrm{d}u)$ este

$$g(u)\,\mathrm{d}u = nc(1 - \mathrm{e}^{-cu})^{n-1}\,\mathrm{e}^{-cu}\,\mathrm{d}u.$$

II. 13. *Presupunem că verigile de un anumit tip folosite pentru alcătuirea unor lanțuri este astfel încît rezistența la rupere* X *a verigilor are densitatea de repartiție*

$$f(x) = \begin{cases} \dfrac{(m+1)(m+2)}{c^{m+2}}\, x^m(c - x), & 0 < x < c, \\ 0 & \text{în rest}, \end{cases} \tag{II.4}$$

unde c *și* m *sînt constante pozitive.*

Dacă un lanț este alcătuit din n *verigi de acest tip luate la întîmplare, care este densitatea de repartiție a rezistenței la rupere a acestui lanț?*

S o l u ț i e. Deoarece rezistența de rupere a lanțului este egală cu rezistența la rupere a celei mai slabe verigi, problema pusă se reduce la găsirea probabilității elementare a celui mai mic element $x_{(1)}$ al selecției de volum n din populația în care rezistența la rupere are densitatea (II.4).

Funcţia de repartiţie a rezistenţei la rupere este

$$F(x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}x = \frac{(m+1)(m+2)}{c^{m+2}} \int_0^x x^m(c-x)\,\mathrm{d}x =$$

$$= (m+2)\left(\frac{x}{c}\right)^{m+1} - (m+1)\left(\frac{x}{c}\right)^{m+2}$$

şi cum

$$G(v) = P(t_{(1)} < v) = 1 - [1 - F(v)]^n,$$

obţinem

$$g(v) = \frac{\mathrm{d}G(v)}{\mathrm{d}v} = \frac{n(m+1)(m+2)\,v^m}{c^{m+2}} \times$$

$$\times \left[1 - (m+2)\left(\frac{v}{c}\right)^{m+1} + (m+1)\left(\frac{v}{c}\right)^{m+2}\right]^{n-1} (c-v), \quad 0 < v < c.$$

II. 14. *Fie* $x_1, x_2, \ldots, x_n$ *o selecţie de volum* n *dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare* X *repartizată uniform pe* $[0, 1]$ *şi* $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$ *statisticile de ordine corespunzătoare:*

a) *Să se găsească* $M(x_{(1)})$ *şi* $D^2(x_{(1)})$.

b) *Să se găsească repartiţia limită a vectorului* (U, V) *unde*

$$U = (n+1)\,x_{(1)} \quad \text{şi} \quad V = (n+1)(1 - x_{(n)}).$$

Soluţie. *a*) Avem

$$P(x_{(1)} < x) = 1 - P(x_{(1)} \geqslant x)$$

şi cum

$$P(x_{(1)} \geqslant x) = P(\min_i x_i \geqslant x) = \prod_1^n P(x_i \geqslant x) = \prod_1^n [1 - P(x_i < x)],$$

$$P(x_i < x) = \int_0^x \mathrm{d}x = x,$$

rezultă că

$$P(x_{(1)} < x) = 1 - (1-x)^n.$$

Prin urmare densitatea de repartiţie a variabilei aleatoare $x_{(1)}$ este $n(1-x)^{n-1}$.

Deci

$$M(x_{(1)}) = n\int_0^1 x(1-x)^{n-1}\,\mathrm{d}x = \frac{1}{n+1},$$

$$M(x_{(1)}^2) = n\int_0^1 x^2(1-x)^{n-1}\,\mathrm{d}x = \frac{2}{(n+1)^2(n+2)},$$

şi

$$D^2(x_{(1)}) = M(x_{(1)}^2) - M^2(x_{(1)}) = \frac{n}{(n+1)^2(n+2)}.$$

b) Probabilitatea elementară a perechii $(x_{(1)}, x_{(n)})$ este

$$n(n-1)\,(y-x)^{n-2}\,\mathrm{d}x\mathrm{d}y.$$

Ţinînd seama că

$$x = \frac{u}{n+1}, \qquad y = 1 - \frac{v}{n+1},$$

rezultă probabilitatea elementară a vectorului (U, V):

$$\frac{n}{n-1}\left(1 - \frac{u+v}{n+1}\right)^{n-2}\mathrm{d}u\frac{n-1}{n+1}\,\mathrm{d}v.$$

Repartiţia limită a vectorului (U, V) (pentru $n \to \infty$) este caracterizată prin probabilitatea elementară

$$\mathrm{e}^{-(u+v)}\,\mathrm{d}u\,\mathrm{d}v = \mathrm{e}^{-u}\,\mathrm{d}u\,\mathrm{e}^{-v}\,\mathrm{d}v,$$

ceea ce arată că variabilele aleatoare U şi V sînt asimptotic independente.

II. 15. *Fie $x_1, x_2, \ldots, x_n$ o selecţie de volum n dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X avînd densitatea de repartiţie* $\mathrm{e}^{-x}(0 < x < \infty)$ şi $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$, *statisticile de ordine corespunzătoare.*

a) *Să se determine probabilitatea elementară a statisticilor $x_{(1)}$ şi $x_{(n)}$.*

b) *Statisticile de ordine determină pe axa ox n segmente consecutive. Fie $u_1, u_2, \ldots, u_n$ lungimile acestor segmente. Să se determine probabilitatea elementară a vectorului $(u_1, u_2, \ldots, u_n)$.*

Soluţie. *a)* Avem

$$P(x_{(1)} \geqslant x) = \prod_1^n P(x_i \geqslant x) = \prod_1^n [1 - P(x_i < x)]$$

şi cum

$$P(x_i < x) = \int_0^x e^{-t}\,dt = 1 - e^{-x}, \tag{II.5}$$

obţinem

$$P(x_{(1)} < x) = 1 - [1 - (1 - e^{-x})]^n.$$

Deci densitatea de repartiţie a statisticii $x_{(1)}$ este ne^{-nx}. Ţinînd seama de (II.5) obţinem

$$P(x_{(n)} < x) = P(\max_i x_i < x) = \prod_1^n P(x_i < x) = (1 - e^{-x})^n,$$

de unde rezultă că densitatea de repartiţie a statisticii $x_{(n)}$ este

$$n(1 - e^{-x})^{n-1}\,e^{-x}.$$

b) Probabilitatea elementară a vectorului aleator $(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ este

$$n!\,e^{-x_{(1)}}\,e^{-x_{(2)}}\ldots e^{-x_{(n)}}\,dx_{(1)}\,dx_{(2)}\ldots dx_{(n)},$$

deoarece statisticile de ordine se pot repartiza pe axa x-lor în $n!$ moduri.

Urmează că probabilitatea elementară a vectorului $(u_1, u_2, \ldots, u_n)$ este

$$n!\,e^{-u_1}e^{-(u_1+u_2)}\ldots e^{-(u_1+u_2+\ldots+u_n)}\,du_1\,du_2\ldots du_n$$

sau

$$n\,e^{-nu_1}\,du_1(n-1)\,e^{-(n-1)u_2}\,du_2\ldots 2e^{-2u_{n-1}}\,du_{n-1}\,e^{-u_n}\,du_n$$

ceea ce arată că variabilele $u_1, u_2, \ldots, u_n$ sînt independente, u_1 avînd probabilitatea elementară $n\,e^{-nu_1}\,du_1$, iar u_k probabilitatea elementară

$$(n-k+1)\,e^{-(n-k+1)u_k}\,du_k.$$

II.16. *Fie $x_1, x_2, \ldots, x_n$ o selecţie de volum n dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X avînd funcţia de repartiţie $F(x)$ şi densitatea $f(x)$.*

Dacă $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ sînt statisticile de ordine corespunzătoare acestei selecţii să se determine probabilitatea elementară

a) *a statisticii $x_{(r)}$,*

b) *a vectorului $(x_{(r)}, x_{(s)})$.*

S o l u ţ i e. *a*) Pentru ca valoarea $x_{(r)}$ să se găsească într-un interval de mărime δ, centrat în x, trebuie ca elementele selecţiei să se repartizeze ca în figura II.1.

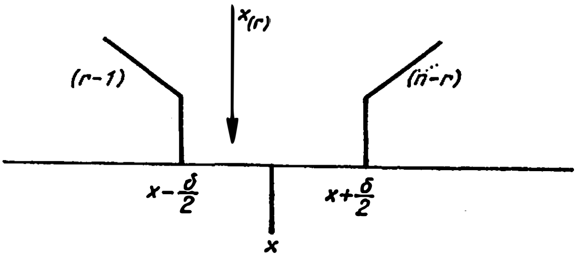

Fig. II. 1.

Probabilitatea ca

$$x_{(1)}, x_{(2)}, \ldots, x_{(r-1)} < x - \frac{\delta}{2}$$

este

$$\left(\int_{-\infty}^{x_{(r)}} f(x)\,\mathrm{d}x\right)^{r-1} = [F(x_{(r)})]^{r-1},$$

(într-o succesiune dată)

$$x_{(r)} \in \left(x - \frac{\delta}{2},\ x + \frac{\delta}{2}\right) \quad \text{este} \quad f(x_{(r)})\ \delta,$$

$$x_{(r+1)}, \ldots, x_{(n)} > x + \frac{\delta}{2} \quad \text{este} \quad \left(\int_{x_{(r)}}^{\infty} f(x)\,\mathrm{d}x\right)^{n-r} = [1 - F(x_{(r)})]^{n-r}$$

(într-o succesiune dată).

Urmează că probabilitatea elementară a statisticii $x_{(r)}$ este

$$\frac{n!}{(r-1)!\,1!\,(n-r)!}[F(x_{(r)})]^{r-1}[1 - F(x_{(r)})]^{n-r} f(x_{(r)})\ \mathrm{d}x_{(r)}$$

sau

$$\frac{1}{B(r, n-r+1)}[F(x_{(r)})]^{r-1}[1 - F(x_{(r)})]^{n-r} f(x_{(r)})\ \mathrm{d}x_{(r)}. \tag{II.6}$$

b) Avem situaţia din fig. II.2.

Deci probabilitatea elementară a vectorului $(x_{(r)}, x_{(s)})$ $r < s$ este

$$\frac{[F(x_{(r)})]^{r-1}[F(x_{(s)}) - F(x_{(r)})]^{s-r-1}[1 - F(x_{(s)})]^{n-s}}{B(r, s-r)\, B(s, n-s+1)} f(x_{(r)})\, f(x_{(s)}) \cdot \mathrm{d}x_{(r)}\, \mathrm{d}x_{(s)} \cdot$$

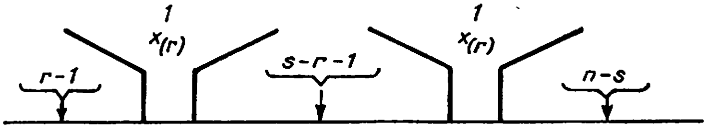

Fig. II. 2

II.17. *Să se determine repartiţia asimptotică pentru $n \to \infty$ a statisticii $x_{(r)}$.*

S o l u ţ i e. În (II.6) facem transformarea $y = nF(x)$; obţinem

$$\mathrm{d}H(y) = \frac{1}{B(r, n-r+1)} \left(\frac{y}{n}\right)^{r-1} \left(1 - \frac{y}{n}\right)^{n-r} \mathrm{d}\left(\frac{y}{n}\right).$$

Dacă $n \to \infty$, r rămînînd fix

$$\lim_{n\to\infty} \mathrm{d}H(y) = \left(\frac{1}{\Gamma(r)}\, y^{r-1}\, \mathrm{e}^{-y}\, \mathrm{d}y\right) \lim_{n\to\infty} \frac{\Gamma(n+1)}{\Gamma(n-r+1)\,(n-y)_r} \cdot$$

Cum

$$\frac{\Gamma(n+1)}{\Gamma(n-r+1)\,(n+y)_r} = \frac{\Gamma(N+r)}{\Gamma(N)\,(N-(y+1-r))^r} \quad \text{dacă} \quad n+1 = N+r$$

şi

$$\lim_{n\to\infty} \frac{\Gamma(N+r)}{\Gamma(N)\, N^r} = 1,$$

obţinem

$$\lim_{n\to\infty} \mathrm{d}H(y) = \lim_{n\to\infty} \mathrm{d}H_r(nF(x)) = \mathrm{d}\left(\frac{1}{\Gamma(r)} \int_0^y t^{r-1}\, \mathrm{e}^{-t}\, \mathrm{d}t\right) \cdot$$

II.18. *Să se determine repartiţia asimptotică a cuantilei empirice de ordin p.*

S o l u ţ i e. Cuantila empirică de ordin p este definită de $F(x) = p$ şi în cazul în care np este întreg, este $x_{(np+1)}$.

Din (II.6) urmează că $x_{(np+1)}$ are probabilitatea elementară

$$\frac{[F(x_{(np+1)})]^{np}[1 - F(x_{(np+1)})]^{nq-1}}{B(np+1, nq)} f(x_{(np+1)})\,\mathrm{d}x_{(np+1)}, p + q = 1 \quad \text{(II.8)}$$

Dacă în (II.8) punem $y = F(x_{(np+1)})$, iar apoi $u = \dfrac{y-p}{\sqrt{pq/n}}$ obţinem succesiv

$$\frac{n!}{(np)!\,(nq-1)!}\, y^{np}(1-y)^{nq-1}\,\mathrm{d}y,$$

$$\frac{n!\,nq}{(np)!\,(nq)!}\left[p + u\sqrt{\frac{pq}{n}}\right]^{np}\left[1 - p - u\sqrt{\frac{pq}{n}}\right]^{nq-1}\sqrt{\frac{pq}{n}}\mathrm{d}u$$

sau

$$\frac{n!}{(np)!\,(nq)!}p^{np}\left[1 + u\sqrt{\frac{q}{np}}\right]^{np} q^{nq}\left[1 - u\sqrt{\frac{p}{nq}}\right]^{nq} \frac{\sqrt{npq}}{1 - u\sqrt{\dfrac{p}{nq}}}\,\mathrm{d}u. \quad \text{(II.9)}$$

Deoarece

$$\left[1 + u\sqrt{\frac{q}{np}}\right]^{np} = \exp\left[np \lg\left(1 + u\sqrt{\frac{q}{np}}\right)\right] =$$

$$= \exp\left[np\left(u\sqrt{\frac{q}{np}} - \frac{u^2}{2}\frac{q}{np} + u^2\varepsilon_1\right)\right],$$

$$\left[1 - u\sqrt{\frac{p}{nq}}\right]^{nq} = \exp\left[nq \lg\left(1 - u\sqrt{\frac{p}{nq}}\right)\right] =$$

$$= \exp\left[nq\left(-u\sqrt{\frac{p}{nq}} - \frac{u^2}{2}\frac{p}{nq} - u^2\varepsilon_2\right)\right]$$

şi după formula lui Stirling (pentru $p \to \infty$)

$$\frac{n!}{(np)!\,(nq)!} \simeq \frac{1}{p^{np}q^{nq}\sqrt{2\pi npq}}$$

din (II.9) urmează că u urmează o repartiţie N (0,1) (fig. II.3).

Fie x_p cuantila de ordin p

$$F(x_p) = p.$$

După formula lui Taylor găsim

$$F(x_{(np+1)}) = F(x_p) + (x_{(np+1)} - x_p) f(x_p)$$

sau

$$x_{(np+1)} - x_p = \frac{y - p}{f(x_p)},$$

ceea ce arată că $x_{(np+1)}$ este repartizată $N\left(x_p, \frac{1}{f(x_p)} \sqrt{\frac{pq}{n}}\right)$.

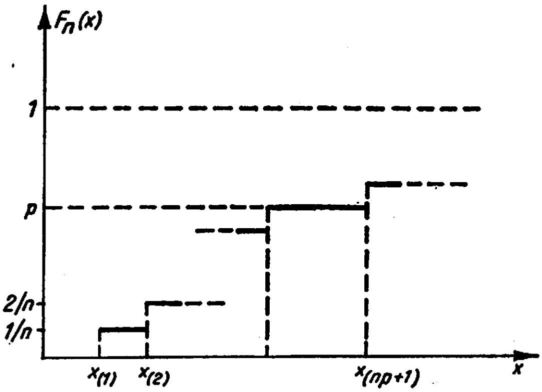

Fig. II. 3

II.19. *Să se determine reapartiţia asimptotică a vectorului bidimensional avînd ca componente cuantilele empirice de ordinul p_1 şi $p_2 (p_2 > p_1)$.*

Soluţie. Din problema II.18 se ştie că $x_{(np_1+1)}$ este asimptotic repartizată $N\left(x_{p_1}, \frac{1}{f(x_{p_1})} \sqrt{\frac{p_1 q_1}{n}}\right)$, $x_{(np_2+1)}$ este asimptotic repartizată $N\left(x_{p_2}, \frac{1}{f(x_{p_2})} \sqrt{\frac{p_2 q_2}{n}}\right)$.

Dacă în formula (II.7) din problema II.16 punem $y_r = F(x_{(r)})$, $y_s = F(x_{(s)})$ găsim pentru vectorul (y_r, y_s), probabilitatea elementară

$$\frac{y_r^{r-1}(y_s - y_r)^{s-r-1}(1 - y_s)^{n-s}}{B(r, s - r)\, B(s, n - s + 1)} dy_r dy_s, \quad 0 < y_r < y_s < 1.$$

Se observă că repartiţia acestui vector este independentă de repartiţia lui x.

Deoarece probabilitatea elementară a statisticii y_r este

$$\frac{y_r^{r-1}(1-y_r)^{n-r}}{B(r, n-r+1)}\, dy_r,$$

urmează că probabilitatea elementară a variabilei $y_s \mid y_r$ este

$$\frac{B(r, n-r+1)}{B(r, s-r)\, B(s, n-s+1)} \left(\frac{y_s - y_r}{1-y_r}\right)^{s-r-1} \left(\frac{1-y_s}{1-y_r}\right)^{n-s} d\left(\frac{y_s}{1-y_r}\right). \tag{II.10}$$

Cum

$$\frac{B(r, n-r+1)}{B(r, s-r)\, B(s, n-s+1)} = \frac{1}{B(s-r)(n-s+1)},$$

$$\frac{1-y_s}{1-y_r} = 1 - \frac{y_s - y_r}{1-y_r}.$$

din (II.10) urmează că variabila $\left.\frac{y_s - y_r}{1-y_r}\right|_{y_r}$ este repartizată $B(s-r, n-s+1)$; deci $y_s \mid y_r$ este repartizată $y_r + (1-y_r)\, B(s-r, n-s+1)$.

Asemănător, găsim că $y_r \mid y_s$ este repartizată $y_s B(r, s-r)$.

Dacă punem $r = np_1 + 1$, $s = np_2 + 1$, urmează că $y_{np_1+1} \mid y_{np_2+1}$ este repartizată ca $y_{np_2+1} B(np_1 + 1, np_2 - np_1)$.

Pe de altă parte

$$y_{np_2+1} \text{ este asimptotic repartizată } \rightsquigarrow N\left(p_2, \sqrt{\frac{p_2 q_2}{n}}\right) \tag{II.11}$$

şi deci

$$B(np_2 + 1, nq_2) \rightsquigarrow N\left(p_2, \sqrt{\frac{p_2 q_2}{n}}\right),$$

$$B(np_1 + 1, np_2 - np_1) \rightsquigarrow N\left(\frac{p_1}{p_2}, \sqrt{\frac{\frac{p_1}{p_2}\left(1 - \frac{p_1}{p_2}\right)}{n}}\right),$$

$$y_{np_1+1} \mid y_{np_2+1} \rightsquigarrow N\left(\frac{p_1}{p_2} y_{np_2+1}, \frac{p_1}{p_2}\, y_{np_2+1} \sqrt{\frac{p_2 - p_1}{np_2 p_1}}\right) \tag{II.12}$$

De asemenea

$$y_{np_1+1} \rightsquigarrow \cdot N\left(p_1, \sqrt{\frac{p_1 q_1}{n}}\right). \tag{II.13}$$

Din relațiile (II.11) — (II.13), urmează că vectorul aleator (y_{np_1+1}, y_{np_2+1}) are, asimptotic, densitatea de repartiție:

$$\frac{1}{2\pi\sigma_{np_1+1}\sigma_{np_2+1}\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{\bar{y}^2_{np_1+1}}{\sigma^2_{np_1+1}}-\right.\right.$$

$$\left.\left.-2\rho\frac{\bar{y}_{np_1+1}\bar{y}_{np_2+1}}{\sigma_{np_1+1}\sigma_{np_2+1}}+\frac{\bar{y}^2_{np_2+1}}{\sigma^2_{np_2+1}}\right)\right],$$

unde

$$\bar{y}_{np_1+1} = y_{np_1+1} - p_1, \qquad \bar{y}_{np_2+1} = y_{np_2+1} - p_2)$$

$$M(y_{np_1+1}) = p_1, \qquad M(y_{np_2+1}) = p_2,$$

$$\sigma^2_{np_1+1} = \frac{p_1q_1}{n}, \quad \sigma^2_{np_2+1} = \frac{p_2q_2}{n}, \quad \rho = \sqrt{\frac{p_1q_2}{p_2q_1}}.$$

Dacă x_{p_1} și x_{p_2} sînt cuantilele de ordin p_1, respectiv p_2, folosind formula lui Taylor găsim

$$x_{(np_1+1)} - x_{p_1} \sim \frac{\bar{y}_{np_1+1}}{f(x_{p_1})},$$

$$x_{(np_2+1)} - x_{p_2} \sim \frac{\bar{y}_{np_2+1}}{f(x_{p_2})}.$$

Deci

$$\begin{pmatrix} x_{(np_1+1)} - x_{p_1} \\ x_{(np_2+1)} - x_{p_2} \end{pmatrix} \rightsquigarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right),$$

unde Σ, matricea de covarianțe, este

$$\Sigma = \begin{pmatrix} \dfrac{1}{f^2(x_{p_1})}\dfrac{p_1q_1}{n} & \dfrac{1}{f(x_{p_1})f(x_{p_2})}\dfrac{p_1q_2}{n} \\ \dfrac{1}{f(x_{p_1})f(x_{p_2})}\dfrac{p_1q_1}{n} & \dfrac{1}{f^2(x_{p_2})}\dfrac{p_2q_2}{n} \end{pmatrix}.$$

II.20. *Se consideră o selecție de volum trei : x_1, x_2, x_3 dintr-o populație de repartiție arbitrară și statistica ordinei corespunzătoare $x_{(1)}, x_{(2)}, x_{(3)}$. Notînd $x^m_{(j)}(x_{i_1}, x_{i_2}, \ldots, x_{i_m})$ $1 \leqslant j \leqslant m$ cel de al j-lea număr în ordine nedescrescătoare dintre $x_{i_1}, x_{i_2}, \ldots, x_{i_m}$ să se exprime în funcție de x_1, x_2, x_3:*

$$X^3_{(1)}(x_1, x_2, x_3); \quad X^3_{(2)}(x_1, x_2, x_3), \quad X^3_{(3)}(x_1, x_2, x_3)$$

S o l u ţ i e. Vom considera mai întîi cazul unei selecţii de volum $n = 2$. Atunci

$$X^2_{(1)}(x_1, x_2) = \min(x_1, x_2) = \frac{x_1 + x_2}{2} - \frac{|x_1 - x_2|}{2},$$

$$X^2_{(2)}(x_1, x_2) = \max(x_1, x_2) = \frac{x_1 + x_2}{2} + \frac{|x_1 - x_2|}{2}.$$

De aici rezultă că

$$X^3_1(x_1, x_2, x_3) = \min\ (X^2_{(1)}(x_1, x_2),\ X^2_{(1)}(x_2, x_3)) =$$

$$= \min(\min(x_1, x_2), \min(x_2, x_3)) =$$

$$= \frac{X^2_{(1)}(x_1, x_2) + X^2_{(1)}(x_2, x_3)}{2} - \frac{|X^2_{(1)}(x_1, x_2) - X^2_{(1)}(x_2, x_3)|}{2} =$$

$$= \frac{\frac{x_1 + x_2}{2} - \frac{|x_1 - x_2|}{2} + \frac{x_2 - x_3}{2} - \frac{|x_2 - x_3|}{2}}{2} -$$

$$- \frac{\left|\frac{x_1 + x_2}{2} - \frac{|x_1 - x_2|}{2} - \frac{x_2 + x_3}{2} + \frac{|x_2 - x_3|}{2}\right|}{2} =$$

$$= \frac{1}{4}\{x_1 + 2x_2 + x_3 - |x_1 - x_2| - |x_2 - x_3| -$$

$$- |x_1 + x_3 - |x_1 - x_2| + |x_2 - x_3||\}$$

Să găsim acum expresia variabilei $X^3_{(3)}(x_1, x_2, x_3)$.

Putem scrie pe baza celor de mai sus

$$X^3_{(3)}(x_1, x_2, x_3) = \max\left(X^2_{(2)}(x_1, x_2), X^2_{(2)}(x_2, x_3)\right) =$$

$$= \max\left(\max(x_1, x_2), \max(x_2, x_3)\right) =$$

$$= \frac{X^2_{(2)}(x_1, x_2) + X^2_{(2)}(x_2, x_3)}{2} + \frac{|X^2_{(2)}(x_1, x_2) - X^2_{(2)}(x_2, x_3)|}{2} =$$

$$= \frac{\frac{x_1 + x_2}{2} + \frac{|x_1 - x_2|}{2} + \frac{x_2 + x_3}{2} + \frac{|x_2 - x_3|}{2}}{2} +$$

$$+ \frac{\left|\frac{x_1 + x_2}{2} + \frac{|x_1 - x_2|}{2} - \frac{x_2 + x_3}{2} - \frac{|x_2 - x_1|}{2}\right|}{2} =$$

$$= \frac{1}{4}\{x_1 + 2x_2 + x_3 + |x_1 - x_2| + |x_2 - x_3| + |x_1 - x_3 +$$

$$+ |x_1 - x_2| - |x_2 - x_3||\}$$

În fine

$$X^3_{(2)}(x_1, x_2, x_3) = X^3_{(1)}(X^2_{(2)}(x_1, x_2), X^2_{(2)}(x_1, x_3), X^2_{(2)}(x_3, x_1)) =$$

$$= \frac{\min(X^2_{(2)}(x_1, x_2), X^2_{(2)}(x_2, x_3))}{2} + \frac{\min(X^2_{(2)}(x_2, x_3), X^2_{(2)}(x_3, x_1))}{2} +$$

$$|\min(X^2_{(2)}(x_1, x_2), X^2_{(2)}(x_2, x_3)) - \min(X^2_{(2)}(x_2, x_3), X^2_{(2)}(x_3, x_1))|$$

$$)(x_1, x_2) + X^2_{(2)}(x_2, x_3) - |X^2_{(2)}(x_1, x_2) - X^2_{(2)}(x_2, x_3)| + X^2_{(2)}(x_2, x_3) + X^2_{(2)}(x_3, x_1) - |X^2_{(2)}(x_2, x_3) - X^2_{(2)}(x_3, x_1)|$$

Făcînd înlocuirile corespunzătoare şi simplificările care apar obţinem

$$X^3_{(2)}(x_1, x_2, x_3) = \frac{1}{8}\{2x_1 + 3x_2 + 3x_3 + |x_1 - x_2| + 2|x_2 - x_3| +$$

$$+ |x_3 - x_1| - |x_1 - x_3 + |x_1 - x_2| - |x_2 - x_3|| - |x_2 - x_1 +$$

$$+ |x_2 - x_3| - |x_3 - x_1|| - |x_2 - x_3 + |x_1 - x_2| - |x_1 - x_3| -$$

$$- |x_1 - x_3 + |x_1 - x_2| - |x_2 - x_3|| + |x_2 - x_1 +$$

$$+ |x_2 - x_3| - |x_1 - x_3|||\}.$$

Observaţie. În general pentru o selecţie de volum n: $x_1, x_2, \ldots, x_n$

dintr-o populaţie arbitrară putem scrie

$$X_{(1)}^{n}(x_1, x_2, ..., x_n) = X_{(1)}^{2}(X_{(1)}^{n-1}(x_1, x_2, ..., x_{n-1}); X_{(1)}^{n-1}(x_2, ..., x_n)),$$

$$X_{(n)}^{n}(x_1, x_2, ..., x_n) = X_{(2)}^{2}(X_{(n-1)}^{n-1}(x_1, x_2, ..., x_{n-1}); X_{(n-1)}^{n-1}(x_2, ..., x_n))$$

şi

$$X_{(j)}^{n}(x_1, x_2, ..., x_n) = X_{(n)}^{n}(X_{(j-1)}^{n-1}(x_1, x_2, ..., x_{n-1}), ..., X_{(j-1)}^{n-1}(x_n, x_1, ..., x_{n-2}))$$

$$(j = 2, 3, ..., n-1)$$

sau echivalent

$$X_{(j)}^{n}(x_1, x_2, ..., x_n) = X_{(1)}^{n}(X_{(j)}^{n-1}(x_1, x_2, ..., x_{n-1}), ..., X_{(j)}^{n-1}(x_n, x_1, ..., x_{n-2}))$$

II.21. *$y_1, y_2, ..., y_n$ urmează o repartiţie multinomială cu funcţia de frecvenţă.*

$$f(y_1, y_2, ..., y_n) = \frac{m!}{\prod_{1}^{n} y_i!} \prod_{1}^{n} p_i^{y_i} \tag{II.14}$$

unde

$$p_i = 1/n, \qquad 1 \leqslant i \leqslant n.$$

Dacă

$$\bar{y} = \frac{1}{n} \sum_{1}^{n} y_i \quad \text{şi} \quad s^2 = \frac{1}{n-1} \sum_{1}^{n} (y_i - \bar{y})^2,$$

să se arate că

$$M(s_n^2) = m/n.$$

Soluţie. Avem

$$M(s^2) = \frac{1}{n-1} M\left(\sum_{1}^{n} y_i^2 - n\bar{y}^2\right) = \frac{1}{n-1}\left[\sum_{1}^{n} M(y_i^2) - n\left(\frac{m}{n}\right)^2\right] =$$

$$= \frac{1}{n-1}\left[n\left(\frac{m(n-1)}{n^2} + \frac{m^2}{n^2}\right) - \frac{m^2}{n}\right] = \frac{m}{n},$$

II.22. *Fie $x_1, x_2, ..., x_n$ o selecţie dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X repartizată Poisson de medie λ.*

Să se arate că:

a) $M[s^2(\bar{x})^\delta] = M[(\bar{x})^{\delta+1}]$ pentru toţi $\delta \geqslant 0$.

b) $M\left[s^2(\bar{x})^\delta \,\middle|\, \sum_1^n x_i \geqslant 1\right] = M\left[(\bar{x})^{\delta+1} \,\middle|\, \bar{x} \geqslant \frac{1}{n}\right] =$

$= n^{-(\delta+1)} M(U^{\delta+1})$ *pentru toţi* $\delta \leqslant 0$.

c) $M[s^2(\bar{x})^\delta] = ce^{-n\lambda} + n^{-(\delta+1)} \sum_{U=1}^{\infty} \frac{e^{-n\lambda}(n\lambda)^U}{U!} U^{\delta+1} \qquad \delta < 0.$

unde

$$\bar{x} = \frac{1}{n}\sum_1^n x_i, \qquad s^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2$$

şi U este o variabilă aleatoare Poisson trunchiată cu funcţia de frecvenţă

$$g(U) = \frac{e^{-n\lambda}(n\lambda)^U}{U!(1 - e^{-n\lambda})}; \qquad U = 1, 2, 3, \dots$$

Soluţie. Avem

$$M[s^2(\bar{x})^\delta] = M_m\left[M\left\{s^2(\bar{x})^\delta \,\middle|\, \sum_1^n x_i = m\right\}\right] =$$

$$= M_m\left[\left(\frac{m}{n}\right)^\delta M\left(s^2 \,\middle|\, \sum_1^n x_i = m\right)\right].$$

Pentru o selecţie $x_1, x_2, \dots, x_n$ dintr-o populaţie în care caracteristica sub cercetare este o variabilă repartizată Poisson, repartiţia vectorului $(x_1, x_2, \dots, x_n)$ dată fiind $\sum_1^n x_i = m$ este multinomială cu funcţia de frecvenţă dată de (II. 14). Urmează că

$$M\left(s^2 \,\middle|\, \sum_1^n x_i = m\right) = M(s^2) = \frac{m}{n}. \tag{II.15}$$

Deci

$$M[s^2(\bar{x})^\delta] = M_m\left[\left(\frac{m}{n}\right)^\delta \left(\frac{m}{n}\right)\right] = M_m\left(\frac{m}{n}\right)^{\delta+1} = M[(\bar{x})^{\delta+1}].$$

b) Avem

$$M[s^2(\bar{x})^\delta \,|\, \sum_1^n x_i \geqslant 1] = M_m\left[\left\{M\left(s^2(\bar{x})^\delta \,\middle|\, \sum_1^n x_i = m\right)\right\} \,\middle|\, m \geqslant 1\right] =$$

$$= M_m\left[\left\{\left(\frac{m}{n}\right)^\delta M\left(s^2 \,\middle|\, \sum_1^n x_i = m\right)\right\} \,\middle|\, m \geqslant 1\right].$$

Ţinînd seama de (II.15), găsim

$$M\left[s^2(\bar{x})^\delta \,\middle|\, \sum_1^n x_i \geqslant 1\right] = M_m\left[\left(\frac{m}{n}\right)^{\delta+1} \,\middle|\, m \geqslant 1\right] =$$

$$= n^{-(\delta+1)} M(m^{\delta+1} \mid m \geqslant 1)$$

şi cum m are o repartiţie Poisson de medie $n\lambda$, obţinem b).

c) O dificultate în evaluarea $M[s^2(\bar{x})^\delta]$ pentru $\delta < 0$, apare atunci cînd toţi x_i sînt zero, deoarece în acest caz $s^2(\bar{x})^\delta$ este de forma $0/0$. Dacă definim $s^2(\bar{x})^\delta = c$ (o constantă arbitrară), putem scrie

$$M[s^2(\bar{x})^\delta] = c\,\mathrm{e}^{-n\lambda} + (1 - \mathrm{e}^{-n\lambda})\, M\left(s^2(\bar{x})^\delta \,\middle|\, \sum_1^n x_i \geqslant 1\right)$$

sau ţinînd seama de b)

$$M[s^2(\bar{x})^\delta] = c\,\mathrm{e}^{-n\lambda} + (1 - \mathrm{e}^{-n\lambda})\, M\left(\bar{x}^{\delta+1} \,\middle|\, \bar{x} \geqslant \frac{1}{n}\right) =$$

$$= c\,\mathrm{e}^{-n\lambda} + (1 - \mathrm{e}^{-n\lambda})\, n^{-(\delta+1)} M(U^{\delta+1}), \qquad \delta < 0. \qquad \text{(II.16)}$$

Observaţie. Pentru $\delta = -1$ (II.16) devine $M(s^2/\bar{x}) = c\mathrm{e}^{-n\lambda} + (1 - \mathrm{e}^{-n\lambda})$.

II. 23. *Fie* $(x_1, y_1) \ldots (x_n, y_n)$ *o selecţie dintr-o populaţie în care caracteristica sub cercetare este un vector aleator repartizat normal de medii* (μ_1, μ_2), *dispersii* (σ_1^2, σ_2^2) *şi coeficient de corelaţie* ρ.

Să se arate că primele patru momente ale statisticii

$$r = \sum_1^n (x_i - \bar{x})(y_i - \bar{y}) \Big/ \left\{\sum_1^n (x_i - \bar{x})^2 \sum_1^n (y_i - \bar{y})^2\right\}^{1/2}$$

față de origine sînt

$$-c_n\rho F\left(\frac{1}{2},\ \frac{1}{2},\ \frac{1}{2}(n+1);\ \rho^2\right),$$

$$-1-\frac{(n-2)(1-\rho^2)}{n-1}F\left(1,1,\frac{1}{2}(n+1);\rho^2\right),$$

$$-c_n\rho F\left(\frac{1}{2},\ \frac{1}{2},\ \frac{n+1}{2};\ \rho^2\right)-\frac{1}{\rho}\,c_n(n-1)(n-2)\times$$

$$\times\left[F\left(\frac{1}{2},\ \frac{1}{2},\ \frac{n-1}{2};\ \rho^2\right)-F\left(\frac{1}{2},\ \frac{1}{2},\ \frac{n+1}{2};\ \rho^2\right)\right],$$

$$-1+\frac{(n-2)(n-4)(1-\rho^2)}{2(n-1)}F\left(1,1,\frac{n+1}{2};\ \rho^2\right)-$$

$$-\frac{n(n-2)(1-\rho^2)}{4\rho^2}\left[F\left(1,1,\ \frac{n+1}{2};\ \rho^2\right)-1\right],$$

unde

$$c_n=\frac{2}{n-1}\left[\Gamma\left(\frac{n}{2}\right)\Big/\Gamma\left(\frac{n-1}{2}\right)\right]^2$$

și

$$F(\alpha,\beta,\delta;\rho^2)=\frac{\Gamma(\delta)}{\Gamma(\alpha)\Gamma(\beta)}\sum_{j=0}^{\infty}\frac{\Gamma(\alpha+j)\Gamma(\beta+j)}{\Gamma(\delta+j)}\cdot\frac{(\rho^2)^j}{j!}.$$

II.24. *Dintr-un lot în care fracția p a pieselor defecte este necunoscută, considerăm o selecție de volum n. Presupunem că p este o variabilă aleatoare cu densitatea*

$$f(p)=\begin{cases}1 \text{ dacă } 0\leqslant p<1,\\ 0 \text{ în rest.}\end{cases}$$

În această selecție se observă m piese defecte. Care este probabilitatea ca p să satisfacă inegalitatea $p\leqslant 0,1$?

Soluție. Fiecărei extracții îi asociem o variabilă aleatoare care poate lua valoarea 1 sau 0 după cum piesa este defectă sau bună. Dacă X reprezintă frecvența pieselor defecte în selecțiile de n elemente, atunci X este o variabilă aleatoare discretă cu funcția de frecvență

$$P\left(X=\frac{m}{n}\right)=C_n^m p^m(1-p)^{n-m}\qquad (m=0,1,\ldots,n).$$

Parametrul p este o variabilă aleatoare continuă. Formula lui Bayes dă:

$$g\left(p \,\middle|\, \frac{m}{n}\right) = \frac{C_n^m p^m (1-p)^{n-m}}{\int_0^1 C_n^m p^m (1-p)^{n-m}\,\mathrm{d}p} = \frac{p^m(1-p)^{n-m}}{B(m+1,\, n-m+1)}.$$

Deci probabilitatea căutată este egală cu

$$\int_0^{0,1} g\left(p \,\middle|\, \frac{m}{n}\right)\mathrm{d}p,$$

$\int_0^{0,1} p^m(1-p)^{n-m}\,\mathrm{d}p$ găsindu-se în tabelele cu funcția B incompletă.

II.25. *Variabila aleatoare X are repartiția $N(\mu, 1)$ unde parametrul μ este o variabilă aleatoare. Presupunem că:*

$$f(\mu) = \begin{cases} \frac{1}{2}a & \text{dacă } \mu \in (-a, a), \\ 0 & \text{dacă } \mu \notin (-a, a), \end{cases}$$

Să se găsească repartiția a posteriori a variabilei aleatoare μ pe baza selecției $(x_1, x_2, \ldots, x_n)$ din populația în care caracteristica sub cercetare este X.

Soluție. Media de selecție $\bar{x}_n$ are repartiția $N\left(\mu, \frac{1}{\sqrt{n}}\right)$. Teorema lui Bayes dă

$$f(\mu \mid \bar{x}) = \frac{\frac{\sqrt{n}}{\sqrt{2\pi}}\exp\left[-\frac{n(\bar{x}-\mu)^2}{2}\right]}{\int_{-a}^{a}\frac{\sqrt{n}}{\sqrt{2\pi}}\exp\left[-\frac{n(\bar{x}-\mu)^2}{2}\right]\mathrm{d}\mu}.$$

Pentru $n \to \infty$, obținem

$$f(\mu \mid \bar{x}) = \frac{\sqrt{n}}{\sqrt{2\pi}}\exp\left[-\frac{n(\bar{x}-\mu)^2}{2}\right],$$

de unde rezultă că μ este repartizată $N(\bar{x}, 1/\sqrt{n})$.

II.26. *Să presupunem că în urma unei experienţe poate rezulta unul dintre evenimentele $E_1, E_2, \ldots, E_k$ şi că probabilitatea de realizare a evenimentului E_i este p_i, $1 \leqslant i \leqslant k$. De asemenea să presupunem că în n repetări ale acestei experienţe, evenimentul E_i, s-a realizat de n_i ori* $\left(\sum_1^k n_i = n\right)$. *Fie*

$$V' = \left(\frac{n_1 - np_1}{\sqrt{p_1}}, \ldots, \frac{n_k - np_k}{\sqrt{p_k}}\right)$$

şi

$$\Phi' = (\sqrt{p_1}, \ldots, \sqrt{p_k}).$$

Să se arate că, asimptotic, funcţia liniară $b'V$ unde $b' = (b_1, \ldots, b_k)$ are repartiţia normală cu media zero şi dispersia $b'b - (b'\Phi)^2 = b'(I - \Phi\Phi')\,b$.

S o l u ţ i e. Considerăm variabila aleatoare X care are repartiţii

$$\begin{pmatrix} \dfrac{b_1}{\sqrt{p_1}} & \dfrac{b_2}{\sqrt{p_2}} \cdots & \dfrac{b_k}{\sqrt{p_k}} \\ p_1 & p_2 \cdots & p_k \end{pmatrix}.$$

deci variabila aleatoare care ia valoarea $\dfrac{b_i}{\sqrt{p_i}}$ cînd se realizează evenimentul E_i, $1 \leqslant i \leqslant k$.

Avem

$$M(X) = \sum_1^k \frac{b_i}{\sqrt{p_i}}\, p_i = \sum_1^k b_i \sqrt{p_i} = b'\Phi,$$

$$M(X^2) = \sum_1^k \frac{b_i^2}{p_i}\, p_i = \sum_1^k b_i^2 = b'b,$$

de unde

$$D^2(X) = M(X^2) - M^2(X) = b'b - (b'\Phi)^2.$$

Media aritmetică a celor n valori observate este

$$\bar{X}_n = \frac{1}{n}\left(\frac{b_1}{\sqrt{p_1}} n_1 + \ldots + \frac{b_k}{\sqrt{p_k}} n_k\right).$$

Teorema limită centrală, ne spune că

$$\sqrt{n}(\bar{X}_n - b'\Phi) \rightarrow Y \sim N(0, b'b - (b'\Phi)^2),$$

şi cum $\sqrt{n}(\bar{X}_n - b'\Phi) = b'V$, obţinem rezultatul dorit.

În cele ce urmează se dau repartiţiile asimptotice ale funcţiilor liniare şi formelor pătratice de V (adică frecvenţele $n_1, n_2, \ldots, n_k$ sînt variabile aleatoare).

II.27. *Să se arate că repartiţia asimptotică a funcţiei liniare $B'V$, unde B este o matrice de ordinul $k \times p$ şi de rang p, este o repartiţie normală p-dimensională de vector valoare medie zero şi matrice de covarianţe $B'(I - \Phi\Phi')B$.*

Soluţie Funcţia liniară $\lambda'B'V$ în V, conform problemei II.26. converge către o variabilă repartizată

$$N[0, \lambda'B'(I - \Phi\Phi')B\lambda], \tag{II.17}$$

Fie Z un vector aleator normal p-dimensional de vector valoare medie 0 şi matrice de covarianţe $B'(I - \Phi\Phi')B$. Funcţia liniară $\lambda'Z$ are aceeaşi repartiţie ca şi $\lambda'B'V$, ceea ce ne permite să afirmăm, în baza teoremei limită centrală multidimensională, că $B'V$ converge către un vector aleator normal p-dimensional $N(0, B'(I - \Phi\Phi')B)$.

II.28. *Fie A o matrice de ordin $k \times (k-1)$ astfel încît matricea $(\Phi \vdots A)$ de ordin $k \times k$ este ortogonală. Să se arate că repartiţia asimptotică a $(k-1)$ funcţii liniare $G = A'V$ este aceeaşi cu a $(k-1)$ variable aleatoare independente, fiecare repartizată $N(0,1)$.*

Soluţie. Din problema II.27 şi din condiţia $A'\Phi = 0$, rezultă că matricea de covarianţe corespunzătoare repartiţiei asimptotice a funcţiei $A'V$ este

$A'(I - \Phi\Phi')A = A'A = I$, care este o matrice de ordin $(k-1) \times (k-1)$. Astfel $G \rightarrow Y \sim N(0, I)$.

II.92. *Să se arate că o condiţie suficientă ca forma pătratică să aibă asimptotic o repartiţie χ^2 este ca*

$$C^2 = C \quad \text{şi} \quad C\Phi = \alpha\Phi,$$

unde α este o constantă, în care caz χ^2 are $R(C)$ grade de libertate dacă $\alpha = 0$ şi $[R(C) - 1]$ dacă $\alpha \neq 0$.

Soluţie. Considerăm ecuaţiile:

$$O = \Phi' V,$$

$$G = A' V,$$

unde A este definită ca în problema II.28. Soluţia acestui sistem de ecuaţii în raport cu V este

$$V = (\Phi \vdots A) \begin{pmatrix} O \\ \cdots \\ G \end{pmatrix} = AG,$$

deoarece $(\Phi : A)$ este o matrice ortogonală. Prin urmare

$$V'CV = G'A'CAG.$$

Repartiţia asimptotică a formei pătratice $V'CV$ este aceeaşi ca a lui $G'A'CAG$, care este o funcţie continuă de G. Deci repartiţia asimptotică a lui $G'A'CAG$ poate fi determinată din repartiţia asimptotică a lui G. Însă repartiţia asimptotică a lui G este aceeaşi ca a $(k - 1)$ variabile aleatoare independente fiecare cu media zero şi dispersia unitate. Condiţia necesară şi suficientă ca $G'A'CAG$ să aibă repartiţie χ^2 este ca $A'CA$ să fie idempotentă, adică

$$A'CAA'CA = A'CA. \tag{II.18}$$

Deoarece matricea $(\Phi \vdots A)$ este ortogonală,

$$(\Phi \vdots A) \begin{pmatrix} \Phi' \\ \cdots \\ A' \end{pmatrix} = \Phi\Phi' + AA' = I.$$

Substituind AA' în (II.18), avem

$$A'C(I - \Phi\Phi')\,CA = A'CA, \tag{II.19}$$

pentru toţi A astfel că $(\Phi \vdots A)$ este o matrice ortogonală. O condiţie suficientă pentru ca II.19. să aibă loc este

$$C^2 = C \quad \text{şi} \quad C\Phi = \alpha\Phi, \tag{II.20}$$

adică C este idempotentă şi Φ este vectorul propriu a lui C. Numărul gradelor de libertate ale repartiţiei χ^2 este $R(A'CA)$, rangul lui $A'CA$. Rezultă

$$R(C) = R\left\{\begin{pmatrix}\Phi' \\ \cdots \\ A'\end{pmatrix} C\,(\Phi \mid A)\right\} = R\begin{pmatrix}\Phi'C\Phi & \Phi'CA \\ A'C\Phi & A'CA\end{pmatrix} =$$

$$= R(A'CA) \quad \textit{dacă} \quad C\Phi = 0,$$

$$= R(A'CA) + 1, \; \textit{dacă} \; C\Phi = \alpha\Phi, \quad \alpha \neq 0.$$

II.30. *Să se arate că condiţia necesară şi suficientă ca $V'CV$ să aibă asimptotic repartiţia χ^2 este ca*

$$(I - \Phi\Phi')C(I - \Phi\Phi')C(I - \Phi\Phi') =$$

$$= (I - \Phi\Phi')C(I - \Phi\Phi'). \tag{II.21}$$

Soluţie. În problema (II.29) s-a arătat că condiţia necesară şi suficientă este

$$A'C(I - \Phi\Phi')\,CA = A'CA. \tag{II,22}$$

Însă $\mathfrak{M}(A) = \mathfrak{M}(I - \Phi\Phi')$. Prin urmare în relaţia II.22, A poate fi înlocuită prin $I - \Phi\Phi'$, care dă II.21.

II.31. *Fie $(x_1, x_2, \ldots, x_n)$ o selecţie dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X avînd media μ şi dispersia σ^2, μ şi σ^2 fiind finite. Dacă $g(x)$ este o funcţie care admite derivată într-o vecinătate a punctului $x = \mu$, astfel încît $g'(\mu) \neq 0$, să se arate că $g(\bar{x}_n)$, pentru valori mari ale lui n, este repartizată $N(g(\mu), [\sigma g'(\mu)]^2/n])$.*

Soluţie. Fie $V(\mu)$ vecinătatea punctului $x = \mu$ pentru care $g'(x)$ există, oricare ar fi $x \in V(\mu)$. Din enunţ urmează că pentru orice $\varepsilon > 0$, există $n(\varepsilon)$ astfel încît pentru $n > n(\varepsilon)$

$$P(\bar{x} \in V(\mu)) > 1 - \varepsilon.$$

Pentru orice punct $x = \bar{x}$ din $V(\mu)$ putem scrie

$$g(\bar{x}) = g(\mu) + g'(x^*)\,(\bar{x} - \mu), \tag{II.23}$$

unde x^* este o variabilă aleatoare astfel încît

$$|x^* - \mu| \leqslant |\bar{x} - \mu|,$$

(II.23) se mai poate scrie

$$(g(\bar{x}) - g(\mu))\sqrt{n} = g'(x^*)(\bar{x} - \mu)\sqrt{n} \qquad \text{(II.24)}$$

şi probabilitatea ca egalitatea (II.24) să aibă loc pentru toţi $n > n(\varepsilon)$ este mai mare decît $1 - \varepsilon$.

Fie

$$u_n = (g(\bar{x}) - g(\mu))\sqrt{n} \quad \text{şi} \quad v_n = g'(x^*)(\bar{x} - \mu)\sqrt{n}.$$

Deoarece σ^2 este finită, şirul de variabile aleatoare $\sqrt{n}(\bar{x} - \mu)$, $n = 1, 2, \ldots$ converge în repartiţie către variabila $N(0, \sigma^2)$. Deoarece $g'(x)$ există în toate punctele din $V(\mu)$, este continuă în $V(\mu)$ şi deci în $x = \mu$. Urmează că $g'(x^*)$ converge în probabilitate către $g'(\mu)$.

Prin urmare

$$\lim_{n\to\infty} P(V_n < w) = P(g'(\mu)s < w), \qquad \text{(II.25)}$$

unde s este o variabilă $N(0, \sigma^2)$, (II.25) implică

$$\lim_{n\to\infty} P(U_n < w) = P(t < w),$$

unde t are o repartiţie $N(0, (\sigma g'(\mu))^2)$.

Deci $g(\bar{x})$ pentru valori mari ale lui n, are o repartiţie $N(g(\mu), [\sigma g'(\mu)]^2/n)$.

II.32. *Dacă $(x_1, x_2, \ldots, x_n)$ este o selecţie dintr-o populaţie în care caracteristica sub cercetare X este o variabilă $P_0(\mu)$, să se arate că pentru valori mari ale lui n, $2\sqrt{\bar{x}}$ este o variabilă $N\left(2\sqrt{\mu}, \frac{1}{n}\right)$.*

S o l u ţ i e. X fiind o variabilă $P_0(\mu)$, urmează că

$$M(X) = \mu, \quad D^2(X) = \mu.$$

Funcţia $g(x) = 2\sqrt{x}$ este derivabilă şi $g'(x) = \frac{1}{\sqrt{x}}$. Avem $g'(\mu) = \frac{1}{\sqrt{\mu}} \neq 0$ şi cum

$$g(\mu) = 2\sqrt{\mu}, \ [D(X)g'(\mu)]^2 = 1,$$

urmează că asimptotic, $g(\bar{x}) = 2\sqrt{\bar{x}}$, este repartizată $N\left(2\sqrt{\mu}, \frac{1}{n}\right)$.

II.33. *Dacă* $(x_1, \ldots, x_n)$ *este o selecţie dintr-o populaţie în care caracteristica sub cercetare* X *este o variabilă* $R\left(\frac{1}{2}\theta, \theta\right)$, *să se arate că asimptotic (pentru valori mari ale lui* n*)* $\sqrt{12}\lg(2\bar{x})$ *este o variabilă* $N\left(\sqrt{12}\lg\theta, \frac{4}{n}\right)$.

S o l u ţ i e. X fiind o variabilă $R\left(\frac{1}{2}\theta, \theta\right)$ urmează că

$$M(X) = \frac{\theta}{2}, \quad D^2(X) = \frac{\theta^2}{12},$$

Funcţia $g(x) = \sqrt{12}\lg 2x$ este derivabilă şi $g'(x) = \frac{\sqrt{12}}{x}$. Avem

$$g'\left(\frac{\theta}{2}\right) = \frac{2\sqrt{12}}{\theta} \neq 0$$

şi cum

$$g\left(\frac{\theta}{2}\right) = \sqrt{12}\lg\theta, \quad D^2(X)\,g'\left(\frac{\theta}{2}\right) = \frac{\theta}{\sqrt{12}}\,\frac{2\sqrt{12}}{\theta} = 2,$$

urmează că asimptotic $\sqrt{12}\lg(2\bar{x})$ este o variabilă $N\left(\sqrt{12}\lg\theta, \frac{4}{n}\right)$.

II.34. *Dacă* $(x_1, \ldots, x_n)$ *este o selecţie dintr-o populaţie în care caracteristica sub cercetare* X *(este o variabilă* $Bi\,(1, p)$, *să se arate că asimptotic* $\arcsin(2\bar{x} - 1)$ *este o variabilă* $N\left(\arcsin(2p - 1), \frac{1}{n}\right)$.

S o l u ţ i e. X fiind o variabilă $Bi(1, p)$,

$$M(X) = p, \quad D^2(X) = qp, \quad p + q = 1.$$

Funcţia $g(x) = \arcsin(2x - 1)$ este derivabilă şi

$$g'(\mu) = g'(p)\,\frac{1}{\sqrt{p(1-p)}} \neq 0.$$

Cum

$$g(\mu) = g(p) = \arcsin(2p - 1),$$

$$[D(X)\,g'(\mu)]^2 = \left(\sqrt{pq}\,\frac{1}{\sqrt{pq}}\right)^2 = 1,$$

urmează că asimptotic $g(\bar{x}) = \arcsin(2\bar{x} - 1)$ este repartizată $N\left(\arcsin(2p - 1),\ \frac{1}{n}\right)$.

II.35. *Dacă* $(x_1, x_2, \ldots, x_n)$ *este o selecție dintr-o populație caracterizată de o variabilă aleatoare* X *avînd funcția de frecvență*

$$qp^{x-1}, \quad x = 1, 2, \ldots,$$

unde $0 < p < 1$, $p + q = 1$, *să se arate că asimptotic*, $\lg\left[\bar{x}\left(1 + \sqrt{1 - \frac{1}{x}}\right) - \frac{1}{2}\right]$ *este o variabilă*

$$N\left(\lg\left[\frac{1 + \sqrt{p}}{q} - \frac{1}{2}\right],\ \frac{1}{n}\right).$$

Soluție. Avem

$$M(X) = \sum_{x=1}^{\infty} xqp^{x-1} = q\left(\sum_{x=1}^{\infty} p^x\right)' = \frac{q}{(1-p)^2} = \frac{1}{q},$$

$$M(X^2) = \sum_{x=1}^{\infty} x^2qp^{x-1} = q[p\Sigma x(x-1)\, p^{x-2} + \Sigma xp^{x-1}] =$$

$$= qp(\Sigma p^x)'' + q(\Sigma p^x)' =$$

$$= \frac{2qp}{q^3} + \frac{q}{q^2} = \frac{2p}{q^2} + \frac{1}{q},$$

de unde

$$D^2(X) = \frac{2p}{q^2} + \frac{1}{q} - \frac{1}{q^2} = \frac{2p + q - 1}{q^2} = \frac{p}{q^2}.$$

Funcția

$$g(x) = \lg\left[x\left(1 + \sqrt{1 - \frac{1}{x}}\right) - \frac{1}{2}\right]$$

este derivabilă și

$$g'(x) = \frac{1}{x\left(1 + \sqrt{1 - \frac{1}{x}}\right) - \frac{1}{2}}\left[1 + \sqrt{1 - \frac{1}{x}} + \frac{1}{2x\sqrt{1 - \frac{1}{x}}}\right].$$

Avem

$$g'(\mu) = g'\left(\frac{1}{q}\right) = \frac{q}{\sqrt{p}} \neq 0,$$

şi cum

$$g(\mu) = g\left(\frac{1}{q}\right) = \lg\left[\frac{1}{q}\left(1 + \sqrt{1-q}\right) - \frac{1}{2}\right] = \lg\left(\frac{1+\sqrt{p}}{q} - \frac{1}{2}\right),$$

$$D(X)\, g'(\mu) = \frac{\sqrt{p}}{q} \cdot \frac{q}{\sqrt{p}} = 1,$$

urmează că asimptotic, pentru valori mari ale lui n,

$$g(\bar{x}) = \lg\left[\bar{x}\left(1 + \sqrt{1 - \frac{1}{\bar{x}}}\right) - \frac{1}{2}\right]$$

este o variabilă

$$N\left(\lg\left[\frac{1+\sqrt{p}}{q} - \frac{1}{2}\right], \quad \frac{1}{n}\right).$$

II.36. *Fie* $T_n = T_{1n}, \ldots, T_{kn})$ *o statistică k-dimensională, astfel încît vectorul aleator k-dimensional*

$$(\sqrt{n}(T_{1n} - \theta_1), \ldots, \sqrt{n}(T_{kn} - \theta_k)) \tag{II.26}$$

are asimptotic o repartiţie $N(0, (\sigma_{ij}))$.

Fie g o funcţie de k variabile care este total diferenţiabilă. Să se arate că

$$\sqrt{n}u_n = \sqrt{n}[g(T_{1n}, \ldots, T_{kn}) - g(\theta_1, \ldots, \theta_k)]$$

are asimptotic o reptariţie normală de medie zero şi dispersie

$$D^2(\theta) = \Sigma\Sigma\sigma_{ij} \frac{\partial g}{\partial \theta_i} \frac{\partial g}{\partial \theta_j},$$

cu condiţia $D^2(\theta) \neq 0$.

Soluţie. Deoarece g este o funcţie total diferenţiabilă, atunci

$$g(T_{1n}, \ldots, T_{kn}) - g(\theta_1, \ldots, \theta_k) = \Sigma(T_{in} - \theta_i)\left(\frac{\partial g}{\partial \theta_i}\right) + \varepsilon_n \| T_n - \theta \|,$$

unde $\varepsilon \to 0$ cînd $T_{in} \to \theta_i$. Cînd $n \to \infty$, $\varepsilon_n \xrightarrow{P} 0$ şi deoarece

$$\sqrt{n}\| T_n - \theta \| = [\Sigma n^2(T_{in} - \theta_i)^2]^{1/2}$$

are o repartiţie asimptotică

$$\left| \sqrt{n}\, u_n - \sqrt{n}\Sigma(T_{in} - \theta_i) \frac{\partial g}{\partial \theta_i} \right| \xrightarrow{P} 0.$$

Însă repartiţia asimptotică a variabilei $\sqrt{n}\Sigma(T_{in} - \theta_i) \frac{\partial g}{\partial \theta_i}$, fiind o funcţie liniară de variabile aleatoare normale, este normală cu media zero şi dispersia (II.26).

Observaţie. **Dacă σ_{ij} şi derivatele parţiale ale funcţiei g sînt funcţii continue de θ**

$$(\sqrt{n}\, u_n / D(T_n)) \to X \sim N(0,1),$$

unde $D^2(T_n)$ este valoarea lui $D^2(\theta)$ în punctul $\theta = T_n$.

II.37. *Fie variabilele aleatoare independente*

$$y_i \sim N(x_{i1}\beta_1 + \ldots + x_{im}\beta_m, \sigma^2)_{1 \leqslant i \leqslant n},$$

unde x_{ij} sînt coeficienţi cunoscuţi şi β_i sînt parametri necunoscuţi. În notaţie matriceală, dacă Y este vectorul coloană de componente y_i, β vectorul de componente β_j şi $X = (x_{ij})$ matricea coeficienţilor, atunci

$$\Sigma(y_i - x_{i1}\beta_1 - \cdots - x_{im}\beta_m)^2 = (Y - X\beta)'(Y - X\beta).$$

Prin urmare densitatea lui $y_1, y_2, \ldots, y_n$ poate fi scrisă

$$c\, e^{-(Y - X\beta)'(Y - X\beta)/2\sigma^2}.$$

Fie $R_0^2 = \min\limits_{\beta} (Y - X\beta)'(Y - X\beta)$. Să se arate că $R_0^2 \sim \sigma^2\chi^2(n - r)$, unde r este rangul matricei X.

Soluţie. Fie F o matrice de ordinul $(n \times r)$, astfel ca coloanele sale să constituie o bază ortonormală a lui $\mathcal{M}(X)$. Completăm matricea F prin G de ordinul $(n \times n - r)$, astfel ca $(F \vdots G)$ să fie o matrice ortogonală. Prin construcţie

$$F'G = 0, \quad X'G = 0, \quad G'X = 0,$$

şi transformarea de la Y la $Z = \binom{Z_1}{Z_2} = \binom{F'}{G'} Y$, *ce dă*

$$Z_1 = F'Y, \quad Z_2 = G'Y$$

este ortogonală. Vectorul $X\beta$ se transformă în

$$\zeta_1 = F' X \beta, \quad \zeta_2 = G' X \beta = 0.$$

Cum transformările ortogonale păstrează distanţele

$$(Y - X\beta)'(Y - X\beta) = (Z_1 - \zeta_1)'(Z_1 - \zeta_1) + Z_2'Z_2.$$

Prin urmare densitatea lui Z este

$$c \exp\{-[(Z_1 - F'X\beta)'(Z_1 - F'X\beta) + Z_2'Z_2]/2\sigma^2\},$$

care arată că Z_1 şi Z_2 sînt independente şi că repartiţia lui Z_2 este caracterizată prin densitatea

$$c \exp(-Z_2'Z_2/2\sigma^2). \tag{II.27}$$

Însă Z_2 este un vector cu $(n-r)$ componente, care după (II.27) sînt toate independente şi repartizate $N(0, \sigma^2)$. Prin urmare

$$Z_2'Z_2 \sim \sigma^2\chi^2(n-r).$$

Acum

$$R_0^2 = \min_\beta (Y - X\beta)'(Y - X\beta) = \min_\beta [(Z_1 - \zeta_1)'(Z_1 - \zeta_1) + Z_2'Z_2] =$$

$$= [\min_\beta (Z_1 - F'X\beta)'(Z_1 - F'X\beta)] + Z_2'Z_2 = Z_2'Z_2,$$

dacă alegem β astfel ca $Z_1 = F'X\beta$, ceea ce este posibil deoarece numărul de ecuaţii este r, care este de asemenea şi rangul lui $F'X$,

$$r = \text{rang } X' = \text{rang } X'(F \mid G) =$$

$$= \text{rang } (X'F \mid X'G) = \text{rang } (X'F \mid 0) = \text{rang } X'F.$$

II. 38. *Fie H o matrice de ordinul $(m \times k)$ şi de rang k, astfel* ca

$$\mathcal{M}(H) \subset (\mathcal{M} X')$$

şi

$$R_1^2 = \min (Y - X\beta)'(Y - X\beta)$$

supus la condiţia $H'\beta = \xi$ (dat). Să se arate că:

1° R_0^2 şi $R_1^2 - R_0^2$ *sînt independente.*

2° $R_0^2 \sim \sigma^2\chi^2(n-r)$ şi $R_1^2 - R_0^2 \sim$ *ca un χ^2 necentrat cu k grade de libertate*

3° *Dacă $H'\beta = \xi$ este adevărată, atunci* $R_1^2 - R_0^2 \sim \sigma^2\chi^2(k)$

şi

$$\frac{R_1^2 - R_0^2}{k} \div \frac{R_0^2}{n-r} \sim F(k, n-r).$$

S o l u ţ i e. Deoarece $\mathfrak{M}(X') = \mathfrak{M}(X'X)$, condiţia $\mathfrak{M}(H) \subset \mathfrak{M}(X') \Rightarrow$ $\Rightarrow \mathfrak{M}(H) \subset \mathfrak{M}(X'X)$ şi prin urmare există o matrice C astfel ca $H = X'XC$ şi rang $XC = k$. Facem transformarea

$$Z_1 = D'Y, \quad Z_2 = G'Y, \quad Z_3 = C'X'Y,$$

unde G este definită ca în problema anterioară şi D' este aleasă astfel ca $D'D = I$ şi ortogonală la G şi XC. Numărul de variabile în Z_1, Z_2 şi Z_3 sînt $r - k$, $n - r$ şi respectiv k. Avem

$$\zeta_1 = D'X\beta, \quad \zeta_2 = G'X\beta = 0, \quad \zeta_3 = C'X'X\beta = H'\beta.$$

Deoarece submatricele în transformare sînt ortogonale avem

$$(Y - X\beta)'(Y - X\beta) = (Z_1 - D'X\beta)'(Z_1 - D'X\beta) +$$

$$+ Z_2'Z_2 + (Z_3 - H'\beta)'(C'X'XC)^{-1}(Z_3 - H'\beta).$$

Scriind densitatea lui Z_1, Z_2, Z_3, găsim că ele sînt independente. Z_3 are densitatea.

$$c \exp[-(Z_3 - H'\beta)'(C'X'XC)^{-1}(Z_3 - H'\beta)/2\sigma^2]$$

care este aceeaşi cu a k funcţii liniare de variabile normale.

Forma pătratică

$$Q = (Z_3 - \xi)'(C'X'XC)^{-1}(Z_3 - \xi) \sim \sigma^2\chi^2(k,\lambda)$$

al cărei parametru de necentricitate este

$$\lambda = \sigma^{-2}(H'\beta - \xi)'(C'X'XC)^{-1}(H'\beta - \xi),$$

$R_0^2 = Z_2'Z_2 \sim \sigma^2\chi^2(n - r)$ ca în problema precedentă şi este independentă de Q. Cînd $H'\beta = \xi$ este adevărată, $\lambda = 0$, în care caz $Q \sim \sigma^2\chi^2(k)$. Rezultă 1° şi 2°, iar 3° va fi dovedit dacă putem să arătăm că $Q = R_1^2 - R_0^2$.

Avem

$$R_1^2 = \min_{H'\beta=\xi}(Y - X\beta)'(Y - X\beta) =$$

$$= [\min_{H'\beta=\xi}(Z_1 - D'X\beta)'(Z_1 - D'X\beta)] + R_0^2 + Q = R_0^2 + Q,$$

cu condiţia să existe β astfel ca $Z_1 - D'X\beta = 0$ şi $H'\beta = \xi$. Această existenţă este posibilă deoarece numărul de ecuaţii este r şi de asemenea $r = \text{rang}\ X' = \text{rang}\ X'(D \vdots G \vdots XC) = \text{rang}\ (X'D \vdots O \vdots H) = \text{rang}(X'D \vdots H)$.

II.39. *Fie* $y_i \sim N(\mu_i, \sigma^2)$, $i = 1, 2, ..., n$ *variabile aleatoare independente şi fie*

$$Y' = (y_1, y_2, ..., y_n) \quad \text{şi} \quad \mu' = (\mu_1, \mu_2, ..., \mu_n).$$

Să se arate că o condiţie suficientă pentru ca $Y'AY$ *şi* $Y'BY$ *să fie independente este ca* $AB = BA = O$.

S o l u ţ i e. Descompunerea spectrală a matricelor A şi B este

$$A = \lambda_1 P_1 P_1' + \lambda_2 P_2 P_2' + ... + \lambda_r P_r P_r', \; \lambda_i \neq 0, \quad i = 1, ..., r$$

$$B = \nu_1 Q_1 Q_1' + \nu_2 Q_2 Q_2' + ... + \nu_s Q_s Q_s', \quad \nu_j \neq 0, \quad j = 1, 2, ..., s,$$

unde $r = \text{rang } A$, $s = \text{rang } B$, P_i şi Q_i sînt ortonormale;

$$AB = 0 = \Sigma\Sigma\lambda_i P_i P_i' Q_j Q_j'.$$

Înmulţind la stînga şi la dreapta prin P_r' şi Q_s găsim

$$P_r' Q_s = 0,$$

astfel că P_i şi Q_j sînt toate ortonormale. Urmează că $P_i'Y$ şi $Q_j'Y$ sînt toate independente şi normale. Suficienţa condiţiei urmează observînd că

$$Y'AY = \lambda_1 (P_1'Y)^2 + ... + \lambda_r (P_r'Y)^2$$

şi

$$Y'BY = \nu_1 (Q_1'Y)^2 + ... + \nu_s (Q_s'Y)^2$$

se bazează exclusiv pe mulţimi de variabile independente.

Se observă că nu se cere ca $Y'AY$ şi $Y'BY$ să fie repartizate X^2.

II.40. *Fie* BY, m *funcţii liniare de* Y, *adică* B *este o matrice* $m \times n$. *Să se arate că o condiţie suficientă ca* BY *să fie independentă de forma pătratică* $Y'AY$ *este ca* $BA = 0$.

S o l u ţ i e. Descompunerea spectrală a matricei A, împreună cu condiţia $BA = 0$, implică $BP_r = 0$, $i = 1, ..., r$. Aceasta înseamnă că BY este independentă de funcţiile lineare $P_1'Y, ..., P_r'Y$. Însă

$$Y'AY = \sum_{i=1}^{n} \lambda_i (P_i'Y)^2$$

ceea ce demonstrează o afirmaţie făcută.

II.41. *Să se arate că* (în condiţiile problemei II.37.):
a) $X = X(X'X)^{-} X'X$.

b) *Matricea* $I - X(X'X)^{-}X'$ *este idempotentă (am notat prin* $(X'X)^{-}$ *inversa generalizată a matricei* $X'X$. *Inversa generalizată a unei matrice* A *de ordin* $m \times n$ *şi de orice rang este o matrice de ordin* $n \times m$, *notată prin* A^{-}, *astfel că pentru orice vector* Y *pentru care* $AX = Y$ *este o ecuaţie consistentă,* $X = A^{-}Y$ *este o soluţie).*

Soluţie. a) Fie $G = X[I - (X'X)^{-}(X'X)]$. Cum $G'G = 0$, rezultă că $G = 0$, ceea ce trebuia demonstrat.

b) Pentru a arăta că matricea $I - X(X'X)^{-}X'$ este idempotentă, considerăm pătratul acesteia şi folosim rezultatul de la punctul a).

II.42. *Fie* H *o matrice de ordin* $m \times k$ *astfel ca* $\mathfrak{M}(H) = \mathfrak{M}(X')$. *Să se arate că* $Z = H'\hat{\beta}$ *şi* $R_0^2 = (Y - X\hat{\beta})(Y - X\beta)$ *sînt independente, unde* $\hat{\beta} = (X'X)^{-}X'Y$ *este soluţia ecuaţiei* $X'X\beta = X'Y$.

Soluţie. Deoarece $\mathfrak{M}(H) = \mathfrak{M}(X')$, există o matrice C astfel ca $H = X'C$. Prin urmare

$$H'\hat{\beta} = C'X(X'X)^{-}X'Y = BY$$

şi

$$R_0^2 = Y'Y - Y'X(X'X)^{-}X'Y = Y'[I - X(X'X)^{-}X']Y = Y'AY.$$

Trebuie să stabilim independenţa mulţimii de funcţii liniare şi funcţii pătratice în Y. Găsim că

$$BA = [C'X(X'X)][I - X(X'X)^{-}X'] = 0,$$

deoarece produsul primilor doi termeni este zero după punctul a) al problemei II.41. Din condiţia suficientă stabilită în problema II.40 rezultă problema.

II.43. *Să se arate că* $Z = H'\hat{\beta}$ *are o repartiţie normală* k *dimensională de vector valoare medie* $H'\beta$ *şi matrice de covarianţe* $\sigma^2 D$. *unde* $D = H'(X'X)^{-}H$ *şi* $R_0^2 \sim \sigma^2\chi^2(n - r)$.

Soluţie. $Z = C'X(X'X)^{-}X'Y = BY$ este o funcţie liniară de variabile normale. Urmează că Z are o repartiţie normală k dimensională. Fie

$$R_0^2 = Y'[I - X(X'X)^{-}X']Y.$$

Matricea $I - X(X'X)^{-}X'$ este idempotentă. Deci $R_0^2 \sim \sigma^2\chi^2(n - r)$ unde $n - r = \mathrm{Ur} I - \mathrm{Ur}\, X(X'X)^{-}X' = n - \mathrm{Ur}(X'X)^{-}X'X = n -$ $- \operatorname{rang} X'X = n - \operatorname{rang} X$.

Aplicînd condiţia suficientă din problema II.40 rezultă că Z şi R_0^2 sînt independente.

II.44. *Fie* β^*, λ^* *soluţii ale ecuaţiilor*

$$X'X\beta + H\lambda = X'Y$$

$$H'\beta = \theta_0 (\text{o valoare dată}),$$

care sînt obţinute derivînd $(Y - X\beta)'(Y - X\beta)$ *în raport cu* β *cu condiţia* $H'\beta = \theta_0$. *Să se arate că*

$$R_1^2 = (Y - X\beta^*)'(Y - X\beta^*) =$$

$$= (Z - \theta_0) D^{-1}(Z - \theta_0) + R_0^2,$$

unde Z *şi* D *sînt definite ca în problema precedentă.*

Soluţie. Se observă că

$$R_1^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (\hat{\beta} - \beta^*)'X'X(\hat{\beta} - \beta^*),$$

şi este suficient să se arate că al doilea termen este

$$(Z - \theta_0) D^{-1}(Z - \theta_0).$$

Deoarece

$$\mathcal{M}(H) \subset \mathcal{M}(X'X),\ H = X'XC.$$

Din ecuaţiile pentru

$$\beta^*,\ \lambda^*,\ X'X(\hat{\beta} - \beta^*) = H\lambda^* = X'XC\lambda^*,$$

astfel că

$$(\hat{\beta} - \beta^*)'X'X(\hat{\beta} - \beta^*) = \lambda^{*\prime}C'X'XC\lambda^*.$$

De asemenea din ecuaţiile

$$C'X'X\beta^* - H'\beta^* + C'H\lambda^* = C'X'Y - \theta_0,$$

sau

$$CH'\lambda = Z - \theta_0, \quad C'X'XC\lambda^* = Z - \theta_0$$

$$\Rightarrow \lambda^* = D^{-1}(Z - \theta_0) \quad \text{unde} \quad D = C'X'XC.$$

Capitolul III

CORELAŢIE ŞI REGRESIE

III.1. *O urnă conţine 8 bile albe şi 4 bile negre. Din această urnă se efectuează o selecţie de volum* $n = 5$, *fără a pune bila extrasă înapoi în urnă. Notîndu-se cu* X_j, $j = 1, 2, \ldots, 5$ *variabila aleatoare definită astfel:*

$$X_j = \begin{cases} 1, \text{ dacă la extracţia de rang } j \text{ bila extrasă a fost albă,} \\ 0, \text{ dacă la extracţia de rang } j \text{ bila extrasă a fost neagră.} \end{cases}$$

Să se calculeze

$$M(X_2); \quad M(X_3); \quad D^2(X_2); \quad D^2(X_3), \ \mathrm{cov}\,(X_1,X_2); \quad \mathrm{cov}\,(X_2,X_3),$$
$$\rho(X_1,X_2), \quad \rho(X_2,X_3).$$

S o l u ţ i e. Din datele problemei rezultă că variabilele aleatoare X_1, X_2 şi X_3 au repartiţiile

$$X_1: \begin{pmatrix} 1 & 0 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}, \quad X_2: \begin{pmatrix} 1 & 0 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}, \quad X_3: \begin{pmatrix} 1 & 0 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

De aici obţinem

$$M(X_2) = \frac{2}{3}; \quad D^2(X_2) = M(X_2^2) - M^2(X_2) = \frac{2}{3} - \frac{4}{9} = \frac{2}{9},$$

$$M(X_3) = \frac{2}{3}; \quad D^2(X_2) = \frac{2}{9}; \qquad \mathrm{cov}\,(X_1,X_2) = \frac{2}{3}\cdot\frac{7}{11} - \frac{4}{9} = \frac{2}{99};$$

$$\mathrm{cov}\,(X_2,X_3) = \frac{2}{99}, \quad \rho(X_1, X_2) = \frac{1}{11}; \quad \rho(X_2,X_3) = \frac{1}{11}.$$

III.2. *O urnă conţine 12 bile, dintre care 8 albe şi 4 negre. Se face o extracţie din această urnă, se notează culoarea bilei după care se pune bila la loc în urnă. Totodată se pun în urnă încă două bile de aceeaşi culoare cu bila extrasă. Procesul de extracţie se repetă pînă ce s-au efectuat 5 extracţii. Dacă notăm cu* X_j, $j = 1, 2, ..., 5$, *variabila aleatoare definită mai jos*

$$X_j = \begin{cases} 1 \text{ dacă la extracţia de rang } j \text{ a apărut o bilă albă,} \\ 0 \text{ dacă la extracţia de rang } j \text{ a apărut o bilă neagră.} \end{cases}$$

Să se calculeze

$$M(X_3),\ D^2(X_3),\ \operatorname{cov}(X_2, X_3),\ \rho(X_2, X_3).$$

Soluţie. Variabilele aleatoare X_2 şi X_3 au repartiţiile

$$X_2: \begin{pmatrix} 1 & 0 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix} \quad X_3: \begin{pmatrix} 1 & 0 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

Deci

$$M(X_3) = \frac{2}{3}, \quad D^2(X_3) = \frac{2}{9},$$

$$\operatorname{cov}(X_2, X_3) = \frac{5}{14} - \frac{4}{9} = -\frac{11}{126},$$

$$\rho(X_2, X_3) = \frac{-\frac{11}{126}}{\frac{2}{9}} = -\frac{99}{252} = -\frac{11}{28}.$$

III.3. *Într-un experiment se urmăreşte realizarea sau nerealizarea evenimentelor A, B. Se fac n observaţii independente şi se notează cu* α_n, *respectiv* β_n *numărul de apariţii ale evenimentului A, respectiv B.*

Să se arate că coeficientul de corelaţie al variabilelor aleatoare α_n *şi* β_n: $\rho(\alpha_n, \beta_n)$ *este independent de n.*

Soluţie. Introducem variabilele aleatoare:

$$X_i = \begin{cases} 1 \text{ dacă se realizează evenimentul } A \text{ în experienţa de rang } i, \\ 0 \text{ dacă nu se realizează evenimentul } A \text{ în experienţa de rang } i. \end{cases}$$

$$Y_i = \begin{cases} 1 \text{ dacă se realizează evenimentul } B \text{ în experienţa de rang } i, \\ 0 \text{ dacă nu se realizează evenimentul } B \text{ în experienţa de rang } i. \end{cases}$$

Atunci se constată că

$$\alpha_n = \sum_1^n X_i \quad \text{şi} \quad \beta_n = \sum_1^n Y_i.$$

Cum

$$M(X_i) = P(A) \quad \text{şi} \quad M(Y_i) = P(B)$$

rezultă că

$$M(\alpha_n) = nM(X_i) = nP(A),$$

$$M(\beta_n) = nM(Y_i) = nP(B).$$

Mai departe

$$M(\alpha_n\beta_n) = M\left(\sum_1^n X_iY_i + \sum_{i=1}^n \sum_{k\neq i} X_iY_k\right) = \sum_1^n M(X_iY_i) +$$

$$\sum_{i=1}^n \sum_{k\neq i} M(X_i)\ M(Y_k) = nP(A\cap B) + n(n-1)\ P(A)\ P(\mathrm{B})$$

deoarece observaţiile rezultate în experienţele de ordin i şi k $(i \neq k)$ sînt independente, de unde rezultă că

$$\operatorname{cov}(\alpha_n, \beta_n) = M(\alpha_n\beta_n) - M(\alpha_n)\ M(\beta_n) =$$

$$= n[P(A\cap B) - P(A)\ P(B)].$$

Cum

$$D^2(X_i) = P(A)\ P(A^c), \quad i = 1, 2, \ldots, n,$$

$$D^2(Y_i) = P(B)\ P(B^c), \quad i = 1, 2, \ldots, n$$

şi observaţiile în experienţa de ranguri diferite sînt independente atunci

$$D^2(\alpha_n) = \sum_1^n D^2(X_i) = nP(A)\ P(A^c),$$

$$D^2(\beta_n) = \sum_1^n D^2(Y_i) = nP(B)\ P(B^c)$$

şi de aici rezultă acum că

$$\rho(\alpha_n, \beta_n) = \frac{\operatorname{cov}(\alpha_n, \beta_n)}{D(\alpha_n)\, D(\beta_n)} = \frac{n[P(A\cap B) - P(A)\, P(B)]}{\sqrt{nP(A)\, P(A^c)\, nP(B)P(B^c)}} =$$

$$= \frac{P(A \cap B) - P(A)\, P(B)}{\sqrt{P(A)\ \ P(A^c)\ \ P(B)\ \ P(B^c)}}.$$

Observaţie. Dacă ţinem seama de relaţiile:

$$P(A) = P(A \cap B) + P(A \cap B^c),$$

$$P(A^c) = P(A^c \cap B) + P(A^c \cap B^c),$$

$$P(B) = P(A \cap B) + P(A^c \cap B),$$

$$P(B^c) = P(A \cap B^c) + P(A^c \cap B^c),$$

$$P(A \cap B) + P(A \cap B^c) + P(A^c \cap B) + P(A^c \cap B^c) = 1,$$

atunci se constată că

$$P(A)\, P(B) = [P(A \cap B) + P(A \cap B^c)]\, [P(A \cap B) + P(A^c \cap B)]$$

şi

$$P(A \cap B) - P(A)\, P(B) = P(A \cap B)\, [1 - P(A \cap B) - P(A \cap B^c) -$$

$$- P(A^c \cap B)] - P(A \cap B^c)\, P(A^c \cap B) =$$

$$= P(A \cap B)\, P(A^c \cap B^c) - P(A \cap B^c)\, P(A^c \cap B)$$

şi de aici

$$\rho(\alpha_n, \beta_n) = \frac{P(A \cap B)\, P(A^c \cap B^c) - P(A \cap B^c)\, P(A^c \cap B)}{\sqrt{P(A)\, P(A^c)\, P(B)\, P(B^c)}}.$$

III. 4. *Dacă $X_1, X_2, \ldots, X_n$ sînt n variabile aleatoare independente totale avînd aceeaşi dispersie σ^2, să se calculeze*

$$\rho(U, V), \quad \rho(V, V - U) \quad \text{şi} \quad \rho V(, U_1 - U_2),$$

unde

$$V = \frac{1}{n}\sum_{1}^{n} X_i, \quad U = \frac{1}{m}\sum_{1}^{m} X_i, \quad m < n, \quad U_1 = \frac{1}{m_1}\sum_{1}^{m_1} X_i,$$

$$U_2 = \frac{1}{m_2}\sum_{1}^{m_2} X_i. \quad m_1, m_2 < n,$$

S o l u ţ i e. Deoarece

$$D^2(V) = \frac{\sigma^2}{n}, \quad D^2(U) = \frac{\sigma^2}{m}$$

şi

$$\operatorname{cov}(U, V) = m\sigma^2/mn = \sigma^2/n \tag{III.1}$$

urmează că

$$\rho(U, V) = \frac{\sigma^2/n}{\sqrt{\frac{\sigma^2}{n}\cdot\frac{\sigma^2}{m}}} = \sqrt{\frac{m}{n}}, \tag{III.2}$$

ceea ce arată că coeficientul de corelaţie între media tuturor variabilelor şi media unei părţi a acestora este egală cu rădăcina pătrată a numărului de variabile din partea acesteia şi numărul tuturor variabilelor.

Din proprietăţile corelaţiei şi din (III.1)

$$\operatorname{cov}(V, V - U) = D^2(V) - \operatorname{cov}(U, V) = 0,$$

de unde

$$\rho(V, V - U) = 0,$$

ceea ce arată că coeficientul de corelaţie dintre media tuturor variabilelor şi diferenţa dintre media tuturor variabilelor şi a unei părţi a acestora este zero.

Analog, se arată că

$$\rho(V, U_1 - U_2) = 0.$$

III.5. *Se consideră vectorul aleator (X_1, X_2), despre ale cărui componente presupunem că există $M(X_1^2)$, $M(X_2^2)$, $M(X_1X_2)$.*

Să se arate în ce condiţii există variabilele aleatoare Y_1 şi Y_2 combinaţii liniare de X_1, X_2 şi necorelate.]

S o l u ţ i e. Va trebui să vedem în ce condiţii există a_{11}, a_{12}, a_{21}, a_{22}, astfel încît

$$Y_1 = a_{11}X_1 + a_{12}X_2,$$

$$Y_2 = a_{21}X_1 + a_{22}X_2$$

şi $\operatorname{cov}(Y_1, Y_2) = 0$.

Din egalităţile date avem

$$M(Y_1) = a_{11}M(X_1) + a_{12}M(X_2),$$

$$M(Y_2) = a_{21}M(X_1) + a_{22}M(X_2)$$

şi

$$\operatorname{cov}(Y_1, Y_2) = a_{11}a_{21}D^2(X_1) + (a_{11}a_{22} + a_{12}a_{21}) \operatorname{cov}(X_1, X_2) + \\ + a_{12}a_{22}D^2(X_2)$$

sau

$$a_{11}a_{21}D^2(X_1) + (a_{11}a_{22} + a_{12}a_{21}) \rho(X_1, X_2) D(X_1) D(X_2) + \\ + a_{12}a_{22}D^2(X_2) = 0.$$

De aici rezultă că

$$\begin{vmatrix} a_{11}a_{21} & \frac{1}{2}(a_{11}a_{22} + a_{12}a_{21}) \rho(X_1, X_2) \\ \frac{1}{2}(a_{11}a_{22} + a_{12}a_{21}) & \rho(X_1, X_2) a_{12}a_{22} \end{vmatrix} = 0.$$

II.6. *Se consideră două evenimente A şi B pentru care se cunosc probabilităţile* $P(A) = \frac{1}{4}$, $P(B \mid A) = \frac{1}{2}$, $P(A \mid B) = \frac{1}{4}$.

Definim variabilele aleatoare X şi Y după cum urmează.

$$X = \begin{cases} 1 \text{ dacă a apărut evenimentul } A, \\ 0 \text{ dacă a apărut evenimentul } A^c, \end{cases}$$

$$Y = \begin{cases} 1 \text{ dacă a apărut evenimentul } B, \\ 0 \text{ dacă a apărut evenimentul } B^c. \end{cases}$$

Să se calculeze coeficientul de corelaţie $\rho(X, Y)$.

S o l u ţ i e. Deoarece $P(A \cap B) = P(B \mid A)P(A) = \frac{1}{8}$,

$$P(B) = \frac{P(A \cap B)}{P(A \mid B)} = \frac{\frac{1}{8}}{\frac{1}{4}} = \frac{1}{2},$$

$$P(A^c \cap B) = P(A^c \mid B)\, P(B) = \frac{3}{4} \cdot \frac{1}{2},$$

şi

$$P(A \cap B^c) = P(A) - P(A \cap B) = \frac{1}{8},$$

urmează că vectorul aleator (X,Y) are repartiţia

$Y \backslash X$	1	0
1	$\frac{1}{8}$	$\frac{3}{8}$
0	$\frac{1}{8}$	$\frac{3}{8}$

De aici rezultă că

$$M(X) = \frac{1}{4}, \quad M(Y) = \frac{1}{2},$$

$$D^2(X) = \frac{3}{16}; \quad D^2(Y) = \frac{1}{4};$$

$$M(XY) = \frac{1}{8}, \quad \text{cov}\,(X, Y) = \frac{1}{8} - \frac{1}{8} = 0.$$

Deci $\rho(X, Y) = 0$, ceea ce înseamnă că variabilele aleatoare sînt necorelate.

III.7. *Se consideră două monede perfecte, fiecare dintre ele avînd imprimat pe o faţă numărul 1 iar pe cealaltă numărul 2 şi se aruncă independent cu aceste monede.*

Se introduce variabila aleatoare X, egală cu suma numerelor obţinute şi variabila aleatoare Y maximul dintre aceste numere (dacă numerele sînt egale se ia valoarea comună a lor).

Să se afle coeficientul de corelaţie al acestor variabile aleatoare.

Soluţie. Obţinem pentru vectorul aleator următorul tabel de valori indicate într-o căsuţă, primul număr fiind suma iar al doilea maximul

$Y \backslash X$	1	2
1	(2,1)	(3,2)
2	(3,2)	(4,2)

De aici urmează

$$M(X)=\frac{12}{4}=3;\quad M(Y)=\frac{7}{4};\quad M(X^2)=\frac{38}{4};$$

$$M(Y^2)=\frac{13}{4};\quad D^2(X)=\frac{1}{2};\quad D^2(Y)=\frac{3}{16};$$

$$M(XY)=\frac{22}{4};\quad \operatorname{cov}(X, Y)=\frac{22}{4}-3\cdot\frac{7}{4}=\frac{1}{4}$$

şi deci

$$\rho(X, Y)=\sqrt{\frac{2}{3}}.$$

III.8. *Se aruncă două zaruri (fiecare cu feţele numerotate de la* 1 la 6) *perfecte şi se notează cu X suma punctelor obţinute şi cu Y numărul maxim de puncte (dacă numerele sînt egale se ia valoarea comună).*

Să se afle coeficientul de corelaţie al variabilelor X şi Y, $\rho(X, Y)$.

Soluţie. Considerăm vectorul aleator (X, Y) cu valorile componentelor indicate în tabelul III.1.

Tabelul III.1

	1	2	3	4	5	6
1	(2,1)	(3,2)	(4,3)	(5,4)	(6,5)	(7,6)
2	(3,2)	(4,2)	(5,3)	(6,4)	(7,5)	(8,6)
3	(4,3)	(5,3)	(6,3)	(7,4)	(8,5)	(9,6)
4	(5,4)	(6,4)	(7,4)	(8,4)	(9,5)	(10,6)
5	(6,5)	(7,5)	(8,5)	(9,5)	(10,5)	(11,6)
6	(7,6)	(8,6)	(9,6)	(10,6)	(11,6)	(12,6)

Fiecare dintre combinaţii apare cu probabilitatea 1/36.
Deci

$$M(X)=\frac{252}{36}=7;\quad M(Y)=\frac{161}{36};\quad M(X^2)=\frac{1990}{36};$$

$$M(Y^2)=\frac{791}{36};\quad D^2(X)=\frac{226}{36};\quad D^2(Y)=\frac{2555}{36^2};$$

$$M(XY)=\frac{1132}{36},\quad \operatorname{cov}(X, Y)=\frac{5}{36};$$

$$\rho(X, Y)=\frac{\frac{5}{36}}{\frac{1}{36}\sqrt{\frac{226}{36}\cdot 2555}}=\sqrt{\frac{90}{57\,743}}.$$

III.9. *Se dau variabilele aleatoare* $X = U + V$ *şi* $Y = U + W$ *unde* U, V, W *sînt variabilele aleatoare independente.*
Să se arate că

$$\rho(X, Y) = \left\{\left(1 + \frac{D^2(V)}{D^2(U)}\right)\left(1 + \frac{D^2(W)}{D^2(U)}\right)\right\}^{-\frac{1}{2}}.$$

Soluţie. Deoarece

$$\operatorname{cov}(X, Y) = M((U + V)(U + W)) - M(U + V)\, M(U + W) =$$

$$= M(U^2) + M(U)\, M(W) + M(U)\, M(V) + M(V)\, M(W) - M^2(U) -$$

$$- M(U)\, M(W) - M(V)\, M(U) - M(V)\, M(W) =$$

$$= M(U^2) - M^2(U) = D^2(U),$$

$$D^2(X) = D^2(U + V) = D^2(U) + D^2(V),$$

şi

$$D^2(Y) = D^2(U + W) = D^2(U) + D^2(W),$$

U, V, W fiind variabile aleatoare independente, rezultă

$$\rho(X, Y = \left\{\left(1 + \frac{D^2(V)}{D^2(U)}\right)\left(1 + \frac{D^2(W)}{D^2(U)}\right)\right\}^{-\frac{1}{2}}.$$

III.10. *Se consideră variabilele aleatoare independente* X_i, $i = 1, 2$ *repartizate normal,* $N(\mu_i, \sigma_i^2)$, $i = 1, 2$ *şi se construiesc variabilele aleatoare* Y_i, $i = 1, 2$ *astfel:*

$$Y_1 = X_1 \cos\alpha + X_2 \sin\alpha,$$

$$Y_2 = X_2 \cos\alpha - X_1 \sin\alpha.$$

Să se calculeze pătratul coeficientului de corelaţie $[\rho^2(Y_1, Y_2)]$ *şi să se stabilească limite pentru pătratul acestui coeficient de corelaţie.*

S o l u ț i e. Avem

$$\operatorname{cov}(Y_1, Y_2) = M[(X_1\cos\alpha + X_2\sin\alpha)(X_2\cos\alpha - X_1\sin\alpha)] -$$
$$- M(X_1\cos\alpha + X_2\sin\alpha)\, M(X_2\cos\alpha - X_1\sin\alpha) =$$
$$= (\cos^2\alpha - \sin^2\alpha)\, M(X_1X_2) - \sin\alpha\cos\alpha[M(X_1^2) - M(X_2^2)] -$$
$$- (\mu_1\cos\alpha + \mu_2\sin\alpha)(\mu_2\cos\alpha - \mu_1\sin\alpha) =$$
$$= (\cos^2\alpha - \sin^2\alpha)\,\mu_1\mu_2 - \sin\alpha\cos\alpha[\sigma_1^2 + \mu_1^2 - \sigma_2^2 - \mu_2^2] -$$
$$- (\mu_1\cos\alpha + \mu_2\sin\alpha)(\mu_2\cos\alpha - \mu_1\sin\alpha) =$$
$$= -\sin\alpha\cos\alpha\,(\sigma_1^2 - \sigma_2^2).$$

Mai departe

$$D^2(Y_1) = M[(X_1\cos\alpha + X_2\sin\alpha)^2] - [M(X_1\cos\alpha + X_2\sin\alpha)]^2 =$$
$$= \cos^2\alpha(\sigma_1^2 + \mu_1^2) + 2\sin\alpha\cos\alpha\,\mu_1\mu_2 + \sin^2\alpha\,(\sigma_2^2 + \mu_2^2) -$$
$$- \mu_1^2\cos^2\alpha - 2\mu_1\mu_2\sin\alpha\cos\alpha - \mu_1^2\sin^2\alpha = \cos^2\alpha\;\sigma_1^2 + \sin^2\alpha\;\sigma_2^2.$$

Analog

$$D^2(Y_2) = M[X_2\cos\alpha - X_1\sin\alpha)^2] - [M(X_2\cos\alpha - X_1\sin\alpha)]^2 =$$
$$= \cos^2\alpha\,\sigma_2^2 + \sin^2\alpha\,\sigma_1^2.$$

De aici rezultă că

$$D^2(Y_1)\,D^2(Y_2) = (\cos^2\alpha\;\sigma_1^2 + \sin^2\alpha\;\sigma_2^2)(\cos^2\alpha\;\sigma_2^2 + \sin^2\alpha\;\sigma_1^2) =$$
$$= \sigma_1^2\sigma_2^2(\cos^4\alpha + \sin^4\alpha) + (\sigma_1^4 + \sigma_2^4)\sin^2\alpha\cos^2\alpha.$$

Deci

$$\rho^2(Y_1, Y_2) = \frac{(\sigma_1^2 - \sigma_2^2)^2\sin^2\alpha\cos^2\alpha}{\sigma_1^2\sigma_2^2(\sin^4\alpha + \cos^4\alpha) + (\sigma_1^4 + \sigma_2^4)\sin^2\alpha\cos^2\alpha} =$$

$$= \frac{(\sigma_1^2 - \sigma_2^2)^2\sin^2 2\alpha}{4\left(1 - \dfrac{\sin^2 2\alpha}{2}\right)\sigma_1^2\sigma_2^2 + (\sigma_1^4 + \sigma_2^4)\sin^2 2\alpha} = \frac{(\sigma_1^2 - \sigma_2^2)^2\sin^2 2\alpha}{4\sigma_1^2\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)^2\sin^2 2\alpha}$$

sau încă

$$\rho^2(Y_1, Y_2) = 1 - \frac{4\sigma_1^2\sigma_2^2}{4\sigma_1^2\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)^2\sin^2 2\alpha}.$$

Pentru a obține limitele valorilor coeficientului $\rho^2(Y_1, Y_2)$ procedăm în modul următor.

Pornim de la expresia

$$\rho^2(Y_1, Y_2) = \frac{(\sigma_1^2 - \sigma_2^2)^2 \sin^2 2\alpha}{4\sigma_1^2\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)^2 \sin^2 2\alpha}$$

și observăm că

$$\frac{(\sigma_1^2 - \sigma_2^2)^2 \sin^2 2\alpha}{4\sigma_1^2\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)^2 \sin^2 2\alpha} \leqslant \frac{(\sigma_1^2 - \sigma_2^2)^2 \sin^2 2\alpha}{4\sigma_1^2\sigma_2^2 \sin^2 2\alpha + (\sigma_1^2 - \sigma_2^2) \sin^2 2\alpha} =$$

$$= \left(\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)^2.$$

Deci

$$0 \leqslant \rho^2(Y_1, Y_2) \leqslant \left(\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)^2.$$

III.11. *Fie X și Y două variabile aleatoare identic repartizate,*

$$U = X - Y$$

și

$$V = X + Y.$$

Să se arate că

a) cov $(U, V) = 0$.

b) *Există variabile aleatoare X, Y independente și totuși U* și *V dependente.*

c) *Există variabile aleatoare X, Y dependente, identic repartizate pentru care coeficientul de corelație al variabilelor U și V*: $\rho(U, V) = 0$ *și totodată variabilele U și V sînt independente.*

S o l u ț i e. a) Deoarece X și Y sînt identic repartizate, rezultă că

$$M(U) = M(X) - M(Y) = 0$$

și de aici urmează că

$$\text{cov}\,(U, V) = M((X - Y)(X + Y)) = M(X^2) - M(Y^2) = 0,$$

adică variabilele U și V sînt necorelate.

b) Considerăm două zaruri omogene care se aruncă și notăm X numărul de puncte ce apare pe primul zar, iar cu Y numărul de puncte ce apar pe al doilea zar. Deoarece valorile variabilei V depind de valorile variabilei U, așa de exemplu

$$P(V = 4 \mid U = 3) = 0,$$

$$P(V = 4) = \frac{3}{36},$$

rezultă că variabilele aleatoare U şi V sînt dependente cu toate că X şi Y sînt independente iar $\rho(U, V) = 0$.

c) Considerăm variabilele aleatoare X şi Y, fiecare repartizate normal cu media zero şi dispersia unu şi cu coeficienţii de corelaţie ρ, adică au densitatea de repartiţie comună

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)}.$$

Atunci

$$F_{U,V}(u, v) = \iint\limits_{x-y<u,\ x+y<v} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)} dxdy.$$

Ţinînd seama că

$$f_{U,V}(u, v) = \frac{\partial^2 F(u, v)}{\partial u \partial v}$$

obţinem

$$f_{U,V}(u, v) = \frac{1}{\sqrt{2\pi}\sqrt{2(1-\rho)}} e^{-\frac{u^2}{4(1-\rho)}} \frac{1}{\sqrt{2\pi}\sqrt{2(1+\rho)}} e^{-\frac{v^2}{4(1+\rho)}},$$

ceea ce probează faptul că variabilele aleatoare U şi V sînt repartizate normal cu media zero şi cu dispersia $2(1 - \rho)$ şi $2(1 + \rho)$ respective.

III.12. *Să se arate că dacă X şi Y au fiecare o repartiţie Bernoulli atunci $\rho(X, Y) = 0$, dacă şi numai dacă variabilele X şi Y sînt independente.*

S o l u ţ i e. Fie variabilele X şi Y cu repartiţiile indicate mai jos:

$$X: \begin{pmatrix} x_1 & x_2 \\ p_1 & 1 - p_1 \end{pmatrix} \quad (x_1 \neq x_2)$$

$$Y: \begin{pmatrix} y_1 & y_2 \\ p_2 & 1 - p_2 \end{pmatrix} \quad (y_1 \neq y_2)$$

şi să notăm

$$p_{ij} = P(X = x_i,\ Y = y_j) \qquad i = 1, 2,\ j = 1, 2.$$

Putem alege $x_1 = 0$ fără a restrînge cu nimic generalitatea problemei.

Atunci $\text{cov}\,(X, Y) = M(X\,Y) - M(X)\,M(Y)$ ne spune că dacă X şi Y sînt independente avem

$$\rho(X, Y) = 0.$$

Reciproc să presupunem că $\rho(X, Y) = 0$.

Rezultă de aici că

$$\text{cov}\,(X, Y) = p_{11}x_1y_1 + p_{12}x_1y_2 + p_{21}x_2y_1 + p_{22}x_2y_2 -$$
$$- [x_1p_1 + x_2(1 - p_1)]\,[y_1p_2 + y_2(1 - p_2)] = 0.$$

Cum am ales $x_1 = 0$, obţinem

$$x_2y_1p_{21} + x_2y_2p_{22} - x_2y_1(1 - p_1)\,p_2 - x_2y_2(1 - p_1)\,(1 - p_2) = 0,$$

sau

$$y_1[p_{21} - (1 - p_1)\,p_2] + y_2[p_{22} - (1 - p_1)\,(1 - p_2)] = 0.$$

Însă

$$p_{21} = 1 - p_1 - p_{22}$$

şi deci

$$y_1[1 - p_1 - p_{22} - (1 - p_1)\,p_2] + y_2[p_{22} - (1 - p_1)\,(1 - p_2)] = 0$$

adică

$$(y_2 - y_1)\,[p_{22} - (1 - p_1)\,(1 - p_2)] = 0.$$

Cum $y_2 \neq y_1$ rezultă

$$p_{22} = (1 - p_1)\,(1 - p_2),$$

adică

$$P(X = x_2, \quad Y = y_2) = P(X_2)\,P(Y = y_2)$$
$$p_{21} = 1 - p_1 - p_{22} = 1 - p_1(1 - p_1)\,(1 - p_2) =$$
$$= (1 - p_1)\,[1 - 1 + p_2] = (1 - p_1)\,p_2$$

deci

$$P(X = x_2, \quad Y = y_1) = P(X = x_2)\,P(Y = y_1).$$

Mai departe din

$$p_2 = p_{11} + p_{21}$$

rezultă

$$p_{11} = p_2 - p_{21} = p_2 - (1 - p_1)\,p_2 = p_2(1 - 1 + p_1) = p_1p_2$$

şi la fel

$$p_{12} = p_1(1 - p_2)$$

ceea ce dovedeşte afirmaţia făcută.

III.13. *Fie X_1, X_2, X_3 şi X_4 patru variabile aleatoare şi*

$$Z_1 = \frac{X_1}{X_3}, \quad Z_2 = \frac{X_2}{X_3} \quad \text{şi} \quad U = \frac{X_2}{X_4}.$$

Să se arate că

$$r(Z_1, Z_2) = \frac{v_3^2 + r_{12}v_1v_2 - r_{13}v_1v_3 - r_{23}v_2v_3 + P}{[v_1^2 - 2r_{13}v_1v_3 + v_3^2]^{1/2}[v_2^2 - 2r_{23}v_2v_3 + v_3^2]^{1/2}},$$

$$r(Z_1, U) = \frac{r_{12}v_1v_2 - r_{23}v_2v_3 - r_{14}v_1v_4 + r_{34}v_3v_4}{[v_1^2 - 2r_{13}v_1v_3 + v_3^2]^{1/2}\,[v_2^2 - 2r_{24}v_2v_4 + v_4^2]^{1/2}},$$

unde: r_{ij} este coeficientul de corelaţie dintre X_i şi X_j.
$v_i\left[= \dfrac{D(X_i)}{M(X)}\right]$ *este coeficientul de variaţie al variabilei X_i*

$$P = r_{23}v_2v_3^3 + r_{13}v_1v_3^3 - r_{13}r_{23}v_1v_2v_3^3 - v_3^4.$$

III.14. *Fie $(X_1, X_2, ..., X_{k-1})$ un vector aleator caracterizat prin funcţia de frecvenţă*

$$P(X_1 = x_1, \quad X_2 = x_2, ..., X_{k-1} = x_{k-1}) = \prod_1^k p_i^{x_i} \quad \textit{pentru} \quad x_i = 0{,}1$$

$$(i = 1, ..., k), \quad \sum_1^k x_i = 1, \quad \sum_1^k p_i = 1$$

şi unde

$$x_k = 1 - \sum_1^{k-1} x_i, \quad p_k = 1 - \sum_1^{k-1} p_i.$$

Fie

$$(X_{11}, ..., X_{k-1,\,1}), (X_{12}, ..., X_{k-1,\,2}), ..., (X_{1n}, ..., X_{k-1,\,n})$$

o selecţie de volum n din această populaţie şi $Y_i = \sum\limits_{m=1}^{n} X_{im}$.

Să se determine funcţia de frecvenţă a covarianţei de selecţie S_{12} dintre X_1 şi X_2 definiţie după cum urmează

$$S_{12} = \frac{1}{n-1}\sum_{m=1}^{n}(X_{1m} - \bar{X}_1)(X_{2m} - \bar{X}_2) = -\frac{1}{n(n-1)}Y_1Y_2 .$$

unde

$$\bar{X}_i = Y_i/n, \quad (i = 1, 2).$$

Să se studieze cazul $n = 3$.

Soluţie. S_{12} este o estimaţie nedeplasată pentru

$$\operatorname{cov}(X_1, X_2) = M[(X_1 - p_1)(X_2 - p_2)] = -p_1p_2$$

şi valorile sale sînt de forma

$s_{12} = -\dfrac{ij}{n(n-1)}$ pentru perechile (i, j) cu $i = 1, 2, ..., n$ şi $j = 0, 1, ...,$ $..., \min(i, n-i), (i + j < n)$.

Dacă punem $r = ij$, atunci $s_{12} = -r/n(n-1)$ pentru $r = 0, 1, ...$ $..., \left[\dfrac{n^2}{4}\right]$ unde $[x]$ reprezintă partea întreagă a lui x.

Fie $\varnothing_r$ mulţimea tuturor valorilor r cu i şi j distincte, $j \leqslant i$. Urmează că funcţia de frecvenţă a statisticii S_{12} poate fi scrisă ca

$$P(S_{12} = s_{12}) = \Sigma_{\varnothing_r} Q(i, j),$$

unde pentru $j < i$ (adică pentru $i = 1, 2, ..., n$ şi $j = 0, 1, ..., \min(i - 1, n - i)$)

$$Q(i, j) = P(Y_1 = i, Y_2 = j \text{ sau } Y_1 = j, Y_2 = i) = \frac{n!p_1^j p_2^j(1 - p_1 - p_2)^{n-i-j}(p_1^{i-j} + p_2^{i-j})}{i!j!(n-i-j)!} \tag{III.3}$$

şi pentru $j = i$ $(= 0, 1, ..., [n/2])$

$$Q(i, j) = P(Y_1 = Y_2 = i) = \frac{n!p_1^i p_2^i(1 - p_1 - p_2)^{n-2i}}{(i!)^2(n-2i)!}.$$

Numărul perechilor (i, j) cu $i = 0, 1, ..., n$ şi $j = 0, 1, ..., \min(i, n - i)$ este

$$\sum_{i=0}^{n}\min(i + 1, n - i + 1) = \left[\frac{(n+2)^2}{4}\right]. \tag{III.4}$$

Pentru $n = 3$ vectorul (Y_1, Y_2) poate lua valorile $(0, 0)$, $(1, 0)$, $(1, 1)$, $(2, 0)$, $(2, 1)$, $(3, 0)$ cu probabilităţile (date de III.3) şi (III.4)) $(1 - p_1 - p_2)^3$, $3(1 - p_1 - p_2)^2 (p_1 + p_2)$, $6p_1p_2(1 - p_1 - p_2)$, $3(1 - p_1 - p_2)(p_1^2 + p_2^2)$, $3p_1p_2(p_1 + p_2)$ şi respectiv $(p_1^3 + p_2^3)$.

Astfel S_{12} poate lua valorile 0, $-\frac{1}{6}$ şi $-\frac{1}{3}$ cu probabilităţile $1 - 6p_1p_2 + 3p_1^2p_2 + 3p_1p_2^2$, $6p_1p_2(1 - p_1 - p_2)$ şi respectiv $3p_1p_2(p_1 + p_2)$.

III.15. *Fie (x_i, y_i) $i = 1, 2, \ldots, n$ o selecţie de volum n dintr-o populaţie în care caracteristica sub cercetare este un vector aleator (X, Y) normal bidimensional, pentru care*

$$M(X) = \mu_x, \quad M(Y) = \mu_y, \quad D^2(X) = D^2(Y) = \sigma^2$$

şi

$$cov\,(X, Y) = \sigma^2\rho.$$

Să se determine densitatea de repartiţie a statisticii

$$R = \frac{2S_{xy}}{S_x^2 + S_y^2},$$

unde am folosit notaţiile

$$(n - 1)\,S_{xy} = \sum_1^n (x_i - \bar{x}_n)(y_i - \bar{y}_n),$$

$$(n - 1)\,S_x^2 = \sum_1^n (x_i - \bar{x}_n)^2, \quad (n - 1)\,S_y^2 = \sum_1^n (y_i - \bar{y}_n)^2.$$

Soluţie. Considerăm transformarea

$$\begin{pmatrix} U \\ V \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}.$$

Deoarece vectorul aleator $(X, Y)'$ este repartizat normal bidimensional de matrice de covarianţe

$$\sigma^2\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

urmează că vectorul aleator $(U, V)'$ este de asemenea repartizat normal cu matricea de covarianţe

$$\frac{1}{2}\,\sigma^2\begin{pmatrix} 1 + \rho & 0 \\ 0 & 1 - \rho \end{pmatrix}.$$

Prin urmare variabilele aleatoare U şi V sînt independente Deoarece

$$X = U + V \quad \text{şi} \quad Y = U - V,$$

avem

$$S_{xy} = S_u^2 - S_v^2, \quad S_x^2 = S_u^2 + S_v^2 + 2S_{uv}$$

şi

$$S_y^2 = S_u^2 + S_v^2 - 2S_{uv}$$

Obţinem

$$R = \frac{S_u^2 - S_v^2}{S_u^2 + S_v^2} = \frac{1 - aF}{1 + aF},$$

unde $a = \dfrac{1-\rho}{1+\rho}$, iar $F = \dfrac{1+\rho}{1-\rho} \cdot \dfrac{S_v^2}{S_u^2}$ este repartizată $F(n-1,\ n-1)$.

Densitatea de repartiţie a variabilei F este

$$f(F) = \frac{\Gamma(n-1)}{\Gamma^2\left(\frac{n-1}{2}\right)} F^{\frac{n-3}{2}}(1+F)^{-(n-1)}, \quad 0 < F < \infty.$$

Jacobianul transformării de la F la R este $\dfrac{2}{a(1+R^2)}$.

Prin urmare densitatea de repartiţie a lui R este

$$h(R) = \frac{\Gamma(n-1)}{\Gamma^2\left(\frac{n-1}{2}\right)} \left(\frac{1}{a} \cdot \frac{1-R}{1+R}\right)^{\frac{n-3}{2}} \times$$

$$\times \left(1 + \frac{1-R}{1+R} \cdot \frac{1}{a}\right)^{-(n-1)} \frac{1}{a} \cdot \frac{2}{(1+R)^2}, \quad -1 < R < 1.$$

Exprimînd pe a în funcţie de ρ, obţinem

$$h(R) = \frac{\Gamma(n-1)}{\Gamma^2\left(\frac{n-1}{2}\right)} \cdot \frac{(1-\rho^2)^{\frac{n-1}{2}}}{2^{n-2}} \times$$

$$\times (1 - \rho R)^{-(n-1)} (1 - R^2)^{\frac{n-3}{2}}$$

Ţinînd seama că

$$\Gamma(n-1)=\frac{2^{n-2}}{\sqrt{\pi}}\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n}{2}\right),$$

obţinem

$$h(R)=\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)}(1-\rho^2)^{\frac{n-1}{2}}(1-\rho R)^{-(n-1)}(1-R^2)^{\frac{n-3}{2}},$$

$$-1<R<1.$$

III.16. *Fie (X, Y) un vector aleator care are o repartiţie uniformă într-un cerc de rază 1 şi centrul în punctul valoarea medie a vectorului aleator considerat.*

Să se arate că variabilele aleatoare X şi Y sînt necorelate şi cu toate acestea dependente.

Soluţie. Vectorul aleator (X, Y) fiind uniform repartizat în cercul de rază 1 şi cu centrul în punctul valoare medie a vectorului aleator dat, rezultă că

$$f(x, y)=\begin{cases}\dfrac{1}{\pi} & \text{dacă } 0\leqslant x^2+y^2\leqslant 1,\\ 0 & \text{în rest.}\end{cases}$$

Atunci

$$\mu_{11}=\iint xyf(x, y)\,\mathrm{d}x\,\mathrm{d}y=\frac{1}{\pi}\iint xy\,\mathrm{d}x\,\mathrm{d}y=$$

$$=\frac{1}{\pi}\int_0^1 x\left[\int_{-(1-x^2)^{1/2}}^{(1-x^2)^{1/2}}\mathrm{d}y\right]\mathrm{d}x=\frac{1}{\pi}\int_0^1 x\,\frac{y^2}{2}\bigg|_{-(1-x^2)^{1/2}}^{(1-x^2)^{1/2}}\mathrm{d}x=0.$$

Deci variabilele aleatoare X şi Y sînt necorelate.

Pe de altă parte se constată imediat că variaţia mărimii fiecăreia dintre cele două variabile, depinde de valoarea celeilalte, ceea ce înseamnă că variabilele sînt dependente.

III.17. *Se consideră o selecţie de volum n, $x_1, x_2, \ldots, x_n$ dintr-o populaţie oarecare şi un sistem de ponderi $w_1, w_2, \ldots, w_n$.*

Să notăm cu M_w media ponderată a variabilelor de selecţie $x_1, x_2, \ldots, \ldots, x_n$, cu $M=\frac{1}{n}\sum_1^n x_i$ *cu σ_x, σ_w respectiv abaterea medie pătratică corespunzător selecţiei $x_1, x_2, \ldots, x_n$ şi ponderilor $w_1, w_2, \ldots, w_n$ şi cu r coeficientul de corelaţie al variabilelor X şi w.*

Să se arate că

$$M_w = M + \frac{r\sigma_x\sigma_w}{\bar{w}}.$$

S o l u ţ i e. Din definiţia mediei ponderate, rezultă

$$M_w = \frac{\sum_1^n w_i x_i}{\sum_1^n w_i}.$$

Această expresie mai poate fi scrisă

$$M_w = \frac{1}{n\bar{w}}\sum_1^n w_i x_i = \frac{1}{n\bar{w}}\sum_1^n (w_i - \bar{w} + \bar{w})(x_i - M + M) =$$

$$= \frac{1}{n\bar{w}}\left\{\sum_1^n (w_i - \bar{w})(x_i - M) + \bar{w}\sum_1^n (x_i - M) + \right.$$

$$\left. + M\sum_1^n (w_i - \bar{w}) + n\bar{w}M\right\} =$$

$$= M + \frac{1}{\bar{w}}\left[\frac{1}{n}\sum_1^n (w_i - \bar{w})(x_i - M)\right].$$

Cum însă

$$\operatorname{cov}(X, W) = \frac{1}{n}\sum_1^n (w_i - \bar{w})(x_i - M) = r\sigma_x\sigma_w$$

rezultă că

$$M_w = M + \frac{r\sigma_x\sigma_w}{\bar{w}}.$$

III.18. *Se consideră variabilele aleatoare de medie zero $X_j, 1 \leqslant j \leqslant n$ a căror matrice de covarianţă este*

$$R = (r_{jl}) = (\operatorname{cov}(X_j, X_l))$$

şi variabilele aleatoare Y_i, $1 \leqslant i \leqslant m$ definite cu ajutorul variabilelor aleatoare X_j prin relaţiile

$$Y_i = \sum_{j=1}^n a_{ij} X_j,$$

Să se arate că dacă

$$P = (p_{ik}) = (\mathrm{cov}\,(Y_i, Y_k))$$

şi

$$A = (a_{ij}), \quad A' = (a_{ji})$$

atunci

$$P = ARA'$$

S o l u ţ i e. Cum $M(X_j) = 0$, $1 \leqslant j \leqslant n$ rezultă că $M(Y_i) = 0, 1 \leqslant \leqslant i \leqslant n$ şi deci

$$\mathrm{cov}\,(X_j, X_l) = M(X_jX_l) = r_{jl} \quad \text{şi} \quad \mathrm{cov}\,(Y_i, Y_k) = M(Y_iY_k) = p_{ik}$$

$$M(Y_iY_k) = M\left[\left(\sum_{j=1}^{n} a_{ij}X_j\right)\left(\sum_{l=1}^{n} a_{kl}X_l\right)\right] =$$

$$= M\left(\sum_{j=1}^{n}\sum_{l=1}^{n} a_{ij}X_jX_la_{kl}\right) =$$

$$= \sum_{j=1}^{n}\sum_{l=1}^{n} a_{ij}M(X_jX_l)a_{kl} = \sum_{j=1}^{n}\sum_{l=1}^{n} a_{ij}r_{jl}a_{kl}.$$

Ori acesta nu este altceva decît elementul situat pe linia i şi coloana k din matricea

$$ARA'$$

şi deci

$$P = ARA'.$$

III.19. *Fie* X, Y *două variabile aleatoare cu* $M(X) = M(Y) = 0$,

$$D^2(X) = \sigma_x^2, \quad D^2(Y) = \sigma_y^2 \quad \text{şi} \quad \rho(X, Y) = r.$$

Se construiesc variabilele aleatoare

$$U = aX + bY,$$

$$V = bX - aY$$

şi necorelate

Să se arate că

$$\sigma_u\sigma_v = (a^2 + b^2)\,\sigma_x\sigma_y(1 - r^2)^{\frac{1}{2}}.$$

S o l u ţ i e. Din problema precedentă rezultă că

$$\begin{pmatrix} D^2(U) & \operatorname{cov}(U, V) \\ \operatorname{cov}(U, V) & D^2(V) \end{pmatrix} = \begin{pmatrix} a & b \\ b & -a \end{pmatrix} \begin{pmatrix} D^2(X) & \operatorname{cov}(X, Y) \\ \operatorname{cov}(X, Y) & D^2(Y) \end{pmatrix} \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$$

De aici dacă ţinem seama de condiţiile în care lucrăm, rezultă că putem scrie

$$\begin{vmatrix} \sigma_U^2 & 0 \\ 0 & \sigma_V^2 \end{vmatrix} = \begin{vmatrix} a & b \\ b & -a \end{vmatrix}^2 \begin{vmatrix} \sigma_X^2 & r\sigma_X\sigma_Y \\ r\sigma_X\sigma_Y & \sigma_Y^2 \end{vmatrix}$$

sau

$$\sigma_U^2\sigma_V^2 = (a^2 + b^2)^2 \sigma_X^2\sigma_Y^2(1 - r^2)$$

şi deci

$$\sigma_U\sigma_V = (a^2 + b^2)\, \sigma_X\sigma_Y(1 - r^2)^{1/2}.$$

III.20. *Se consideră vectorul aleator* (X, Y) *normal bidimensional cu* $M(X) = M(Y) = 0$, $D^2(X) = D^2(y) = 1$ şi $\rho(X, Y) = \rho$.

d) *Să se calculeze coeficientul de corelaţie al variabilelor aleatoare* X^2 *şi* Y^2.

b) *Să se calculeze regresiile* $M(X^2 | y^2)$ *şi* $M(Y^2, | x^2)$.

S o l u ţ i e. a) Vom utiliza funcţiile caracteristice pentru a calcula momentele care intervin în expresia coeficientului de corelaţie.

Din definiţia funcţiei caracteristice bidimensionale rezultă că

$$\varphi(t, u) = \iint e^{itx^2 + iuy^2} f(x, y)\, dx\, dy +$$

$$+ \frac{1}{2\pi(- \rho^2)^{1/2}} \iint e^{i tx^2 + iuy^2}\, e^{-\frac{1}{2(1-\rho^2)}[x^2 - 2\rho xy + y^2]}\, dx dy =$$

$$= \frac{1}{2\pi(1-\rho^2)^{1/2}} \iint e^{-\frac{1}{2(1-\rho^2)}\{x^2[1 - 2(1-\rho^2)\, it] - 2\rho xy + y^2[1 - 2(1-\rho^2)\, iu]\}}\, dx\, dy$$

Dacă facem schimbarea de variabile

$$x = (1 - \rho^2)^{1/2} \left\{ \frac{1}{[1 - 2(1 - \rho^2)\,\mathrm{i}\,t]^{1/2}} + \right.$$

$$\left. + \rho \frac{z_2}{\left[1 - 2(1 - \rho^2)\,\mathrm{i}\,u - \dfrac{\rho^2}{1 - 2\mathrm{i}\,t(1 - \rho^2)}\right]^{1/2}} \right\}$$

$$y = (1 - \rho^2)^{1/2} \frac{z_2}{\left[1 - 2(1 - \rho^2)\,\mathrm{i}\,u - \dfrac{\rho^2}{1 - 2(1 - \rho^2)\,\mathrm{i}\,t}\right]^{1/2}},$$

$$\begin{vmatrix} \dfrac{\partial x}{\partial z_1} & \dfrac{\partial x}{\partial z_2} \\ \dfrac{\partial y}{\partial z_1} & \dfrac{\partial y}{\partial z_2} \end{vmatrix} =$$

$$= (1 - \rho^2)\left[\{1 - 2(1 - \rho^2)\,\mathrm{i}\,t\}\,\{1 - 2(1 - \rho^2)\,\mathrm{i}\,u\} - \rho^2\right]^{-\frac{1}{2}},$$

atunci

$$\varphi(t, u) = \frac{1}{2\pi(1 - \rho^2)^{1/2}} \iint e^{-\frac{1}{2(1-\rho^2)}\{x^2[1-2(1-\rho^2)\,\mathrm{i}\,t] - 2\rho xy + y^2[1-2(1-\rho^2)\,\mathrm{i}\,u]\}}\, \mathrm{d}x\, \mathrm{d}y =$$

$$= \frac{(1 - \rho^2)\,[(1 - 2(1 - \rho^2)\,\mathrm{i}\,t)\,(1 - 2(1 - \rho^2)\,\mathrm{i}\,u) - \rho^2]^{-\frac{1}{2}}}{2\pi(1 - \rho^2)^{1/2}} \iint e^{-\frac{1}{2}(z_1^2 + z_2^2)}\, \mathrm{d}z_1\, \mathrm{d}z_2;$$

Deci

$$\varphi(t, u) = (1 - \rho^2)^{1/2}[(1 - 2(1 - \rho^2)\,\mathrm{i}\,t)\,(1 - 2(1 - \rho^2)\,\mathrm{i}\,u) - \rho^2]^{-\frac{1}{2}} =$$

$$= [(1 - 2\mathrm{i}\,t)\,(1 - 2\mathrm{i}\,u) + 4\rho^2 tu]^{-\frac{1}{2}}.$$

De aici

$$M(X^2Y^2) = \frac{1}{i^2}\frac{\partial^2\varphi(t, u)}{\partial t\partial u}\bigg|_{t=u=0} = 1 + 2\rho^2,$$

$$M(X^2) = 1, \qquad M(Y^2) = 1,$$

$$D^2(X^2) = 2,$$

$$D^2(Y^2) = 2$$

şi deci

$$\rho(X^2, Y^2) = \rho^2.$$

b) Pornim de la relaţia

$$D^2(X \mid y) = M(X^2 \mid y) - M^2(X \mid y).$$

Putem scrie deci

$$M(X^2 \mid y) = D^2(X \mid y) + M^2(X \mid y).$$

Însă

$$M(X \mid y) = \rho y$$

şi

$$D^2(X \mid y) = 1 - \rho^2.$$

Prin urmare

$$M(X^2 \mid y) = 1 - \rho^2 + \rho^2 y^2.$$

Cum fiecărei valori y^2 ale variabilei Y^2 îi corespund două valori $-y$ şi y cu probabilităţi egale cu $\frac{1}{2}$, rezultă că

$$M(X^2 \mid y^2) = \frac{1}{2} M(X^2 \mid y) + \frac{1}{2} M(X^2 \mid y) =$$

$$= 1 - \rho^2 + \rho^2 y^2$$

şi analog

$$M(Y^2 \mid x^2) = 1 - \rho^2 + \rho^2 x^2.$$

III.21. *Se consideră variabilele aleatoare X, Y, Z de valoare medie zero şi abatere medie pătratică unu şi astfel încît*

$$M(YX) = M(ZX) = \frac{1}{2}.$$

Să se determine limitele între care variază coeficientul de corelaţie al variabilelor X şi Y, $r = M(XY)$.

S o l u ţ i e. Ştim că întotdeauna coeficientul de corelaţie a două variabile aleatoare este cuprins între -1 şi $+1$ şi problema care se pune este de a vedea dacă aceste valori sînt atinse sau nu. Pentru a obţine valorile extreme ale coeficientului r vom porni de la inegalitatea

$$M[(\lambda X + \mu Y + \nu Z)^2] \geqslant 0,$$

valabilă oricare ar fi numerele reale λ, μ, ν.

Inegalitatea de mai sus, ţinînd cont de condiţiile problemei este echivalentă cu

$$f(\lambda, \mu, \nu) = \lambda^2 + \mu^2 + \nu^2 + \mu\nu + \nu\lambda + 2r\lambda\mu \geqslant 0,$$

adică o formă pătratică pozitiv definită. Deci r trebuie astfel ales încît să avem satisfăcută această condiţie, ceea ce revine la a spune că cele trei valori proprii ale matricei simetrice

$$\begin{pmatrix} 1 & r & \frac{1}{2} \\ r & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix}$$

sînt pozitive sau echivalent, că forma pătratică $f(\lambda, \mu, \nu)$ este suma pătratelor a trei forme liniare.

Forma pătratică $f(\lambda, \mu, \nu)$ o putem scrie astfel

$$f(\lambda, \mu, \nu) \equiv \left(\lambda + r\mu + \frac{\nu}{2}\right)^2 + (1 - r^2)\,\mu^2 + (1 - r)\,\mu\nu + \frac{3}{4}\,\nu^2.$$

Expresia

$$(1 - r^2)\,\mu^2 + (1 - r)\,\mu\nu + \frac{3}{4}\,\nu^2$$

constituie o formă pătratică de două variabile. Pentru ca această formă pătratică să fie pozitivă este necesar şi suficient ca discriminantul ei să fie negativ, adică

$$(1 - r)^2 - 3(1 - r^2) < 0$$

sau

$$(1-r)[1-r-3(1+r)]<0.$$

Cum $1-r>0$ rezultă că

$$1-r-3(1+r)<0$$

sau

$$r>-\frac{1}{2}.$$

Dacă $r=-\frac{1}{2}$, atunci

$$f(\lambda, \mu, \nu)=M[(\lambda X+\mu Y+\nu Z)^2]=$$

$$=\left(\lambda-\frac{\mu}{2}+\frac{\nu}{2}\right)^2+\frac{3}{4}(\mu+\nu)^2$$

şi $f(\lambda, \mu, \nu)$ se anulează dacă

$$\lambda-\frac{\mu}{2}+\frac{\nu}{2}=0,$$

$$\mu+\nu=0,$$

adică

$$\nu=-\mu; \quad \lambda=\frac{\mu}{2}-\frac{\nu}{2}=\mu,$$

ceea ce revine la a spune că între variabilele aleatoare X, Y, Z are loc aproape sigur relaţia

$$X+Y-Z=0.$$

Observaţie. Dacă se consideră λ, μ, ν drept coordonatele unui punct într-un sistem de axe oblice în spaţiu, vectorii bazei sînt unitari însă fac între ei unghiurile α, β, γ ale căror cosinusuri sînt $\frac{1}{2}$, $\frac{1}{2}$, r. Atunci $f(\lambda, \mu, \nu)$ reprezintă pătratul lungimii vectorului de componente λ, μ, ν. Din problemă rezultă că trebuie impusă condiţia

$$\alpha+\beta+\gamma<2\pi$$

sau

$$\gamma<2\pi-(\alpha+\beta)$$

ceea ce conduce la

$$\cos\gamma>\cos(\alpha+\beta).$$

Dar $\alpha = \beta$ şi deci

$$\cos(\alpha + \beta) = \cos 2\alpha = 2\cos^2\alpha - 1 = \frac{2}{4} - 1 = -\frac{1}{2}$$

ceea ce revine la

$$r > -\frac{1}{2},$$

rezultat care s-a obţinut fără a face apel la interpretarea geometrică.

III.22. *Fie $X_1, X_2, \ldots, X_n$, n variabile aleatoare astfel încît $M(X_i) = 0$, $D^2(x_i) = 1$, $1 \leqslant i \leqslant n$. Notăm cu $\Gamma_{ij} = M(X_iX_j)$ (care este tocmai coeficientul de corelaţie al variabilelor aleatoare X_i şi X_j) şi punem*

$$\Delta_n = \begin{vmatrix} 1 & \Gamma_{12} & \Gamma_{13} \ldots \Gamma_{1n} \\ \Gamma_{21} & 1 & \Gamma_{22} \ldots \Gamma_{2n} \\ \Gamma_{31} & \Gamma_{32} & 1 \;\ldots\; \Gamma_{3n} \\ \vdots & \vdots & \vdots \quad\;\; \vdots \\ \Gamma_{n1} & \Gamma_{n2} & \Gamma_{n3} \ldots 1 \end{vmatrix}$$

a) *Să se arate că*

$$\Gamma^{nn} = \frac{\Delta_{n-1}}{\Delta_n} \text{ şi că } 0 \leqslant \Delta_n \leqslant \Delta_{n-1}$$

unde

$$\begin{pmatrix} \Gamma^{11} & \Gamma^{12} \ldots \Gamma^{1n} \\ \Gamma^{21} & \Gamma^{22} \ldots \Gamma^{2n} \\ \vdots & \vdots \quad\;\; \vdots \\ \Gamma^{n1} & \Gamma^{n2} \ldots \Gamma^{nn} \end{pmatrix} = \begin{pmatrix} 1 & \Gamma_{12} \ldots \Gamma_{1n} \\ \Gamma_{21} & 1 \;\ldots\; \Gamma_{2n} \\ \vdots & \vdots \quad\;\; \vdots \\ \Gamma_{n1} & \Gamma_{n2} \ldots \Gamma_{nn} \end{pmatrix}^{-1}.$$

b) *Să se calculeze Δ_n cînd variabilele aleatoare $X_1, X_2, \ldots, X_n$ sînt necorelate cu excepţia variabilelor X_1 şi X_2.*

c) *Să se calculeze Δ_n cînd $\Gamma_{ij} = \Gamma$, $1 \leqslant i, j \leqslant n$.*

Soluţie. a) Să presupunem că forma pătratică $\sum_{i,j} \Gamma_{ij}X_iY_j$ reprezintă pătratul distanţei de la originea axelor la punctul de coordonate $(X_1, X_2, \ldots, X_n)$, într-un spaţiu euclidian avînd ca tensor metric pe Γ_{ij}.

În acest caz raportul $\frac{\Delta n}{\Delta_{n-1}}$ va apărea ca pătratul cosinusului unghiului a doi vectori.

Cum am pus

$$\Gamma_{ij} = M(X_i X_j), \qquad \Gamma_{ii} = M(X_i^2) = 1,$$

$$\Delta_n = \begin{vmatrix} 1 & \Gamma_{12} \dots & \Gamma_{1n} \\ \Gamma_{21} & 1 & \dots \Gamma_{2n} \\ \vdots & \vdots & \vdots \\ \Gamma_{n1} & \Gamma_{n2} \dots & 1 \end{vmatrix}$$

se constată că Δ_{n-1} este minorul tensorului extrem $\Gamma_{nn} = 1$ din Δ_n. Să considerăm acum un spaţiu euclidian real raportat la o bază de vectori $e_1, e_2, \dots, e_n$ normată, însă nu ortogonală. Fie Γ_{ij} cosinusul unghiului dintre vectorii e_i şi e_j.

Pătratul distanţei de la origine la $M(x_1, \dots, x_n)$ este egal cu

$$OM^2 = \Sigma \Gamma_{ij} X_i X_j = \Sigma X_i^2 + 2 \sum_{i<j} \Gamma_{ij} X_i X_j.$$

Dacă notăm cu Γ^{ij} elementele matricei

$$\begin{pmatrix} \Gamma^{11} & \Gamma^{12} \dots & \Gamma^{1n} \\ \vdots & \vdots & \vdots \\ \Gamma^{n1} & \Gamma^{n2} \dots & \Gamma^{nn} \end{pmatrix},$$

atunci putem scrie

$$\sum_{i=1}^{n} \Gamma_{ij} \Gamma^{ik} = \delta_{jk}$$

şi

$$\Gamma^{nn} = \frac{\Delta_{n-1}}{\Delta_n},$$

deoarece Δ_{n-1} este minorul elementului Γ_{nn} din determinantul Δ_n. În particular

$$\sum_{i=1}^{n} \Gamma_{ij} \Gamma^{in} = 0 \quad \text{dacă} \quad j \neq n.$$

Condiţia de ortogonalitate a doi vectori X şi Y este

$$\sum_{i,j} \Gamma_{ij} X_i Y_j = 0.$$

Condiţia de ortogonalitate a vectorului $(X_1, X_2, ..., X_n)$ cu vectorul bazei $e_j(0, 0, ..., 1, 0, ..., 0)$ este în particular

$$\sum_{i=1}^{n} \Gamma_{ij} X_i = 0, \quad j \text{ fixat.}$$

Rezultă că vectorul de componente $u_1 = \Gamma^{1n}$, $u_2 = \Gamma^{2n}$, ..., $u_n = \Gamma^{nn}$ este ortogonal cu toţi vectorii bazei cu excepţia lui e_n.

Să calculăm unghiul θ pe care vectorul U îl face cu vectorul $e_n(0, 0, ..., 0, 1)$.

Pentru aceasta punem în evidenţă produsul scalar al acestor doi vectori care este egal cu

$$\sum_{i=1}^{n} \Gamma_{in} u_i = \sum_{i=1}^{n} \Gamma_{in} \Gamma^{in} = 1.$$

Pătratele lungimilor vectorilor sînt egale cu Γ^{nn} şi

$$\sum_{i,j} \Gamma_{ij} \Gamma^{in} \Gamma^{jn} = \sum_{i} \Gamma^{in} (\sum_{j} \Gamma_{ij} \Gamma^{jn}) = \sum_{i} \delta_{ij} \Gamma^{in} = \Gamma^{nn}.$$

Deci

$$\cos^2 \theta = \frac{1}{\Gamma_{nn} \Gamma^{nn}} = \frac{1}{\Gamma^{nn}} = \frac{\Delta}{\Delta_{n-1}},$$

de unde rezultă că $0 \leqslant \Delta_n \leqslant \Delta_{n-1}$.

Cum

$$\Delta_1 = 1 \quad \text{şi} \quad \Delta_2 = 1 - \Gamma_{12}^2 \leqslant 1,$$

rezultă că şirul $\{\Delta_n\}$ este descrescător şi termenii săi sînt toţi cuprinşi în intervalul $[0, 1]$.

b) Dacă toate variabilele aleatoare $X_1, X_2, ..., X_n$ sînt necorelate cu excepţia variabilelor X_1, X_2 cu $\Delta_2 = 1 - \Gamma_{12}^2$, atunci

$$\Delta_n = \begin{vmatrix} 1 & \Gamma_{12} & 0 \dots 0 \\ \Gamma_{21} & 1 & 0 \dots 0 \\ 0 & 0 & 1 \dots 0 \\ \dots & \dots & \dots \\ 0 & 0 & 0 \dots 1 \end{vmatrix} = 1 - \Gamma_{12}^2 = \Delta_2.$$

c) Dacă toţi $\Gamma_{ij} = \Gamma$, $i, j = 1, 2, \ldots, n$, atunci

$$\Delta_n = \begin{vmatrix} 1 & \Gamma & \Gamma \ldots \Gamma \\ \Gamma & 1 & \Gamma \ldots \Gamma \\ \Gamma & \Gamma & 1 \ldots \Gamma \\ \vdots & \vdots & \vdots \quad \vdots \\ \Gamma & \Gamma & \Gamma \ldots 1 \end{vmatrix}$$

Pentru a calcula acest determinant procedăm în felul următor. Se scade din fiecare linie, linia următoare şi se obţine

$$\Delta_n = \begin{vmatrix} 1-\Gamma & \Gamma-1 & 0 \ldots 0 & 0 \\ 0 & 1-\Gamma & \Gamma-1 \ldots 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 \ldots 1-\Gamma & \Gamma-1 \\ \Gamma & \Gamma & \Gamma \ldots \Gamma & 1 \end{vmatrix} =$$

$$= (1-\Gamma)^{n-1} \begin{vmatrix} 1 & -1 & 0 \ldots 0 & 0 \\ 0 & 1 & -1 \ldots 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 \ldots 1 & -1 \\ \Gamma & \Gamma & \Gamma \ldots \Gamma & 1 \end{vmatrix}.$$

Adunînd acum toate coloanele determinantului la coloana întîia şi dezvoltînd apoi după prima coloană se obţine

$$\Delta_n = (1-\Gamma)^{n-1} \begin{vmatrix} 0 & -1 & 0 \ldots 0 & 0 \\ 0 & 1 & -1 \ldots 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 \ldots 1 & -1 \\ 1+(n-1)\Gamma & \Gamma & \Gamma \ldots \Gamma & 1 \end{vmatrix} =$$

$$= (1-\Gamma)^{n-1}(-1)^{n-1}[1+(n-1)\Gamma] \begin{vmatrix} -1 & 0 & 0 \ldots 0 & 0 \\ 1 & -1 & 0 \ldots 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 \ldots 1 & -1 \end{vmatrix} =$$

$$= (1-\Gamma)^{n-1}(-1)^{n-1}[1+(n-1)\Gamma](-1)^{n-1}.$$

Deci

$$\Delta_n = (1 - \Gamma)^{n-1} [(n - 1)\Gamma + 1]$$

şi de aici

$$\frac{\Delta_{n+1}}{\Delta_n} = (1 - \Gamma) \frac{n\Gamma + 1}{(n - 1)\Gamma + 1}$$

raport care (o funcţie omografică) tinde către $1 - \Gamma$ cînd $n \to \infty$.

Observaţie. Calculul se poate face mai simplu în modul următor: se consideră matricea simetrică

$$\begin{pmatrix} \Gamma & \Gamma & \Gamma \dots \Gamma \\ \Gamma & \Gamma & \Gamma \dots \Gamma \\ \Gamma & \Gamma & \Gamma \dots \Gamma \\ \vdots & \vdots & \vdots \quad \vdots \\ \Gamma & \Gamma & \Gamma \dots \Gamma \end{pmatrix}$$

Vectorii proprii ai acestei matrice sînt soluţii ale sistemului liniar

$$\Gamma s = \lambda x_i \qquad i = 1, 2, \dots, n,$$

$$s = \sum_1^n x_i.$$

Valoarea proprie simplă pentru vectorul propriu $x_1 = x_2 = \dots = x_n$ este $\lambda = n\Gamma$ iar $\lambda = 0$ este valoarea proprie multiplă de ordinul $n - 1$ pentru subspaţiul propriu $x_1 + x_2 + \dots + x_n = 0$.

Urmează că polinomul caracteristic este egal cu

$$(-1)^n \lambda^{n-1}(\lambda - n\Gamma).$$

De aici rezultă că făcînd pe $\lambda = \Gamma - 1$, determinantul caracteristic se reduce la Δ_n şi regăsim pe această cale valoarea calculată mai înainte.

III.23. *Repartiţia vectorului aleator* (X, Y) este dată în tabelul III.2.

Tabelul III.2

y_k \ x_k	1	2	3	4	5	$P(Y = y_k)$
1	$\frac{1}{12}$	$\frac{1}{24}$	0	$\frac{1}{24}$	$\frac{1}{30}$	$\frac{1}{5}$
2	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{30}$	$\frac{1}{5}$
3	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$	0	$\frac{1}{30}$	$\frac{1}{5}$
4	$\frac{1}{12}$	0	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{30}$	$\frac{1}{5}$
5	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{30}$	$\frac{1}{5}$
$P(X = x_k)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Să se afle dreptele de regresie şi unghiul dintre ele.

Soluţie. Dreapta de regresie a lui Y asupra lui X este

$$y - m_{01} = \rho \frac{\sigma_2}{\sigma_1} (x - m_{10}),$$

iar dreapta de regresie a lui X asupra lui Y este

$$y - m_{01} = \rho^{-1} \frac{\sigma_2}{\sigma_1} (x - m_{10}).$$

Din repartiţia vectorului (X, Y) se obţine

$$m_{10} = M(X) = \sum_k x_k P(X = x_k) = \frac{16}{6},$$

$$m_{01} = M(Y) = \sum_k y_k P(Y = y_k) = \frac{15}{5} = 3,$$

$$\sigma_1 = \sqrt{M[(X - m_{10})^2]} = 1{,}49,$$

$$\sigma_2 = \sqrt{M[(Y - m_{01})^2]} = 1{,}41,$$

$$\rho = \frac{M[(X - m_{10})(Y - m_{01})]}{\sigma_1\sigma_2} = 0{,}06.$$

Urmează că dreapta de regresie a lui Y asupra lui X este:

$$y - 3 = 0{,}056\,(x - 2{,}66),$$

iar dreapta de regresie a lui X în raport cu Y este

$$y - 3 = 15{,}7\,(x - 2{,}66).$$

Dacă notăm cu θ unghiul dintre dreptele de regresie atunci

$$\operatorname{tg}\theta = \frac{\frac{\sigma_2}{\sigma_1}\left(\frac{1}{\rho} - \rho\right)}{1 + \frac{\sigma_2^2}{\sigma_1^2}} = \frac{1 - \rho^2}{\rho}\,\frac{\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}.$$

Înlocuind cu valorile numerice determinate mai sus, obţinem

$$\operatorname{tg}\theta = \frac{15{,}6440}{0{,}8792}$$

şi

$$\theta = \operatorname{arc\,tg}\frac{15{,}6440}{0{,}8792} \simeq 86°50'$$

III.24. *Să se arate că unghiul* θ *dintre dreptele de regresie este dat de expresia:*

$$\theta = \operatorname{arc\,tg}\left(\frac{1 - \rho^2}{\rho}\,\frac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}\right).$$

Soluţie. Dacă regresiile corespunzătoare vectorului aleator (X, Y) sînt liniare, atunci ecuaţiile dreptelor de regresie sînt

$$y - M(Y) = \rho \frac{\sigma_y}{\sigma_x}[x - M(X)],$$

$$y - M(Y) = \frac{1}{\rho}\frac{\sigma_y}{\sigma_x}[x - M(X)],$$

punctul lor de intersecţie fiind $P(M(X), M(Y))$.

Dacă punem

$$m_1 = \frac{1}{\rho}\frac{\sigma_y}{\sigma_x} \quad \text{şi} \quad m_2 = \rho\frac{\sigma_y}{\sigma_x}$$

atunci

$$\operatorname{tg}\theta = \frac{m_1 - m_2}{1 + m_1 m_2} = \frac{\left(\frac{\sigma_y}{\sigma_x}\right)\left[\frac{1}{\rho} - \rho\right]}{1 + \frac{\sigma_y^2}{\sigma_x^2}} =$$

$$= \frac{1 - \rho^2}{\rho}\frac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}.$$

Observaţie. Pentru $\rho = 0$ dreptele de regresie sînt perpendiculare iar pentru $\rho = 1$ sînt paralele.

III.25. *Se dă densitatea de repartiţie a vectorului aleator* (X, Y):

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2(1 - \rho^2)}\left[\left(\frac{x}{\sigma_1}\right)^2 - 2\rho\frac{xy}{\sigma_1\sigma_2} + \left(\frac{y}{\sigma_2}\right)^2\right]\right\}.$$

a) *Să se afle dreapta de regresie a lui X în raport cu Y şi valoarea medie condiţionată* $M(X/y)$.

b) *Se consideră evenimentele* $B = \{Y < y_0\}$, $B' = \{Y > y_0\}$, *se notează cu* m, m', p *şi* q *expresiile*

$$m = M(X/B), \quad m' = M(X/B'), \quad p = P(B), \quad q = P(B')$$

şi se pune

$$u_0 = \frac{y_0}{\sigma_2}, \quad g(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right).$$

Să se calculeze mp şi m'q şi să se arate că

$$\rho = \frac{m' - m}{\sigma_1} \cdot \frac{pq}{g(u)}.$$

c) *Fie evenimentele*

$$B_1 = \{Y < y_1\}; \quad B_2 = \{Y > y_2\}$$

cu

$$y_1 > y_2 \quad \text{şi} \quad m_1 = M(X/B_1), \quad m_2 = M(X/B_2),$$
$$p_1 = P(B_1), \quad p_2 = P(B_2).$$

Să se arate că

$$\rho = \frac{m_2 - m_1}{\sigma_1} \cdot \frac{1}{Z},$$

unde

$$Z = \frac{g(u_1)}{p_1} + \frac{g(u_2)}{p_2}; \quad u_1 = y_1/\sigma_1; \quad u_2 = y_2/\sigma_2.$$

Soluţie. a) Avem

$$M(X/y) = \rho \frac{\sigma_1}{\sigma_2} y,$$

iar dreapta de regresie este

$$y = \frac{1}{\rho} \frac{\sigma_2}{\sigma_1} x.$$

b) Densitatea de repartiţie a vectorului aleator (X, Y) condiţionată de evenimentul B este

$$\frac{f(x, y)}{P(B)},$$

de unde

$$M(X/B) = \frac{1}{P(B)} \iint_B x f(x, y) \, dx \, dy$$

sau cu notaţiile menţionate

$$mp = \int_{-\infty}^{\infty} \int_{-\infty}^{y_0} x f(x, y) \, dx \, dy =$$

$$= \frac{1}{\sigma_2 \sqrt{2\pi}} \int_{-\infty}^{y_0} \exp\left(-\frac{1}{2\sigma_2^2} y^2\right) dy \int_{-\infty}^{\infty} x \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{x - \rho \frac{\sigma_1}{\sigma_2} y}{\sigma_2}\right]^2 dx\right.$$

de unde urmează că

$$mp = \frac{\rho\sigma_1}{\sqrt{2\pi}} \int_{-\infty}^{u_0} u\, e^{-\frac{u^2}{2}}\, du,$$

adică $mp = -\rho\sigma_1 g(uo)$.

Analog obţinem

$$m'q = \rho\sigma_1 g(u_0).$$

Din ultimele două relaţii obţinem prin înmulţire cu $-q$ respectiv cu p şi adunînd

$$(m' - m)\, pq = (q + q)\, \rho\sigma_1 g(u_0),$$

de unde scoatem coeficientul de corelaţie

$$\rho = \frac{m' - m}{\sigma_1} \frac{pq}{g(u_0)}.$$

c) Urmărind acelaşi raţionament ca la punctul b) găsim că

$$m_1 p_1 = -\rho\sigma_1 g(u_1),$$

$$m_2 p_2 = \rho\sigma_1 g(u_2),$$

de unde

$$m_2 - m_1 = \rho\sigma_1 \left[\frac{g(u_1)}{p_1} + \frac{g(u_2)}{p_2}\right]$$

sau

$$\rho = \frac{m_2 - m_1}{\sigma_1} \cdot \frac{1}{Z}.$$

III.26. *Să se calculeze ecuaţiile curbelor de regresie ale variabilei X, asupra variabilei Y, în cazul cînd densitatea de repartiţie a vectorului aleator (X, Y) este*

a) $$f(x, y) = \begin{cases} \frac{4}{5}(x + 3y)\, e^{-x-2y} & x, y \geqslant 0, \\ 0 \text{ în rest.} \end{cases}$$

b) $$f(x, y) = \begin{cases} \frac{y}{(1 + x)^4}\, e^{-\frac{y}{1+x}} & x, y \geqslant 0, \\ 0 \text{ în rest.} \end{cases}$$

c) $$f(x, y) = \begin{cases} \frac{9}{2} \frac{1 + x + y}{(1 + x)^4 (1 + y)^4}, & x, y \geqslant 0, \\ 0 \text{ în rest.} \end{cases}$$

S o l u ţ i e. a). Ecuaţia curbei de regresie a variabilei X faţă de Y este dată de $M(X/y)$.
Însă

$$M(X/y) = \frac{\frac{4}{5}\int_0^\infty x(x+3y)\,e^{-x-2y}\,dx}{\frac{4}{5}\int_6^\infty (x+3y)\,e^{-x-2y}\,dx} = \frac{\int_0^\infty x(x+3y)\,e^{-x}\,dx}{\int_0^\infty (x+3y)\,e^{-x}\,dx}.$$

Deci

$$x = \frac{2+3y}{1+3y},$$

deoarece

$$\int_0^\infty x(x+3y)\,e^{-x}\,dx = \Gamma(3) + 3y\,\Gamma(2) = 2 + 3y,$$

şi

$$\int_0^\infty (x+3y)\,e^{-x}\,dx = \Gamma(2) + 3y(-e^{-x}|_0^\infty) = 1 + 3y.$$

b) $$M(X/y) = \frac{\int_0^\infty \frac{xy}{(1+x)^4} e^{-\frac{y}{1+x}}\,dx}{\int_0^\infty \frac{y}{(1+x)^4} e^{-\frac{y}{1+y}}\,dx} = \frac{\int_0^\infty \frac{x}{(1+x)^4} e^{-\frac{y}{1+x}}\,dx}{\int_0^\infty \frac{1}{(1+x)^4} e^{-\frac{y}{1+x}}\,dx}.$$

Cum

$$\int_0^\infty \frac{1}{(1+x)^4} e^{-\frac{y}{1+x}}\,dx = \frac{2}{y^3} - \frac{e^{-y}}{y^3}(y^2+2y+2)$$

şi

$$\int_0^\infty \frac{x}{(1+x)^4} e^{-\frac{y}{1+x}}\,dx = \int_0^\infty \frac{1}{(1+x)^3} e^{-\frac{y}{1+x}}\,dx -$$

$$-\int_0^\infty \frac{1}{(1+x)^4} e^{-\frac{y}{1+x}}\,dx = \frac{y+2}{y^3}(e^{-y}+1),$$

rezultă că

$$M(X/y) = \frac{(y+2)(e^{-y}+1)}{2 - e^{-y}(y^2+2y+2)},$$

adică

$$x = \frac{(y+2)\,(e^{-y}+1)}{2 - e^{-y}(y^2+2y+2)}.$$

b) Din cauza simetriei funcției $f(x, y)$ în raport cu x și y rezultă că

$$M(X/y) = \frac{\frac{9}{2}\int_0^\infty \frac{x(1+x+y)}{(1+x)^4(1+y)^4}\,dx}{\frac{9}{2}\int_0^\infty \frac{1+x+y}{(1+x)^4(1+y)^4}\,dx} = \frac{\int_0^\infty \frac{x(1+x+y)}{(1+x)^4}\,dx}{\int_0^\infty \frac{1+x+y}{(1+x)^4}\,dx} = \frac{y+3}{2y+3}.$$

Deci

$$x = \frac{y+3}{2y+3}.$$

III.27. *Să se calculeze ecuațiile curbelor de regresie ale variabilei Y asupra variabilei X, pentru cazul cînd densitatea de repartiție a vectorului aleator (X, Y) este:*

a) $$f(x, y) = \begin{cases} \frac{4}{5}(x+3y)\,e^{-x-2y}, & x, y \geqslant 0, \\ 0 & \textit{în rest} \end{cases}$$

b) $$f(x, y) = \begin{cases} \frac{y}{(1+x)^4}\,e^{-\frac{y}{1+x}}, & x, y \geqslant 0; \\ 0 & \textit{în rest.} \end{cases}$$

c) $$f(x, y) = \begin{cases} \frac{9}{2}\cdot\frac{(1+x+y)}{(1+x)^4(1+y)^4}, & x, y \geqslant 0, \\ 0 & \textit{în rest.} \end{cases}$$

Soluție. a) Ecuația curbei de regresie a variabilei X asupra variabilei Y este dată de expresia $M(Y/x)$.

Dar

$$M(Y/x) = \frac{\frac{4}{5}\int_0^\infty y(x+3y)\,e^{-x-2y}\,dy}{\frac{4}{5}\int_0^\infty (x+3y)\,e^{-x-2y}\,dy} = \frac{\int_0^\infty y(x+3y)\,e^{-2y}\,dy}{\int_0^\infty (x+3y)\,e^{-2y}\,dy}.$$

Însă

$$\int_0^\infty (x+3y)\, e^{-2y}\, dy = \frac{2x+3}{4}$$

şi

$$\int_0^\infty y(x+3y)\, e^{-2y}\, dy = \frac{x+3}{4}.$$

Deci ecuaţia curbei de regresie este

$$y = \frac{x+3}{2x+3},$$

b) $$M(Y/x) = \frac{\displaystyle\int_0^\infty \frac{y^2}{(1+x)^4}\, e^{-\frac{y}{1+x}}\, dy}{\displaystyle\int_0^\infty \frac{y}{(1+x)^4}\, e^{-\frac{y}{1+x}}\, dy} = \frac{\displaystyle\int_0^\infty y^2 e^{-\frac{y}{1+x}}\, dy}{\displaystyle\int_0^\infty y\, e^{-\frac{y}{1+x}}\, dy}.$$

Însă

$$\int_0^\infty y\, e^{-\frac{y}{1+x}}\, dy = (1+x)^2$$

şi

$$\int_0^\infty y^2\, e^{-\frac{y}{1+x}}\, dy = 2(1+x)^3.$$

Deci ecuaţia curbei de regresie este

$$y = 2(1+x) \text{ (o dreaptă).}$$

c) $$M(Y/x) = \frac{\displaystyle\frac{9}{2}\int_0^\infty \frac{y(1+x+y)}{(1+x)^4(1+y)^4}\, dy}{\displaystyle\frac{9}{2}\int_0^\infty \frac{1+x+y}{(1+x)^4(1+y)^4}\, dy} = \frac{\displaystyle\int_0^\infty \frac{y(1+x+y)}{(1+y)^4}\, dy}{\displaystyle\int_0^\infty \frac{1+x+y}{(1+y)^4}\, dy}.$$

Dar

$$\int_0^\infty \frac{1+x+y}{(1+y)^4}\, dy = \frac{2x+3}{6}$$

şi

$$\int_0^\infty \frac{y(1+x+y)}{(1+y)^4}\, dy = \frac{x+3}{6}.$$

Deci ecuaţia curbei de regresie este

$$y = \frac{x+3}{2x+3}.$$

III.28. *Fie X şi Y două variabile aleatoare a căror densitate de repartiţie comună este $f(x, y)$ şi $m(x) = M(Y/x)$.*

Să se arate că

a) $M[m(X)] = M(Y)$,

$M[Ym(X)] = M[m(X)]^2$,

$M[XY] = M\ [Xm(X)]$,

unde $m(X)$ este variabila aleatoare obţinută prin înlocuirea lui x cu X în funcţia $m(x)$.

b) *Dacă se notează cu σ_1^2, respectiv σ_2^2, dispersiile variabilelor aleatoare X şi Y, cu r coeficientul de corelaţie al variabilelor X şi Y şi cu*

$$c^2 = \frac{M[((m(X))^2]}{\sigma_2^2}$$

pătratul raportului de corelaţie, să se arate că

$$\sigma_2^2(1 - c^2) = M[Y - m(X)]^2,$$

$$\sigma_2^2(1 - r^2) = M\left[Y - r\frac{\sigma_2}{\sigma_1}X\right]^2.$$

c) *Să se arate că $0 \leqslant r^2 \leqslant c^2 \leqslant 1$.*

Să se arate că $c^2 = 1$, dacă şi numai dacă X şi Y sînt legate printr-o relaţie funcţională şi că dacă $r^2 = c^2$, atunci Y este o funcţie liniară de X.

S o l u ţ i e. a) Valoarea medie a variabilei aleatoare Y condiţionată de faptul că $X = x$ este definită ca

$$m(x) = \frac{\int_{-\infty}^{\infty} yf(x, y)\,dy}{\int_{-\infty}^{\infty} f(x, y)\,dy},$$

de unde

$$m(x)\int_{-\infty}^{\infty} f(x, y)\,dy = \int_{-\infty}^{\infty} yf(x, y)\,dy. \qquad \text{(III.5)}$$

Integrînd în raport cu x în ambii membri, obținem

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m(x) f(x, y)\, dx\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x, y)\, dx\, dy,$$

adică

$$M[m(X)] = M(Y).$$

Dacă acum înmulțim relația (III.5) cu $m(x)$, apoi cu x și după aceea integrăm în raport cu x se obține

$$M[(m(X))^2] = M[Ym(X)],$$

$$M[Xm(X)] = M(XY).$$

b) Din definiția raportului de corelație

$$c^2 = \frac{M[m(X)]^2}{\sigma_2^2}$$

rezultă

$$\sigma_2^2(1 - c^2) = \sigma_2^2 - M[m(X)]^2 = M\{Y^2 - [m(X)]^2\}.$$

Însă

$$M\{Y^2 - [m(X)]^2\} = M\{Y^2 - 2Ym(X) + $$
$$+ [m(X)]^2\} = M[Y - m(X)]^2.$$

Prin urmare

$$\sigma_2^2(1 - c^2) = M\{[Y - m(X)]^2\}$$

ceea ce arată $c^2 \leqslant 1$.

Mai departe

$$M\left(Y - r\frac{\sigma_2}{\sigma_1} X\right)^2 = M\left(Y^2 - 2r\frac{\sigma_2}{\sigma_1} XY + r^2\frac{\sigma_2^2}{\sigma_1^2} X^2\right) =$$
$$= \sigma_2^2 - 2r^2\sigma_2^2 + r^2\sigma_2^2 = \sigma_2^2(1 - r^2)$$

adică

$$M\left(Y - r\frac{\sigma_2}{\sigma_1} X\right)^2 = \sigma_2^2(1 - r^2).$$

c) Relațiile puse în evidență ne arată faptul că r^2 este coeficientul asociat funcției liniare $r\frac{\sigma_2}{\sigma_1} X$ iar c^2 funcției $m(X)$.

Expresia $r\frac{\sigma_2}{\sigma_1}x$ poate fi considerată ca aproximație liniară a curbei de regresie $y = m(x)$.

În cazul regresiei liniare, funcțiile $r\frac{\sigma_2}{\sigma_1}x$ și $m(x)$ coincid și atunci $c^2 = r^2$.

Dacă acum aplicăm inegalitatea lui Schwartz expresiei $M[Xm(X)]$ obținem

$$\{M[Xm(X)]\}^2 \leqslant M(X^2)\, M[m(X)]^2 = \sigma_1^2\sigma_2^2c^2$$

și deci

$$\sigma_1^2\sigma_2^2r^2 \leqslant \sigma_1^2\sigma_2^2c^2,$$

de unde rezultă că

$$0 \leqslant r^2 \leqslant c^2 \leqslant 1.$$

$c^2 = 1$, implică $Y = m(X)$ aproape sigur; există o relație funcțională între X și Y reprezentată grafic prin linia de regresie $y = m(X)$. Dacă în particular $r^2 = 1$, această relație este liniară. Egalitatea $r^2 = c^2$, implică faptul că între X și Y există o relație liniară; regresia lui Y asupra lui X este liniară.

Egalitatea $c = 0$, implică $m(X) = 0$ aproape sigur, însă aceasta nu antrenează în mod necesar independența variabilelor aleatoare X și Y cu toate că $c = 0$ constituie o condiție mai puternică decît $r = 0$.

III.29. *Dintr-o populație oarecare caracterizată de variabila aleatoare X cu $M(X) = a$, se face o selecție de volum n, obținîndu-se valorile $X_1, X_2, \ldots, X_n$, apoi încă o selecție de volum $m - k$ obținîndu-se valorile $X'_{k+1}, X'_{k+2}, \ldots, X'_m$.*

Dacă se pune

$$Y = X_1 + X_2 + \ldots + X_k + X_{k+1} + \ldots + X_n,$$

$$Z = X_1 + X_2 + \ldots + X_k + X'_{k+1} + \ldots + X'_m,$$

să se arate că regresia variabilei Y asupra variabilei Z este dată de

$$y = \frac{k}{m}z + (n-k)\,a.$$

S o l u ț i e. Regresia variabilei Y asupra variabilei Z este

$$y - M(Y) = r_{Y,Z}\frac{\sigma_Y}{\sigma_Z}(Z - M(Z)) = \frac{\mu_{11}}{\sigma_Z^2}(Z - M(Z)).$$

Însă

$$M(Y) = \sum_{i=1}^{n} M(X_i) = na,$$

$$M(Z) = \sum_{j=1}^{k} M(X_j) + \sum_{j=k+1}^{m} M(X_j') = ma.$$

$$\sigma_Z^2 = M(X_1 + \dots + X_k + X_{k+1}' + \dots + X_m')^2 - m^2a^2 =$$

$$= \sum_{i=1}^{k} M(X_i^2) + \sum_{i=k+1}^{m} M(X_i'^2) + 2 \sum_{1=i<j}^{k} M(X_iX_j) + 2 \sum_{k+1=i<j}^{m} M(X_i'X_j') +$$

$$+ 2\sum_{i=1}^{k} \sum_{j=k+1}^{m} M(X_iX_j') = m\sigma_X^2 - m(m-1)a^2 +$$

$$+ k(k-1)a^2 + (m-k)(m-k-1)a^2 + 2k(m-k)a^2 = m\sigma_X^2.$$

$$\mu_{11} = M(YZ) - M(Y)M(Z) =$$

$$= M(X_1 + \dots + X_k + X_{k+1} + \dots + X_n)(X_1 + \dots + X_k +$$

$$+ X_{k+1}' + \dots + X_m') - nma^2 = M(X_1^2 + \dots + X_k^2) +$$

$$+ 2 \sum_{1=i<j}^{k} M(X_iX_j) + \sum_{i=1}^{k} \sum_{j=k+1}^{m} M(X_iX_j') + \sum_{i=k+1}^{n} \sum_{j=1}^{k} M(X_iX_j) +$$

$$+ \sum_{i=k+1}^{n} \sum_{j=k+1}^{m} M(X_iX_j') - nma^2 = k\sigma_X^2 + ka^2 + k(k-1)a^2 +$$

$$+ k(m-k)a^2 + (n-k)ka^2 + (n-k)(m-k)a^2 - nm^2.$$

Deci

$$\mu_{11} = k\sigma_X^2.$$

Aşadar

$$y - na = \frac{k\sigma_X^2}{m\sigma_X^2}(z - ma),$$

de unde urmează că

$$y = \frac{k}{m} z + (n-k)a,$$

adică tocmai relaţia pe care ne-am propus s-o dovedim.

III.30. *Se consideră variabilele aleatoare W_1 şi W_2 definite ca*

$$W_1 = X_1 + X_2 + \dots + X_n + U_1 + U_2 + \dots + U_a,$$

$$W_2 = X_1 + X_2 + \dots + X_n + V_1 + V_2 + \dots + V_b,$$

unde variabilele aleatoare X_i, U_j, V_k sînt necorelate şi

$$M(X_i) = M(U_j) = M(V_k) = 0, \quad D^2(X_i) = D^2(U_j) = D^2(V_k) = 1.$$

a) *Să se arate că*

$$r_{w_1, w_2} = \frac{n}{\sqrt{(n+a)(n+b)}}.$$

b) *Utilizîndu-se acest rezultat, să se arate că dacă*

$$Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6.$$

$$Z_1 = X_1; \; Z_2 = X_2 + X_3; \; Z_3 = X_4 + X_5 + X_6,$$

unde X_i, $1 \leqslant i \leqslant 6$ satisfac condiţiile impuse mai sus, atunci regresia multiplă a variabilei Y asupra variabilelor Z_1, Z_2, Z_3 este dată de

$$y = z_1 + z_2 + z_3.$$

c) *Să se calculeze regresia multiplă a lui X faţă de Y_1, Y_2, Y_3, Y_4 unde*

$$X = \sum_{i=1}^{8} X_i; \qquad Y_1 = X_1 + X_2; \qquad Y_2 = X_3 + X_4 + X_5;$$

$$Y_3 = X_3 + X_4 + X_5 + X_6; \qquad Y_4 = X_4 + X_5 + X_6 + X_7 + X_8,$$

iar

$$(X_i)_{1 \leqslant i \leqslant 8}$$

cu proprietăţile indicate mai sus.

S o l u ţ i e. *a)* Întrucît variabilele X_i, U_j, V_k sînt necorelate şi de dispersie 1, rezultă că

$$D^2(W_1) = D^2(X_1) + \dots + D^2(X_n) + D^2(U_1) + \dots + D^2(U_a) = n + a,$$

$$D^2(W_2) = D^2(X_1) + \dots + D^2(X_n) + D^2(V_1) + \dots + D^2(V_b) = n + b,$$

$$\operatorname{cov}(W_1, W_2) = M(W_1 W_2) - M(W_1)\, M(W_2) = M(X_1^2 + \dots + X_n^2) = n.$$

Deci

$$r_{W_1; W_2} = \frac{n}{\sqrt{(n+a)(n+b)}}.$$

b) Din rezultatul obținut avem

$$\operatorname{cov}(Z_1, Z_2) = 1; \quad \operatorname{cov}(Z_1, Z_2) = 0; \quad \operatorname{cov}(Z_1, Z_3) = 0;$$
$$\operatorname{cov}(Z_1, Y) = 1; \quad \operatorname{cov}(Z_2, Z_2) = 2; \quad \operatorname{cov}(Z_2, Z_3) = 0;$$
$$\operatorname{cov}(Z_2, Y) = 2, \quad \operatorname{cov}(Z_3, Z_3) = 3; \quad \operatorname{cov}(Z_3, Y) = 3.$$

Regresia multiplă satisface relația

$$\begin{vmatrix} \sigma_{11} & \sigma_{12} & \dots \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \dots \sigma_{2k} \\ \dots & \dots & \dots \\ \sigma_{k-1\,1} & \sigma_{k-2\,2} & 2 \dots \sigma_{k-1k} \\ x_1 - \mu_1 & x_2 - \mu_2 & \dots x_k - \mu_k \end{vmatrix} = 0$$

rezultă că în cazul nostru vom avea

$$\begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & 0 & 2 \\ 0 & 0 & 3 & 3 \\ z_1 & z_2 & z_3 & y \end{vmatrix} = 0,$$

adică

$$6y - 6z_3 - 6z_2 - 6z_1 = 0$$

sau

$$y = z_1 + z_2 + z_3.$$

c) Procedînd întocmai ca la punctul *b)*, regresia multiplă a variabilei X asupra variabilelor Y_1, Y_2, Y_3, Y_4 este dată de ecuația

$$\begin{vmatrix} 2 & 1 & 0 & 0 & 2 \\ 1 & 3 & 2 & 1 & 3 \\ 0 & 2 & 4 & 3 & 4 \\ 0 & 1 & 3 & 5 & 5 \\ y_1 & y_2 & y_3 & y_4 & x \end{vmatrix} = 0.$$

După efectuarea tuturor calculelor se obţine

$$x = \frac{23}{31} y_1 + \frac{8}{31} y_2 - \frac{5}{31} y_3 + \frac{24}{31} y_4.$$

III.31. *Două staţii hidrometeorologice fac 100 de observaţii de-a lungul unui an (1 octombrie — 31 septembrie anul următor) asupra nivelului mediu al apelor unui rîu şi se obţin date trecute în tabelul* III.3.

Tabelul III.3

x	3,5	3,6	3,7	3,8	3,9	4,0	4,1	4,2	4,3	4,4	4,5	$n_{.j}$
3,2	1	1										2
3,3		1	2		1							4
3,4			2	2	1	1						6
3,5			3	5	2	1	1	1				13
3,6			1	4	4	4	1	1	1			16
3,7			1	1	12	1	2					17
3,8				1	1	10	3		1			16
3,9		1			1	12	1		1		1	17
4,0										1	1	2
4,1										2	3	5
4,2								1		1		2
$n_{h.}$	1	3	9	13	22	29	8	3	3	4	5	100

a) *Să se determine mediile şi dispersiile variabilelor x şi y.*

b) *Să se determine mediile condiţionate.*

c) *Să se determine liniile de regresie, utilizînd metoda celor mai mici pătrate.*

d) *Să se calculeze coeficienţii de corelaţie şi coeficienţii de regresie.*

Soluţie. *a*) Pentru uşurarea calculelor necesare determinării mediei şi dispersiei variabilei x întocmim tabelul III.4.

Tabelul III.4

x_h	$n_{h\cdot}$	$x_h - a$	$n_{h\cdot}(x_h - a)$	$n_{h\cdot}(x_g - a)^2$
3,5	1	−0,5	−0,5	0,25
3,6	3	−0,4	−1,2	0,48
3,7	9	−0,3	−2,7	0,81
3,8	13	−0,2	−2,6	0,52
3,9	22	−0,1	−2,2	0,22
a=4,0	29	0,0	0,0	0,00
4,1	8	0,1	0,8	0,08
4,2	3	0,2	0,6	0,12
4,3	3	0,3	0,9	0,27
4,4	4	0,4	1,4	0,64
4,5	5	0,5	2,5	1,25
Total	100	—	−0,3	4,44

De aici rezultă că

$$\bar{x} = \frac{\sum\limits_h n_{h\cdot}(x_h - a)}{\sum n_{h\cdot}} + a = 4 - \frac{3}{100} = 3,97,$$

$$\bar{\sigma}_x^2 = \frac{\sum\limits_h n_{h\cdot}(x_h - a)^2}{\sum n_{h\cdot}} - (\bar{x} - a)^2 = \frac{4,44}{100} - (3,97 - 4,0)^2 = 0,0435,$$

de unde $\bar{\sigma}_x = 0,208$.

Pentru calcularea mediei și dispersiei variabilei y avem tabelul III.5

Tabelul III.5

y_k	$n_{k\cdot}$	$y_k - b$	$n_{\cdot k}(y - b)$	$n_{\cdot k}(y_k - b)^2$
3,2	2	−0,5	−1,0	0,50
3,3	4	−0,4	−1,6	0,64
3,4	6	−0,3	−1,8	0,54
3,5	13	−0,2	−2,6	0,52
3,6	16	−0,1	−1,6	0,16
b=3,7	17	0,0	0,0	0,00
3,8	16	0,1	1,6	0,16
3,9	17	0,2	3,4	0,68
4,0	2	0,3	0,6	0,18
4,1	5	0,4	2,0	0,80
4,2	2	0,2	1,0	0,50
Total	100		0,0	4,68

De aici rezultă că

$$\bar{y} = 3{,}7.$$

$$\overline{\sigma}_y^2 = \frac{\sum_k n_{\cdot k}(y_k - b)^2}{\sum_k n_{\cdot k}} - (\bar{y} - b)^2 = \frac{4{,}68}{100} = 0{,}0468$$

de unde

$$\overline{\sigma}_y = 0{,}216.$$

b) Pentru determinarea mediilor condiţionate utilizăm relaţiile

$$\bar{x}_{y_i} = \frac{\sum_h n_{h_i} x_h}{n_{\cdot i}},$$

şi

$$\bar{y}_{x_h} = \frac{\sum_i n_{h_i} y_i}{n_{h\cdot}}.$$

Dacă acum punem $\bar{x}_y$ pentru valoarea medie condiţionată a variabilei x, pentru y dat, pe baza relaţiilor de mai sus se obţine tabelul III.6.

Tabelul III.6

y	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	4,0	4,1	4,2
$\bar{x}_y$	3,550	3,725	3,817	3,861	3,944	3,912	4,019	4,024	4,450	4,460	4,300

iar în cazul mediei condiţionate $\bar{y}_x$, se obţine pentru diferitele valori ale lui x tabelul III.7.

Tabelul III.7

x	3,5	3,6	3,7	3,8	3,9	4,0	4,1	4,2	4,3	4,4	4,5
$\bar{y}_x$	3,200	3,466	3,467	3,554	3,645	3,786	3,725	3,767	3,767	4,100	4,040

c) Ecuațiile liniilor de regresie sînt date de relațiile

$$\bar{y}_h = ax_h + b$$

şi

$$\bar{x}_h = a'y_k + b',$$

unde constantele a, b, și a', b' se determină prin metoda celor mai mic pătrate.

Ecuațiile normale pentru determinarea coeficienților a și b sînt:

$$a\sum_h x_h + nb = \sum_h \bar{y}_h,$$

$$a\sum_h x_h^2 + b\sum_h x_h = \sum x_h\bar{y}_h$$

cu soluția

$$a = \frac{\begin{vmatrix} \sum\limits_h \bar{y}_h & n \\ \sum\limits_h x_h\bar{y}_h & \sum\limits_h x_h \end{vmatrix}}{\begin{vmatrix} \sum\limits_h x_h & n \\ \sum\limits_h x_h^2 & \sum\limits_h x_n \end{vmatrix}}; \quad b = \frac{\begin{vmatrix} \sum\limits_h x_h & \sum\limits_h \bar{y}_h \\ \sum\limits_h x_h^2 & \sum\limits_h x_h\bar{y}_h \end{vmatrix}}{\begin{vmatrix} \sum\limits_h x_h & n \\ \sum\limits_h x_n^2 & \sum\limits_h x_h \end{vmatrix}},$$

$$a = \frac{-14505,472}{-15974} \simeq 0,908; \qquad b = \frac{-5,09737}{-15974}\ 0,000,$$

iar cele pentru determinarea coeficienților a' și b' sînt

$$a'\sum_k y_k + nb' = \sum_k \bar{x}_k,$$

$$a'\sum_k y_k^2 + b'\sum_k y_k = \sum_k \bar{x}_k y_k$$

cu soluţia

$$a' = \frac{\begin{vmatrix} \sum_k \bar{x}_k & n \\ \sum_k \bar{x}_k y_k & \sum_k y_k \end{vmatrix}}{\begin{vmatrix} \sum_k y_k & n \\ \sum_k y_k^2 & \sum_k y_k \end{vmatrix}}; \quad b' = \frac{\begin{vmatrix} \sum_k y_k & \sum_k \bar{x}_k \\ \sum_k y_k^2 & \sum_k \bar{x}_k y_k \end{vmatrix}}{\begin{vmatrix} \sum_k y_k & n \\ \sum_k y_k^2 & \sum_k y_k \end{vmatrix}}$$

$$a' = \frac{154{,}0394}{-3512{,}51} \simeq 0{,}044; \qquad b' = \frac{-11{,}8789}{-3512{,}51} \simeq 0{,}003.$$

Deoarece în cazul de faţă obţinem

$$\sum_h x_h = 44; \quad \sum x_h^2 = 177{,}10; \quad \sum_h \bar{y}_h = 40{,}517;$$

$$(\sum_h x_h)^2 = 1936; \qquad \sum_h x_h \bar{y}_h = 162{,}8822;$$

$$\sum_k \bar{x}_k = 44{,}062; \quad \sum_k y_k = 40{,}7; \quad \sum_k y_k^2 = 151{,}69;$$

$$\sum_k y_k \bar{x}_k = 163{,}9284.$$

d) Covarianţa variabilelor x şi y, μ_{11} se calculează cu ajutorul relaţiei:

$$\mu_{11} = \frac{1}{n} \sum_{h,k} n_{h.k}(x_h - a)(y_k - b) - (\bar{x} - a)(\bar{y} - b)$$

sau încă

$$\mu_{11} = \frac{1}{n} \sum_h (x_h - a)(\bar{y}_h - b) - (\bar{x}' - a)(\bar{y} - b),$$

unde $\bar{y}_h$ este media condiţionată a lui y pentru x dat, iar $a = 4{,}0$; $b = 3{,}7$.

Pentru calculul covarianţei μ_{11}, întocmim tabelul III.8.

Tabelul III. 8

n_h	$x_h - a$	$\bar{y}_h$	$\bar{y}_h - b$	$n_h \cdot (x_h - a)(\bar{y}_h - b)$
1	−0,5	3,200	−0,500	0,2500
3	−0,4	3,466	−0,234	0,2808
9	−0,3	3,467	−0,233	0,6291
13	−0,2	3,554	−0,146	0,3796
22	−0,1	3,645	−0,055	0,1210
29	0,0	3,786	0,086	0,0000
8	0,1	3,725	0,025	0,0200
3	0,2	3,767	0,067	0,0402
3	0,3	3,767	0,067	0,0603
4	0,4	4,100	0,400	0,6400
5	0,5	4,040	0,340	0,8500
				3,2710

Deci

$$\mu_{11} = \frac{3,2710}{100} = 0,032710.$$

Atunci coeficientul de corelaţie r are valoarea

$$r = \frac{0,032710}{0,208 \cdot 0,216} \simeq 0,728.$$

Dacă notăm cu b_{xy} şi b_{yx} coeficienţii de regresie, atunci

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}; \quad b_{yx} = r\frac{\sigma_y}{\sigma_x},$$

sau încă

$$b_{xy} = \frac{\mu_{11}}{\sigma_y^2}; \quad b_{yx} = \frac{\mu_{11}}{\sigma_x^2}.$$

În cazul exemplului numeric obţinem

$$b_{xy} \simeq 0,7; \; b_{yx} \simeq 0,752.$$

III. 32. *Doi observatori fac 70 de măsurători asupra unei variabile continue. Rezultatele observaţiilor sînt trecute în tabelul* III.9.

Tabelul III.9

y \ x	$0-4$	-8	-12	-16	-20	-24	-28	-32	-36	n_i
$0-4$	1	3								4
-8	1	1	4							6
-12		5	6	1						12
-16			6	6	2					14
-20				4	3	5				12
-24					4	3	2			9
-28						1	1	3		5
-32							2	2	2	6
-36								1	1	2
$n_{k.}$	2	9	16	11	9	9	5	6	3	70

Să se calculeze coeficientul de corelaţie, coeficienţii de regresie şi dreptele de regresie.

Soluţie. Deoarece în expresiile coeficientului de corelaţie şi ai coeficienţilor de regresie intră dispersiile, le vom calcula mai întîi pe acestea

$$\bar{\sigma}_x^2 = \frac{\sum_h n_{h.}(x_h - a)^2}{n} - (\bar{x} - a)^2,$$

$$\bar{\sigma}_y^2 = \frac{\sum_k n_{k.}(y_k - b)^2}{n} - (\bar{y} - b)^2,$$

unde am pus

$$\bar{x} = \frac{\sum_h n_{h.}(x_h - a)}{n} + a; \qquad \bar{y} = \frac{\sum_k n_{.k}(y_k - b)}{n} + b,$$

$$n = \sum_h n_{h.} = \sum_k n_{.k},$$

şi x_h, y_k mijloacelor intervalelor pe care se înregistrează observaţiile.

Pentru calcularea valorilor statisticilor $\bar{x}$, $\bar{\sigma}_x$, $\bar{y}$, $\bar{\sigma}_y$ avem tabelele:

Tabelul III.10

$n_{h\cdot}$	x_h	$x_h - a$	$n_{h\cdot}(x_h - a)$	$n_{h\cdot}(x_h - a)^2$
2	2	−16	− 32	512
9	6	−12	−108	1 296
16	10	− 8	−128	1 024
11	14	− 4	− 44	176
9	$a =$ 18	0	0	0
9	22	4	36	144
5	26	8	40	320
6	30	12	72	864
3	34	16	48	768
70			−116	5 104

de unde

$$\bar{x} = 18 - \frac{116}{70} = 16{,}343; \qquad \bar{\sigma}_x^2 = \frac{5\,104}{70} + \frac{4}{12} - \left(\frac{116}{70}\right)^2 = 70{,}50$$

$$\bar{\sigma}_x \simeq 8{,}4$$

şi

Tabelul III.11

$n_{\cdot k}$	y_k	$y_k - b$	$n_{\cdot k}(y_k - b)$	$n_{\cdot k}(y_k - b)^2$
4	2	−16	−64	1 024
6	6	−12	−72	864
12	10	− 8	−96	768
14	14	− 4	−56	224
12	$b=$18	0	0	0
9	22	4	36	144
5	26	8	40	320
6	30	12	72	864
2	34	16	32	512
70			−108	4 720

de unde

$$\bar{y} = 18 - \frac{108}{70} \simeq 16{,}458; \quad \overline{\sigma}_y^2 = \left(\frac{4\,720}{70} + \frac{4}{12}\right) - \left(\frac{108}{70}\right)^2 = 65{,}3815$$

$$\overline{\sigma}_y \simeq 8{,}08.$$

Observație. În expresiile $\overline{\sigma}_x^2$, $\overline{\sigma}_y^2$ termenul $\frac{4}{12}$ reprezintă corecția de continuitate a lui Sheppard.

Pentru a calcula coeficienții de regresie b_{xy} și b_{yx}, vom calcula momentul mixt de ordinul unu, m_{11}.

În cazul unei repartiții bidimensionale cu densitatea de repartiție $f(x,y)$, momentul mixt de ordinul unu este dat de

$$m_{11} = \frac{\iint\limits_D xy\, f(x, y)\, \mathrm{d}x\mathrm{d}y}{\iint\limits_D f(x, y)\, \mathrm{d}x\mathrm{d}y}.$$

În cazul nostru D (fig. III.1) este pătratul de laturi [0; 36] și [0;36] pe care punctele de diviziune $0 < 4 < 8 < \ldots < 36$ îl împart în pătrate d_{ij} și în care $f(x, y)$ ia valoarea n_{ij}.

Atunci

$$m_{11} = \frac{\sum\limits_{i,j} \iint\limits_{d_{ij}} xyn_{ij}\, \mathrm{d}x\, \mathrm{d}y}{\sum\limits_{i,j} \iint\limits_{d_{ij}} n_{ij}\, \mathrm{d}x\, \mathrm{d}y} = \frac{\sum\limits_{i,j} d_{ij} \dfrac{x_{i+1} + x_i}{2} \cdot \dfrac{y_{j+1} + y_j}{2} n_{ij}}{\sum\limits_{i,j} n_{ij} d_{ij}},$$

unde d_{ij} reprezintă chiar aria pătratului notat astfel.

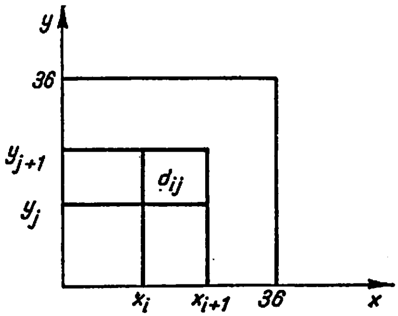

Fig. III.1

Cum toate subpătratele au aceeaşi arie, d_{ij} se simplifică şi obţinem:

$$m_{11} = \frac{\sum\limits_{i,j} \frac{x_{i+1}+x_i}{2} \frac{y_{j+1}+y_j}{2} n_{ij}}{\sum\limits_{i,j} n_{ij}}.$$

Urmează de aici că m_{11} se calculează cu şi în cazul variabilelor de repartiţie discretă cu masa plasată în mijloacele intervalelor.

Repartiţia discretă corespunzătoare este dată în tabelul III.12.

Tabelul III.12

x / y	$x-a$ / $y-b$	2 / −16	6 / 12	10 / −8	14 / −4	$a=18$ / 0	22 / 4	26 / 8	30 / 12	34 / 16	$n_{\cdot i}$
2	−16	1 256	3 192								4
6	−12	1 192	1 144	4 96							6
10	− 8		5 96	6 64	1 32						12
14	− 4			6 32	6 16	2 0					14
$b=18$	0				4 0	3 0	5 0				12
22	4					4 0	3 16	2 32			9
26	8						1 32	1 64	3 96		5
30	12							2 96	2 144	2 192	6
34	16								1 192	1 256	2
$n_{h\cdot}$		2	9	16	11	9	9	5	6	3	70
$\sum\limits_i n_{hi}(x_h-a)(y_i-b)$		448	1 200	860	128	0	80	320	768	640	4 444

În căsuţele separate printr-o linie punctată, deasupra sînt trecute frecvenţele iar sub linia punctată produsele $(x_i - a)\,(y_j - b)$.

Pentru $\operatorname{cov}(x, y) = \mu_{11}$ avem

$$\operatorname{cov}(x, y) = m_{11} - \bar{x}\bar{y} = \frac{4\,444}{70} - (\bar{x} - 18)\,(\bar{y} - 18) = \frac{4\,444}{70} -$$

$$- \frac{116 \cdot 108}{70^2} = 60{,}9290,$$

adică

$$\mu_{11} = \operatorname{cov}(x, y) = 60{,}9290$$

şi de aici coeficientul de corelaţie

$$r = \frac{\mu_{11}}{\bar{\sigma}_x \bar{\sigma}_y} \simeq 0{,}90.$$

Coeficienţii de regresie sînt $b_{xy} = \frac{\mu_{11}}{\bar{\sigma}^2_y}$; $b_{yx} = \frac{\mu_{11}}{\bar{\sigma}^2_x}$ care în cazul nostru numeric au valorile

$$b_{xy} = \frac{60{,}9290}{65{,}3815} = 0{,}93; \qquad b_{yx} = \frac{60{,}9290}{70{,}50} \simeq 0{,}864,$$

iar ecuaţiile liniilor de regresie

$$\bar{y}_x - 16{,}458 = 0{,}93(x - 16{,}343),$$

$$\bar{x}_y - 16{,}343 = 0{,}864(y - 16{,}458).$$

III.33. *Nivelul apelor unui rîu se măsoară în trei puncte diferite A_0, A_1, A_2. Notîndu-se observaţiile în punctul A_0 cu y, în punctul A_1 cu x_1 şi în punctul A_2 cu x_2, s-au obţinut datele din tabelul* III.13

Tabelul III.13

y	x_1	x_2	n	y	x_1	x_2	n
3	5	10	3	10	15	18	10
3	5	10	2	10	15	18	10
5	5	10	1	12	18	18	5
6	7	15	7	12	18	20	1
9	10	15	4	15	20	20	1
10	15	15	5	15	20	20	1

Să se determine ecuaţia planului de regresie a lui y în raport cu (x_1, x_2).

S o l u ţ i e. Dacă avem trei variabile (y, x_1, x_2) atunci ecuaţia planului de regresie a lui y faţă de x_1, x_2 este dată de relaţia

$$y - \bar{y} = - \frac{\sigma_0}{\sigma_1} \frac{D_{01}}{D_{00}} (x_1 - \bar{x}_1) - \frac{\sigma_0}{\sigma_2} \frac{D_{02}}{D_{00}} (x_2 - \bar{x}_2),$$

unde

$$D = \begin{vmatrix} 1 & r_{01} & r_{02} \\ r_{10} & 1 & r_{12} \\ r_{20} & r_{21} & 1 \end{vmatrix}$$

şi

$$D_{00} = \begin{vmatrix} 1 & r_{12} \\ r_{21} & 1 \end{vmatrix} = 1 - r_{12}^2, \quad D_{01} = - \begin{vmatrix} r_{01} & r_{12} \\ r_{20} & 1 \end{vmatrix} = - (r_{01} - r_{12} r_{20}),$$

$$D_{02} = \begin{vmatrix} r_{01} & r_{02} \\ 1 & r_{12} \end{vmatrix} = r_{01} r_{12} - r_{02} = - (r_{02} - r_{01} r_{12}).$$

Pr in urmare

$$y - \bar{y} = \frac{\sigma_0}{\sigma_1} \frac{r_{01} - r_{12} r_{20}}{1 - r_{12}^2} (x_1 - \bar{x}_1) + \frac{\sigma_0}{\sigma_1} \frac{r_{02} - r_{21} r_{10}}{1 - r_{12}^2} (x_2 - \bar{x}_2).$$

Rămîne acum să calculăm efectiv mărimile ce intervin în ecuaţia acestui plan de regresie: $\bar{y}$, $\bar{x}_1$, $\bar{x}_2$, σ_0, σ_1, σ_2, r_{12}, r_{01}, r_{02}. Pentru uşurinţa calculelor vom nota

$$u = y - 9; \quad v_1 = x_1 - 10; \quad v_2 = x_2 - 15.$$

Obţinem tabelul III.14.

T a b e l u l III.14

u	v_1	v_2	n	nu	nv_1	nv_2	nu^2	nv_1^2	nv_2^2	nuv_1	nuv_2	nv_1v_2
−6	−5	−5	3	−18	−15	−15	108	75	75	90	90	75
−6	−5	−5	2	−12	−10	−10	72	50	50	60	60	50
−4	−5	−5	1	− 4	− 5	− 5	16	25	25	20	20	25
−3	−3	0	7	−21	−21	0	63	63	0	63	0	0
0	0	0	4	0	0	0	0	0	0	0	0	0
1	5	0	5	5	25	0	5	125	0	25	0	0
1	5	0	10	10	50	0	10	250	0	50	0	0
1	5	0	10	10	50	0	10	250	0	50	0	0
3	8	3	5	15	40	15	45	320	45	120	45	120
3	8	3	1	3	8	3	9	64	9	24	9	24
6	10	5	1	6	10	5	36	100	25	60	30	50
6	10	5	1	6	10	5	36	100	25	60	30	50
			50	0	142	−2	410	1422	254	622	284	394

Rezultă că

$$\bar{u} = 0; \quad \bar{v}_1 = \frac{124}{50} = 2{,}84; \quad \bar{v}_2 = -\frac{5}{50} = -0{,}1$$

şi de aici

$$\bar{y} = 9; \quad \bar{x}_1 = 12{,}84; \quad \bar{x}_2 = 14{,}9,$$

$$\bar{\sigma}_u^2 = \frac{410}{50} = 8{,}2; \quad \bar{\sigma}_{v_1}^2 = \frac{1422}{50} - (2{,}84)^2 = 28{,}440 - 8{,}0666 = 20{,}3744,$$

$$\bar{\sigma}_{v_2}^2 = \frac{254}{50} - 0{,}01 = 5{,}07.$$

Mai departe avem

$$\frac{\sigma_0}{\sigma_1} = \sqrt{\frac{\bar{\sigma}_u^2}{\bar{\sigma}_{v_1}^2}} = \sqrt{\frac{8{,}2}{20{,}3744}} = \sqrt{0{,}40246} \simeq 0{,}63,$$

$$\frac{\sigma_0}{\sigma_2} = \sqrt{\frac{\bar{\sigma}_u^2}{\bar{\sigma}_{v_2}^2}} = \sqrt{\frac{8{,}2}{5{,}07}} = \sqrt{1{,}61735} = 1{,}27,$$

$$\operatorname{cov}(u_1 v_1) = \frac{622}{50} - \bar{u}\bar{v}_1 = 12{,}44;$$

$$\operatorname{cov}(u, v_2) = \frac{284}{50} - \bar{u}\bar{v}_2 = 5{,}68,$$

$$\operatorname{cov}(v_1, v_2) = \frac{394}{50} - \bar{v}_1\bar{v}_2 = 7{,}88 + 0{,}284 = 8{,}164,$$

$$r_{01} = r_{uv_1} = \frac{12{,}44}{2{,}86 \cdot 4{,}51} = 0{,}964;$$

$$r_{02} = r_{uv_2} = \frac{5{,}68}{2{,}86 \cdot 2{,}25} = 0{,}883,$$

$$r_{12} = r_{v_1 v_2} = \frac{8{,}164}{4{,}51 \cdot 2{,}25} = 0{,}804.$$

Coeficienţii ecuaţiei de regresie au acum valorile:

$$\frac{r_{01} - r_{02} r_{12}}{1 - r_{12}^2} \cdot \frac{\sigma_0}{\sigma_1} = 0{,}46; \quad \frac{r_{02} - r_{01} r_{12}}{1 - r_{12}^2} \cdot \frac{\sigma_0}{\sigma_2} = 0{,}36.$$

Cu aceste valori, ecuaţia planului de regresie devine

$$y - 9 = 0{,}46(x_1 - 12{,}84) + 0{,}36(x_2 - 14{,}9).$$

III.34. *Se consideră variabilele aleatoare X_i, $1 \leqslant i \leqslant n$, independente şi repartizate normal $N(\mu_i, \sigma)$, $1 \leqslant i \leqslant n$. Dacă a_i, b_i, $1 \leqslant i \leqslant n$ sînt constante, să se calculeze:*

a) $\operatorname{cov}\left[\sum\limits_{j=1}^{n} a_j X_j^2, \left(\sum\limits_{j=1}^{n} b_j X_j\right)^2\right].$

b) $\operatorname{cov}\left[\left(\sum\limits_{=1}^{n} a_j X_j\right)^2, \left(\sum\limits_{j=1}^{n} b_j X_j\right)^2\right].$

c) $\operatorname{cov}\left[\sum\limits_{j=1}^{n} a_j X_j, \sum\limits_{j=1}^{n} b_j X_j^2\right].$

d) $\operatorname{cov}\left[\sum\limits_{j=1}^{n} a_j X_j, \left(\sum\limits_{j=1}^{n} b_j X_j\right)^2\right].$

S o l u ţ i e. Deoarece X_j, $1 \leqslant j \leqslant n$ sînt variabile $N(\mu_j, \sigma)$ rezultă că pentru orice $j \in \{1, 2, \ldots, n\}$ avem

$$\varphi_{X_j}(t) = \mathrm{e}^{\mathrm{i}\, t\mu_j - \frac{t^2\sigma^2}{1}}$$

şi deci

$$\frac{1}{i} \frac{\partial \varphi_{X_j}(t)}{\partial t}\bigg|_{t=0} = \mu_j,$$

$$\frac{1}{i^2} \frac{\partial^2 \varphi_{X_j}(t)}{\partial t^2}\bigg|_{t=0} = \sigma^2 + \mu_j^2,$$

$$\frac{1}{i^3} \frac{\partial^3 \varphi_{X_j}(t)}{\partial t^3}\bigg|_{t=0} = 3\mu_j\sigma^2 + \mu_j^3,$$

$$\frac{1}{i^4} \frac{\partial^4 \varphi_{X_j}(t)}{\partial t^4}\bigg|_{t=0} = 3\sigma^4 + 6\sigma^2\mu_j^2 + \mu_j^4.$$

Să evaluăm acum covarianţele amintite:

$$a)\ \operatorname{cov}\left[\sum_{j=1}^{n} a_j X_j^2, \left(\sum_{j=1}^{n} b_j X_j\right)^2\right] = M\left[\left(\sum_{j=1}^{n} a_j X_j^2\right)\left(\sum_{j=1}^{n} b_j X_j\right)^2\right] -$$

$$- M\left(\sum_{j=1}^{n} a_j X_j^2\right) M\left(\left(\sum_{j=1}^{n} b_j X_j\right)^2\right) =$$

$$= M\left[\left(\sum_{j=1}^{n} a_j X_j^2\right)\left(\sum_{k=1}^{n} b_k^2 X_k^2 + \sum_{k\neq l} b_k b_l X_k X_l\right) -\right.$$

$$-\left(\sum_{j=1}^{n} a_j M(X_j^2)\right)\left[\sum_{k=1}^{n} b_k^2 M(X_k^2) + \sum_{k\neq l} b_k b_l M(X_k) M(X_l) =\right.$$

$$= \sum_{j=1}^{n} a_j b_j^2 M(X_j^4) + \sum_{j\neq k} a_j b_k^2 M(X_j^2)\, M(X_k^2) +$$

$$+ \sum_{k\neq l} a_k b_k b_l M(X_k^2)\, M(X_k)\, M(X_l) + \sum_{j\neq k\neq l} a_j b_k b_l M(X_j^2)\, M(X_k)\, M(X_l) -$$

$$- \left[\sum_{j=1}^{n} a_j b_j^2 M^2(X_j^2) + \sum_{j\neq k} a_j b_k^2 M(X_j^2)\, M(X_k^2) +\right.$$

$$+ \sum_{j\neq k} a_j b_j b_k M(X_j^2)\, M(X_j)\, M(X_k) + \sum_{j\neq k\neq l} a_j b_k b_l M(X_j^2)\, M(X_k)\, M(X_l)].$$

Făcîndu-se înlocuirile se obţine:

$$\operatorname{cov}\left[\sum_{j=1}^{n} a_j X_j^2, \left(\sum_{k=1}^{n} b_k X_k\right)^2\right] = \sum_{j=1}^{n} a_j b_j^2 (3\sigma^4 + 6\sigma^2\mu_j^2 + \mu_j^4) -$$

$$- \sum_{j=1}^{n} a_j b_j^2 (\sigma^2 + \mu_j^2)^2 = 2\sigma^4 \sum_{j=1}^{n} a_j b_j^2.$$

$$b)\ \operatorname{cov}\left[\left(\sum_{j=1}^{n} a_j X_j\right)^2, \left(\sum_{k=1}^{n} b_k X_k\right)^2\right] =$$

$$= M\left[\left(\sum_{j=1}^{n} a_j X_j\right)^2\left(\sum_{k=1}^{n} b_k X_k\right)^2\right] - \left(M\left(\sum_{j=1}^{n} a_j X_j\right)^2 M\left(\sum_{k=1}^{n} b_k X_k\right)^2\right) =$$

$$= M\left(\sum_{j=1}^{n} a_j^2 X_j^2 + \sum_{j\neq 1} a_j a_l X_j X_l\right)\left(\sum_{k=1}^{n} b_k^2 X_k^2 + \sum_{k\neq p} b_k b_p X_k X_p\right) -$$

$$-\left[\sum_{j=1}^{n} a_j^2 M(X_j^2) + \sum_{j\neq 1} a_j a_l M(X_j)\ M(X_l)\right]\left[\sum_{k=1}^{n} b_k^2 M(X_k^2) +\right.$$

$$\left. + \sum_{k\neq p} b_k b_p M(X_k)\ M(x_p)\right] = \sum_{j=1}^{n} a_j^2 b_j^2 M(X_j^4) + \sum_{j\neq k} a_j^2 b_k^2 M(X_j^2)\ M(X_k^2) +$$

$$+ \sum_{k\neq p} a_k^2 b_k b_p M(X_k^3)\ M(X_p) + \sum_{j\neq k\neq p} a_j^2 b_k b_p M(X_j^2)\ M(X_k)\ M(X_p) +$$

$$+ \sum_{j\neq l} a_j a_l b_l^2 M(X_j)\ M(X_l^3) + \sum_{j\neq l\neq k} a_j a_l b_k^2 M(X_j)\ M(X_l)\ M(X_k^2) +$$

$$+ \sum_{j\neq 1} a_j a_l b_j b_l M(X_j^2)\ M(X_l^2) + \sum_{j\neq l\neq k} a_j a_l b_k b_l \times$$

$$\times M(X_j)\ M(X_l^2) M(X_k) + \sum_{j\neq l\neq k\neq p} a_j a_l b_k b_p M(X_j)\ M(X_l)\ M(X_k) M(X_p) -$$

$$- \sum_{j=1}^{n} a_j^2 b_j^2 M^2(X_j^2) - \sum_{j\neq k} a_j^2 b_k^2 M(X_j^2)\ M(X_k^2) -$$

$$- \sum_{j\neq k} a_j^2 b_j b_k M(X_j^2)\ M(X_j)\ M(X_k) - \sum_{j\neq k\neq p} a_j^2 b_k b_p M(X_j^2)\ M(X_k)\ M(X_p) -$$

$$- \sum_{j\neq l} a_j a_l b_l^2 M(X_j)\ M(X_l)\ M(X_l^2) - \sum_{j\neq l\neq k} a_j a_l b_k^2 M(X_j)\ M(X_l)\ M(X_k^2) -$$

$$- \sum_{j\neq l} a_j a_l b_j b_l M^2(X_j)\ M^2(X_l) - \sum_{j\neq l\neq k} a_j a_l b_l b_k M(X_j)\ M(X_l^2)\ M(X_k) -$$

$$- \sum_{j\neq l\neq k\neq p} a_j a_l b_k b_p M(X_j)\ M(X_l)\ M(X_k)\ M(X_p)$$

Se fac acum înlocuirile şi reducerea termenilor asemenea.

c) $$\operatorname{cov}\left[\sum_{i=1}^{n} a_i X_i,\ \sum_{i=1}^{n} b_i X_i^2\right] = M\left(\sum_{i=1}^{n} a_i X_i\right)\left(\sum_{i=1}^{n} b_i X_i^2\right) -$$

$$- M\left(\sum_{i=1}^{n} a_i X_i\right) M\left(\sum_{i=1}^{n} b_i X_i^2\right) =$$

$$= M\left(\sum_{i=1}^{n} a_i b_i X_i^3 + \sum_{i \neq j} a_i b_j X_i X_j^2\right) - \left(\sum_{i=1}^{n} a_i M(X_i)\right)\left(\sum_{j=1}^{n} b_j M(X_j^2)\right) =$$

$$= \sum_{i=1}^{n} a_i b_i M(X_i^3) + \sum_{i \neq j} a_i b_j M(X_i)\ M(X_j^2) - \left(\sum_{i=1}^{n} a_i M(X_i)\right)\left(\sum_{j=1}^{n} b_j M(X_j^2)\right) =$$

$$= \sum_{i=1}^{n} a_i b_i (3\mu_i \sigma^2 + \mu_i^3) + \sum_{i \neq j} a_i b_j \mu_i (\sigma^2 + \mu_j^2) - \left(\sum_{i=1}^{n} a_i \mu_i\right)\left(\sum_{j=1}^{n} b_j (\sigma^2 + \mu_j^2)\right) =$$

$$= 3\sigma^2 \sum_{i=1}^{n} a_i b_i \mu_i + \sum_{i=1}^{n} a_i b_i \mu_i^3 + \sigma^2 \sum_{i \neq j} a_i b_j \mu_i + \sum_{i \neq j} a_i b_j \mu_i \mu_j^2 -$$

$$- \left(\sum_{i=1}^{n} a_i \mu_i\right)\left(\sigma^2 \sum_{j=1}^{n} b_j + \sum_{j=1}^{n} b_j \mu_j^2\right) =$$

$$= 3\sigma^2 \sum_{i=1}^{n} a_i b_i \mu_i + \sum_{i=1}^{n} a_i b_i \mu_i^3 + \sigma^2 \sum_{i \neq j} a_i b_j \mu_i + \sum_{i \neq j} a_i b_j \mu_i \mu_j^2 -$$

$$- \sigma^2 \sum_{i=1}^{n} a_i b_i \mu_i - \sigma^2 \sum_{i \neq j} a_i b_j \mu_i - \sum_{i=1}^{n} a_i b_i \mu_i^3 - \sum_{i \neq j} a_i b_j \mu_i \mu_j^2 =$$

$$= 2\sigma^2 \sum_{i=1}^{n} a_i b_i \mu_i.$$

d) $$\operatorname{cov}\left[\sum_{i=1}^{n} a_i X_i,\ \left(\sum_{i=1}^{n} b_i X_i\right)^2\right] = M\left[\left(\sum_{i=1}^{n} a_i X_i\right)\left(\sum_{i=1}^{n} b_i X_i\right)^2\right] -$$

$$- M\left(\sum_{i=1}^{n} a_i X_i\right) M\left[\left(\sum_{i=1}^{n} b_i X_i\right)^2\right] =$$

$$= M\left[\left(\sum_{i=1}^{n} a_i X_i\right)\left(\sum_{i=1}^{n} b_i^2 X_i^2 + \sum_{i \neq j} b_i b_j X_i X_j\right)\right] -$$

$$-\left(\sum_{i=1}^{n} a_i\mu_i\right)\left(\sum_{i=1}^{n} b_i^2(\sigma^2+\mu_i^2)+2\sum_{i\neq j} b_ib_j\mu_i\mu_j\right)=$$

$$=M\left(\sum_{i=1}^{n} a_ib_i^2X_i^3+\sum_{i\neq j}a_ib_j^2X_iX_j^2+\sum_{i\neq j}a_ib_ib_jX_i^2X_j+\sum_{i\neq j\neq k}a_ib_jb_kX_iX_jX_k-\right.$$

$$-\left(\sum_{i=1}^{n} a_i\mu_i\right)\left(\sigma^2\sum_{i=1}^{n} b_i^2+\sum_{i=1}^{n} b_i^2\mu_i^2+\sum_{i\neq j}b_ib_j\mu_i\mu_j\right)=$$

$$=\sum_{i=1}^{n} a_ib_i^2(3\sigma^2\mu_i+\mu_i^3)+\sum_{i\neq j}a_ib_j^2\mu_i(\mu_j^2+\sigma^2)+$$

$$+\sum_{i\neq j}a_ib_ib_j(\mu_i^2+\sigma^2)\,\mu_j+\sum_{i\neq j\neq k}a_ib_jb_k\mu_i\mu_j\mu_k-$$

$$-\sigma^2\sum_{i=1}^{n}a_ib_i^2\mu_i-\sigma^2\sum_{i\neq j}a_ib_j^2\mu_i-\sum_{i=1}^{n}a_ib_i^2\mu_i^3-$$

$$-\sum_{i\neq j}a_ib_j^2\mu_i\mu_j^2-\sum_{i\neq j}a_ib_ib_j\mu_i^2\mu_j-\sum_{i\neq j\neq k}a_ib_jb_k\mu_i\mu_j\mu_k=$$

$$=2\sigma^2\sum_{i=1}^{n}a_ib_i^2\mu_i+\sigma^2\sum_{i\neq j}a_ib_ib_j\mu_j.$$

III.35. *Se consideră vectorul aleator* (X_1, X_2) *a cărui funcție generatoare de momente este dată de*

$$\psi_{X_1,X_2}(t_1, t_2)=a(e^{t_1+t_2}+1)+b(e^{t_1}+e^{t_2})$$

cu a, b constante astfel încît $2a+2b=1$. Să se calculeze

$$M(X_i),\ D^2(X_i),\quad i=1,2;\quad \operatorname{cov}(X_1, X_2)\ \text{ și }\ \rho(X_1, X_2).$$

Să se afle între ce limite variază $\rho(X_1, X_2)$.

S o l u ţ i e. Din legătura dintre momente şi funcţia generatoare de momente, obţinem

$$M(X_1) = \frac{\partial\psi_{X_1, X_2}(t_1, t_2)}{\partial t_1}\bigg|_{t_1=t_2=0} = a + b = \frac{1}{2},$$

$$M(X_2) = \frac{1}{2},$$

$$D^2(X_1) = \frac{1}{4}; \qquad D^2(X_2) = \frac{1}{4},$$

$$\operatorname{cov}(X_1, X_2) = a - \frac{1}{4}; \qquad \rho(X_1, X_2) = \frac{a - \frac{1}{4}}{\frac{1}{4}} = 4a - 1.$$

Din datele problemei $0 \leqslant a \leqslant \frac{1}{2}$, $0 \leqslant b \leqslant \frac{1}{2}$ şi deci $-1 \leqslant \rho(X_1, X_2) \leqslant 0$.

III.36. *Se consideră vectorul aleator* (X_1, X_2) *cu funcţia generatoare de momente*

$$\psi_{X_1, X_2}(t_1, t_2) = [a(e^{t_1+t_2} + 1) + b(e^{t_1+t_2})]^2$$

cu a, b *constante pozitive astfel încît* $2a + 2b = 1$.

Să se afle

$$M(X_i),\ D^2(X_i),\ i = 1, 2; \quad \operatorname{cov}(X_1, X_2) \text{ şi } \rho(X_1, X_2).$$

S o l u ţ i e.

$$M(X_1) = \frac{\partial\psi_{X_1, X_2}(t_1, t_2)}{\partial t_1}\bigg|_{t_1=t_2=0} = 4(a + b)^2 = 1,$$

$$M(X_2) = 1,$$

$$D^2(X_1) = M(X_1^2) - M^2(X_1).$$

Cum

$$M(X_1^2) = \frac{\partial^2\psi_{X_1, X_2}(t_1, t_2)}{\partial t_1^2}\bigg|_{t_1+t_2=0} = 6(a + b)^2 = \frac{3}{2},$$

rezultă

$$D^2(X_1) = \frac{3}{2} - 1 = \frac{1}{2}.$$

Analog

$$D^2(X_2) = \frac{1}{2},$$

$$\operatorname{cov}(X_1, X_2) = M(X_1X_2) - M(X_1)\, M(X_2)$$

cum

$$M(X_1X_2) = \left.\frac{\partial^2\psi_{X_1,X_2}(t_1, t_2)}{\partial t_1 \partial t_2}\right|_{t_1=t_2=0} = 2a + \frac{1}{2},$$

rezultă că

$$\operatorname{cov}(X_1, X_2) = 2a - \frac{1}{2}$$

şi

$$\rho(X_1, X_2) = \frac{2a - \frac{1}{2}}{\frac{1}{2}} = 4a - 1.$$

Observaţie. Coeficientul de corelaţie este cuprins între aceleaşi limite ca şi în cazul precedent, consideraţiile fiind absolut aceleaşi.

III.73. *Se consideră vectorul aleator* (X_1, X_2), *cu funcţia generatoare de momente* $\psi_{X_1,X_2}(t_1, t_2)$ *satisfăcînd relaţia:*

$$\ln \psi_{X_1,X_2}(t_1, t_2) = \nu \int_{-\infty}^{\infty} \mathrm{d}u \int_{-\infty}^{\infty} \mathrm{d}y\, f_Y(y)\, \{\mathrm{e}^{y[t_1w_1(u)+t_2w_2(u)]} - 1\},$$

unde Y *este o variabilă aleatoare cu densitatea de repartiţie* $f_Y(.\,)$ *iar* $w_1(.)$ *şi* $w_2(.)$ *sînt funcţii cunoscute;* $\nu > 0$.

a) *Să se arate că*

$$M(X_i) = \nu M(Y) \int_{-\infty}^{\infty} w_i(u)\, \mathrm{d}u,$$

$$D^2(X_i) = \nu M(Y^2) \int_{-\infty}^{\infty} w_i^2(u)\, \mathrm{d}u, \qquad i = 1, 2,$$

și

$$\operatorname{cov}(X_1, X_2) = \nu M(Y^2) \int_{-\infty}^{\infty} w_1(u)\, w_2(u)\, du.$$

b) *Să se calculeze* $M(X_i)$, $D^2(X_i)$, $i = 1, 2$ *precum și* $\operatorname{cov}(X_1, X_2)$, $\rho(X_1, X_2)$, *în cazul cînd*

$$w_i(u) = \begin{cases} e^{-(u-a_i)}, & u \geqslant a_i, \\ 0 & , \quad u < a_i, \end{cases}$$

cu a_1, a_2 *constante care satisfac relația* $0 < a_1 < a_2$, *variabila aleatoare* Y *este uniform repartizată pe intervalul* $[-1, 1]$ *iar* $\nu = 4$.

Soluție *a)* Din relația dată rezultă că

$$\psi_{X_1, X_2}(t_1, t_2) = e^{\nu \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dy f_Y(y) \{e^{y[t_1 w_1(u) + t_2 w_2(u)]} - 1\}}.$$

Pe de altă parte

$$M(X_1) = \left.\frac{\partial \psi_{X_1, X_2}(t_1, t_2)}{\partial t_1}\right|_{t_1 = t_2 = 0}.$$

Dar

$$\left.\frac{\partial \psi_{X_1, X_2}(t_1, t_2)}{\partial t_1}\right|_{t_1 = t_2 = 0} = \nu \int_{-\infty}^{\infty} du \left(\int_{-\infty}^{\infty} dy\, f_Y(y)\, y w_1(u)\right) =$$

$$= \nu \int_{-\infty}^{\infty} y f_Y(Y)\, dy \int_{-\infty}^{\infty} w_1(u)\, du = \nu M(Y) \int_{-\infty}^{\infty} w_1(u)\, du.$$

În mod cu totul analog rezultă

$$\left.\frac{\partial \psi_{X_1, X_2}(t_1, t_2)}{\partial t_2}\right|_{t_1 = t_2 = 0} = \nu \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dy\, f_Y(y)\, y w_2(u) =$$

$$= \nu M(Y) \int_{-\infty}^{\infty} w_2(u)\, du.$$

Mai departe,

$$M(X_1^2) = \left.\frac{\partial^2 \psi_{X_1, X_2}(t_1, t_2)}{\partial t_1^2}\right|_{t_1 = t_2 = 0} = \nu \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dy\, f_Y(y)\, y^2 w_1^2(u) +$$

$$+ \left(\nu \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dy\, f_Y(y)\, y w_1(u)\right)^2 =$$

$$= \nu M(Y^2) \int_{-\infty}^{\infty} w_1^2(u)\, du + \left[\nu M(Y) \int_{-\infty}^{\infty} w_1(u)\, du\right]^2.$$

De aici rezultă imediat că

$$D^2(X_1) = \nu M(Y^2) \int_{-\infty}^{\infty} w_1^2(u)\, \mathrm{d}u$$

şi în mod cu totul analog

$$M(X_2^2) = \left.\frac{\partial^2 \psi_{X_1, X_2}(t_1, t_2)}{\partial t_2^2}\right|_{t_1=t_2=0} =$$

$$= \nu \int_{-\infty}^{\infty} \mathrm{d}u \int_{-\infty}^{\infty} \mathrm{d}y\, f_Y(y)\, y^2 w_1^2(u) + \left(\nu \int_{-\infty}^{\infty} \mathrm{d}u \int_{-\infty}^{\infty} \mathrm{d}y\, f_Y(y)\, y w_2(u)\right)^2$$

şi de aici

$$D^2(X_2) = \nu M(Y^2) \int_{-\infty}^{\infty} w_2^2(u)\, \mathrm{d}u.$$

Pentru cov (X_1, X_2) vom calcula mai întîi $M(X_1, X_2)$.

$$M(X_1 X_2) = \left.\frac{\partial^2 \psi_{X_1, X_2}(t_1, t_2)}{\partial t_1 \partial t_2}\right|_{t_1=t_2=0} =$$

$$= \nu \int_{-\infty}^{\infty} \mathrm{d}u \int_{-\infty}^{\infty} \mathrm{d}y\, f_Y(y)\, y^2 w_1(u)\, w_2(u) + \left(\nu \int_{-\infty}^{\infty} \mathrm{d}u \int_{-\infty}^{\infty} \mathrm{d}y\, f_Y(y)\, y w_1(u)\right) \times$$

$$\times \left(\nu \int_{-\infty}^{\infty} \mathrm{d}u \int_{-\infty}^{\infty} \mathrm{d}y\, f_Y(y)\, y w_2(u)\right)$$

sau

$$M(X_1 X_2) = \nu M(Y^2) \int_{-\infty}^{\infty} w_1(u)\, w_2(u)\, \mathrm{d}u +$$

$$+ \nu^2 M^2(Y) \left(\int_{-\infty}^{\infty} w_1(u)\, \mathrm{d}u\right) \left(\int_{-\infty}^{\infty} w_2(u)\, \mathrm{d}u\right)$$

şi de aici

$$\mathrm{cov}\,(X_1, X_2) = \nu M(Y^2) \int_{-\infty}^{\infty} w_1(u)\, w_2(u)\, \mathrm{d}u.$$

b) În cazul particular considerat obţinem

$$M(Y) = \frac{1}{2}; \qquad M(Y^2) = \frac{2}{3},$$

$$\int_{-\infty}^{\infty} w_1(u)\,\mathrm{d}u = \int_{a_1}^{\infty} \mathrm{e}^{-(u-a_1)}\,\mathrm{d}u = 1.$$

şi deci

$$M(X_1) = 4 \cdot \frac{1}{2} = 2.$$

Deoarece

$$\int_{-\infty}^{\infty} w_2(u)\,\mathrm{d}u = \int_{a_2}^{\infty} \mathrm{e}^{-(u-a_2)}\,\mathrm{d}u = 1$$

avem şi

$$M(X_2) = 2.$$

Să calculăm acum

$$\int_{-\infty}^{\infty} w_1^2(u)\,\mathrm{d}u = \int_{a_1}^{\infty} \mathrm{e}^{-2(u-a_1)}\,\mathrm{d}u = \frac{1}{2},$$

$$\int_{-\infty}^{\infty} w_2^2(u)\,\mathrm{d}u = \int_{a_1}^{\infty} \mathrm{e}^{-2(u-a_1)}\,\mathrm{d}u = \frac{1}{2}$$

şi

$$\int_{-\infty}^{\infty} w_1(u)\, w_2(u)\,\mathrm{d}u = \int_{a_2}^{\infty} \mathrm{e}^{-(u-a_1)}\, \mathrm{e}^{-(u-a_2)}\,\mathrm{d}u = \frac{1}{2}\,\mathrm{e}^{-(a_2-a_1)}.$$

Cu aceste rezultate obţinem

$$D^2(X_1) = D^2(X_2) = 4 \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{4}{3},$$

$$\operatorname{cov}(X_1, X_2) = 4 \cdot \frac{2}{3} \cdot \frac{1}{2}\,\mathrm{e}^{-(a_2-a_1)} = \frac{4}{3}\,\mathrm{e}^{-(a_2-a_1)}.$$

În fine

$$\rho(X_1, X_2) = \mathrm{e}^{-(a_2-a_1)}.$$

De aici se vede că

$$0 < \rho(X_1, X_2) < 1.$$

III. 38. *Se consideră vectorul aleator* (X_1, X_2, X_3) *repartizat normal, tridimensional, nedegenerat cu vectorul valoare medie* (0, 0, 0) *şi matricea de covarianţă.*

$$C = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix}.$$

a) *Să se afle coeficientul parţial de corelaţie al variabilelor* X_1 *şi* X_2 *pentru* $X_3 = x_3$ *dat.*

b) *Să se calculeze* $\rho_{12\cdot 3}$, $\rho_{23\cdot 1}$ şi $\rho_{13\cdot 2}$ *în cazul cînd*

$$\rho_{12} = \frac{1}{2}; \quad \rho_{13} = \frac{1}{2^2}; \quad \rho_{23} = \frac{1}{2^3}.$$

S o l u ţ i e. *a)* Deoarece densitatea de repartiţie a vectorului aleator (X_1, X_2, X_3) este dată de

$$f(X_1, X_2, X_3) = \frac{1}{(2\pi)^{3/2}\sqrt{|C|}} \exp\left\{\frac{1}{2C} \sum_{i,j=1}^{3} C_{ij}x_i x_j\right\},$$

unde C_{ij} este complementul algebric corespunzător elementului de la intersecţia liniei i cu coloana j din determinantul

$$|C| = \begin{vmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{vmatrix},$$

atunci densitatea de repartiţie condiţionată comună a variabilelor X_1, X_2 pentru $X_3 = x_3$ este dată de

$$g(x_1, x_2 \mid x_3) = k \exp\left\{-\frac{1}{2}(C^{11}x_1^2 + 2C^{12}x_1x_2 + C^{22}x_2^2 + 2C^{13}x_1x_3 + \right.$$

$$\left. + C^{23}x_2x_3)\right\},$$

unde

$$C^{ij} = \frac{C_{ij}}{|C|}.$$

Densitatea de repartiţie $g(x_1,x_2/x_3)$ se poate pune sub forma

$$g(x_1, x_2/x_3) = k \exp\left\{-\frac{1}{2} C^{11}(x_1 - u_1)^2 + 2C^{12}(x_1 - u_1)(x_2 - u_2) + \right.$$

$$\left. + C^{22}(x_2 - u_2)^2]\right\}$$

unde

$$C^{11}u_1 + C^{12}u_2 = - C^{13}x_3,$$

$$C^{12}u_1 + C^{22}u_2 = C^{23}x_3.$$

Pe expresia densităţii de repartiţie la care am ajuns constatăm că pentru x_3 dat X_1, X_2 sînt variabile aleatoare repartizate normal bidimensional cu coeficientul de corelaţie

$$\rho_{12\cdot 3} = - \frac{C^{12}}{(C^{11}C^{22})^{1/2}}.$$

Ţinînd cont de faptul că

$$C^{ij} = \frac{C_{ij}}{|C|},$$

obţinem

$$\rho_{12\cdot 3} = - \frac{C_{12}}{(C_{11}C_{22})^{1/2}}.$$

Din expresia determinantului $|C|$ constatăm că

$$C_{11} = 1 - \rho_{23}^2; \quad C_{22} = 1 - \rho_{13}^2; \quad C_{12} = -(\rho_{12} - \rho_{13}\rho_{23})$$

şi deci

$$\rho_{12\cdot 3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{[(1 - \rho_{13}^2)(1 - \rho_{23}^2)]^{1/2}}.$$

Un raţionament analog ne permite să scriem şi coeficientul de corelaţie parţială al variabilelor X_2, X_3 pentru $X_1 = x_1$ sau coeficientul de corelaţie parţială al variabilelor X_1, X_3 pentru $X_2 = x_2$ şi anume

$$\rho_{23\cdot 1} = \frac{\rho_{23} - \rho_{12}\rho_{13}}{[(1-\rho_{12}^2)(1-\rho_{13}^2)]^{1/2}}; \qquad \rho_{13\cdot 2} = \frac{\rho_{13} - \rho_{12}\rho_{23}}{[(1-\rho_{12}^2)(1-\rho_{13}^2])^{1/2}}.$$

b)

$$\rho_{12\cdot 3} = \frac{\frac{1}{2} - \frac{1}{2^2}\cdot\frac{1}{2^3}}{\left[\left(1-\frac{1}{2^4}\right)\left(1-\frac{1}{2^6}\right)\right]^{1/2}} = \sqrt{\frac{15}{63}};$$

$$\rho_{13\cdot 2} = \frac{\frac{1}{2^2} - \frac{1}{2}\cdot\frac{1}{2^3}}{\left[\left(1-\frac{1}{2^2}\right)\left(1-\frac{1}{2^6}\right)\right]^{1/2}} = \sqrt{\frac{3}{63}};$$

$$\rho_{23\cdot 1} = \frac{\frac{1}{2^3} - \frac{1}{2}\cdot\frac{1}{2^2}}{\left[\left(1-\frac{1}{2^2}\right)\left(1-\frac{1}{2^4}\right)\right]^{1/2}} = 0.$$

Capitolul IV

TEORIA ESTIMAŢIEI

IV.1. *Să se precizeze care dintre estimaţiile:*

$$\hat{\theta}_a = X_1 - X_2, \qquad \hat{\theta}_h = \frac{1}{n}\sum_{1}^{n} |X_j|,$$

$$\hat{\theta}_b = \frac{X_1 - X_2}{X_3}, \qquad \hat{\theta}_i = |X_1| - |X_2| + |X_3|,$$

$$\hat{\theta}_c = X_1^2 - X_2^2, \qquad \hat{\theta}_j = \overline{X}_n / S_n^2,$$

$$\hat{\theta}_d = \frac{X_1^2 - X_2^2}{X_3^2}, \qquad \hat{\theta}_k = \overline{X}_n / S_n,$$

$$\hat{\theta}_e = \sin^2 X_1 + \sin^2 X_2, \qquad \hat{\theta}_l = \overline{X}_n^2 / S_n^2,$$

$$\hat{\theta}_f = X_1 + X_2 - X_3, \qquad \hat{\theta}_m = M_n^2,$$

$$\hat{\theta}_g = \frac{X_1^2 X_2}{X_3}, \qquad \hat{\theta}_n = M_n^2 / S_n^2,$$

sînt: a) *local invariante*; b) *invariante de scară*; c) *invariante*; d) *local invariante de primul ordin*; e) *invariante de scară de primul ordin, de al doilea ordin*; f) *invariante reflexive*; g) *invariante la permutare.*

Soluţie. a) Estimaţia $\hat{\theta}(X_1, X_2, ..., X_n)$ este local invariantă dacă $\hat{\theta}(X_1 + b, X_2 + b, ..., X_n + b) = \hat{\theta}(X_1, X_2, ..., X_n)$.

Deoarece

$$(X_1 + b) - (X_2 + b) = X_1 - X_2,$$

$$\frac{1}{n}\sum_1^n\left[(X_i + b) - \frac{1}{n}\sum_1^n (X_i + b)\right]^2 = \frac{1}{n}\sum_1^n\left(X_i - \frac{1}{n}\sum_1^n X_i\right)^2,$$

$$\frac{1}{n-1}\sum_1^n\left[(X_i + b) - \frac{1}{n}\sum_1^n (X_i + b)\right]^2 = \frac{1}{n-1}\sum_1^n\left(X_i - \frac{1}{n}\sum_1^n X_i\right)^2,$$

urmează că $\hat{\theta}_a$, $\hat{\theta}_m$ şi $\hat{\theta}_n$ sînt estimaţii local invariante.

b) Estimaţia $\hat{\theta}(X_1, X_2, \ldots, X_n)$ este invariantă de scară dacă

$$\hat{\theta}(aX_1, aX_2, \ldots, aX_n) = \hat{\theta}(X_1, X_2, \ldots, X_n).$$

Deoarece

$$\frac{aX_1 - aX_2}{aX_3} = \frac{X_1 - X_2}{X_3},$$

$$\frac{a^2X_1^2 - a^2X_2^2}{a^2X_3^2} = \frac{X_1^2 - X_2^2}{X_3^2},$$

$$\frac{\left(\frac{1}{n}\sum_1^n aX_i\right)^2}{\frac{1}{n-1}\sum_1^n\left[aX_i - \frac{1}{n}\sum_1^n aX_i\right]^2} = \frac{\frac{1}{n}\sum_1^n X_i}{\frac{1}{n-1}\sum_1^n\left[X_i - \frac{1}{n}\sum_1^n X_i\right]^2},$$

$$\frac{\frac{1}{n}\sum_1^n a_iX_i}{\sqrt{\frac{1}{n-1}\sum_1^n\left[aX_i\ \frac{1}{n}\sum_1^n aX_i\right]^2}} = \frac{\frac{1}{n}\sum_1^n X_i}{\sqrt{\frac{1}{n-1}\sum_1^n\left(X_i - \frac{1}{n}\sum_1^n X_i\right)^2}},$$

$$\frac{\frac{1}{n}\sum_1^n\left(aX_i - \frac{1}{n}\sum_1^n aX_i\right)^2}{\frac{1}{n-1}\sum_1^n\left(aX_i - \frac{1}{n}\sum_1^n aX_i\right)^2} = \frac{\frac{1}{n}\sum_1^n\left(X_i - \frac{1}{n}\sum_1^n X_i\right)^2}{\frac{1}{n-1}\sum_1^n\left(X_i - \frac{1}{n}\sum_1^n X_i\right)^2},$$

urmează că $\hat{\theta}_b$, $\hat{\theta}_a$, $\hat{\theta}_k$, $\hat{\theta}_l$, $\hat{\theta}_n$ sînt estimaţii invariante de scară.

c) Estimaţia $\hat{\theta}(X_1, X_2, \ldots, X_n)$ este invariantă, dacă

$$\hat{\theta}(aX_1 + b,\ aX_2 + b, \ldots, aX_n + b) = \hat{\theta}(X_1, X_2, \ldots, X_n).$$

Deoarece

$$\frac{1}{n}\sum_{1}^{n}\left[\left(aX_i + b\right) - \frac{1}{n}\sum_{1}^{n}(aX_i + b)\right]^2 = a^2\frac{1}{n}\sum_{1}^{n}\left(X_i - \frac{1}{n}\sum_{1}^{n}X_i\right)^2,$$

$$\frac{1}{n-1}\sum_{1}^{n}\left[(aX_i + b) - \frac{1}{n}\sum_{1}^{n}(aX_i + b)\right]^2 =$$

$$= a^2\frac{1}{n-1}\sum_{1}^{n}\left(X_i - \frac{1}{n}\sum_{1}^{n}X_i\right)^2,$$

urmează că M_n^2/S_n^2 este o estimaţie invariantă.

d) Estimaţia $\hat{\theta}(X_1, X_2. ..., X_n)$ este local invariantă de ordinul k dacă

$$\hat{\theta}(X_1 + b, \quad X_2 + b, ..., X_n + b) = \hat{\theta}(X_1, X_2, ..., X_n) + kb.$$

Deoarece

$$(X_1 + b) + (X_2 + b) - (X_3 + b) = X_1 + X_2 + X_3 + b,$$

urmează că $\hat{\theta}_f$ este local invariantă de ordinul 1.

e) Estimaţia $\hat{\theta}(X_1, X_2, ..., X_n)$ este invariantă de scară de ordinul k dacă

$$\hat{\theta}(aX_1, aX_2, ..., aX_n) = a^k\hat{\theta}(X_1, X_2, ..., X_n).$$

Deoarece

$$\frac{(aX_1)^2\, aX_2}{aX_3} = a^2\frac{X_1X_2}{X_3}$$

$$\frac{1}{n}\sum_{1}^{n}\left(aX_i - \frac{1}{n}\sum_{1}^{n}aX_i\right)^2 = a^2\frac{1}{n}\sum_{1}^{n}\left(X_i - \frac{1}{n}\sum_{1}^{n}X_i\right)^2,$$

urmează că $\hat{\theta}_g$ şi M_n^2 sînt invarianţi de scară de ordinul doi.

f) Estimaţia $\hat{\theta}(X_1, X_2, ..., X_n)$ este invariant reflexivă dacă

$$\hat{\theta}(\pm X_1, \pm X_2, ..., \pm X_n) = \hat{\theta}(X_1, X_2, ..., X_n)$$

(adică folosind pe X_i cu semnul plus sau minus, valoarea estimaţiei nu se schimbă).

Deoarece

$$(\pm X_1)^2 - (\pm X_2)^2 = X_1^2 - X_2^2,$$

$$\frac{(\pm X_1)^2 - (\pm X_2)^2}{(\pm X_3)^2} = \frac{X_1^2 - X_2^2}{X_3^2},$$

$$\sin^2(\pm X_1) + \sin^2(\pm X_2) = \sin^2 X_1 + \sin^2 X_2,$$

$$\frac{1}{n}\sum_1^n |X_i| = \frac{1}{n}\sum_1^n |\pm X_i|,$$

$$|X_1| - |X_2| + |X_3| = |\pm X_1| - |\pm X_2| + |\pm X_3|,$$

urmează că $\hat{\theta}_c$, $\hat{\theta}_d$, $\hat{\theta}_l$, $\hat{\theta}_h$, şi $\hat{\theta}_i$ sînt invariante reflexive.

g) Estimaţia $\hat{\theta}(X_1, X_2, ..., X_n)$ este invariantă la permutare dacă

$$\hat{\theta}(X_{i_1}, X_{i_2}, ..., X_{i_n}) = \hat{\theta}(X_1, X_2, ..., X_n),$$

unde $i_1, i_2, ..., i_n$ este orice permutare a indicilor 1, 2, ..., n. (Aceasta arată că estimaţia depinde numai de valorile variabilelor aleatoare şi nu de ordinea lor).

Deoarece

$$\sin^2 X_1 + \sin^2 X_2 = \sin^2 X_2 + \sin^2 X_1,$$

$$\frac{1}{n}\sum_{i_k=1}^n |X_{i_k}| = \frac{1}{n}\sum_1^n |X_j|,$$

$$\left.\frac{\frac{1}{n}\sum\limits_{i_k=1}^n X_{i_k}}{\sqrt{\frac{1}{n-1}\sum\limits_{i_k=1}^n\left(X_{i_k} - \frac{1}{n}\sum\limits_{j_k=1}^n X_{j_k}\right)^2}}\right|^2 = \left.\frac{\frac{1}{n}\sum\limits_1^n X_j}{\frac{1}{n-1}\sum\limits_{1}^n\left(X_j - \frac{1}{n}\sum\limits_1^n{}_j X\right)^2}\right|^2,$$

$$\frac{1}{n}\sum_{i_k=1}^n\left(X_{i_k} - \frac{1}{n}\sum_{j_k=1}^n X_{j_k}\right)^2 = \frac{1}{n}\sum_{i=1}^n\left(X_i - \frac{1}{n}\sum_1^n X_i\right)^2,$$

urmează că $\hat{\theta}_e$, $\hat{\theta}_h$, $\hat{\theta}_j$, $\hat{\theta}_k$, $\hat{\theta}_l$, $\hat{\theta}_m$, şi $\hat{\theta}_n$ sînt invariante la permutare.

IV.2. *Dintr-o populaţie conţinînd D articole defecte şi N—D corespunzătoare se consideră o selecţie de volum n. Fie d numărul de articole defecte găsite în selecţie.*

a) *Să se arate că $\frac{d}{n}$ este o estimaţie nedeplasată pentru fracţia defectelor $\frac{D}{N}$ din populaţie şi să se calculeze dispersia acestei estimaţii.*

b) *Să se arate că $\frac{d}{n}$ este o estimaţie nedeplasată pentru fracţia defectelor $\frac{D-d}{N-n}$ în populaţia din care am considerat selecţia de volum n şi să se calculeze dispersia acestei estimaţii.*

Soluţie. *a)* Eroarea de selecţie e este dată de diferenţa dintre estimaţie şi cantitatea de estimat.

Repartiţia erorii de selecţie are media

$$M(e) = M\left[\frac{d}{n} - \frac{D}{N}\right] = 0,$$

ceea ce arată că $\frac{d}{n}$ este o estimaţie nedeplasată pentru $\frac{D}{N}$.

Dispersia repartiţiei erorii de selecţie este

$$D^2(e) = M\left[\frac{d}{n} - \frac{D}{N}\right]^2 = \frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \frac{D}{N}\left[1 - \frac{D}{N}\right]. \qquad \text{(IV.1)}$$

b) În acest caz eroarea de selecţie e^* este

$$e^* = \frac{d}{n} - \frac{D-d}{N-n} = \frac{N}{N-n}\left[\frac{d}{n} - \frac{D}{N}\right]$$

şi

$$M(e^*) = \frac{N}{N-n} M\left[\frac{d}{n} - \frac{D}{N}\right] = 0,$$

ceea ce arată că $\frac{d}{n}$ este o estimaţie nedeplasată pentru $\frac{D-d}{N-n}$.

Repartiţia erorii de selecţie are dispersia

$$D^2(e^*) = M\left[\frac{N}{N-n}\left(\frac{d}{n} - \frac{D}{N}\right)\right]^2 = \left(\frac{N}{N-n}\right)^2 D^2(e). \qquad \text{(IV.2)}$$

Deci $D^2(e^*) > D^2(e)$.

În cele mai multe situaţii în care măsurătorile sînt distructive, fracţia de selecţie $\frac{n}{N}$ este mică, valorile date de (IV.1) şi (IV.2) fiind foarte apropiate.

IV.3. *Presupunem că o selecţie de volum n este luată dintr-o populaţie formată din N articole cu valorile* $Y_1, Y_2, \ldots, Y_N$.

a) *Să se arate că* $\bar{y}$*—media aritmetică a selecţiei este o estimaţie nedeplasată pentru* $\bar{Y} = \frac{Y_1 + Y_2 + \ldots + Y_N}{N}$ *= media populaţiei şi să se calculeze dispersia acestei estimaţii.*

b) *Să se arate că* $\bar{y}$ *este o estimaţie nedeplasată pentru* $(Y - n\bar{y})/(N-n)$, *unde* $Y = \sum_{1}^{N} Y_j$ *şi să se calculeze dispersia acestei estimaţii.*

Soluţie. *a)* Repartiţia erorii de selecţie ε are media

$$M(\varepsilon) = M(\bar{y} - \bar{Y}) = 0,$$

ceea ce arată că $\bar{y}$ este o estimaţie nedeplasată pentru $\bar{Y}$.

Avem

$$D^2(\varepsilon) = M(\bar{y} - \bar{Y})^2 = \left[1 - \frac{n}{N}\right]\frac{S^2}{n},$$

unde

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2.$$

b) În acest caz eroarea de selecţie ε^* este

$$\varepsilon^* = \bar{y} - \frac{Y - n\bar{y}}{N-n} = \frac{N}{N-n}(\bar{y} - \bar{Y}).$$

Avem

$$M(\varepsilon^*) = \frac{N}{N-n} M(\bar{y} - \bar{Y}) = 0,$$

ceea ce arată că $\bar{y}$ este o estimaţie nedeplasată pentru $\frac{Y - n\bar{y}}{N-n}$,

$$D^2(\varepsilon^*) = M\left[\frac{N}{N-n}(\bar{y} - \bar{Y})\right]^2 =$$

$$= \left(\frac{N}{N-n}\right)^2\left(1 - \frac{n}{N}\right)\frac{S^2}{n} = \left(\frac{N}{N-n}\right)^2 D^2(e).$$

Observaţie. 1) Rezultatele anterioare pot fi folosite la compararea estimaţiei $N\bar{y}$ pentru Y care are dispersia $N^2D^2(\varepsilon)$ cu estimaţia $(N-n)\bar{y}$ pentru $Y - n\bar{y}$, cînd măsurătorile sînt distructive, cu dispersia $(N-n)^2 D^2(\varepsilon^*) = N^2D^2(\varepsilon)$. Aici cele două dispersii sînt egale.

2) O estimaţie nedeplasată pentru $D^2(\varepsilon^*)$ este

$$\left(\frac{N}{N-n}\right)^2\left(1-\frac{n}{N}\right)\frac{s^2}{n}$$

unde

$$s^2 = \frac{1}{n-1}\sum_{1}^{n}(y_i-\bar{y})^2.$$

IV.4. *Fie* $(x_1, x_2, \ldots, x_n)$ *o selecţie dintr-o populaţie cu media* μ *şi dispersia* σ^2. *Dacă* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} w_i x_i$ *unde* $w_i \geqslant 0$ *şi* $\sum_{i=1}^{n} w_i = 1$, *să se arate că*

$$\hat{D}^2(\bar{X}) = \left(\sum_{i=1}^{n} w_i^2 x_i^2 - \bar{x}^2\sum_{i=1}^{n} w_i^2\right)\Big/\left(1-\sum_{i=1}^{n} w_i^2\right)$$

este o estimaţie nedeplasată a lui $D^2(\bar{X})$.

S o l u ţ i e. Suma $\sum_{i=1}^{n} w_i^2 x_i^2$ o exprimăm după cum urmează

$$\sum_{i=1}^{n} w_i^2 x_i^2 = \sum_{i=1}^{n} w_i^2[(x_i-\mu)+\mu]^2 = \sum_{i=1}^{n} w_i^2(y_i-\mu)^2 +$$

$$+2\mu\sum_{i=1}^{n} w_i^2(y_i-\mu)+\mu^2\sum_{i=1}^{n} w_i^2 =$$

$$= M^{-1}\left[\sum_{i=1}^{n} w_i^2 D^2(x_i) + 0 + \mu^2\sum_{i=1}^{n} w_i^2\right],$$

unde M^{-1} este operatorul invers operatorului valoare medie.

Egalitatea obţinută mai sus o putem scrie astfel:

$$\sum_{i=1}^{n} w_i^2 x_i^2 = M^{-1}\left[D^2(\bar{x}) + \mu^2\sum_{i=1}^{n} w_i^2\right].$$

În mod analog avem

$$\bar{x}^2 = [(\bar{x}-\mu)+\mu]^2 = (\bar{x}-\mu)^2 + 2\mu(\bar{x}-\mu)+\mu^2 =$$

$$= M^{-1}[D^2(\bar{x}) + 0 + \mu^2],$$

deoarece $M(\bar{x}) = \mu$.

Acum obţinem

$$\sum_{i=1}^{n} w_i^2 x_i^2 - \bar{x}^2 \sum_{i=1}^{n} w_i^2 = M^{-1} \left\{ \left(1 - \sum_{i=1}^{n} w_i^2\right) D^2(\bar{x}) \right\}$$

sau

$$M^{-1}(D^2(\bar{x})) = \frac{\sum\limits_{i=1}^{n} w_i^2 x_i^2 - \bar{x}^2 \sum\limits_{i=1}^{n} w_i^2}{1 - \sum\limits_{i=1}^{n} w_i^2}$$

de unde rezultă

$$M \left(\frac{\sum\limits_{i=1}^{n} w_i^2 x_i^2 - \bar{x}^2 \sum\limits_{1=1}^{n} w_i^2}{1 - \sum\limits_{i=1}^{n} w_i^2} \right) = D^2(\bar{x}),$$

ceea ce dovedeşte faptul că statistica

$$\frac{\sum\limits_{i=1}^{n} w_i^2 x_i^2 - \bar{x}^2 \sum\limits_{i=1}^{n} w_i^2}{1 - \sum\limits_{i=1}^{n} w_i^2}$$

este o estimaţie nedeplasată a dispersiei $D^2(\bar{x})$.

IV.5. *Fie* $(x_1, x_2, \ldots, x_n)$ $n = v + 1$ *o selecţie dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare* X *repartizată* $N(\mu, \sigma^2)$.

a) *Să se găsească o estimaţie nedeplasată pentru* σ^p *(unde* $v + p > 0$).

b) *Să se găsească estimaţia de eroare medie pătratică minimă pentru*

$$\sigma^p\left(p > -\frac{1}{2}(n-1)\right).$$

S o l u ţ i e. *a)* Dacă $S^2 = \sum\limits_{1}^{n} (x_i - \bar{x}_n)$ atunci S/σ este repartizată $\chi^2(v)$. Avem

$$M(S^p) = 2^{\frac{p}{2}} \sigma^p \Gamma\left(\frac{v+p}{2}\right) \Big/ \Gamma\left(\frac{v}{2}\right),$$

ceea ce arată că $t_p = c_p S^p$, unde

$$c_p = 2^{-\frac{p}{2}} \Gamma\left(\frac{v}{2}\right) \Big/ \Gamma\left(\frac{v+p}{2}\right) \cdot$$

este o estimaţie nedeplasată pentru σ^p.

Observaţie. Coeficientul c_p cu care trebuie multiplicată S^p pentru a obţine o estimaţie nedeplasată pentru $\sigma(\sigma^2)$ este

$$c_1 = 2^{-\frac{1}{2}} \Gamma\left(\frac{1}{2}\right) \Big/ \Gamma\left(\frac{v+1}{2}\right) [c_2 = v^{-1}] \cdot$$

b) Găsim coeficientul a_p cu care înmulţim pe t_p astfel încît $u_p = a_p t_p$ să aibe eroarea medie pătratică minimă.

Avem

$$M(a_p t_p - \sigma^p)^2 = a_p^2 D^2(t_p) + \sigma^{2p}(a_p - 1).$$

Derivînd această egalitate în raport cu a_p, găsim că aceasta este minimizată unic cînd

$$a_p = 1/[1 + \sigma^{-2p} D^2(t_p)].$$

Urmează că

$$D^2(t_p) = M(t_p^2) - M^2(t_p) = c_p^2 M(S^{2p}) - \sigma^{2p} =$$

$$= \sigma^{2p} \left[\frac{\Gamma\left(\frac{v}{2}\right) \Gamma\left(\frac{v+2p}{2}\right)}{\Gamma^2\left(\frac{v+p}{2}\right)} - 1 \right]; \quad v + 2p > 0.$$

de unde

$$a_p = \frac{\Gamma^2\left(\frac{v+p}{2}\right)}{\Gamma\left(\frac{v}{2}\right) \Gamma\left(\frac{v+2p}{2}\right)} \cdot$$

Deci

$$u_p = a_p t_p = 2^{-\frac{p}{2}} \frac{\Gamma\left(\frac{v+p}{2}\right)}{\Gamma\left(\frac{v+2p}{2}\right)} S^p,$$

ceea ce arată că coeficientul lui S^p în estimaţia de cea mai mică dispersie pentru σ^p $\left(p > -\frac{1}{2}(n-1)\right)$ într-o selecţie de volum n este acelaşi cu coeficientul lui S^p în estimaţia de eroare medie pătratică minimă pentru σ^p într-o selecţie de volum $(n+p)$.

IV.6. a) *Să se arate că*

$$\hat{\mu}^m = \frac{(n-m)!}{n!} \Sigma x_{i_1} . x_{i_2}, \ldots, x_{i_m}; \qquad i_1 \neq i_2 \neq \ldots \neq i_m, \quad m \leqslant n,$$

unde suma se extinde asupra tuturor permutărilor de m observaţii dintr-o selecţie de volum n, este o estimaţie nedeplasată pentru μ^m, şi să se determine $D^2(\hat{\mu}^m)$.

b) *În ce caz $\hat{\mu}^m$ este preferată statisticii $\bar{x}^m$.*

c) *Să se găsească o estimaţie pentru $D^2(\hat{\mu}^m)$.*

Soluţie. Avem

$$M(\Sigma x_{i_1} x_{i_2} \ldots x_{i_m}) = \frac{n!}{(n-m)!} M(x_{i_1} x_{i_2} \ldots x_{i_m}) =$$

$$= \frac{n!}{(n-m)!} M(x_{i_1}) M(x_{i_2}) \ldots M(x_{i_m}) = \frac{n!}{(n-m)!} \mu^m,$$

ceea ce arată că $M(\hat{\mu}^m) = \mu^{(m)}$.

Prin definiţie avem

$$D^2(\hat{\mu}^m) = M(\hat{\mu}^m - \mu^m)^2 = M(\hat{\mu}^m)^2 - \mu^{2m} =$$

$$= \left(\frac{(n-m)!}{n!}\right)^2 M(\Sigma x_{i_1} x_{i_2} \ldots x_{i_m})^2 - \mu^{2m} (i_1 \neq i_2 \neq \ldots \neq i_m).$$

Putem scrie

$$(\Sigma x_{i_1} x_{i_2} \ldots x_{i_m})^2 = \Sigma x_{i_1} x_{i_2} \ldots x_{i_{2m}} (i_1 \neq i_2 \neq \ldots \neq i_m;$$

$$i_{m+1} \neq i_{m+2} \neq \ldots \neq i_{2m})$$

$$= \sum_{\alpha=0}^{m} \frac{(m!)^2}{((m-\alpha)!)^2 \alpha!} \Sigma x_{i_1}^2 x_{i_2}^2 \ldots x_{i_\alpha}^2 x_{i_{\alpha+1}} \ldots x_{i_{2m-\alpha}}$$

$$(i_1 \neq i_2 \neq \ldots \neq i_{2m-\alpha}).$$

Coeficientul $\frac{(m!)^2}{((m-\alpha)!)^2\alpha!}$ apare deoarece sînt $(C_m^\alpha)^2$ combinări de α perechi cu indice egal, C_m^α de la fiecare mulţime de m şi acestea trebuie înmulţite cu $\alpha!$ numărul modurilor în care α perechi pot fi ajustate. Deci

$$M(\Sigma x_{i_1}x_{i_2}\ldots x_{i_m})^2 = M\sum_{\alpha=0}^{m}\frac{(m!)^2}{((m-\alpha)!)^2\alpha!}\Sigma x_{i_1}^2 x_{i_2}^2\ldots x_{i_\alpha}^2 x_{i_{\alpha+1}}\ldots x_{i_{2m-\alpha}} =$$

$$= \sum_{\alpha=0}^{m}\frac{(m!)^2}{((m-\alpha)!)^2\alpha!}\,\frac{n!}{(n-2m-\alpha)!}[M(x_i^2)]^\alpha[M(x_i)]^{2m-2\alpha}$$

şi cum

$$M(x_i) = \mu \text{ şi } M(x_i^2) = \sigma^2 + \mu^2;$$

$$M(\Sigma x_{i_1}x_{i_2}\ldots x_{i_m})^2 =$$

$$= \sum_{\alpha=0}^{m}\frac{(m!)^2\, n!}{((m-\alpha)!)^2\,\alpha!\,(n-2m-\alpha)!}\left\{\sum_{\beta=0}^{\alpha}C_\alpha^\beta\sigma^{2\beta}\mu^{2\alpha-2\beta}\right\}\mu^{2m-2\alpha}.$$

Prin urmare

$$D^2(\hat{\mu}^m) = \sum_{\alpha=1}^{m}\frac{C_m^\alpha C_{n-\alpha}^{n-m}}{C_n^m}\sigma^{2\alpha}\mu^{2m-2\alpha}.$$

Observaţii

1) Pentru $m = 1$, $D^2(\hat{\mu}) = \frac{\sigma^2}{n}$.

2) Pentru $m = 2$, $D^2(\hat{\mu}^2) = \frac{4\mu^2\sigma^2}{n} + \frac{2\sigma^4}{n(n-1)}$.

3) Pentru $m = 3$,

$$D^2(\hat{\mu}^3) = \frac{9\mu^4\sigma^2}{n} + \frac{18\mu^2\sigma^4}{n(n-1)} + \frac{6\sigma^6}{n(n-1)(n-2)}.$$

b) Aproximativ, avem

$$M(\bar{x}^m - \mu^m)^2 = M(\bar{x}^{2m}) - M^2(\bar{x}^m) + [M(\bar{x}^m) - \mu^m]^2 =$$

$$= \frac{m^2\mu^{2m-2}\sigma^2}{n} + \frac{m^2(m-1)\,\mu^{2m-3}\mu_3 + \frac{1}{2}\,m^2(m-1)\,(3m-5)\,\mu^{2m-4}\sigma^4}{n^2} +$$

$$+ \ldots + \frac{\frac{1}{4}\,m^2(m-1)^2\,\mu^{2m-4}\sigma^4}{n^2} + \ldots =$$

$$= \frac{m^2\mu^{2m-2}\sigma^2}{n} + \frac{m^2(m-1)\,\mu^{2m-3}\mu_3 + \frac{1}{4}\,m^2(m-1)\,(7m-11)\,\mu^{2m-4}\sigma^4}{n^2},$$

unde μ_3 este momentul centrat de ordinul 3 al populaţiei. Acest rezultat face abstracţie de termenii în n^{-3}. $\hat{\mu}^m$ poate fi preferat statisticii $\bar{x}^m$ dacă

$$D^2(\hat{\mu}^m) < M(\bar{x}^m - \mu^m)^2,$$

care cere ca

$$4\mu\mu_3 + (5m-9)\,\sigma^4 > 0, \quad (m > 1; \quad \mu \neq 0).$$

c) O estimaţie nedeplasată pentru

$$[M(x_i^2)]^\alpha\,[M(x_i)]^\beta \; (\alpha + \beta \leqslant n)$$

este

$$\frac{(n-\alpha-\beta)!}{n!}\,\Sigma x_{i_1}^2 x_{i_2}^2 \ldots x_{i_\alpha}^2 x_{i_{\alpha+1}} \ldots x_{i_{\alpha+\beta}}, \quad i_1 \neq i_2 \neq \ldots \neq i_{\alpha+\beta},$$

ceea ce face ca o estimaţie nedeplasată pentru $\hat{\mu}^m (2m \leqslant n)$ să fie

$$(C_n^m)^{-2}\left[\sum_{\alpha=0}^{m} \frac{1}{\alpha!\,((m-\alpha)!)^2}\,\Sigma x_{i_1}^2 x_{i_2}^2 \ldots x_{i_\alpha}^2 x_{i_{\alpha+1}} \ldots x_{i_{2m-\alpha}} -\right.$$

$$\left. - \frac{(n-2m)!}{n!}\,\Sigma x_{i_1} x_{i_2} \ldots x_{i_{2m}}, \quad i_1 \neq i_2 \neq \ldots \neq i_{2m}. \right] \qquad \text{(IV.3)}$$

Observații. Din (VI.3) pentru

$$m = 1, \quad \widehat{D^2}(\hat{\mu}) = \frac{1}{n-1}\left(s_2 - \frac{1}{n} s_1^2\right).$$

$$m = 2, \ \widehat{D^2}(\hat{\mu}) = \frac{2}{n(n-1)(n-2)(n-3)}\left\{(2s_2s_1^2 - s_2^2 - 4s_3s_1 + 3s_4) - \right.$$

$$\left. - \frac{2n-3}{n(n-1)}(s_1^2 - s_2^2)^2\right\}.$$

IV.7. *Din m populații se extrag independent m selecții (cu întoarcere) de volum n_1, n_2, ..., respectiv n_m.*

a) *Să se arate că statistica*

$$\hat{\pi} = \Sigma x_{i_1} x_{i_2} \ldots x_{i_m} / \Pi n_j, \qquad i_j = 1, 2, \ldots, n_j,$$

unde x_{i_j} este observația i din populația j, este o estimație nedeplasată pentru

$$\pi = \mu_1 \mu_2 \ldots \mu_m,$$

unde μ_i este valoarea medie a variabilei X_i care reprezintă caracteristica sub cercetare în populația i.

b) *Să se arate că*

$$D^2(\hat{\pi}) = \prod_{i=1}^{m}\left\{\mu_i^2 + \frac{\sigma_i^2}{n_i}\right\} - \pi^2.$$

Dezvoltarea produsului include 2^m termeni dintre care unul este $\mu_1^2\mu_2^2 \ldots \mu_m^2 = \pi^2$ și aceasta se reduce cu al doilea termen. Fiecare dintre cei $2^m - 1$ termeni include α valori $\frac{\sigma_i^2}{n_i}$ și $(m - \alpha)$ valori $\mu_i^2 (\alpha = 1, 2, \ldots, m)$.

c) *Să se arate că statistica*

$$\prod_{i=1}^{m} \bar{x}_i^2 - \prod_{i=1}^{m}\left\{\bar{x}_i^2 - \frac{s_i^2}{n_i - 1}\right\}$$

este o estimație nedeplasată pentru $D^2(\hat{\pi})$.

Produsul include $2^m - 1$ termeni. Fiecare are α valori $\frac{s_i^2}{n_i - 1}$ și $m - \alpha$ valori $\bar{x}_i^2 (\alpha = 1, 2, \ldots, m)$. Deci termenii sînt pozitivi sau negativi după cum α este impar sau par.

IV.8. *Fie $x_1, x_2, \ldots, x_n$ o selecţie de volum n dintr-o populaţie în care caracteristica sub cercetare este o variabilă X repartizată uniform pe intervalul* $\left(\theta - \frac{1}{2},\ \theta + \frac{1}{2}\right)$.

a) *Să se arate că $\bar{x}$ şi $w = \frac{x_{(1)} + x_{(n)}}{2}$ sînt estimaţii nedeplasate pentru* θ.

b) *Să se calculeze dispersiile celor două estimaţii şi să se compare.*

Soluţie: a) Avem

$$f(x_{(1)}, x_{(n)}) = n(n-1)\,(x_{(n)} - x_{(1)})^{n-2} = n(n-1)\,(v-u)^{n-2}$$

de unde

$$M(w) = \int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} \int_{0}^{\theta+\frac{1}{2}-u} \frac{u+v}{2}\, n(n-1)\,(v-u)^{n-2}\,du dv = \theta, \qquad \text{(IV.4)}$$

deoarece

$$\frac{2\theta - 1}{2} \leqslant \frac{u+v}{2} \leqslant \frac{2\theta + 1}{2}.$$

Prin simetrie $\theta - x_{(1)}$ şi $x_{(n)} - \theta$ au aceeaşi repartiţie, astfel că

$$0 = M\{(\theta - x_{(1)}) - (x_{(n)} - \theta)\} = 2\theta - M\{x_{(1)} + x_{(n)}\}$$

şi rezultă relaţia (IV.4).

Observaţie. Deoarece $M(X) = \theta$ şi $M(\bar{x}) = \frac{1}{n}\sum_{1}^{n} M(x_i) = M(X) = \theta$, urmează că şi $\bar{x}$ este o estimaţie nedeplasată pentru θ.

b) Cum $D^2(X) = \frac{1}{12}$, urmează că

$$D^2(\bar{x}) = \frac{1}{n^2}\, nD^2(X) = \frac{1}{12n}.$$

Avem

$$M(x_{(1)}^2) = \int_0^1 \int_{x_{(1)}}^1 x_{(1)}^2 n(n-1)\,(x_{(n)} - x_{(1)})^{n-2}\, dx_{(1)}\, dx_{(n)} = \frac{2}{(n+1)(n+2)} \cdot$$

$$M(x_{(1)}x_{(n)}) = \int_0^1 \int_{x_{(1)}}^1 x_{(1)}\, x_{(n)} n(n-1)\,(x_{(n)} - x_{(1)})^{n-2}\, dx_{(1)}\, dx_{(n)} = \frac{1}{n+2},$$

şi

$$M(x_{(n)}^2) = \int_0^1 \int_{x_{(1)}}^1 x_{(n)}^2\, n(n-1)\,(x_{(n)} - x_{(1)})^{n-2} dx_{(1)}\, dx_{(n)} = \frac{n}{n+2},$$

de unde

$$D^2(w) = \frac{1}{4}\,(M(x_{(1)}^2) - 2M(x_{(1)}x_{(n)}) + M(x_{(n)}^2)) - \left(\frac{1}{2}\right)^2 =$$

$$= \frac{1}{2(n+1)(n+2)} \cdot$$

Pentru $n \geqslant 3$, $D^2(w) < D^2(\bar{x})$, ceea ce arată că estimaţia nedeplasată w, care este indiferentă de comportarea celor $n-2$ observaţii dintre $x_{(1)}$ şi $x_{(2)}$, are dispersia mai mică decît a estimaţiei nedeplasate $\bar{x}$ care depinde de toate observaţiile.

Observaţie. Acest rezultat este valabil numai în acest caz particular şi nu în general. Mai mult, chiar şi în acest caz compararea nu este hotărîtoare nici w şi nici $\bar{x}$ nu sînt repartizate normal pentru n finite şi numai $\bar{x}$, pentru valori mari ale lui n, este repartizat normal.

IV.9. *Să se arate că*

a) $$\hat{\sigma} = \frac{2\sqrt{\pi}}{n(n-1)} \sum_{i=1}^{n} \left[i - \frac{n+1}{2}\right] X_{i,n},$$

unde $X_{1,n} \leqslant X_{2,n} \leqslant \ldots \leqslant X_{n,n}$ *reprezintă statisticile de ordine ale selecţiei* $X_1\ X_2. \ldots, X_n$ *ce provine dintr-o populaţie normală, este o estimaţie nedeplasată pentru* σ.

b) $D^2(\hat{\sigma}) = \sigma^2 A(n)$,

unde

$$A(n) = \frac{1}{n(n-1)}\left[n\left(\frac{\pi}{3} + 2(3)^{1/2} - 4\right) + \left(6 - 4(3)^{1/2} + \frac{\pi}{3}\right)\right].$$

c) $a(n)\,\hat{\sigma} = \frac{b(n)}{n(n-1)}\sum_{i=1}^{n}\left[i - \frac{n+1}{2}\right] X_{i,n}$

unde

$$b(n) = \frac{2\sqrt{\pi}}{A(n)+1}.$$

este estimaţia de eroare medie pătratică minimă pentru σ şi că eroarea medie pătratică este

$$\sigma^2 \frac{A(n)}{A(n)+1}.$$

d) $2\sum_{i=1}^{n}\left[i - \frac{n+1}{2}\right] X_{i,n} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} |X_i - X_j|.$

IV.10. *Să se arate că mediana de selecţie este o estimaţie nedeplasată pentru parametrul θ, din repartiţia Cauchy ce are densitatea de repartiţie*

$$f(x) = \frac{1}{\pi}\cdot\frac{1}{1+(x-\theta)^2}.$$

Soluţie. Mediana de selecţie pentru o selecţie de volum $n = 2r + 1$ are densitatea de repartiţie

$$f^*(m_e) = \frac{(2r+1)!}{(r!)^2}\left[\int_{-\infty}^{m_e}\frac{1}{\pi}\cdot\frac{dx}{1+(x-\theta)^2}\right]^r\left[\int_{m_e}^{\infty}\frac{1}{\pi}\cdot\frac{dx}{1+(x-\theta)^2}\right]^r f(m_e),$$

unde

$$f(m_e) = \frac{1}{\pi}:\frac{1}{1+(m_e-\theta)^2}.$$

Avem

$$M(m_e - \theta) = \int_R (m_e - \theta)\, f^*(m_e)\, dm_e =$$

$$= \frac{(2r+1)!}{(r!)^2}\int_R\left\{\frac{1}{4} - \left[\frac{1}{\pi}\operatorname{arctg}(m_e - \theta)\right]^2\right\}^r \frac{m_e - \theta}{1+(m_e-\theta)^2}\, dm_e =$$

$$= \frac{(2r+1)!}{\pi(r!)^2}\int_R\sum_{j=1}^{r}\frac{r!}{j!(r-j)!}\left(\frac{1}{4}\right)^{r-j} \times$$

$$\times\left\{-\left[\frac{1}{4}\operatorname{arctg}(m_e - \theta)\right]^2\right\}^j \frac{m_e-\theta}{1+(m_e-\theta)^2}\, dm_e$$

şi făcînd schimbarea tg $y = m_e - \theta$, obţinem

$$M(m_e - \theta) = 0,$$

ceea ce arată că m_e este o estimaţie nedeplasată pentru θ.

IV.11. *Fie $x_1, x_2, \ldots, x_n$ o selecţie de volum n dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X repartizată $N(\mu, \sigma^2)$ (μ — cunoscută). Să se arate că*

$$v = \sqrt{\frac{\pi}{2}} \sum_{i=1}^{n} |x_i - \mu| / n$$

este o estimaţie nedeplasată pentru σ şi să se calculeze eficienţa acestei estimaţii.

Soluţie. Avem

$$M(v) = \frac{1}{n}\sqrt{\frac{\pi}{2}} \sum_{1}^{n} M(|x_i - \mu|) = \sqrt{\frac{\pi}{2}}\, M(|X - \mu|)$$

şi cum

$$M(|X - \mu|) = \int_{-\infty}^{\infty} |x - \mu|\, n(x;\ \mu, \sigma^2)\, dx =$$

$$= -\int_{-\infty}^{0} (x - \mu)\, n\,(x;\ \mu, \sigma^2)\, dx +$$

$$+ \int_{0}^{\infty} (x - \mu)\, n(x;\ \mu, \sigma^2)\, dx =$$

$$= 2\int_{0}^{\infty} (x - \mu)\, n(x;\ \mu, \sigma^2)\, dx = \sqrt{\frac{2}{\pi}}\, \sigma,$$

urmează că v este o estimaţie nedeplasată pentru σ.

Avem

$$D^2(v) = D^2\left(\frac{1}{n}\sqrt{\frac{\pi}{2}} \sum_{1}^{n} (x_i - \mu)\right) = \frac{1}{n}\,\frac{\pi}{2}\, D^2(x_i - \mu) =$$

$$= \frac{\pi}{2n}[M(x_i - \mu)^2 - M^2(|x_i - \mu|)] = \frac{\pi}{2n}\left(\sigma^2 - \frac{2}{\pi}\,\sigma^2\right)$$

sau

$$D^2(v) = \frac{\pi - 2}{2n}\,\sigma^2.$$

Deci, eficienţa lui v este

$$e = \frac{\sigma^2}{n} \bigg/ \frac{\pi - 2}{2n} \sigma^2 = \frac{1}{\pi - 2} \simeq 0{,}876.$$

IV.12. *Folosind inegalitatea lui Rao-Cramer şi presupunînd că $x_1, x_2, \ldots, x_n$ este o selecţie dintr-o populaţie caracterizată de o variabilă aleatoare X avînd:*

a) *Funcţia de frecvenţă*

$$e^{-\theta} \frac{\theta^x}{x!}, \quad x = 0, 1, \ldots; \Theta = (0, \infty).$$

b) *Densitatea de repartiţie*

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x - \theta)^2\right]; \quad \begin{cases} \Theta = (-\infty, \infty), \\ -\infty < x < \infty. \end{cases}$$

c) *Densitatea de repartiţie*

$$\frac{1}{\sqrt{2\pi}\sqrt{\theta}} \exp\left[-\frac{1}{2\theta} x^2\right]; \quad \begin{cases} \Theta = (0, \infty), \\ -\infty < x < \infty. \end{cases}$$

d) *Densitatea de repartiţie*

$$\frac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\theta}; \quad \begin{cases} \Theta = (0, \infty), \\ 0 < x < \infty. \end{cases}$$

e) *Densitatea de repartiţie*

$$\frac{1}{\theta}, \; 0 < x < \infty; \Theta = (0, \infty).$$

să se găsească estimaţia nedeplasată de minimă dispersie pentru θ.

S o l u ţ i e. *a)* Funcţia de verosimilitate este

$$P = \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{x_i}}{x_i!} = e^{-n\theta} \frac{\theta^{\Sigma x_i}}{\prod_{1}^{n} (x_i)!}.$$

Pentru această familie condiţiile de regularitate necesare aplicării inegalităţii lui Rao-Cramer:
1° Θ este un interval al dreptei reale;

2°. $\frac{\partial}{\partial\theta} F'_\theta < \infty$ pentru toţi $\theta \in \Theta$ şi toţi $x_1, ..., x_n$;

3°. $\frac{\partial}{\partial\theta}\int dF_\theta = \int \frac{\partial}{\partial\theta} dF_\theta$ pentru toţi $\theta \in \Theta$;

4°. pentru orice estimaţie $\hat{\theta}$ nedeplasată;

$$\frac{\partial}{\partial\theta}\int \hat{\theta}\, dF_\theta = \int \frac{\partial}{\partial\theta}\hat{\theta}\, dF_\theta = 1;$$

5°. $M\left[\frac{\partial}{\partial\theta} \ln F'_\theta\right]^2 > 0$ pentru toţi $\theta \in \Theta$

sînt satisfăcute.

Avem

$$\ln P = -n\theta + \Sigma x_i \ln \theta - \Sigma \ln x_i!$$

de unde

$$\frac{\partial}{\partial\theta} \ln P = -n + \frac{1}{\theta}\Sigma x_i = n\left(\frac{\bar{x}_n - \theta}{\theta}\right)$$

şi

$$M\left(\frac{\partial}{\partial\theta} \ln P\right)^2 = \frac{n^2}{\theta^2} D^2(\bar{x}_n) = \frac{n}{\theta}.$$

Prin urmare

$$D^2(\hat{\theta}) \geqslant \frac{\theta}{n}.$$

Dar $D^2(\bar{x}_n) = \frac{\theta}{n}$, ceea ce arată că $\bar{x}_n$ este o estimaţie nedeplasată de minimă dispersie pentru θ.

b) Funcţia de verosimilitate este

$$P = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_i - \theta)^2\right].$$

Condiţiile de regularitate sînt satisfăcute. Avem

$$\frac{\partial}{\partial\theta}\ln P = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\theta) = \frac{\bar{x}_n-\theta}{\sigma^2/n};$$

$$M\left(\frac{\bar{x}_n-\theta}{\sigma^2/n}\right)^2 = \frac{n^2}{\sigma^4}\cdot\frac{\sigma^2}{n} = n/\sigma^2$$

de unde

$$D^2(\hat{\theta}) \geqslant \frac{\sigma^2}{n}.$$

Dar $D^2(\bar{x}_n) = \sigma^2/n$, ceea ce arată că $\bar{x}_n$ este o estimaţie nedeplasată de minimă dispersie pentru θ.

Observaţie. **Marginea inferioară Rao-Cramer fiind atinsă de estimaţia $\hat{\theta}$ dacă şi numai dacă ea poate fi exprimată sub forma**

$$\hat{\theta} = \frac{Z}{D^2(Z)} + \theta = \frac{\dfrac{\partial}{\partial\theta}\ln F'_\theta}{M\left(\dfrac{\partial}{\partial\theta}\ln F'_\theta\right)^2} + \theta, \tag{IV.5}$$

procedeul de găsire a estimaţiei nedeplasate de minimă dispersie pentru θ *este următorul*:

1° se verifică condiţiile de regularitate;

2° se formează ecuaţia (IV.5);

3° se verifică că $\hat{\theta}$ este o statistică. Dacă acestea sînt îndeplinite, $\hat{\theta}$ este o estimaţie nedeplasată de minimă dispersie pentru θ.

În cazul de faţă (IV.5) se scrie

$$\hat{\theta} = \frac{\dfrac{\bar{x}_n-\theta}{\sigma^2/n}}{n/\sigma^2} + \theta = \bar{x}_n - \theta + \theta = \bar{x}_n.$$

c) Funcţia de verosimilitate este

$$P = \frac{1}{(2\pi)^{\frac{n}{2}}\theta^{\frac{n}{2}}}\exp\left(-\frac{1}{2\theta}\sum_{1}^{n}x_i^2\right).$$

Condiţiile de regularitate fiind satisfăcute, obţinem

$$\frac{\partial}{\partial\theta}\ln P = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{1}^{n}x_i^2 = \frac{\left(\sum_{1}^{n}x_i^2/n\right)-\theta}{2\theta^2/n}.$$

Deoarece

$$M\left(\frac{1}{n}\sum_1^n x_i^2\right)=\theta \quad \text{şi} \quad D^2\left(\frac{1}{n}\sum_1^n x_i^2\right)=\frac{2\theta^2}{n},$$

obţinem

$$M\left(\frac{\partial}{\partial\theta}\ln P\right)^2=\frac{n^2}{4\theta^4}\cdot\frac{2\theta^2}{n}=\frac{n}{2\theta^2}.$$

Avem

$$D^2(\hat{\theta})\geqslant\frac{2\theta^2}{n}.$$

Dar $D^2\left(\frac{1}{n}\sum_1^n x_i^2\right)=\frac{2\theta^2}{n}$, ceea ce arată că $\frac{1}{n}\sum_1^n x_i^2$ este o estimaţie de minimă dispersie pentru θ.

d) Funcţia de verosimilitate este

$$P=\frac{1}{\theta^{n\alpha}\Gamma^n(\alpha)}\prod_{i=1}^n x_i^{\alpha-1}\,e^{-\frac{1}{\theta}\sum_1^n x_i}.$$

Condiţiile de regularitate fiind satisfăcute, avem

$$\frac{\partial}{\partial\theta}\ln P=\frac{-n\alpha}{\theta}+\frac{1}{\theta^2}\sum_1^n x_i=\frac{\bar{x}_n-\alpha\theta}{\theta^2/n}$$

Avem $M(\bar{x}_n)=\alpha\theta$ şi $D^2(\bar{x}_n)=\alpha\theta^2/n$, ceea ce permite să scriem

$$M\left(\frac{\partial}{\partial\theta}\ln P\right)^2=\frac{n^2}{\theta^4}\cdot\frac{\alpha\theta^2}{n}=n\alpha/\theta^2.$$

Prin urmare

$$D^2(\hat{\theta})\geqslant\frac{\theta^2}{n\alpha}.$$

Considerăm estimaţia nedeplasată $\hat{\theta}=\bar{x}_n/\alpha$ pentru care $D^2(\hat{\theta})=\theta^2/n\alpha$ şi prin urmare $\hat{\theta}$ este o estimaţie nedeplasată de minimă dispersie pentru θ.

e) Funcţia de verosimilitate este

$$P = \frac{1}{\theta^n}.$$

Avem

$$\frac{\partial}{\partial\theta}\int_0^\theta \int_0^\theta \cdots \int_0^\theta \frac{1}{\theta^n}\, \mathrm{d}x_1 \mathrm{d}x_2 \dots \mathrm{d}x_n = 0$$

însă

$$\int_0^\theta \int_0^\theta \cdots \int_0^\theta \frac{\partial}{\partial\theta}\frac{1}{\theta^n}\, \mathrm{d}x_1 \mathrm{d}x_2 \dots \mathrm{d}x_n =$$

$$= \int_0^\theta \int_0^\theta \cdots \int_0^\theta \frac{-n}{\theta^{n+1}}\, \mathrm{d}x_1 \mathrm{d}x_2 \dots \mathrm{d}x_n = -\frac{n}{\theta} \neq 0,$$

care implică că pentru această familie condiţiile de regularitate nu sînt satisfăcute.

IV.13. *Fie $x_1, x_2, \dots, x_n$ o selecţie dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X avînd densitatea de repartiţie*

$$f^*(x;\,\theta) = \begin{cases} \dfrac{1}{\theta} & \text{dacă } 0 < x < \infty, \\ 0 & \text{în rest.} \end{cases}$$

Folosind inegalitatea lui Chapman-Robbins să se determine o margine inferioară pentru o estimaţie a parametrului θ.

S o l u ţ i e. $Z(\theta + h) \subset Z(\theta)$ ori de cîte ori $-\theta < h < \theta$ ceea ce permite să scriem

$$D^2(\hat{\theta}) \geqslant \sup_{-\theta<h<0} \frac{h^2}{M\left(\dfrac{f_{\theta+h}}{f_\theta}\right)^2 - 1}$$

Avem

$$\frac{f_{\theta+h}}{f_\theta} = \frac{\theta^n}{(\theta+h)^n} \quad \text{pentru} \quad 0 < x_i < \theta + h$$

şi prin urmare

$$M\left(\frac{f_{\theta+h}}{f_\theta}\right)^2 = \frac{\theta^{2n}}{(\theta+h)^{2n}} \int_0^{\theta+h}\int_0^{\theta+h}\dots\int_0^{\theta+h} \frac{1}{\theta^n}\,\mathrm{d}x_1 \dots \mathrm{d}x_n =$$

$$= \frac{\theta^{2n}(\theta+h)^n}{(\theta+h)^{2n}\theta^n} = \frac{\theta^n}{(\theta+h)^n} \qquad (\text{deoarece } -\theta < h < 0)$$

ceea ce dă

$$D^2(\hat{\theta}) \quad \sup_{-\theta<h<0} \frac{h^2}{\dfrac{\theta^n}{(\theta+h)^n} - 1} = \sup_{-\theta<h<0} \frac{h^2(\theta+h)^n}{\theta^n - (\theta+h)^n}.$$

IV.14. *Fie* $(x_1, x_2, \dots, x_n)$ *o selecţie dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare.*

a) X *astfel încît*

$$P(X = x) = \theta^x(1-\theta)^{1-x}; \quad x = 0, 1.$$

b) X *astfel încît*

$$P(X = x) = \frac{\theta^x}{x!}\,\mathrm{e}^{-\theta}; \qquad x = 0, 1, \dots$$

c) X caracterizată prin densitatea de repartiţie

$$f(x;\ \theta) = \frac{1}{\theta}, \quad 0 < x < \theta.$$

Folosind definiţia statisticii suficiente, să se găsească o estimaţie suficientă pentru θ.

S o l u ţ i e. Mai întîi reamintim definiţia statisticii suficiente. O statistică suficientă pentru θ este orice statistică $\Phi(x_1, \dots, x_n)$ astfel încît

$$P(x_1, \dots, x_n;\, \theta \mid \Phi) = g(x_1, \dots, x_n) \text{ pentru toţi } \theta,$$

unde $g(.)$ este o funcţie care nu conţine pe θ.

a) Funcţia de verosimilitate este

$$P(x_1, \dots, x_n;\, \theta) = \prod_1^n \theta^{x_j}(1-\theta)^{1-x_j} = \theta^{\sum_1^n x_j}(1-\theta)^{n-\sum_1^n x_j}.$$

Considerăm statistica $\Phi = \sum_1^n x_j$, pentru care

$$P(\Phi ; \theta) = C_n^\Phi \theta^\Phi (1 - \theta)^{n-\Phi}.$$

Urmează că

$$P(x_1, ..., x_n ; \theta \mid \Phi) = \frac{\theta^\Phi (1-\theta)^{n-\Phi}}{C_n^\Phi \theta^\Phi (1-\theta)^{n-\Phi}} = \frac{1}{C_n^\Phi},$$

care este independentă de θ.

Prin urmare Φ este o statistică suficientă pentru θ.

b) În acest caz funcţia de verosimilitate este

$$P(x_1, ..., x_n ; \theta) = \prod_1^n \frac{\theta^{x_j}}{x_j!} e^{-\theta} = e^{-n\theta} \theta^{\sum_1^n x_j} \prod_1^n \left(\frac{1}{x_j!}\right)$$

Considerăm statistica $\Phi = \sum_1^n x_j$ pentru care

$$P(\Phi ; \theta) = \frac{e^{-n\theta}(n\theta)^\Phi}{\Phi!}; \quad \Phi = 0, 1, ...$$

Urmează că

$$P(x_1, ..., x_n ; \theta \mid \Phi) = \frac{e^{-n\theta} \theta^{\sum_1^n x_j} \prod_{j=1}^n (1/x_j!)}{e^{-n\theta}(n\theta)^\Phi / \Phi!} = \frac{\Phi!}{n^\Phi \prod_{j=1}^n x_j!},$$

care este independentă de θ. Prin urmare Φ este o statistică suficientă pentru θ.

c) Funcţia de verosimilitate este

$$P(x_1, ..., x_n ; \theta) = \frac{1}{\theta^n}, \quad 0 < x_j < \theta.$$

Considerăm statistica $\Phi = \max(x_1, ..., x_n)$ pentru care

$$F(z) = P(\Phi < z) = \prod_1^n P(x_j < z) = \frac{z^n}{\theta^n}, \quad 0 < z < \theta.$$

Urmează că

$$f(z) = \frac{dF(z)}{dz} = \frac{nz^{n-1}}{\theta^n}, \quad 0 < z < \theta,$$

de unde

$$P(x_1, \ldots, x_n; \theta \mid \Phi) = \frac{1/\theta^n}{n\Phi^{n-1}/\theta^n} = \frac{1}{n\Phi^{n-1}}, \quad 0 < x_i < \Phi,$$

care este independentă de θ. Prin urmare Φ este o statistică suficientă pentru θ.

4.15. *Fie $(x_1, \ldots, x_n)$ o selecţie dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X caracterizată prin densitatea de repartiţie*

a) $n(x; \theta, \sigma^2)$, b) $f(x; \theta) = \frac{1}{\theta}$, $0 < x < \theta$.

Folosind criteriul de factorizare a lui Neyman să se găsească o estimaţie suficientă pentru θ.

Soluţie. Reamintim criteriul de factorizare a lui Neyman. Statistica Φ este suficientă pentru θ dacă şi numai dacă

$$P(x_1, \ldots, x_n; \theta) = g(x_1, \ldots, x_n)\, h(\theta; \Phi)$$

pentru toţi θ unde $g(.)$ este o funcţie care nu depinde de θ şi unde $h(\theta; \Phi)$ este funcţie numai de θ şi $\Phi(x_1, \ldots, x_n)$.

a) Funcţia de verosimilitate este

$$P(x_1, \ldots, x_n, \theta) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_j - \theta)^2\right], -\infty < x_j < \infty$$

şi ţinînd seama că

$$\exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_j - \theta)^2\right] =$$

$$= \exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_j - \bar{x}_n)^2\right] \exp\left[-\frac{1}{2\sigma^2} n(\bar{x}_n - \theta)^2\right]$$

urmează că $\bar{x}_n$ este o statistică suficientă pentru θ.

b) În acest caz funcţia de verosimilitate are expresia

$$P(x_1, \ldots, x_n, \theta) = \frac{1}{\theta^n}, \quad 0 < x_j < \theta.$$

Luînd $\Phi = \max(x_1, ..., x_n)$ găsim

$$P(x_1, ..., x_n; \theta) = \frac{1}{\theta^n} \delta(\Phi < \theta), \quad 0 < x_j < \Phi$$

unde

$$\delta(\Phi < \theta) = \begin{cases} 1 & \text{dacă} \quad \Phi < \theta, \\ 0 & \text{dacă} \quad \Phi \geqslant \theta. \end{cases}$$

Prin urmare am factorizat $P(x_1, ..., x_n; \theta)$ într-un produs de două funcţii dintre care una numai de $x_1, ..., x_n$, iar cealaltă de θ şi Φ.

Astfel, Φ este suficientă pentru θ.

4.16. *Fie* $\{F(x_1, ..., x_n; \theta : \theta \in \Theta\}$ *familia de repartiţii caracterizată prin funcţia de frecvenţă (densitatea de repartiţie):*

a) $P(x_1, ..., x_n; \theta) = \prod_1^n \theta^{x_j} (1 - \theta)^{1-x_j}$, $x_j = 0, 1$ *oricare ar fi* j *şi* $\Theta = (0, 1)$;

b) $P(x_1, ..., x_n; \theta) = \prod_1^n \frac{e^{-\theta} \theta^{x_j}}{x_j!}$, $x_j = 0, 1, ...,$; *pentru orice* j *şi* $\Theta = (0, \infty)$;

c) $P(x_1, ..., x_n; \theta) = \frac{1}{\theta^n}$, $0 < x_j < \theta$ *oricare ar fi* j;

$= 0$, *în rest şi*

$\Theta = (0, \infty)$.

Să se arate că aceste familii de repartiţii sînt complete.

S o l u ţ i e. Reamintim definiţia unei familii complete.

Vom spune că familia de repartiţii

$$\{F(x_1, ..., x_n; \theta), \quad \theta \in \Theta\}$$

este completă dacă

$$M[\hat{\theta}(X_1, ..., X_n)] = 0 \text{ pentru toţi } \theta \in \Theta \text{ implică } \hat{\theta} = 0.$$

Dacă Φ este o statistică suficientă pentru θ, şi dacă familia indusă $\{F(\theta; \Phi); \theta \in \Theta\}$ unde

$$F(\theta; \Phi) = \int dF(x_1, ..., x_n),$$

$$\{x_1, ..., x_n; \Phi(x_1, ..., x_n) \leqslant \Phi\}$$

este completă, atunci Φ poartă numele de statistică suficient completă.

a) Am văzut că statistica $\Phi = \sum_{1}^{n} x_j$ este suficientă pentru θ şi că are funcţia de frecvenţă

$$P(\Phi; \theta) = C_n^{\Phi} \theta^{\Phi} (1 - \theta)^{n-\Phi}, \Phi = 0, 1, \ldots, n.$$

Fie $\hat{\theta}(\Phi)$ o estimaţie care este funcţie numai de Φ şi fie $M\hat{\theta} = 0$. Aceasta arată că

$$M\hat{\theta} = \sum_{k=0}^{n} \hat{\theta}(k)\, C_n^k \theta^k (1 - \theta)^{n-k} = (1 - \theta)^n \sum_{k=0}^{n} \hat{\theta}(k) C_n^k \left(\frac{\theta}{1 - \theta}\right)^k$$

sau punînd $\frac{\theta}{1 - \theta} = \alpha$,

$$M\hat{\theta} = (1 - \theta)^n \sum_{k=0}^{n} \hat{\theta}(k)\, C_n^k \alpha^k = 0.$$

Prin urmare

$$\sum_{k=0}^{n} \hat{\theta}(k) C_n^k \alpha^k = 0, \quad 0 < \alpha < \infty,$$

$$\frac{\partial^n}{\partial \alpha^n} \left[\sum_{k=0}^{n} \hat{\theta}(k) C_n^k \alpha^k\right] = 0 \Rightarrow n!\, \hat{\theta}(n) = 0 \quad \text{sau} \quad \hat{\theta}(n) = 0.$$

Astfel

$$\sum_{k=0}^{n-1} \hat{\theta}(k) C_n^k \alpha^k = 0.$$

În continuare

$$\frac{\partial^{n-1}}{\partial \alpha^{n-1}} \left[\sum_{k=0}^{n-1} \hat{\theta}(k) C_n^k \alpha^k\right] = 0 \Rightarrow n!\, \hat{\theta}(n - 1) = 0 \quad \text{sau} \quad \hat{\theta}(n - 1) = 0.$$

În general

$$\frac{\partial^{n-r}}{\partial \alpha^{n-r}} \left[\sum_{k=0}^{n-r} \hat{\theta}(k)\, C_n^k \alpha^k\right] = 0 \Rightarrow \frac{n!}{r!}\, \hat{\theta}(n - r) = 0 \text{ sau } \hat{\theta}(n - r) = 0$$

şi prin urmare

$$\hat{\theta}(k) = 0; \quad k = 0, 1, 2, ..., n.$$

Astfel Φ/n este o statistică suficient completă şi deoarece este o estimaţie nedeplasată pentru θ, trebuie să fie o estimaţie nedeplasată uniform de minimă dispersie.

b) Am văzut că statistica $\Phi = \sum_{1}^{n} X_i$ este suficient pentru θ şi că are funcţia de frecvenţă

$$P(\Phi; \theta) = e^{-n\theta} \frac{(n\theta)^{\Phi}}{\Phi!}, \quad \Phi = 0, 1, 2, ...$$

Dacă $\hat{\theta}(\Phi)$ este astfel încît $M\hat{\theta} = 0$, atunci

$$M\hat{\theta} = \sum_{k=0}^{\infty} \hat{\theta}(k)\, e^{-n\theta} \frac{(n\theta)^k}{k!} = 0$$

conduce la

$$\sum_{k=0}^{\infty} \frac{\hat{\theta}(k)}{k!} (n\theta)^k = 0, \quad 0 < \theta < \infty.$$

Aceasta implică că

$$\hat{\theta}(k) = 0, \ k = 0, 1, 2, ...$$

şi că Φ este o statistică suficient completă.

Dacă facem pe $n = 1$, urmează că familia de repartiţii Poisson este completă.

c) Am văzut că statistica $\Phi = \sum_{1}^{n} X_i$ este suficientă pentru θ şi că are densitatea de repartiţie

$$f(\Phi; \theta) = \begin{cases} \dfrac{n\Phi^{n-1}}{\theta^n}, & 0 < \Phi < \theta, \\ 0 & , \text{ în rest.} \end{cases}$$

Dacă $\hat{\theta}(\Phi)$ este astfel încît $M\hat{\theta} = 0$, atunci

$$M\hat{\theta} = \int_0^{\theta} \hat{\theta}(z) \frac{nz^{n-1}}{\theta^n} dz = 0$$

sau

$$\int_0^{\theta} \hat{\theta}(z) z^{n-1} dz = 0, \quad 0 < \theta < \infty,$$

care implică

$$\hat{\theta}(z) = 0, \quad 0 < z < \theta.$$

IV.17. a) *Să se arate că* $\frac{x}{n}$ *este o estimaţie nedeplasată şi de minimă dispersie pentru parametrul* p *din funcţia de frecvenţă*

$$b(n, x; p) = C_n^x p^x (1-p)^{n-x}.$$

b) *Să se găsească o estimaţie nedeplasată pentru* p^2.

S o l u ţ i e. Reamintim că

$$M(x) = np, \quad D^2(x) = np(1-p); \quad M(x^2) = n^2p^2 + np(1-p).$$

a) Avem

$1°$ $M\left(\frac{x}{n}\right) = \frac{1}{n} M(x) = p,$

ceea ce arată că $\frac{x}{n}$ este o estimaţie nedeplasată pentru p.

$2°$ $\ln P = \ln C_n^x + x \ln p + (n - x) \ln (1 - p),$

$$\frac{\partial \ln P}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x - np}{p(1-p)},$$

$$M\left(\frac{\partial \ln P}{\partial p}\right)^2 = \sum_{x=1}^{n} b(n;\ x,\ p) \frac{(x-np)^2}{p^2(1-p)^2} = \frac{npq}{p^2q^2} = \frac{n}{pq}, \quad q = 1 - p.$$

Deci dispersia minimă este $\frac{pq}{n}$.

Cum

$$D^2\left(\frac{x}{n}\right) = \frac{1}{n^2} D^2(x) = \frac{pq}{n},$$

urmează că $\frac{x}{n}$ este o estimaţie nedeplasată şi de minimă dispersie pentru p.

b) Avem

$$M\left(\frac{x^2}{n^2}\right)=\frac{n^2p^2+np(1-p)}{n^2}=p^2+\frac{p(1-p)}{n}.$$

ceea ce arată că $\frac{x^2}{n^2}$ nu este o estimaţie nedeplasată pentru p^2.

Aplicăm corecţia necesară pentru a obţine deplasarea în $\frac{1}{n^2}$. Dacă punem

$$p_n=\frac{x^2}{n^2},$$

atunci p_{n-1} poate lua două valori

$$\frac{x^2}{(n-1)^2} \quad \text{şi} \quad \frac{(x-1)^2}{(n-1)^2}.$$

Deci

$$\bar{p}_{n-1}=\frac{1}{n}\left[x\left(\frac{x-1}{n-1}\right)^2+(n-x)\left(\frac{x}{n-1}\right)^2\right]=$$

$$=\frac{x^2(n-2)+x}{n(n-1)^2}.$$

Atunci

$$\bar{p}_n'=np_n-(n-1)\,\bar{p}_{n-1}=\frac{x(x-1)}{n(n-1)},$$

$$M(\bar{p}_n')=\frac{M(x^2)-M(x)}{n(n-1)}=\frac{n^2p^2-np^2}{n(n-1)}=p^2,$$

ceea ce arată că $\frac{x(x-1)}{n(n-1)}$ este o estimaţie nedeplasată pentru p^2.

Inegalitatea lui Rao-Cramer pentru p^2 este

$$D^2(p^2)\geqslant\frac{4p^2}{\frac{n}{pq}}=\frac{4p^3(1-p)}{n}.$$

Deci estimaţia de minimă dispersie pentru p^2 va avea dispersia $\frac{4p^3(1-p)}{n}$.

IV.18. *Statisticile u și w sînt caracterizate prin densitatea de repartiție*

$$f(u, w) = \frac{1}{2\pi D(u)\, D(w)\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{u-\theta}{D(u)}\right)^2 - \right.\right.$$

$$\left.\left. - \frac{2\rho(u-\theta)(w-\theta)}{D(u)D(w)} + \left(\frac{w-\theta}{D(w)}\right)^2\right]\right\} \qquad \text{(IV.6)}$$

Care este condiția ca u să fie o statistică suficientă pentru θ în raport cu w?

S o l u ț i e. Dacă fiind dat u, densitatea $g(w \mid u; \theta)$ nu depinde de θ, pentru orice statistică w, statistica u va fi numită suficientă pentru θ. Găsim $g(w \mid u)$ și vom determina condițiile în care densitatea condiționată nu depinde de θ,

$$g(w \mid u) = \frac{f(u, v)}{h(u)} = \frac{f(u, w)}{\int_{-\infty}^{\infty} f(u, w)\, dw}.$$

Cum

$$\int_{-\infty}^{\infty} f(u, w)\, dw = \frac{1}{D(u)\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{u-\theta}{D(u)}\right]^2\right\}.$$

rezultă că

$$g(w \mid u) = \frac{\exp\left\{-\frac{1}{2}\left[\frac{w - \{\theta + \rho(D(w)/D(u))(u-\theta)\}}{D(w)\sqrt{1-\rho^2}}\right]^2\right\}}{D(w)\sqrt{2\pi}\,\sqrt{1-\rho^2}},$$

ceea ce arată că w este repartizată

$$N\left[\theta + \rho\frac{D(w)}{D(u)}(u-\theta),\quad D(w)\sqrt{1-\rho^2}\right].$$

Pentru ca $g(w \mid u)$ să nu depindă de θ este suficient ca

$$1 = \rho\frac{D(w)}{D(u)}, \text{ adică } \frac{D(w)}{D(u)} = \frac{1}{\rho}.$$

Dar $D(w)/D(u)$ este pozitiv, așa că și ρ trebuie să fie pozitiv.

De altminteri, dacă ρ este pozitiv, trebuie să avem $0 \leqslant \rho \leqslant 1$, pentru care $D(u) \leqslant D(w)$.

Dacă funcţia (IV.6) apare adesea în selecţii mari, putem trage concluzia că în interiorul unei clase de statistici asimptotic nedeplasate şi normal repartizate, statisticile suficiente au asimptotic cea mai mică dispersie.

IV.19. *Procesul mişcării Browniene este definit de* $P(X_0 = 0) = 1$ *şi*

$$f(\tau, x, t, y, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi(t-\tau)}} \exp\left\{-\frac{[y - x - \mu(t-\tau)]^2}{2\sigma^2(t-\tau)}\right\}.$$

Ştiind că la momentul t_j, $0 \leqslant j \leqslant n$ $(0 = t_0 < t_1 < \ldots < t_n)$ *procesul se găseşte în starea* x_j, $0 \leqslant j \leqslant n$; $x_0 = 0$, *să se estimeze prin metoda verosimilităţii maxime parametrii* μ *şi* σ^2

Soluţie. Pentru funcţia de verosimilitate

$$P(x, t, \mu, \sigma) = \frac{1}{\sigma^n(2\pi)^{n/2}} \prod_{k=0}^{n-1} \frac{1}{\sqrt{t_{k+1} - t_k}} \exp\left\{-\frac{[x_{k+1} - x_k - \mu(t_{k+1} - t_k)]^2}{2(t_{k+1} - t_k)\sigma^2}\right\},$$

obţinem

$$\ln P = -n\ln\sigma - \frac{n}{2}\ln(2\pi) + \sum_{k=0}^{n-1} - \ln\sqrt{t_{k+1} - t_k} - \frac{[x_{k+1} - x_k - \mu(t_{k+1} - t_k)]^2}{2(t_{k+1} - t_k)\sigma^2}.$$

Ecuaţia $\frac{\partial \ln P}{\partial \mu} = 0$, care în cazul de faţă se scrie

$$\sum_{k=0}^{n-1} \frac{[x_{k+1} - x_k - \mu(t_{k+1} - t_k)]}{(t_{k+1} - t_k)\sigma^2} = 0.$$

conduce la estimaţia

$$\hat{\mu} = \frac{x_n}{t_n}$$

pentru μ.

Analog, ecuaţia

$$\frac{\partial \ln P}{\partial \sigma^2} = 0$$

dă ca estimaţie a parametrului σ^2, statistica

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=0}^{n-1} \frac{[x_{k+1} - x_k - \hat{\mu}(t_{k+1} - t_k)]^2}{t_{k+1} - t_k},$$

IV.20. *Fie $x_1, x_2, \ldots, x_n$ o selecţie dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X avînd densitatea de repartiţie*

$$f(x; \theta) = \frac{1}{\pi} \cdot \frac{1}{1 + (x - \theta)^2}, \quad -\infty < x < \infty.$$

a) *Să se găsească estimaţia de maximă verosimilitate a parametrului θ.*

b) *Să se calculeze dispersia asimptotică a estimaţiei de maximă verosimilitate.*

Soluţie. a) Funcţia de verosimilitate fiind

$$P(x, \theta) = \frac{1}{\pi^n} \prod_1^n \frac{1}{1 + (x_j - \theta)^2},$$

obţinem

$$\ln P(x, \theta) = -n \ln \pi - \sum_1^n \ln [1 + (x_j - \theta)^2].$$

Deci ecuaţia de verosimilitate se scrie

$$\frac{\partial \ln P(x, \theta)}{\partial \theta} = \frac{\mathrm{d} \ln P(x, \theta)}{\mathrm{d}\theta} = \sum_1^n \frac{2(x_j - \theta)}{1 + (x_j - \theta)^2} = 0.$$

Această ecuaţie poate fi rezolvată în raport cu θ numai prin aproximaţii.

b) Avem

$$\ln f(x;\theta) = -\ln \pi - \ln[1 + (x-\theta)^2],$$

$$\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta) = \frac{-2 + 2(x+\theta)^2}{[1+(x-\theta)^2]^2},$$

$$M\left[\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta)\right] = \int_R \frac{-2+2(x-\theta)^2}{[1+(x-\theta)^2]^2}\cdot\frac{1}{\pi}\cdot\frac{1}{1+(x-\theta)^2}\,\mathrm{d}x =$$

$$= -\frac{2}{\pi}\int_R \frac{1-u^2}{(1+u^2)^3}\,\mathrm{d}u$$

de unde

$$M(t-\theta)^2 = 1/nM\left[\frac{\partial^2}{\partial\theta^2}\ln f(x,\theta)\right] = \frac{2}{n}.$$

Observaţie. Estimaţia de maximă verosimilitate a parametrului θ din repartiţia Cauchy, ce poate fi determinat numai aproximativ, are asimptotic dispersia $2/n$.

IV.21. *Fie $x_1, x_2, \ldots, x_n$ o selecţie de volum n dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X cu repartiţie Laplace avînd densitatea de repartiţie*

$$f(x;\theta) = \frac{1}{2}\,e^{-|x-\theta|}, \quad \theta \in R.$$

Să se arate că mediana de selecţie este o estimaţie de verosimilitate maximă a parametrului θ.

Soluţie. Funcţia de verosimilitate corespunzătoare este

$$P(\theta) = \left(\frac{1}{2}\right)^n e^{-\sum_{i=1}^{n}|x_i-\theta|}.$$

Funcţia $P(\theta)$ ia valoarea maximă atunci cînd $\sum_{i=1}^{n}|x_i-\theta|$ ia valoarea minimă. Pentru a arăta că mediana de selecţie este estimaţie de verosimilitate maximă în cazul nostru va fi suficient să arătăm că

$$\sum_{i=1}^{n}|x_i - m_e| \leqslant \sum_{i=1}^{n}|x_i-\theta|,$$

unde m_e este mediana de selecţie.

Pentru aceasta vom considera cazurile n par şi n impar.

a) Fie n par şi $y_1, y_2, \ldots, y_n$ statisticile de ordine corespunzătoare selecţiei $x_1, x_2, \ldots, x_n$. Fie m un număr real situat între cele două valori din mijloc ale statisticilor de ordine, adică

$$y_1 \leqslant y_2 \leqslant \ldots \leqslant y_{\frac{n}{2}} \leqslant m \leqslant y_{\frac{n}{2}+1} \leqslant \ldots \leqslant y_{n-1} \leqslant y_n.$$

În particular m poate fi m_e.

Atunci din şirul de inegalităţi scris mai sus şi din egalitatea triunghiului rezultă că putem scrie

$$|y_i - m| + |m - y_{n+1-i}| = |y_i - y_{n+1-i}| \leqslant |y_i - \theta| + |\theta - y_{n+1-i}|$$
$$\left(i = 1, 2, \ldots, \frac{n}{2};\quad \theta \in R\right)$$

Sumînd aceste inegalităţi membru cu membru obţinem

$$\sum_{i=1}^{\frac{n}{2}} (|y_i - m| + |m - y_{n+1-i}|) \leqslant \sum_{i=1}^{\frac{n}{2}} (|y_i - \theta| + |\theta - y_{n+1-i}|).$$

Dar

$$\sum_{i=1}^{\frac{n}{2}} |m - y_{n+1-i}| = \sum_{i=\frac{n}{2}+1}^{n} |m - y_i|$$

şi

$$\sum_{i=1}^{\frac{n}{2}} |\theta - y_{n+1-i}| = \sum_{i=\frac{n}{2}+1}^{n} |\theta - y_i|.$$

Cu aceasta inegalitatea de mai sus devine

$$\sum_{i=1}^{n} |y_i - m| \leqslant \sum_{i=1}^{n} |y_i - \theta|$$

şi cum

$$\sum_{i=1}^{n} |y_i - m| = \sum_{i=1}^{n} |x_i - m|,$$

$$\sum_{i=1}^{n} |y_i - \theta| = \sum_{i=1}^{n} |x_i - \theta|,$$

obținem

$$\sum_{i=1}^{n} | x_i - m | \leqslant \sum_{i=1}^{n} | x_i - \theta |,$$

unde în particular numărul m poate fi m_e, ceea ce justifică afirmația în acest caz.

b) Fie n impar și $m_e = y_{\frac{n+1}{2}}$.

Procedînd ca în cazul a) putem scrie

$$| y_i - y_{\frac{n+1}{2}} | + | y_{\frac{n+1}{2}} - y_{n+1-i} | = | y_i - y_{n+1-i} | \leqslant$$

$$\leqslant | y_i - \theta | + | \theta - y_{n+1-i} |$$

$$\left(i = 1, 2, ..., \frac{n+1}{2} - 1; \quad \theta \in R\right)$$

și de asemenea

$$0 = | y_{\frac{n+1}{2}} - y_{\frac{n+1}{2}} | \leqslant | y_{\frac{n+1}{2}} - \theta |.$$

Sumînd succesiunea de inegalități pentru $i = 1, 2, ..., \frac{n+1}{2} - 1$ obținem

$$\sum_{i=1}^{\frac{n+1}{2}-1} (| y_i - y_{\frac{n+1}{2}} | + | y_{\frac{n+1}{2}} - y_{n+1-i} |) \leqslant$$

$$\leqslant \sum_{i=1}^{\frac{n+1}{2}-1} (| y_i - \theta | + | \theta - y_{n+1-i} |).$$

Cum însă

$$\sum_{i=1}^{\frac{n+1}{2}-1} | y_{\frac{n+1}{2}} - y_{n+1-i} | = \sum_{i=\frac{n+1}{2}+1}^{n} | y_{\frac{n+1}{2}} - y_i |$$

și

$$\sum_{i=1}^{\frac{n+1}{2}-1} | \theta - y_{n+1-i} | = \sum_{i=\frac{n+1}{2}+1}^{n} | y_i - \theta |$$

obţinem

$$\sum_{i=1}^{n} | y_i - y_{\frac{n+1}{2}} | \leqslant \sum_{i=1}^{n} | y_i - \theta |$$

sau din aceleaşi motive ca la punctul a), avem

$$\sum_{i=1}^{n} | x_i - m_e | \leqslant \sum_{i=1}^{n} | x_i - \theta |,$$

ceea ce dovedeşte că mediana de selecţie este estimaţie de verosimilitate maximă a parametrului θ în cazul repartiţiei cu densitatea

$$f(x; \theta) = \frac{1}{2} e^{-|x-\theta|}, \quad \theta \in R.$$

IV.22. *Repartiţia timpului de funcţionare a unei lămpi electronice este caracterizată prin densitatea*

$$f(t; \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{t}{\theta}} & \text{dacă} \quad t > 0, \\ 0 & \text{dacă} \quad t < 0. \end{cases}$$

a) *Folosind metoda verosimilităţii maxime şi presupunînd că* $t_1, t_2, \ldots, t_n$ *reprezintă duratele de funcţionare a* n *lămpi electronice, să se estimeze* θ.

b) *Să se arate că* $\hat{\theta}$ *este o estimaţie nedeplasată pentru* θ.

S o l u ţ i e. Funcţia de verosimilitate este

$$P = \theta^{-n} e^{-\frac{1}{\theta} \sum_{i=1}^{n} t_i}.$$

Urmează că

$$\ln P = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^{n} t_i,$$

$$\frac{\partial \ln P}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} t_i,$$

şi

$$\frac{\partial \ln P}{\partial \theta} = 0 \Rightarrow \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} t_i.$$

Aşadar media de selecţie este o estimaţie pentru media teoretică.

Avem

$$M(\hat{\theta}) = \frac{1}{n} \sum_{j=1}^{n} M(t_j) = \frac{n\theta}{n} = \theta,$$

şi

$$M_2(\hat{\theta}) = M\left(\frac{1}{n}\sum_{j=1}^{n} t_j\right)^2 = \frac{1}{n^2}\left[\sum_{j=1}^{n} M(t_j^2) + 2\sum_{j<k} M(t_j)\, M(t_k)\right].$$

Cum

$$M(t_j^2) = \frac{1}{\theta}\int_0^{\infty} t^2 e^{-\frac{t}{\theta}}\, dt = \frac{1}{\theta}\cdot\frac{\Gamma(3)}{\left(\frac{1}{\theta}\right)^3} = 2\theta^2$$

urmează că

$$M_2(\hat{\theta}) = \frac{1}{n^2}(2n\theta^2 + n(n-1)\,\theta^2) = \frac{(n+1)\,\theta^2}{n}.$$

Deci

$$D^2(\hat{\theta}) = M_2(\hat{\theta}) - M^2(\hat{\theta}) = \frac{\theta^2}{n},$$

Din faptul că $M(\hat{\theta}) = \theta$ şi $D^2(\hat{\theta}) = \frac{\theta^2}{n} \xrightarrow[n\to\infty]{} 0$, rezultă că $\hat{\theta}$ este o estimaţie nedeplasată a parametrului θ.

VI.23. *Fie X o variabilă aleatoare caracterizată prin:*
a) *Funcţia de frecvenţă*

$$f(x;\theta) = \begin{cases} \frac{1}{3}(5-2\theta)^{(3-x)(2-x)/2}(\theta-1)^{(x-1)(4-x)/2} & \\ \qquad\qquad \textit{pentru } x = 1, 2, 3; \quad 1 \leqslant \theta \leqslant 2{,}5, \\ 0 \qquad\qquad \textit{în rest.} \end{cases}$$

b) *Funcţia de frecvenţă*

$$f(x;\theta) = (2-\theta)\, x^{[\ln 2]^{-1}\ln(\theta-1)/(2-\theta)} \qquad \text{pentru } x = 1, 2, \quad 1 < \theta < 2.$$

c) *Densitatea de repartiţie*

$$f(x;\theta) = \begin{cases} \frac{\theta}{1-\theta}\, x^{(2\theta-(1/(1-\theta))} & \textit{pentru} \quad 0 < x < 1, \quad 1/2 < \theta < 1, \\ 0 & \textit{în rest.} \end{cases}$$

d) *Densitatea de repartiţie*

$$f(x;\theta) = \begin{cases} \frac{1}{\Gamma(\theta)}\, x^{\theta-1} e^{-x} & \textit{pentru} \quad x > 0, \\ 0 & \textit{în rest} \end{cases}$$

e) *Densitatea de repartiţie*

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left[-\frac{1}{2\theta}(x-\theta)^2\right] \qquad \textit{pentru} \quad -\infty < x < \infty, \quad 0 < \theta < \infty.$$

şi $(x_1, \ldots, x_m)$ *o selecţie din populaţia în care caracteristica sub cercetare este această variabilă.*

Ne propunem să găsim estimaţia de maximă verosimilitate a parametrului θ.

Soluţie. a) Avem

$$M(X) = 1 \cdot \frac{1}{3}(5-2\theta) + 2 \cdot \frac{1}{3}(\theta - 1) + 3 \cdot \frac{1}{3}(\theta - 1) = \theta.$$

Pentru funcţia de verosimilitate

$$P = \frac{1}{3}(5-2\theta)^{\frac{1}{2}\Sigma(3-x_j)(2-x_j)}(\theta-1)^{\frac{1}{2}\Sigma(x_j-1)(4-x_j)}$$

$$\frac{\partial \ln P}{\partial \theta} = 0 \Rightarrow 4n\theta = -8n + 15\Sigma x_j - 3\Sigma x_j^2$$

a cărei soluţie este

$$\hat{\theta} = \frac{15}{4}\bar{x}_n - \frac{3}{4}\frac{1}{n}\sum_1^n x_i^2 - 2.$$

b) Avem

$$M(X) = (2-\theta) + 2(2-\theta)\,2^{\frac{1}{\ln 2}\ln\frac{\theta-1}{2-\theta}} =$$

$$= (2-\theta) + 2(2-\theta)\,2^{\frac{1}{\ln_2 2}\ln_2\frac{\theta-1}{2-\theta}} =$$

$$= (2-\theta) + 2(2-\theta)\frac{\theta-1}{2-\theta} = \theta.$$

Pentru funcţia de verosimilitate

$$P = (2-\theta)^n\left(\prod_1^n x_i\right)^{[\ln 2]^{-1}\ln\frac{\theta-1}{2-\theta}},$$

$\frac{\partial \ln P}{\partial \theta} = 0$ se scrie $\frac{n}{2-\theta} = \frac{[\ln 2]^{-1}}{(\theta-1)(2-\theta)}\sum_1^n \ln x_i.$

Cum $0 < \theta < 2$ rezultă

$$\theta - 1 = \frac{1}{\ln 2}\cdot\frac{1}{n}\sum_1^n \ln x_i$$

sau

$$\hat{\theta} = 1 + \frac{1}{\ln 2} \cdot \frac{1}{n} \sum_1^n \ln x_i.$$

c) Pentru funcţia de verosimilitate

$$P = \left(\frac{\theta}{1-\theta}\right)^n \prod_1^n x_i^{(2\theta-1)/(1-\theta)}$$

avem

$$\ln P = n \ln \theta - n \ln (1-\theta) + \sum_1^n (2\theta - 1)/(1-\theta)] \ln x_i.$$

Ecuaţia de verosimilitate

$$\frac{\partial \ln P}{\partial \theta} = 0$$

se scrie

$$\frac{n}{\theta} + \frac{n}{1-\theta} + \sum_1^n \frac{1}{(1-\theta)^2} \ln x_i = 0$$

şi are ca soluţie

$$\hat{\theta} = \left[1 - \frac{1}{n} \sum_1^n \ln x_i\right]^{-1}.$$

Avem

$$M(X) = \frac{\theta}{1-\theta} \int_0^1 x^{\frac{2\theta-1}{1-\theta}+1} \mathrm{d}x = \theta.$$

d) Avem

$$M(X) = \frac{1}{\Gamma(\theta)} \int_0^\infty x^\theta \mathrm{e}^{-x} \mathrm{d}x = \theta.$$

Pentru funcţia de verosimilitate

$$P = \frac{1}{\Gamma^n(\theta)} \prod_1^n x_i^{\theta-1} \mathrm{e}^{-\sum_1^n x_i},$$

obținem ecuația de verosimilitate

$$-n\frac{d\Gamma(\theta)}{d\theta}\cdot\frac{1}{\Gamma(\theta)}+\sum_{1}^{n}\ln x_i=0$$

a cărei soluție $\hat{\theta}\neq\bar{x}_n$.

Observație. **Această problemă arată că există repartiții în care parametrul este media repartiției dar pentru care estimația de maximă verosimilitate a parametrului nu este media de selecție.**

e) Avem

$$M(X)=\frac{1}{\sqrt{2\pi\theta}}\int_{-\infty}^{\infty}x\exp\left[-\frac{1}{2\theta}(x-\theta)^2\right]dx=\theta.$$

Pentru funcția de verosimilitate

$$P=(2\pi\theta)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\theta}\sum_{1}^{n}(x_i-\theta)^2\right]$$

avem

$$\frac{\partial\ln P}{\partial\theta}=0\Rightarrow n\theta^2+n\theta-\sum_{1}^{n}x_i^2=0,$$

ale cărei soluții sînt

$$\hat{\theta}_{1,2}=\frac{-n\pm\sqrt{n^2+4n\Sigma x_i^2}}{2n}=\frac{1}{2}\left(-1\pm\sqrt{1+4\sum_{1}^{n}\frac{x_i^2}{n}}\right)$$

Soluția

$$\hat{\theta}_2=\frac{1}{2}\left(-1-\sqrt{1+4\sum_{1}^{n}\frac{x_i^2}{n}}\right)<0$$

nu este acceptabilă deoarece $\theta>0$.

Deci estimația de maximă verosimilitate a parametrului θ este

$$\hat{\theta}=\frac{1}{2}\left(-1+\sqrt{1+4\sum_{1}^{n}\frac{x_i^2}{n}}\right)$$

IV.24. *Folosind metoda verosimilității maxime și presupunînd că* $(x_1, ..., x_n)$ *este o selecție dintr-o populație în care caracteristica sub cercetare este o variabilă aleatoare X caracterizată prin densitatea*

$$f(x;\lambda,\alpha)=\begin{cases}\dfrac{\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} & \textit{dacă} \quad x>0, \quad \alpha,\lambda>0,\\ 0 & \textit{dacă} \quad x\leqslant 0\end{cases} \tag{IV.7}$$

să se estimeze parametrii λ *și* α.

În ipoteza că timpul de funcționare al unui generator electric este caracterizat prin densitatea de repartiție (IV.7) *și că duratele de funcționare a șapte generatoare electrice sînt* 100, 110, 150, 175, 185, 200 *și* 220 *ore să se estimeze* α *și* λ.

Soluție. Funcția de verosimilitate este

$$P(x_1,x_2,\dots,x_n;\lambda,\alpha)=\frac{\lambda^{n\alpha}\prod_{j=1}^{n}x_j^{\alpha-1}e^{-\lambda\sum_{j=1}^{n}x_j}}{(\Gamma(\alpha))^n}.$$

Urmează că

$$\frac{\partial \ln P}{\partial\alpha}=n\ln\lambda-n\psi(\alpha)+\sum_{j=1}^{n}\ln x_j$$

și

$$\frac{\partial \ln P}{\partial\lambda}=\frac{\alpha n}{\lambda}-\sum_{j=1}^{n}x_j,$$

unde am notat prin

$$\psi(\alpha)=\frac{d\ln\Gamma(\alpha)}{d\alpha}.$$

Soluția ecuațiilor de verosimilitate căpătate, adică $\hat{\alpha}$ și $\hat{\lambda}$ se obține prin metoda aproximațiilor succesive.

În cazul cînd α nu-i prea mic, de exemplu dacă $\alpha\geqslant 2$, atunci pentru funcția $\psi(\alpha)$ se poate da următoarea aproximație

$$\psi(\alpha)\simeq\ln\left(\alpha-\frac{1}{2}\right)+\frac{1}{24\left(\alpha-\frac{1}{2}\right)^2}.$$

În felul acesta deducem

$$\hat{\alpha} = \frac{\hat{\lambda}}{n} \sum_{j=1}^{n} x_j$$

şi

$$\hat{\lambda} = e^{\left[\psi(\hat{\alpha}) - \frac{1}{n} \sum\limits_{j=1}^{n} \lg x_j\right]}.$$

Cea de-a doua ecuaţie se poate simplifica dacă punem:

$$e^{\psi(\hat{\alpha})} \simeq \left(\hat{\alpha} - \frac{1}{2}\right)\left(1 + \frac{1}{24\left(\hat{\alpha} - \frac{1}{2}\right)^2}\right) = \hat{\alpha} - \frac{1}{2} + \frac{1}{24\left(\hat{\alpha} - \frac{1}{2}\right)},$$

unde am ţinut cont că pentru valori mici ale lui x avem

$$e^x \simeq 1 + x.$$

Dacă notăm cu

$$\hat{w} = \hat{\alpha} - \frac{1}{2}$$

atunci putem scrie

$$\hat{w} = -\frac{1}{2} + \frac{\hat{\lambda}}{n} \sum_{j=1}^{n} x_j,$$

$$\hat{\lambda} = \left(\hat{w} + \frac{1}{24\hat{w}}\right)\left(\prod_{j=1}^{n} x_j\right)^{-\frac{1}{n}}.$$

Avem

$$\frac{1}{7} \sum_{j=1}^{7} x_j = 162{,}9$$

şi

$$e^{-\frac{1}{7} \sum\limits_{j=1}^{n} \ln x_j} = \left(\prod_{j=1}^{7} x_j\right)^{-\frac{1}{7}} = 0{,}00637.$$

Prin urmare

$$\hat{w} = \frac{1}{2} + 162{,}9\,\hat{\lambda}$$

şi

$$\hat{\lambda}\left(\hat{w} + \frac{1}{2\hat{w}}\right) \cdot 0{,}00637$$

Folosind o metodă de rezolvare aproximativă aflăm

$$\hat{\lambda} = 0{,}0841, \qquad \hat{w} = 13{,}20$$

de unde avem

$$\lambda = 0{,}0841 \quad \text{şi} \quad \alpha = 13{,}70.$$

IV.25. *Să presupunem că în urma unei experienţe poate rezulta unul dintre evenimentele E_1, E_2, E_3 şi E_4, cu probabilităţile*

$$p_1 = (2 + \theta)/4,$$

$$p_2 = p_3 = (1 - \theta)/4,$$

$$p_4 = \theta/4$$

şi că în urma a n repetări a experienţei $E_1(E_2)$ $[E_3]$ $\{E_4\}$ s-a realizat de a (b) [c] {d} ori. Să se găsească prin aproximaţii succesive estimaţia de maximă verosimilitate a parametrului θ [cuprins în (0,1)]. Să se studieze cazul particular: $a = 1997$; $b = 906$; $c = 904$; $d = 32$.

S o l u ţ i e. Din funcţia de verosimilitate

$$P(a, b, c, d \mid \theta) = (2 + \theta)^a (1 - \theta)^{b+c} \theta^d 4^{-(a+b+c+d)}$$

obţinem

$$\frac{\partial \ln P}{\partial \theta} = \frac{a}{2 + \theta} - \frac{b + c}{1 - \theta} + \frac{d}{\theta}.$$

Rezultă ecuaţia de verosimilitate

$$n\theta^2 + (d + 2b + 2c - a)\,\theta - 2d = 0.$$

Deoarece produsul dintre coeficientul lui θ^2 şi termenul liber este negativ, produsul rădăcinilor acestei ecuaţii trebuie să fie de asemenea negativ, şi numai o rădăcină poate fi pozitivă. Numai această rădăcină pozitivă cade în intervalul permis pentru θ. Această valoare a lui θ este dată de relația

$$2n\hat{\theta} = [a - d - 2(b + c)] + \{[a + 2(b + c) + 3d]^2 - 8a(b + c)\}^{1/2}.$$

Estimația de maximă verosimilitate poate fi evaluată din această formulă.

Ținînd seama de valorile observate, obținem $\hat{\theta} = 0{,}0357$; de asemenea

$$D^2(\hat{\theta}) \sim -\frac{1}{M\left(\frac{\partial^2 \ln P}{\partial \theta^2}\right)} = \frac{2\theta(1-\theta)(2+\theta)}{n(1+2\theta)} = 0{,}0000336,$$

$\hat{\theta}$ fiind înlocuit prin $\hat{\theta}$.

Plecînd de la un estimator ineficient să găsim iterativ pe $\hat{\theta}$. Fisher propune estimatorul ineficient

$$t = \{a + d - (b + c)\}/n$$

care este consistent și are dispersia:

$$D^2(t) = (1 - \theta^2)/n.$$

Pentru t obținem

$$t = \{1997 + 32 - (906 + 904)\}/3839 = 0{,}0570.$$

Aceasta este depărtată de valoarea lui $\hat{\theta}$, 0,0357 pe care o căutăm. Cum asimptotic avem

$$\hat{\theta} = t + \left(\frac{\partial \ln P}{\partial \theta}\right)_t D^2(\hat{\theta}),$$

pentru prima aproximație rezultă

$$\hat{\theta}_1 = 0{,}0570 + \left(\frac{\partial \ln P}{\partial \theta}\right)_{\theta=t} D^2(\hat{\theta})\,|_{\theta=t}.$$

Acum

$$\left(\frac{\partial \ln P}{\partial \theta}\right)_{\theta=0{,}057} = \frac{1\,997}{2(0{,}057)} - \frac{1\,810}{0{,}943} + \frac{32}{0{,}057} = -387{,}1713,$$

$$D^2(\hat{\theta})\,|_{\theta=0{,}057} = \frac{2 \times 0{,}057 \times 0{,}943 \times 2{,}057}{3\,839 \times 1{,}114} = 0{,}00005170678,$$

astfel că estimaţia îmbunătăţită este

$$\hat{\theta}_1 = 0{,}0570 - 387{,}1713 \times 0{,}00005170678 =$$

$$= 0{,}0570 - 0{,}0200 = 0{,}0370,$$

care este destul de apropiat de valoarea căutată 0,0357.

O a doua iteraţie dă:

$$\left(\frac{\partial \ln P}{\partial \theta}\right)_{\theta=0{,}0370} = \frac{1\,997}{2{,}037} - \frac{1\,810}{0{,}963} + \frac{32}{0{,}037} = -34{,}31495,$$

$$D^2(\hat{\theta})\,|_{\theta=0{,}0370} = \frac{2 \times 0{,}037 \times 0{,}963 \times 2{,}037}{3\,839 \times 1{,}074} = 0{,}00003520681,$$

şi prin urmare

$$\hat{\theta}_2 = 0{,}0370 - 34{,}31495 \times 0{,}00003520681 =$$

$$= 0{,}0370 - 0{,}0012 = 0{,}0358.$$

Aceasta este foarte apropiată de valoarea căutată.

IV.26. *Să se estimeze parametrii α, σ, p din repartiţia caracterizată prin densitatea*

$$f(x) = \frac{1}{\sigma^p \Gamma(p)} \left(\frac{x-\alpha}{\sigma}\right)^{p-1} \exp\left\{-\left(\frac{x-\alpha}{\sigma}\right)\right\}$$

$$\alpha \leqslant x < \infty. \quad \alpha > 0; \quad p > 2,$$

pe baza unei selecţii de volum n.

S o l u ţ i e. Pentru funcţia de verosimilitate

$$P = \sigma^{-np}[\Gamma(p)]^{-n} \prod_{i=1}^{n} \left(\frac{x_i - \alpha}{\sigma}\right)^{p-1} \exp\left\{-\sum_{i=1}^{n} \left(\frac{x_i - \alpha}{\sigma}\right)\right\}$$

avem

$$\ln P = -np \ln \sigma - n \ln \Gamma(p) + \sum_{i=1}^{n} (p-1) \ln \frac{x_i - \alpha}{\sigma} -$$

$$- \sum_{i=1}^{n} \left(\frac{x_i - \alpha}{\sigma}\right).$$

Cele trei ecuații de verosimilitate sînt

$$\frac{\partial \ln P}{\partial \alpha} = -(p-1)\sum_{i=1}^{n}\frac{1}{(x_i-\alpha)} + \frac{n}{\sigma} = 0,$$

$$\frac{\partial \ln P}{\partial \sigma} = -\frac{np}{\sigma} + \sum \frac{x_i-\alpha}{\sigma^2} = 0,$$

$$\frac{\partial \ln P}{\partial p} = -n\ln\sigma - n\frac{d}{dp}\ln\Gamma(p) + \sum \ln(x_i-\alpha) = 0.$$

Luînd parametrii în ordinea de mai sus, avem matricea inversă a matricei de covarianțe

$$\sum^{-1} = n\begin{vmatrix} \frac{1}{\sigma^2(p-2)} & \frac{1}{\sigma^2} & \frac{1}{\sigma(p-1)} \\ \frac{1}{\sigma^2} & \frac{p}{\sigma^2} & \frac{1}{\sigma} \\ \frac{1}{\sigma(p-1)} & \frac{1}{\sigma} & \frac{d^2\ln\Gamma(p)}{dp^2} \end{vmatrix}$$

cu determinantul

$$|\Sigma^{-1}|/n^3 = \Delta = \frac{1}{(p-2)\sigma^4}\left\{2\frac{d^2\ln\Gamma(p)}{dp^2} - \frac{2}{p-1} + \frac{1}{(p-1)^2}\right\}.$$

Dispersiile estimaților sînt

$$D^2(\hat{\alpha}) = \frac{1}{n\Delta\sigma^2}\left\{p\frac{d^2\ln\Gamma(p)}{dp^2} - 1\right\}.$$

$$D^2(\hat{\sigma}) = \frac{1}{n\Delta\sigma^2}\left\{\frac{1}{p-2}\cdot\frac{d^2\ln\Gamma(p)}{dp^2} - \frac{1}{(p-1)^2}\right\},$$

$$D^2(\hat{p}) = \frac{2}{n\Delta(p-2)\sigma^4} = \frac{2}{n}\Bigg/\left\{2\frac{d^2\ln\Gamma(p)}{dp^2} - \frac{2}{p-1} + \frac{1}{(p-1)^2}\right\}.$$

Pentru p mare, folosind seria lui Stirling

$$\frac{d^2}{dp^2}\ln\Gamma(1+p) = \frac{d^2}{dp^2}\left\{\frac{1}{2}\ln(2\pi) + \left(p+\frac{1}{2}\right)\ln p - p + \right.$$

$$\left. + \frac{1}{12p} - \frac{1}{360p^2} + \dots\right\}$$

găsim că

$$2\frac{d^2}{dp^2}\ln\Gamma(1+p)-\frac{2}{p}+\frac{1}{p^2}=\frac{1}{3}\left\{\frac{1}{p^3}-\frac{1}{5p^5}+\frac{1}{7p^7}-\ldots\right\}$$

şi prin urmare

$$D^2(\hat{p})=\frac{6}{n}\left\{(p-1)^3+\frac{1}{5}(p-1)\right\}. \qquad \text{(IV.8)}$$

Dacă estimăm parametrii prin egalarea momentelor de selecţie cu momentele teoretice corespunzătoare găsim

$$\alpha+\sigma p=m_1; \qquad \sigma^2 p=m_2; \qquad 2\sigma^3 p=m_3,$$

astfel că

$$b_1=m_3^2/m_2^3=4/p, \qquad \text{(IV.9)}$$

unde b_1 este valoarea de selecţie a lui β_1,

Pentru estimarea prin metoda momentelor

$$D^2(b_1)=\frac{\beta_1}{n}\{4\beta_4-24\beta_2+36+9\beta_1\beta_2-12\beta_3+35\beta_1\},$$

care pentru repartiţia de faţă se reduce la

$$D^2(b_1)=\frac{\beta_1^2}{n}\cdot\frac{6(p+1)(p+5)}{p},$$

Deci, din (IV.9) avem pentru $\tilde{p}$ estimaţia prin metoda momentelor

$$D^2(\tilde{p})=\frac{p^4}{16}D^2(b_1)=\frac{6}{n}p(p+1)(p+5).$$

Eficienţa acestui estimator, pentru p mare, din (IV.8) este

$$\frac{D^2(\hat{p})}{D^2(\tilde{p})}=\frac{\{(p-1)^3+(p-1)/5\}}{p(p+1)(p+5)}.$$

care este mai mică ca unu.

IV.27. a) *Folosind metoda verosimilităţii maxime şi presupunînd că $(x_1, x_2, \ldots, x_n)$ este o selecţie dintr-o populaţie în care caracteristica sub cercetare are densitatea*

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta} & \textit{dacă} \quad 0<x<\theta,\\[2mm] 0 & \textit{dacă} \quad x\notin(0,\theta)\end{cases}$$

să se estimeze θ.

b) *Să se calculeze densitatea de repartiţie a variabilei aleatoare* $\hat{\theta}$.

c) *Să se calculeze* $M(\hat{\theta})$ *şi* $D^2(\hat{\theta})$.

Soluţie. a) Funcţia de verosimilitate va fi în acest caz

$$P(x_1, x_2, \ldots, x_n; \theta) = \left(\frac{1}{\theta}\right)^n,$$

unde

$$0 < x_j < \theta, \quad 1 \leqslant j \leqslant n.$$

Funcţia $P(x_1, x_2, \ldots, x_n; \theta)$ este maximă pentru cea mai mică valoare posibilă a lui θ. Dar θ nu poate fi mai mic decît cea mai mare dintre valorile $x_1, x_2, \ldots, x_n$. Aşadar

$$\hat{\theta} = \max_{1 \leqslant j \leqslant n} x_j.$$

b) Aplicînd definiţia funcţiei de repartiţie avem

$$F_{\hat{\theta}}(x) = P(\hat{\theta} < x) = P(\max_{1 \leqslant j \leqslant n} x_j < x) =$$

$$= P(x_1 < x;\ x_2 < x; \ldots; x_n < x) = \prod_{j=1}^{n} P(X_j < x) =$$

$$= \begin{cases} 0 & \text{dacă} \quad x \leqslant 0, \\ \left(\int_0^x \frac{dt}{\theta}\right)^n & \text{dacă} \quad 0 < x < \theta, \\ 1 & \text{dacă} \quad x \geqslant \theta. \end{cases}$$

Urmează că

$$f_{\hat{\theta}}(x) = \begin{cases} \frac{nx^{n-1}}{\theta^n} & \text{dacă} \quad x \in (0, \theta), \\ 0 & \text{dacă} \quad x \notin (0, \theta). \end{cases}$$

c) Din definiţia momentelor avem

$$M(\hat{\theta}) = \frac{n}{\theta^n} \int_0^{\theta} x^n \, dx = \frac{n}{n+1} \theta,$$

iar

$$M_2(\hat{\theta}) = \frac{n}{\theta^n} \int_0^\theta x^{n+1} \, dx = \frac{n\theta^2}{n+2}.$$

De aici urmează că

$$D^2(\hat{\theta}) = M_2(\hat{\theta}) - M^2(\hat{\theta}) = \frac{n\theta^2}{(n+1)^2 (n+2)}.$$

Observație. Se vede că $\hat{\theta}$ nu mai este o estimație nedeplasată a parametrului θ.

IV.28. a) *Folosind metoda verosimilității maxime și presupunînd că (x_i, y_i), $1 \leqslant i \leqslant n$ este o selecție dintr-o populație în care vectorul aleator (X, Y) are densitatea de repartiție*

$$f(x, y) = \frac{1}{2\pi \sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} (x^2 - 2\rho xy + y^2) \right\}$$

să se estimeze parametrul ρ.

b) *Să se găsească expresia asimptotică a dispersiei estimației astfel obținute.*

Soluție. a) Funcția de verosimilitate este

$$P = \frac{1}{(2\pi)^n (1-\rho^2)^{n/2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} (\Sigma x^2 - 2\rho \Sigma\Sigma xy + \Sigma y^2) \right\}.$$

Avem

$$\ln P = -n \ln (2\pi) - \frac{n}{2} \ln (1 - \rho^2) - \frac{1}{2(1-\rho^2)} (\Sigma x^2 - 2\rho \Sigma\Sigma xy + \Sigma y^2),$$

de unde ecuația de maximă verosimilitate

$$\frac{n\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2} (\Sigma x^2 - 2\rho \Sigma\Sigma xy + \Sigma y^2) + \frac{1}{1-\rho^2} \Sigma xy = 0,$$

ecuația care se mai scrie

$$\rho(1-\rho^2) + (1+\rho^2) \frac{1}{n} \Sigma xy - \rho \left(\frac{1}{n} \Sigma x^2 + \frac{1}{n} \Sigma y^2 \right) = 0.$$

Această ecuație are trei rădăcini, două dintre ele putînd fi complexe. Dacă toate trei sînt reale, cum $|\rho| < 1$, ținînd seama că $P(x, \overline{\theta}) \geqslant P(x, \theta)$, vom alege ca estimație de maximă verosimilitate acea care corespunde la valoarea cea mai mare a funcției de verosimilitate.

Dacă exprimăm ecuaţia cubică în forma

$$\lambda^3 + p\lambda + q = 0$$

cu

$$\lambda = \rho - \frac{1}{3n} \Sigma xy,$$

condiţia ca să avem numai o rădăcină reală, este ca:

$$4p^3 + 27q^2 > 0$$

şi este îndeplinită cînd $p > 0$, unde

$$p = \frac{1}{n} \Sigma x^2 + \frac{1}{n} \Sigma y^2 - \frac{1}{3}\left(\frac{1}{n} \Sigma xy\right)^2 - 1. \qquad \text{(IV.10)}$$

Deoarece, momentele de selecţie în (IV.10) sînt estimaţii consistente ale momentelor teoretice corespunzătoare, p converge în probabilitate către

$$\left(1 + 1 - \frac{1}{3} \rho^2 - 1\right) = 1 - \frac{1}{3} \rho^2 > 0.$$

Astfel, în selecţiile mari rădăcina reală a ecuaţiei de maximă verosimilitate va fi o estimaţie de maximă verosimilitate, rădăcinile complexe fiind inadmisibile.

b) Derivînd $\ln P$ de două ori, obţinem

$$\frac{\partial^2 \ln P}{\partial \rho^2} = \frac{n(1 + \rho^2)}{(1 - \rho^2)^2} - \frac{(1 + 3\rho^2)}{(1 - \rho^2)^3} (\Sigma x^2 - 2\rho\Sigma xy + \Sigma y^2) +$$

$$+ \frac{4\rho}{(1 - \rho^2)^3} \Sigma xy,$$

şi cum $M(x^2) = M(y^2) = 1$ şi $M(x\, y) = \rho$,

$$M\left(\frac{\partial^2 \ln P}{\partial \rho^2}\right) = \frac{n(1 + \rho^2)}{(1 - \rho^2)^2} - \frac{2n(1 + 3\rho^2)}{(1 - \rho^2)^3} + \frac{4n\rho^2}{(1 - \rho^2)^3} = - \frac{n(1 + \rho^2)}{(1 - \rho^2)^2}.$$

Asimptotic avem

$$D^2(\hat{\rho}) = \frac{(1 - \rho^2)^2}{n(1 + \rho^2)}.$$

III.29. *Folosind metoda verosimilității maxime și presupunînd că (x_i, y_i), $1 \leqslant i \leqslant n$ este o selecție dintr-o populație în care vectorul aleator (X, Y) are densitatea*

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \right.\right.$$

$$\left.\left. - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}.$$

$$(-\infty < x, y < \infty, \quad \sigma_1, \sigma_2 > 0, |\rho| < 1)$$

să se estimeze parametrii μ_1, μ_2, σ_1, σ_2 și ρ. Să se evalueze matricea dispersiilor estimațiilor de maximă verosimilitate.

Soluție. Avem

$$\ln P(x, y \mid \mu_1, \mu_2, \sigma_1, \sigma_2) = -n\ln(2\pi) - \frac{n}{2}\left[\ln\sigma_1^2 + \ln\sigma_2^2 + \right.$$

$$\left. + \ln(1-\rho^2)\right] - \frac{1}{2(1-\rho^2)}\sum\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \right.$$

$$\left. - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right],$$

de unde rezultă ecuațiile

$$\left.\begin{aligned} \frac{\partial \ln P}{\partial\mu_1} &= \frac{n}{\sigma_1(1-\rho^2)}\left[\frac{\bar{x}-\mu_1}{\sigma_1} - \rho\frac{\bar{y}-\mu_2}{\sigma_2}\right] = 0,\\ \frac{\partial \ln P}{\partial\mu_2} &= \frac{n}{\sigma_2(1-\rho^2)}\left[\frac{\bar{y}-\mu_2}{\sigma_2} - \rho\frac{\bar{x}-\mu_1}{\sigma_1}\right] = 0, \end{aligned}\right\} \qquad \text{(IV.11)}$$

$$\left.\begin{aligned} \frac{\partial \ln P}{\partial(\sigma_1^2)} &= -\frac{1}{2\sigma_1^2(1-\rho^2)}\left[n(1-\rho^2) - \frac{\sum(x-\mu_1)^2}{\sigma_1^2} + \right.\\ &\left. + \rho\frac{\sum(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right] = 0,\\ \frac{\partial \ln P}{\partial(\sigma_2^2)} &= -\frac{1}{2\sigma_2^2(1-\rho^2)}\left[n(1-\rho^2) - \frac{\sum(y-\mu_2)^2}{\sigma_2^2} + \right.\\ &\left. + \rho\frac{\sum(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right] = 0, \end{aligned}\right\} \qquad \text{(IV.12)}$$

$$\frac{\partial \ln P}{\partial\rho} = \frac{1}{(1-\rho^2)}\left\{n\rho - \frac{1}{(1-\rho^2)}\left[\rho\left(\frac{\sum(x-\mu_1)^2}{\sigma_1^2} + \right.\right.\right.$$

$$\left.\left.\left. + \frac{\sum(y-\mu_2)^2}{\sigma_2^2}\right) - (1+\rho^2)\frac{\sum(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right\} = 0. \qquad \text{(IV.13)}$$

a) Presupunînd că μ_1, μ_2, σ_1, σ_2 sînt cunoscuţi, pentru a estima pe ρ rezolvăm (IV. 13). Se obţine o ecuaţie cubică în ρ, a cărei soluţie este estimaţia de maximă verosimilitate $\hat{\rho}$.

b) μ_1 şi μ_2 fiind cunoscute pentru a găsi estimaţiile de maximă verosimilitate ale parametrilor σ_1^2, σ_2^2 şi ρ vom rezolva sistemul de ecuaţii constituit din cele două ecuaţii din (IV.12) şi din ecuaţia (IV.13). Obţinem

$$\left.\begin{aligned} n(1-\rho^2) &= \frac{\Sigma(x-\mu_1)^2}{\sigma_1^2} - \rho\,\frac{\Sigma(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}, \\ n(1-\rho^2) &= \frac{\Sigma(y-\mu_2)^2}{\sigma_2^2} - \rho\,\frac{\Sigma(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}, \end{aligned}\right\} \tag{IV.14}$$

şi

$$\begin{aligned} n(1-\rho^2) = {} & \frac{\Sigma(x-\mu_1)^2}{\sigma_1^2} + \frac{\Sigma(y-\mu_2)^2}{\sigma_2^2} - \\ & - \frac{1+\rho^2}{\rho}\cdot\frac{\Sigma(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}. \end{aligned} \tag{IV.15}$$

Scăzînd ecuaţia (IV.15) din suma celor două ecuaţii din (IV.14) rezultă

$$n(1-\rho^2) = \frac{1-\rho^2}{\rho}\,\frac{\Sigma(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2},$$

de unde

$$\hat{\rho} = \frac{\Sigma(x-\mu_1)(y-\mu_2)}{n\sigma_1\sigma_2}. \tag{IV.16}$$

Ţinînd seama de (IV.16) în (IV.14), obţinem

$$\begin{aligned} \hat{\sigma}_1^2 &= \frac{1}{n}\,\Sigma(x-\mu_1)^2, \\ \hat{\sigma}_2^2 &= \frac{1}{n}\,\Sigma(y-\mu_2)^2, \end{aligned} \tag{IV.17}$$

şi din (IV.16)

$$\hat{\rho} = \frac{\Sigma(x-\mu_1)(y-\mu_2)}{n\hat{\sigma}_1\hat{\sigma}_2}. \tag{IV.18}$$

Deci în acest caz, estimaţia de maximă verosimilitate este coeficientul de corelaţie de selecţie calculat cînd mediile sînt cunoscute.

Estimațiile de maximă verosimilitate $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\rho}$ date de (IV.17) și (IV.18) sînt suficiente și prin urmare

$$\Sigma_3^{-1} = (\sigma_{rs}^{-1}) = -\left(\frac{\partial^2 \ln P}{\partial\theta_r \partial\theta_s}\right)_{\hat{\theta}=\theta} = \left(\frac{\partial \ln P}{\partial\theta_r}\cdot\frac{\partial \ln P}{\partial\theta_s}\right) =$$

$$= \frac{n}{(1-\rho^2)}\begin{pmatrix} \frac{2-\rho^2}{4\sigma_1^4} & \frac{-\rho^2}{4\sigma_1^2\sigma_2^2} & \frac{-\rho}{2\sigma_1^2} \\ \frac{-\rho^2}{4\sigma_1^2\sigma_2^2} & \frac{2-\rho^2}{4\sigma_2^4} & \frac{-\rho}{2\sigma_2^2} \\ \frac{-\rho}{2\sigma_1^2} & \frac{-\rho}{2\sigma_2^2} & \frac{1+\rho^2}{1-\rho^2} \end{pmatrix}, \tag{IV.19}$$

de unde

$$\Sigma_3 = \frac{1}{n}\begin{pmatrix} 2\sigma_1^4 & 2\rho^2\sigma_1^2\sigma_2^2 & \rho(1-\rho^2)\,\sigma_1^2 \\ 2\rho^2\sigma_1^2\sigma_2^2 & 2\sigma_2^4 & \rho(1-\rho^2)\,\sigma_2^2 \\ \rho(1-\rho^2)\,\sigma_1^2 & \rho(1-\rho^2)\,\sigma_2^2 & (1-\rho^2)^2 \end{pmatrix}. \tag{IV.19'}$$

c) Dacă vrem să estimăm toți parametrii repartiției, rezolvăm sistemul constituit din (IV.11), (IV.12), și (IV.13). (IV.11) se reduce la

$$\frac{(\bar{x}-\mu_1)}{\sigma_1} = \rho\,\frac{(\bar{y}-\mu_2)}{\sigma_2},$$

$$\frac{(\bar{y}-\mu_2)}{\sigma_2} = \rho\,\frac{(\bar{x}-\mu_1)}{\sigma_1},$$

o pereche de ecuații a cărei singură soluție este

$$\bar{x}-\mu_1 = \bar{y}-\mu_2 = 0. \tag{IV.20}$$

(IV.17), (IV.18) și (IV.20) dau estimațiile de maximă verosimilitate

$$\hat{\mu}_1 = \bar{x}, \quad \hat{\sigma}_1^2 = \frac{1}{n}\Sigma(x-\bar{x})^2,$$

$$\hat{\mu}_2 = \bar{y}, \quad \hat{\sigma}_2^2 = \frac{1}{n}\Sigma(y-\bar{y})^2, \tag{IV.21}$$

$$\hat{\rho} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\hat{\sigma}_1\hat{\sigma}_2},$$

Estimațiile de maximă verosimilitate definite în (IV.21) constitui o mulțime suficientă de statistici. Pe lîngă aceasta, matricea Σ_3^{-1} (IV.19

va fi o parte a inversei matricei de covarianţe $\Sigma_5^{-1}(5\times 5)$ pe care o căutam. Scriind parametrii înordinea μ_1, μ_2, σ_1^2, σ_2^2, ρ, (IV.19) va fi minorul principal al lui Σ_5^{-1}. Din (IV.11) obţinem

$$\frac{\partial^2 \ln P}{\partial \mu_1^2} = \frac{-n}{\sigma_1^2(1-\rho^2)}, \qquad \frac{\partial^2 \ln P}{\partial \mu_1 \partial \mu_2} = \frac{n\rho}{\sigma_1\sigma_2(1-\rho^2)},$$

$$\frac{\partial^2 \ln P}{\partial \mu_2^2} = \frac{-n}{\sigma_2^2(1-\rho^2)},$$

în timp ce

$$\frac{\partial^2 \ln P}{\partial \mu_i \partial \sigma_1^2} = \frac{\partial^2 \ln P}{\partial \mu_i \partial \sigma_2^2} = \frac{\partial^2 \ln P}{\partial \mu_i \partial \rho} = 0, \qquad i = 1, 2,$$

pentru $\bar{x} = \mu_1$ $\bar{y} = \mu_2$ Inversa matricei de covarianţe a estimaţiilor de maximă verosimilitate $\hat{\mu}_1$ şi $\hat{\mu}_2$ fiind

$$\Sigma_2^{-1} = \frac{n}{(1-\rho^2)} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix},$$

avem

$$\Sigma_5^{-1} = \begin{pmatrix} \Sigma_2^{-1} & 0 \\ 0 & \Sigma^{-1} \end{pmatrix}.$$

Cum

$$\Sigma_2 = \frac{1}{n} \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

rezultă că

$$\Sigma_5 = \begin{pmatrix} \Sigma_2 & 0 \\ 0 & \Sigma_3 \end{pmatrix},$$

unde Σ_3 este dată de (IV.19′).

Din acest rezultat se vede că mediile de selecţie sînt repartizate independent de dispersiile şi covarianţele de selecţie în repartiţia normală bidimensională şi că corelaţia între mediile de selecţie este ρ; corelaţia între dispersiile de selecţie este ρ^2.

IV.30. *Dacă*

$$\begin{pmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{ki} \end{pmatrix},$$

$1 \leqslant i \leqslant n$ este o selecție de volum dintr-o populație normală k dimensională $N(\mu, \Sigma)$, $(n > k)$, atunci estimația de verosimilitate maximă a vectorului μ este

$$\hat{\mu} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_k \end{pmatrix},$$

iar a matricei de covarianță Σ este

$$\hat{\Sigma} = \frac{S}{n},$$

unde S este matricea de covarianță corespunzătoare selecției considerate.

Soluție. Densitatea de repartiție corespunzătoare repartiției $N(\mu, \Sigma)$ este

$$f(x_1, x_2, \dots, x_n) = \frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{(1/2)k}} e^{-\frac{1}{2}Q(x_1, x_2, \dots, x_n)}$$

unde

$$Q(x_1, x_2, \dots, x_n) = \sum_{i,j=1}^{k} \sigma^{ij}(x_i - \mu_i)(x_j - \mu_j),$$

iar σ^{ij} sînt elementele matricei inverse matricei de covarianță teoretice.

Luînd logaritmi naturali obținem

$$\ln f(x_1, x_2, \dots, x_n) = \frac{1}{2}\ln|\sigma^{ij}| - \frac{1}{2}k\ln 2\pi - \frac{1}{2}Q(x_1, \dots, x_n).$$

Ecuațiile verosimilității maxime

$$\frac{\partial \ln f(x_1, \dots, x_n)}{\partial \mu_i} = 0 \qquad \left(\frac{\partial\left(-\frac{1}{2}Q(x_1, \dots, x_n)\right)}{\partial \mu_i} = 0\right)$$

$$i = 1, 2, \dots, k$$

sînt

$$\sum_{j=1}^{k} \sigma^{1j}(x_{jl} - \mu_j) = 0,$$

$$\sum_{j=1}^{k} \sigma^{2j}(x_{jl} - \mu_j) = 0,$$

$$\dots\dots\dots\dots\dots\dots$$

$$\sum_{j=1}^{k} \sigma^{kj}(x_{kl} - \mu_j) = 0,$$

care este un sistem liniar și omogen în necunoscutele

$$x_{1l} - \mu_1, \quad x_{2l} - \mu_2, \dots, x_{kl} - \mu_k, \quad l = 1, 2, \dots, n,$$

sistem care admite numai soluția banală deoarece

$$\begin{vmatrix} \sigma^{11} & \sigma^{21} \dots \sigma^{k1} \\ \sigma^{12} & \sigma^{22} \dots \sigma^{k2} \\ \dots & \dots \\ \sigma^{1n} & \sigma^{2n} \dots \sigma^{kk} \end{vmatrix} \neq 0$$

(repartiție normală k-dimensională nedegenerată).
Atunci avem

$$x_{jl} = \mu_j, \quad 1 \leqslant j < k, \quad 1 \leqslant l \leqslant n,$$

adică

$$\sum_{l=1}^{n} x_{jl} = n\mu_j \Rightarrow \hat{\mu}_j = \frac{\sum\limits_{l=1}^{n} x_{jl}}{n} = \bar{x}_j.$$

Pentru a afla pe S pornim de la

$$\frac{\partial \ln f(x_1, \dots, x_n)}{\partial \sigma^{ij}} = \frac{\partial \left\{ \frac{1}{2} \ln |\sigma^{ij}| - \frac{1}{2} Q(x_1, \dots, x_n) \right\}}{\partial \sigma^{ij}} = 0$$

care ne conduce la

$$\sigma_{ij} - (x_{il} - \hat{\mu}_i)(x_{jl} - \hat{\mu}_j) = 0,$$

adică

$$n\sigma_{ij} = \sum_{l=1}^{n} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j)$$

sau

$$\hat{\sigma}_{ij} = s_{ij} = \frac{1}{n} \sum_{l=1}^{n} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j).$$

IV.31. a) *Să se arate că estimaţia* $\hat{\theta}$ *obţinută prin metoda celor mai mici pătrate pentru parametrul* θ *din modelul* (IV.22)

$$y = x\theta + \varepsilon, \tag{IV.22}$$

unde

$y' = (y_1, \ldots, y_n)$, $x = (x_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq k}}$ *este de rang* k, $\theta' = (\theta_1, \ldots, \theta_k)$ *şi* $\varepsilon' = (\varepsilon_1, \ldots, \varepsilon_k)$ *cu* $M(\varepsilon) = 0$, $D^2(\varepsilon) = \sigma^2 I$, I *fiind de ordin* $n \times n$ *este nedeplasată.*

b) *Să se calculeze dispersia acestei estimaţii.*

Soluţie. a) Estimaţia parametrului θ prin metoda celor mai mici pătrate este

$$\hat{\theta} = (x'x)^{-1} x'y$$

sau ţinînd seama de (IV.22).

$$\hat{\theta} = (x'x)^{-1} x'(x\theta + \varepsilon) = \theta + (x'x)^{-1} x'\varepsilon.$$

Cum $M(\varepsilon) = 0$, $M(\hat{\theta}) = \theta$ ceea ce arată că θ este o estimaţie nedeplasată a parametrului θ.

b) Prin definiţie

$$D^2(\hat{\theta}) = M[(\hat{\theta} - \theta)(\hat{\theta} - \theta)']$$

şi cum

$$(\hat{\theta} - \theta)(\hat{\theta} - \theta)' = ((x'x)^{-1} x'\varepsilon)((x'x)^{-1} x'\varepsilon)' =$$
$$= (x'x)^{-1} x'\varepsilon\varepsilon' x(x'x)^{-1}.$$

Obţinem

$$D^2(\hat{\theta}) = (x'x)^{-1} x' M(\varepsilon\varepsilon') x(x'x)^{-1} = \sigma^2 (x'x)^{-1}.$$

IV.32. *Să se arate că*

$$\hat{\sigma}^2 = \frac{(y - x\hat{\theta})'(y - x\hat{\theta})}{n - k}$$

este o estimație nedeplasată a parametrului σ^2.

Soluție. Avem $y = x\theta + \varepsilon$ și $\hat{\theta} = (x'x)^{-1}x'y$, ceea ce ne permite să scriem

$$y - x\hat{\theta} = x\theta + \varepsilon - x(x'x)^{-1}x(x\theta + \varepsilon) =$$

$$= x\theta + \varepsilon - x\theta - x(x'x)^{-1}x'\varepsilon = \varepsilon - x(x'x)^{-1}x'\varepsilon$$

sau punînd

$$L = x(x'x)^{-1}x', \qquad I_n - L = K,$$

$$y - x\hat{\theta} = (I_n - L)\,\varepsilon = K\varepsilon.$$

Deoarece

$$[(x'x)^{-1}]' = [(x'x)']^{-1} = (x'x)^{-1}$$

urmează că $L = L'$ și $K' = K$.

Cum

$$L^2 = x(x'x)^{-1}x'x(x'x)^{-1}x' = L, \qquad K^2 = K,$$

ceea ce arată că K este idempotentă.

Putem scrie

$$(y - x\hat{\theta})'(y - x\hat{\theta}) = \varepsilon'K'K\varepsilon = \varepsilon'K\varepsilon =$$

$$= \sum_i k_{ii}\varepsilon_i^2 + \sum_{i \neq j} k_{ij}\varepsilon_i\varepsilon_j,$$

$$\text{Ur } L = \text{Ur}\,(x(x'x)^{-1}x') = \text{Ur}\,(x'x(x'x)^{-1}) = \text{Ur } I_k,$$

$$\text{Ur } K = \text{Ur } I_n - \text{Ur } I_k = n - k,$$

de unde

$$M[(y - x\hat{\theta})'(y - x\hat{\theta})] = \sigma^2 \text{Ur } K = \sigma^2(n - k).$$

Urmează că $\hat{\sigma}^2$ este o estimație nedeplasată pentru σ^2.

IV.33. a) *Să se estimeze prin metoda celor mai mici pătrate parametrul* θ *din modelul*

$$y = x\theta + \varepsilon$$

unde

$$D^2(\varepsilon) = \sigma^2 \begin{pmatrix} \frac{1}{p_1} & & & 0 \\ & \frac{1}{p_2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{p_n} \end{pmatrix} = \sigma^2 P^{-1}$$

b) *Să se găsească valoarea medie şi dispersia acestei estimaţii.*

Soluţie. a) Avem

$$S = \varepsilon' P \varepsilon = (y - x\theta)' P(y - x\theta),$$

$$\frac{\partial S}{\partial \theta} = -2x'Py + 2x'Px\theta,$$

$$\frac{\partial^2 S}{\partial \theta^2} = 2x'Px \text{ sau punînd } x'Px = G, \ \frac{\partial^2 S}{\partial \theta^2} = 2G.$$

$|x'Px|$ fiind un determinant Gram

$$\frac{\partial S}{\partial \theta} = 0 \Rightarrow \hat{\theta} = G^{-1}\, x'Py.$$

b) Cum

$$\hat{\theta} = G^{-1}\, x'Py = G^{-1}x'P(x\theta + \varepsilon) = \theta + G^{-1}x'P\varepsilon$$

urmează că $M(\hat{\theta}) = \theta$, ceea ce arată că şi în acest caz $\hat{\theta}$ este o estimaţie nedeplasată a parametrului θ.

Avem

$$D^2(\hat{\theta}) = M[(\hat{\theta} - \theta)\,(\hat{\theta} - \theta)'] = M[G^{-1}x'P\varepsilon\varepsilon'P'x(G^{-1})']$$

şi cum

$$G = G' \quad \text{şi} \quad (G^{-1})' = (G')^{-1},$$

obţinem

$$D^2(\hat{\theta}) = M[G^{-1}x'P\varepsilon\varepsilon'P'xG^{-1}].$$

Ţinînd seama că $M(\varepsilon\varepsilon') = \sigma^2 P^{-1}$, găsim

$$D^2(\hat{\theta}) = \sigma^2 G^{-1}x'PxG^{-1} = \sigma^2 G^{-1}.$$

IV.34. *Să se arate că dacă vectorul* ε *este normal, atunci vectorii* $(y - x\hat{\theta})$ *şi* $(\hat{\theta} - \theta)$ *sînt independenţi şi repartizaţi normal* $(n - k)$ *-dimensional şi respectiv normal k-dimensional.*

S o l u ţ i e. Avem

$$y - x\hat{\theta} = x\theta + \varepsilon - xG^{-1}x'Py =$$
$$= x\theta + \varepsilon - xG^{-1}x'P(x\theta + \varepsilon) = \varepsilon - xG^{-1}x'P\varepsilon,$$

sau punînd $L = xG'x'P$, $k = I_n - L$, obţinem

$$y - x\hat{\theta} = (I_n - L)\varepsilon = k\varepsilon.$$

Deoarece ε este repartizat $N(0, \sigma^2 P^{-1})$ urmează că $y - x\hat{\theta}$ este repartizat de asemenea normal cu matricea de covarianţe

$$\sigma^2 kP^{-1}k' [N(0, \sigma^2 kP^{-1}k')].$$

Avem

$$kP^{-1}k' = (I_n - L)\, P^{-1}(I_n - L)' = P^{-1}[I_n - PLP^{-1} - L' + PLP^{-1}L']$$

şi cum L' este idempotentă

$$PLP^{-1} = PxG'x' = L',$$
$$kP^{-1}k' = P^{-1}\,[I_n - 2L' + (L')^2] =$$
$$= P^{-1}[I_n - L'].$$

Deci

$$y - x\hat{\theta} \text{ este repartizat } N(0, \sigma^2 P^{-1}(I_n - L')).$$

Vectorul $\hat{\theta} - \theta$ este repartizat normal cu matricea de covarianţe

$$\sigma^2 G^{-1}x'PP^{-1}PxG^{-1} = \sigma^2 G^{-1}.$$

Deoarece

$$k\mathcal{M}_\varepsilon(G^{-1}x'P)' = \sigma^2 kP^{-1}PxG^{-1} = \sigma^2 kxG^{-1} =$$
$$= \sigma^2(I_n - L)\,xG^{-1} =$$
$$= \sigma^2(xG^{-1} - xG^{-1}x'PxG^{-1}) =$$
$$= \sigma^2(xG^{-1} - xG^{-1}) = 0,$$

urmează că vectorii $y - x\hat{\theta}$ şi $\hat{\theta} - \theta$ sînt independenţi.

Matricea $G^{-1}x'P$ fiind o matrice de ordin $k \times n$, urmează că $\hat{\theta} - \theta$ este un vector normal k-dimensional. Se ştie că matricea H, definită prin

$$H\mathcal{M}_\varepsilon(G^{-1}x'P)' = 0,$$

unde

$$\mathcal{M}_\varepsilon = M(\varepsilon\varepsilon')$$

este chiar matricea k. Prin urmare $k\varepsilon$ şi deci $y - x\hat{\theta}$ este $(n - k)$-dimensional.

IV.35. *Să se arate că*

a) $(y - x\hat{\theta})'\,P(y - x\hat{\theta})/\sigma^2$ *este repartizat* $\chi^2(n - k)$.

b) $s^2 = (y - x\hat{\theta})'P(y - x\hat{\theta})/(n - k)$ *este o estimaţie nedeplasată pentru* σ^2.

S o l u ţ i e. a) Deoarece ε este repartizat $N(0, \sigma^2 P^{-1})$ urmează că $\varepsilon' P\varepsilon/\sigma^2$ este repartizată $\chi^2(n)$; deci $(y - x\theta)'P(y - x\theta)/\sigma^2$ este repartizat $\chi^2(n)$.

Dacă facem $\theta = \hat{\theta}$, aceasta revine la a impune k constrîngeri liniare. Prin urmare

$$(y - x\hat{\theta})'\,P(y - x\hat{\theta})/\sigma^2$$

va fi repartizat $\chi^2(n - k)$.

b) Ţinînd seama de rezultatul de la punctul a) găsim

$$M\left[\frac{(y - x\hat{\theta})'\,P(y - x\hat{\theta})}{\sigma^2}\right] = n - k$$

ceea ce trebuia arătat.

IV.36. *Să se arate că dacă $\hat{\theta}$ este estimația parametrului θ obținută prin metoda celor mai mici pătrate, atunci $\psi = \hat{\varphi} = b'\hat{\theta}$ este o estimație nedeplasată și de minimă dispersie pentru*

$$\varphi = b'\theta = \sum_1^k b_j\theta_j.$$

Soluție. Fie

$$\psi = \sum_1^n a_i y_i = a'y,$$

unde $y = x\theta + \varepsilon$.

Dorim ca $M(\psi) = \varphi$ oricare ar fi ψ. Avem

$$M(\psi) = a'M(y) = a'x\theta$$

și

$$\varphi = b'\theta$$

Pentru ca $M(\psi) = \varphi$ oricare ar fi θ, trebuie ca

$$b' = a'x \quad \text{sau} \quad b = x'a.$$

Pe de altă parte

$$D^2(\psi) = \sum_1^n a_i^2 D^2(y_i) = \sigma^2 \sum_1^n a_i^2 = \sigma^2 A.$$

Pentru ca $D^2(\psi)$ să fie minimă, trebuie să minimizăm A, ținînd seama că $b = x'a$.

Dacă h_j sînt multiplicatorii lui Lagrange, trebuie să minimizăm expresia

$$S = A + 2\sum_{j=1}^k h_j\left(b_j - \sum_{i=1}^n x_{ij}a_i\right),$$

$$\frac{\partial S}{\partial a_i} = 0 \Rightarrow a_i = \sum_{j=1}^k x_{ij}h_j \quad \text{sau} \quad a = xh.$$

Avem

$$b = x'a, \qquad a = xh,$$

de unde

$$b = x'xh, \qquad \psi = h'x'y = a'y.$$

Comparăm această ecuaţie cu $\lambda = b'\hat{\theta}$, unde $\hat{\theta}$ este soluţia ecuaţiei

$$x'x\hat{\theta} = x'y.$$

Avem

$$\psi = h'x'x\hat{\theta} = b\ \hat{\theta} = \lambda$$

şi

$$D^2(\psi) = \sigma^2 A = \sigma^2 a'a = \sigma^2 h'x'xh = \sigma^2 b'h.$$

IV.37. *Fie* $(Y_{ij})_{\substack{1\leqslant i\leqslant I\\ 1\leqslant j\leqslant J_i}}$, *variabilele aleatoare necorelate astfel încît*

$$M(Y_{ij}) = \beta_i \quad \text{şi} \quad D^2(Y_{ij}) = \lambda_i^2\sigma^2,$$

unde asupra valorilor medii β_i *nu se impune nici o restricţie,* σ^2 *este un număr pozitiv necunoscut şi* λ_i^2 *sînt numere pozitive cunoscute.*

a) *Să se arate că estimaţia parametrului* σ^2, *obţinută:*

a$_1$) *prin metoda celor mai mici pătrate este*

$$\sum_{i,j}(Y_{ij} - \overline{Y}_{i\cdot})^2.$$

a$_2$) *după metoda Gauss-Markov este*

$$\sum_{i,j}(Y_{ij} - \overline{Y}_i)^2/\lambda_i^2.$$

b) *Să se arate că*

$$\frac{1}{\sum_i[(J_i - 1)\lambda_i^2]}\sum_{i,j}(Y_{ij} - \overline{Y}_{i\cdot})^2$$

şi

$$\frac{1}{\sum_i(J_i - 1)}\sum_{i,j}(Y_{ij} - \overline{Y}_i)^2/\lambda_i^2$$

sînt estimaţii nedeplasate pentru σ^2.

IV.38. *Fie* $(Y_i)_{1\leqslant i\leqslant n}$, *n variabile aleatoare astfel încît*

$$M(Y_i) = \beta, \quad D^2(Y_i) = \sigma^2 \quad \text{şi} \quad \operatorname{cov}(Y_i, Y_{i'}) = \sigma^2\gamma, \quad i \neq i',$$

unde β *este un număr necunoscut*, σ^2 *este un număr pozitiv necunoscut și* γ *este un număr cunoscut ce satisface condiția*

$$[-1/(n-1)] < \gamma < 1.$$

a) *Să se arate că estimația parametrului* σ^2, *obținută prin:*
a_1) *metoda celor mai mici pătrate este*

$$\frac{1}{n-1}\Sigma(Y_i - \overline{Y})^2.$$

a_2) *metoda Gauss-Markov este*

$$\frac{1}{(n-1)(1-\gamma)}\Sigma(Y_i - \overline{Y})^2.$$

b) *Să se arate că estimația parametrului* β *obținută atît prin metoda celor mai mici pătrate cît și după metoda Gauss-Markov este* $\overline{Y}$.

Definiție. *Vom numi statistică ancilară (ancillary*) o statistică a cărei repartiție este aceeași pentru toate valorile posibile ale parametrului necunoscut.*

IV.39. *Fie X și Y două variabile aleatoare a căror densitate de repartiție comună este*

$$f(x, y) = e^{-\theta x - \frac{y}{\theta}}, \quad x > 0, \quad y > 0, \quad \theta > 0.$$

Să se arate că:

a) $F = XY$ *este o statistică ancilară.*

b) $T = \sqrt{Y/X}$ *este o estimație de maximă verosimilitate pentru* θ.

c) *Perechea* (F, T) *este suficientă, cu toate că* T *nu este o statistică suficientă.*

IV.40. *Fie X și Y două variabile aleatoare independente repartizate* $N(\theta, 1)$, θ *necunoscută.*
a) *Să se găsească o statistică ancilară pentru* θ.

---

*) Numele „ancillary" este dat de Fisher (1925) Theory of statistical estimation Proc. Camb. Phil. Soc. 22. 700.

b) *Să se arate că familia de statistici ancilare* $\{F_c\}$ *unde*

$$F_c = F_c(X, Y) = \begin{cases} X - Y & \textit{dacă} \quad X + Y < c \\ Y - X & \textit{dacă} \quad X + Y \geqslant c \end{cases}, \quad c \textit{ constantă fixată}$$

este echivalentă cu întregul spaţiu (X, Y).

S o l u ţ i e. a) Deoarece repartiţia statisticii $X - Y$ este independentă de θ urmează că $X - Y$ este o statistică ancilară.

b) Deoarece $X - Y$ şi $Y - X$ sînt identic repartizate $[N(0,2)]$ şi fiecare este independentă de $X + Y$, urmează că F_c este independentă de $X + Y$ şi are aceeaşi repartiţie ca $X - Y$. Astfel F_c este ancilară pentru fiecare c.

Considerăm familia de statistici ancilare $\{F_c\}$, $-\infty < c < \infty$.

Pentru X şi Y fixate, diferitele valori ale familiei F_c (pentru c variind) sînt fie $X - Y$, fie $Y - X$.

Valoarea c_0 a lui c în care F_c îşi schimbă semnul (F_c nu-şi schimbă semnul decît numai dacă $X - Y = 0$) este valoarea lui $X + Y$.

Astfel, dată $F_c(X, Y)$ pentru toţi c putem găsi $X + Y$ şi $X - Y$.

Deci familia F_c de statistici ancilare este echivalentă cu întregul spaţiu (X, Y).

Observaţie. În condiţii puţin restrictive, orice statistică independentă de o statistică suficientă este ancilară. Reciproca este de asemenea adevărată, statistica suficientă fiind completă.

Statistica $T = X + Y$ este o statistică suficient completă. Deci o statistică F poate fi ancilară dacă şi numai dacă F este independentă de T.

Dacă o statistică este ancilară atunci orice funcţie măsurabilă de F este de asemenea ancilară.

Definiţie. *Statistica* F_2 *include (sau este mai informativă decît) statistica* F_1 *dacă* F_1 *poate fi exprimată ca o funcţie de* F_2 *(şi scriem* $F_2 \supset F_1$ *sau* $F_1 \subset F_2$).

Definiţie. *Două statistici sînt echivalente dacă fiecare poate fi exprimată ca funcţie de cealaltă.*

IV.41. *Fie* $X_1, X_2, \ldots, X_n$ *o selecţie de volum n dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X repartizată* $N(\theta, 1)$.

Să se arate că:

a) *Fiecare dintre statisticile*

$$F_1 = X_1 - X_2, \quad F_2 = (X_1 - X_2, X_1 - X_3), \ldots, F_{n-1} =$$

$$= (X_1 - X_2, X_1 - X_3, \ldots, X_1 - X_n)$$

este ancilară și $F_1 \subset F_2 \subset \ldots \subset F_{n-1}$.

b) *Statisticile ancilare* F_{n-1} și

$$F = (X_2 - X_1, X_2 - X_3, \ldots, X_2 - X_n)$$

sînt echivalente.

Capitolul V

VERIFICAREA IPOTEZELOR STATISTICE

V.1. *Rezultatele a 24 de măsurători asupra diametrului rotactorului de la strungul revolver sînt trecute în primele două coloane ale tabelului V.1.*

Tabelul V.1

x_i	n_i	$n_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
7,90	2	15,80	−0,03	0,0009	0,0018
7,91	3	23,73	−0,02	0,0004	0,0012
7,92	3	23,76	−0,01	0,0001	0,0003
7,93	2	15,86	0	0	0
7,94	7	55,58	0,01	0,0001	0,0007
7,95	2	15,90	0,02	0,0004	0,0008
7,96	2	15,92	0,03	0,0009	0,0018
7,97	0	0	0,04	0,0016	0
7,98	1	7,98	0,05	0,0025	0,0025
7,99	2	15,98	0,06	0,0036	0,0072

În presupunerea că dimensiunea diametrului este o variabilă repartizată normal, să se folosească testul lui Grubbs la un prag de semnificație $P = 0{,}05$, *pentru a decide dacă* 7,99 *este rezultatul unei măsurători greșite.*

Soluție. Din tabelul V.1, găsim

$$\bar{x} = 7{,}93 \qquad s^2 = \frac{1}{n-1}\sum_{1}^{n}(x_i - \bar{x})^2 = 0{,}0006; \qquad s = 0{,}0245,$$

de unde

$$T_n = \frac{x_n - \bar{x}}{s} = 2{,}449.$$

Din tabele găsim $T = 2{,}64$.

Deoarece $T_n < T$ urmează că 7,99 nu este rezultatul unei măsurători greşite.

V.2. *Rezultatele a 40 de măsurători asupra lungimii rotactorului de la strungul revolver, scrise în ordinea mărimii, sînt trecute în primele două coloane ale tabelului V.2.*

Tabelul V.2

x_i	n_i	$x_i n_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$n_i(x_i - \bar{x})^2$
6,47	1	6,47	−0,2	0,0200	0,0200
6,60	1	6,60	−0,07	0,0049	0,0049
6,61	1	6,61	−0,06	0,0036	0,0036
6,65	2	13,30	−0,02	0,0004	0,0008
6,66	2	13,32	−0,01	0,0001	0,0002
6,67	13	86,71	0	0	0
6,68	5	33,40	0,01	0,0001	0,0005
6,69	5	33,45	0,02	0,0004	0,0020
6,70	7	46,90	0,03	0,0009	0,0063
6,71	2	13,42	0,04	0,0016	0,0032
6,73	1	6,73	0,06	0,0036	0,0036

În presupunerea că lungimea rotactorului este o variabilă normală, să se folosească la un prag de semnificaţie $P = 0{,}05$ testul lui Grubbs pentru a decide dacă valorile extreme diferă semnificativ de la celelalte valori.

Soluţie. Din tabelul V.2, găsim

$$\bar{x} = 6{,}67, \quad s^2 = 0{,}0451\,\frac{1}{39} = 0{,}001156; \quad s = 0{,}034$$

de unde

$$T_1 = \frac{\bar{x} - x_1}{s} = 5{,}88, \qquad T_n = \frac{x_n - \bar{x}}{s} = 1{,}76.$$

Din tabele găsim $T = 2{,}87$.

Deoarece $T_1 > T$ şi $T_n < T$ urmează că 6,47 trebuie înlăturată ca o eroare.

V.3. *În coloana 1 a tabelului V.3 sînt trecute valorile ohmice în KΩ a 15 tronsoane de ceramică, acoperite cu carbon. Aceste tronsoane provin de la un cuptor de* 0,5 W; *au fost decaptate în soluţia* 1—4—10 *şi la piroliză s-a folosit amestecul* 1/3 (metan/azot).

Temperatura în cuptor a fost de 1 100°C, *măsurătorile executîndu-se la 20 de minute după scoaterea din cuptor.*

Tabelul V.3

x_i	$(x_i - \bar{x})^2$	$x_{i+1} - x_i$	$(x_{i+1} - x_i)^2$
1,68	0,01850	0,27	0,0729
1,95	0,01796	−0,12	0,0144
1,83	0,00020	−0,27	0,0729
1,56	0,06554	0,38	0,1444
1,94	0,01538	−0,10	0,0100
1,84	0,00058	−0,06	0,0036
1,78	0,00130	0,14	0,0196
1,92	0,01082	−0,06	0,0036
1,86	0,00194	−0,13	0,0169
1,73	0,00740	−0,03	0,0009
1,70	0,01346	0,06	0,0036
1,76	0,00314	0,19	0,0361
1,95	0,01896	−0,05	0,0025
1,90	0,00706	−0,05	0,0025
1,85	0,00116		
27,25	0,18340		0,4039

Folosindu-se testul diferenţelor succesive şi un prag de semnificaţie $\alpha = 0{,}05$, *să se verifice caracterul întîmplător al acestor determinări.*

Soluţie. Avem

$$\bar{x} = 1{,}816; \quad s^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2 = \frac{1}{14}\,0{,}1834 = 0{,}01310;$$

$$q^2 = \frac{1}{2(n-1)}\sum_1^{n-1} (x_{i+1} - x_i)^2 = \frac{1}{28}\,0{,}4039 = 0{,}01442,$$

de unde

$$r = \frac{q^2}{s^2} = 1{,}1$$

Din tabele găsim $r_{0,05;15} = 0,603$.

Deoarece $r > r_{0,05;15}$ urmează că determinările au un caracter întîmplător.

V.4. *Dintr-un lot în care proporția p a articolelor necorespunzătoare este necunoscută se extrage o selecție de volum $n = 30$. Presupunem că în această selecție am găsit 4 articole defecte. Cu un prag de semnificație $\alpha = 0,05$, să se verifice ipoteza H_0: $p = 0,10$, față de ipoteza H_1:$p > 0,10$.*

Soluție. Asociem fiecărui element al selecției o variabilă aleatoare X_i care poate lua valoarea 1 sau 0, după cum articolul este necorespunzător sau corespunzător stasului (variabilele X_i astfel definite sînt independente). Numărul articolelor necorespunzătoare din selecție este dat de statistica

$$X = \sum_{1}^{k} X_i,$$

care urmează o repartiție binomială.

Dacă este adevărată ipoteza H_0, atunci

$$P(X = m) = C_m^{30}(0,1)^m \cdot (0,9)^{30-m} \qquad (m = 0, 1, ..., 30).$$

Prin urmare

$$P(X \geqslant 4) = \sum_{m=4}^{30} C_{30}^{m}(0,1)^m (0,9)^{30-m}.$$

Cum

$$P(X = 0) = (0,9)^{30}; \quad P(X = 2) = \frac{30!}{2!\,28!}(0,1)^2 \cdot (0,9)^{28};$$

$$P(X = 1) = \frac{30!}{29!}(0,1)\,(0,9)^{29};$$

$$P(X = 3) = \frac{30!}{3!\,27!}(0,1)^3 \cdot (0,9)^{27},$$

rezultă că

$$\sum_{m=0}^{3} P(X = m) = 0,6473.$$

Deci $P(X \geqslant 4) = 1 - 0,6473 = 0,3527$ și cum $P(X \geqslant 4) > \alpha$, nu respingem ipoteza H_0.

V.5. *Din experienţa anterioară se ştie că 10% dintre articolele produse de o anumită unitate sînt defecte. Modificîndu-se procesul de producţie într-o selecţie de 250 de articole se găsesc $x = 16$ articole defecte.*

a) *La un prag de semnificaţie $\alpha = 0,05$, să se verifice ipoteza conform căreia în urma modificării procesului de producţie proporţia articolelor defecte a scăzut.*

b) *Să se determine puterea testului pentru $p_1 = 0,05$.*

Soluţie. a) Trebuie verificată ipoteza

$$H_0 : p_0 = 0,1$$

faţă de ipoteza alternativă

$$H_1 : p_1 < 0,1.$$

Pentru marginea superioară x_c a regiunii critice, lema Neyman-Pearson conduce la expresia

$$x_c = np_0 - \frac{1}{2} + \sqrt{np_0(1 - p_0)}\, u_\alpha,$$

a cărei valoare este

$$250\frac{1}{10} - \frac{1}{2} + \sqrt{250\frac{1}{10}\cdot\frac{9}{10}}(-1,64) = 17,04.$$

Deoarece $x < x_c$, respingem ipoteza H_0, s-au altfel spus modificarea procesului de producţie conduce la scăderea proporţiei articolelor defecte.

b) Avem

$$\pi(p_1) = \Phi\left(\sqrt{n}\,\frac{p_0 - p_1}{\sqrt{p_1(1 - p_1)}} + \sqrt{\frac{p_0(1 - p_0)}{p_1(1 - p_1)}}\,u_\alpha\right)$$

şi cum $u_\alpha = -\,1,64$,

$$\pi(p_1) = \Phi\left(\sqrt{256}\,\frac{0,1 - 0,05}{\sqrt{0,05(1 - 0,05)}} + \sqrt{\frac{0,1(1 - 0,1)}{0,05(1-0,05)}}(-1,64)\right) = \Phi(0,58)$$

Din tabele găsim $\Phi(0,58) = 0,719$; deci $\pi\,(0,05) = 0,719$.

V.6. *O întreprindere de montaje a instalat două seturi a cîte 50 de traverse de stejar. Cele două seturi au fost tratate cu creozot prin două procedee diferite. După un număr de ani de folosire s-a constatat că 22 de traverse din primul set şi 18 din al doilea set mai erau în bună stare.*

Dacă $p_1(p_2)$ reprezintă probabilitatea ca o traversă de cale ferată tratată prin primul (al doilea) procedeu să fie în bună stare după un anumit număr de ani de folosire, la un prag de semnificație $\alpha = 0{,}05$, *să se verifice ipoteza*

$$H_0: p_1 = p_2,$$

față de ipoteza alternativă

$$H_1: p_1 \neq p_2.$$

Soluție. Dacă este adevărată ipoteza H_0, atunci $\frac{x_1}{n_1} - \frac{x_2}{n_2}$ pentru valori mari ale lui n_1 și n_2, este repartizată normal de medie zero și dispersie

$$p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right),$$

unde am notat prin p valoarea comună a probabilităților p_1, p_2.

Estimația de maximă verosimilitate a parametrului p este

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2},$$

unde $x_1(x_2)$ reprezintă numărul de traverse din primul (al doilea) set în bună stare după un anumit număr de ani.

Statistica

$$U = \frac{|\hat{p}_1 - \hat{p}_2)|}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

este repartizată $N(0,1)$ $\left(\text{aici } \hat{p}_1 = \frac{x_1}{n_1}, \ \hat{p}_2 = \frac{x_2}{n_2}\right)$ și are valoarea $0{,}08/0{,}98 = 0{,}08$. Din tabele găsim $u_1 - \frac{\alpha}{2} = u_{0,975} = 1{,}96$.

Deoarece $0{,}08 < 1{,}96$ ipoteza H_0 se acceptă.

V.7. *Din trei loturi de articole produse de trei unități diferite s-au considerat selecții, înregistrîndu-se numărul de articole defecte în fiecare selecție. S-au obținut rezultatele din tabelul V.4.*

Tabelul V.4

	Unitatea i			Total
	1	2	3	
Volumul selecției (n_i)	150	120	130	$x = 400$
Nr. de art. def. (x_i)	40	15	20	$x = 75$

La un prag de semnificație $\alpha = 0{,}05$, *să se verifice ipoteza*

$$H_0: p_1 = p_2 = p_3,$$

unde p_i *este proporția de articole defecte din articolele produse de unitatea i.*

Soluție. Statistica

$$T = \frac{n^2}{x(n-x)}\left(\sum_{i=1}^{k} \frac{x_i^2}{n_i} - \frac{x^2}{n}\right) = \frac{\sum_{i=1}^{k} n_i\hat{p}_i^2 - n\hat{p}^2}{\hat{p}(1-\hat{p})}, \qquad k \geqslant 2,$$

unde

$$\hat{p}_i = \frac{x_i}{n_i}, \quad 1 \leqslant i \leqslant k, \quad \hat{p} = \frac{x}{n}, \quad x = \sum_1^k x_i, \quad n = \sum_1^k n_i,$$

este repartizată $\chi^2(k-1)$.

În cazul de față $k = 3$, $n = 400$, $x = 75$.

$$\hat{p}_1 = 0{,}2666, \quad \hat{p}_2 = 0{,}1250, \quad \hat{p}_3 = 0{,}1538, \quad \hat{p} = 0{,}1875.$$

Deoarece

$$T = 10{,}3 > 5{,}99 = \chi^2_{2;0,05}$$

respingem ipoteza H_0, deci ipoteza conform căreia cele trei unități produc articole de aceeași calitate.

V.8. *Pe 12 loturi a câte 90 de plante fiecare s-au înregistrat* 19, 6, 9, 18, 15, 13, 14, 15, 16, 20, 22 *şi respectiv 14 plante infectate.*

Să se verifice la un prag de semnificaţie $\alpha = 0{,}05$ *ipoteza conform căreia proporţia de plante infectate este aceeaşi în toate cele 12 loturi.*

Soluţie. Avem

$$\bar{x} = 15{,}1\,;\ \sum_{1}^{12} (x_1 - \bar{x})^2 = 223,$$

de unde

$$X^2 = \frac{\sum\limits_{1}^{12} (x_i - \bar{x})^2}{\bar{x}\left[1 - \dfrac{\bar{x}}{n}\right]} = 18.$$

În cazul în care proporţia de plante infectate este aceeaşi în toate cele 12 loturi, statistica X^2 este repartizată $\chi^2(11)$.

Din tabele găsim $\chi^2_{11;0,05} = 19{,}7$.

Deoarece $X^2 < \chi^2_{11;0,05}$ nu putem respinge ipoteza egalităţii proporţiilor de plante infectate pe cele 12 loturi.

V.9. *Numărul mediu de accidente pe lună, datorită vitezei, este de* 10. *In prima lună după introducerea unor restricţii de viteză, acest număr scade la 4.*

Numărul de accidente pe lună, după introducerea restricţiilor de circulaţie, poate fi considerat o variabilă Poisson de medie λ. *La un prag de semnificaţie* $\alpha = 0{,}05$, *să se verifice ipoteza*

$$H_0 : \lambda = 10,$$

faţă de ipoteza alternativă

$$H_1 : \lambda < 10.$$

Soluţie. Deoarece $n = 1$ statistica

$$P = \sum_{r=0}^{s} e^{-n\lambda_0} (n\lambda_0)^r / r! = \sum_{r=0}^{4} e^{-10} \frac{10^r}{r!},$$

unde $s = \sum\limits_{1}^{n} x_i = x = 4$, are valoarea 0,029; urmează că ipoteza H_0 se respinge. Deci noile restricţii conduc la micşorarea numărului de accidente.

V.10. *Numărul de morţi, prin accident auto, într-un oraş în două săptămîni consecutive a fost de* 6 *şi respectiv* 3. *Presupunînd că numărul de morţi prin accident auto este o variabilă Poisson de medie* λ, *să se verifice, la un prag de semnificaţie* $\alpha = 0{,}01$, *ipoteza*

$$H_0\colon \lambda_1 = \lambda_2$$

faţă de ipoteza alternativă

$$H_1\colon \lambda_1 > \lambda_2 :$$

a) *Folosind aproximaţia normală.*
b) *Folosind un test exact.*

Soluţie. a) Statistica

$$u = \frac{\left(x_1 - \frac{1}{2}\right) - \left(x_2 + \frac{1}{2}\right)}{x_1 + x_2}$$

(repartizată $N(0,1)$*) are valoarea* 0,66. *Din tabele găsim* $u_{1-\alpha} = u_{0,99} = 2{,}33$.

Deoarece $u = 0{,}66 < 2{,}33 = u_{0,99}$ *nu respingem ipoteza* H_0.

b) *Deoarece* $x_1 = 6 > 3 = x_2$ *statistica*

$$\sum_{r=6}^{9} C_9^r \left(\frac{1}{2}\right)^9$$

are valoarea 0,25. *Deoarece* $P = 0{,}25 > 0{,}01 = \alpha$ *nu respingem ipoteza* H_0.

Acelaşi rezultat se obţine ţinînd seama că

$$\sum_{r=6}^{9} C_9^r \left(\frac{1}{2}\right)^9 = P\left(F_{12,8} < \frac{4}{6}\right),$$

F fiind o variabilă repartizată Snedecor cu 12 *şi* 8 *grade de libertate.*

V.11. *În* 10 *săptămîni într-un oraş s-au înregistrat* 12, 8, 15, 5, 2, 18, 13, 7, 6, 14 *accidente de automobil. Presupunînd că aceste frecvenţe urmează repartiţii Poisson de parametri* λ_i, $1 \leqslant i \leqslant 10$ *să se verifice, la un prag de semnificaţie* $\alpha = 0{,}05$, *ipoteza*

$$\lambda_1 = \lambda_2 = \ldots = \lambda_{10},$$

folosind indicele de dispersie a lui Poisson.

S o l u ţ i e. Avem

$$\bar{x} = 10, \quad \sum_{1}^{10} (x_i - \bar{x})^2 = 336, \quad \sum_{1}^{10} \frac{(x_i - \bar{x})^2}{\bar{x}} = 23{,}6.$$

Din tabele găsim $\chi^2_{9;0,05} = 16{,}919$.

Deoarece

$$\sum_{1}^{10} \frac{(x_i - \bar{x})^2}{\bar{x}} = 23{,}6 > 16{,}919 = \chi^2_{9;0,05}$$

respingem ipoteza egalităţii acestor medii.

V.12. *Într-o experienţă legată de încrucişarea a două specii de mazăre s-a obţinut, referitor la culoarea şi tipul mazării, următoarea împărţire*

Tabelul V.5

Sortul	1	2	3	4
Frecvenţele observat n_i	315	101	108	32

G. Mendel: Experiments in Plant-Hybridisation

Dacă se foloseşte teoria lui Mendel asupra eredităţii, probabilitatea obţinerii unui bob de mazăre din primul, al doilea, al treilea sau al patrulea sort este 9/16; 3/16; 3/16; 1/16.

La un prag de semnificaţie $\alpha = 0{,}05$, *folosind testul* X^2 *să se verifice concordanţa rezultatelor experimentale cu teoria lui Mendel.*

S o l u ţ i e. Prin ipoteză

$$p_1 = \frac{9}{16}, \quad p_2 = \frac{3}{16}, \quad p_3 = \frac{3}{16}, \quad p_4 = \frac{1}{16},$$

ceea ce permite să scriem pentru frecvenţele teoretice

$$np_1 = 312{,}75; \quad np_2 = 104{,}25; \quad np_3 = 104{,}25 \quad \text{şi} \quad np_4 = 34{,}75,$$

$$n = \sum_{1}^{n} n_i = 556.$$

Deoarece $\sum_{1}^{4} (n_i - np_i) = 0$, urmează că statistica

$$X^2 = \sum_{1}^{4} \frac{(n_i - np_i)^2}{np_i}$$

este repartizată $\chi^2(3)$.

Găsim $X^2 = 0{,}47$, iar din tabele $\chi^2_{3;0,05} = 7{,}81$.

Deoarece $X^2 = 0{,}47 < 7{,}81 = \chi^2_{3;0,05}$, tragem concluzia că rezultatele experimentale confirmă ipoteza lui Mendel.

V.13. *Se aruncă* 12 *zaruri simultan de* 26 306 *ori și se consideră ca un eveniment favorabil apariția pe fața unui zar a numărului* 5 *sau a numărului* 6.

Folosind testul X^2, *să se verifice ipoteza conform căreia repartiția observată (coloanele* 1 *și* 2 *ale tabelului V.6) este o repartiție binomială.*

Tabelul V.6

Numărul de succese	Frecvențele observate (n_i)	Frecvențele toretice date de dezvoltarea $26\,306\left(\frac{2}{3}+\frac{1}{3}\right)^{12}$ (np_i)	$\frac{(n_i - np_i)^2}{np_i}$	Frecvențele teoretice date de dezvoltarea 26 306 (0,6623 + + 0,3377)^{12} (np_i^*)	$\frac{(n_i - np_i^*)^2}{np_i^*}$
0	185	203	1,596	187	0,021
1	1149	1217	3,800	1146	0,008
2	3265	3345	1,913	3215	0,778
3	5475	5576	1,829	5465	0,018
4	6114	6273	4,030	6269	3,832
5	5194	5018	6,173	5115	1,220
6	3067	2927	6,696	3043	0,189
7	1331	1254	4,728	1330	0,001
8	403	392	0,309	424	0,040
9	105	87	3,724	96	0,844
≥ 10	18	14	1,143	16	0,250
Total	26306	26306	35,941	26306	8,201

Soluție. Dacă zarurile sînt cuburi perfecte, omogene din punct de vedere al structurii și aruncate pe o suprafață plană, atunci probabilitatea de a obține, la un singur zar, fața cu numărul 5 sau fața cu numărul 6 este $\frac{2}{6} = \frac{1}{3}$.

Deci p_i este dată de termenul de ordin $i + 1$ din dezvoltarea binomului $\left(\frac{2}{3} + \frac{1}{3}\right)^{12}$.

Din coloana a patra a tabelului V.6, găsim

$$X^2 = \sum_{1}^{11} \frac{(n_i - np_i)^2}{np_i} = 35{,}941,$$

iar din tabele $\chi^2_{10;0,001} = 29{,}6$.

Avem 10 grade de libertate, deoarece avem o singură constrîngere şi anume că suma frecvenţelor teoretice este egală cu suma frecvenţelor observate.

Deoarece $X^2 > \chi^2_{10;0,001}$, respingem ipoteza conform căreia repartiţia observată este o repartiţie binomială.

Deoarece repartiţia teoretică este sigură şi numărul de experimente destul de mare, neconcordanţa pusă în evidenţă de testul X^2 este surprinzătoare.

Înlăturînd faptul că în decursul experienţei s-au făcut greşeli, singura cauză căreia i s-ar datora acest lucru ar fi imperfecţiunea unui zar.

W.F.R. Weldon, căruia i se datorează experimentul, a supus unui examen atent zarurile şi a găsit într-adevăr, că unul dintre ele avea unele imperfecţiuni, din care cauză a recalculat X^2, luînd ca repartiţie teoretică în loc de

$$\left(\frac{2}{3} + \frac{1}{3}\right)^{12} \quad \text{pe} \quad (0{,}6623 + 0{,}3377)^{12},$$

în care 0,3377 nu este altceva decît frecvenţa relativă cu care în experiment, pe un zar se obţine faţa cu numărul 5 sau faţa cu numărul 6.

Cu alte cuvinte, Weldon a obţinut valoarea 0,3377 calculînd media

$$p^* = \frac{1}{26\,306 \times 12}[0 \cdot 185 + 1 \cdot 1149 + 2 \cdot 3265 + \ldots +$$

$$+ 10 \cdot 18] = 0{,}3377.$$

Din ultima coloană a tabelului V.6, găsim $X^2 = 0{,}201$.

Deoarece din datele de observaţie am estimat parametrul p^*, urmează că statistica X^2 va fi repartizată $\chi^2(9)$.

Din tabele găsim $\chi^2_{9;0,001} = 27{,}9$.

Deoarece $X^2 = 8{,}201 < 27{,}9 = \chi^2_{9;0,01}$, acceptăm ipoteza conform căreia repartiţia observată este o repartiţie binomială.

V.14. *Folosind testele: a) X^2; b) Kolmogorov, să se verifice la un prag de semnificaţie $\alpha = 0{,}05$ ipoteza conform căreia rezistenţa ohmică (în kΩ) a unor tronsoane de ceramică acoperite cu carbon este o variabilă aleatoare X repartizată normal.*

În urma unei selecţii de volum $n = 124$ s-au obţinut datele din tabelul V.7:

Tabelul V.7

1,97	1,77	1,61	1,61	1,63	1,76	1,80	1,64	1,86	1,67
1,69	1,85	1,75	1,68	1,78	1,93	1,98	1,65	1,85	1,75
1,65	1,77	1,77	1,73	1,83	1,95	1,91	1,81	1,67	2,01
1,69	1,78	2,01	1,95	2,02	1,86	1,82	1,72	1,78	1,87
1,72	1,69	1,74	1,86	1,92	1,90	1,74	1,86	1,80	1,90
1,81	1,68	1,90	1,83	1,77	1,85	1,85	2,03	1,68	1,93
1,70	1,74	1,87	1,61	1,98	1,94	1,74	1,70	1,89	1,76
1,77	1,71	1,76	1,77	1,86	1,97	1,93	1,73	1,71	
1,84	1,69	1,77	1,94	1,62	1,69	1,80	1,84	1,99	
1,81	1,92	1,82	2,02	1,73	1,72	1,79	1,91	1,63	
1,96	1,74	1,86	1,83	1,81	1,90	2,01	1,96	1,92	
1,81	2,03	1,88	1,84	1,70	1,97	1,64	1,99	1,94	
1,87	1,88	1,99	1,82	1,62	1,74	2,02	1,73	1,67	

Aceste tronsoane provin de la un cuptor de 0,5 W; au fost decaptate în soluţia 1—4—10 şi la piroliză s-a folosit amestecul 1/3 (metan/azot). Temperatura în cuptor a fost de 1 100°C, măsurătorile executîndu-se la 15 minute după scoaterea din cuptor.

Soluţie. Deoarece nu cunoaştem nici media şi nici dispersia variabilei aleatoare X, le vom estima cu ajutorul datelor din tabelul V.7. Estimaţiile de maximă verosimilitate ale parametrilor μ şi σ^2 (obţinute înainte de a grupa datele) sînt

$$\bar{x} = \frac{1}{124}\, 225{,}09 = 1{,}81$$

şi respectiv

$$s^2 = \frac{1}{n-1}\sum_{1}^{n}(x_i - \bar{x})^2 = \frac{1}{123}\, 1{,}5867 = 0{,}0129; \qquad s = 0{,}113.$$

a) Din tabelul V.7, obţinem tabelul V.8.

Tabelul V.8

Intervalul i	n_i	x_i	$\frac{x_i - \bar{x}}{s}$	$\Phi\left(\frac{x_i - \bar{x}}{s}\right)$	p_i	np_i	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
$(-\infty;\ 1{,}65]$	11	1,65	−1,41	0,0793	0,0793	9,8332	1,3614224	0,13845
$(1{,}65;\ 1{,}70]$	14	1,70	−0,97	0,1660	0,0867	10,7508	10,55730064	0,98200
$(1{,}70;\ 1{,}75]$	17	1,75	−0,53	0,2981	0,1321	16,3804	0,38390416	0,02344
$(1{,}75;\ 1{,}80]$	17	1,80	−0,09	0,4641	0,1660	20,5840	12,845056	0,62403
$(1{,}80;\ 1{,}85]$	18	1,85	0,35	0,6338	0,1697	21,0428	9,25863184	0,43999
$(1{,}85;\ 1{,}90]$	16	1,90	0,79	0,7852	0,1514	18,7736	7,69285696	0,40977
$(1{,}90;\ 1{,}95]$	13	1,95	1,23	0,8907	0,1055	13,0820	0,006724	0,00051
$(1{,}95;\ 2{,}00]$	10	2,00	1,68	0,9535	0,0628	7,7872	4,89648384	0,62879
$(2{,}00;\ +\infty)$	8	$+\infty$	$+\infty$	1,0000	0,0465	5,7660	4,990756	0,86555
	124							4,11253

Elementele coloanei 6 s-au obţinut cu ajutorul relaţiilor

$$p_1 = P(-\infty < X < x_1) = \Phi\left(\frac{x_1 - \bar{x}}{s}\right),$$

$$p_k = P(x_{k-1} < X \leqslant x_k) = \Phi\left(\frac{x_k - \bar{x}}{s}\right) - \Phi\left(\frac{x_{k-1} - \bar{x}}{s}\right), \qquad 2 \leqslant k \leqslant 8,$$

$$p_9 = P(x_8 < X < \infty) = 1 - \Phi\left(\frac{x_8 - \bar{x}}{s}\right).$$

Din ultima coloană a tabelului V.8, găsim

$$X^2 = \sum_1^9 \frac{(n_i - np_i)^2}{np_i} = 4{,}11253.$$

Deoarece am estimat doi parametri (μ şi σ^2) şi $\sum_1^9 (n_i - np_i)^2 = 0$, urmează că statistica X^2 are repartiţia $X^2(6)$.

Deoarece $X^2 = 4{,}11253 < 12{,}6 = \chi^2_{6;0,05}$ acceptăm ipoteza de normalitate.

b) Completăm coloanele de la 1 la 5 din tabelul V.8 cu coloanele

$\frac{n_i}{n}$	F_n	$\lvert F_n - F \rvert$
0,0887	0,0887	0,0094
0,1129	0,2016	0,0356
0,1371	0,3387	$\boxed{0,0406}$
0,1371	0,4758	0,0117
0,1452	0,6210	0,0128
0,1290	0,7500	0,0352
0,1094	0,8549	0,0358
0,0806	0,9355	0,0180
0,0645	1,0000	0,0000

Avem

$$d_n = \max \lvert F_n - F \rvert = 0{,}0406.$$

Din tabelele pentru

$n = 124$ şi $\alpha = 0{,}05$ $[K(\lambda) = 0{,}95]$ găsim

$$\lambda = 1{,}36.$$

Urmează că

$$\frac{\lambda}{\sqrt{n}} = 0{,}1221.$$

Deoarece $d_n < \frac{\lambda}{\sqrt{n}}$, acceptăm ipoteza conform căreia rezistenţa ohmică este o variabilă normală.

V.15. *Folosind testul lui Massey să se verifice la un prag de semnificaţie* $\alpha = 0{,}05$, *ipoteza conform căreia rezistenţa ohmică (în* kΩ*) a unor tronsoane de ceramică acoperite cu carbon este o variabilă aleatoare repartizată normal.*

Datele de selecţie scrise în ordinea mărimii sînt trecute în prima coloană a tabelului V.9.

Soluţie. În tabelul V.9 se arată modul de lucru cu testul Massey.

Tabelul V.9

$y_{(i)}$	$(y_{(i)} - \bar{y})^2$	$z_{(i)} = \frac{y_{(i)} - \bar{y}}{S}$	Frecv.	Fcţ. emp. de rep. $F_n(z_{(i)})$	$\Phi(z_{(i)})$	$\lvert F_n(z_{(i)}) - \Phi(z_{(i)})\rvert$
1,60	0,040401	−1,78	1	0,10	0,03754	0,06246
1,68	0,014641	−1,07	1	0,20	0,1423	0,05770
1,72	0,006561	−0,72	1	0,30	0,2358	0,06420
1,74	0,003721	−0,54	1	0,40	0,2946	$\boxed{0,10540}$
1,79	0,000121	−0,09	1	0,50	0,4641	0,03590
1,82	0,000361	0,17	1	0,60	0,5675	0,03250
1,85	0,002401	0,43	1	0,70	0,6664	0,03360
1,90	0,009801	0,88	1	0,80	0,8106	0,01060
1,93	0,016641	1,14	1	0,90	0,8729	0,02710
1,98	0,032041	1,59	1	1,00	0,9440	0,05592
18,01	0,126690					

Din primele coloane ale tabelului V.9, găsim

$$\bar{y} = 1,801, \quad S^2 = \frac{1}{10}\sum_{1}^{10}(y_{(i)} - \bar{y})^2 = 0,012669, \quad S = 0,1125.$$

Deoarece $d_{max} = 0,10540 < d^*_{10;0,05} = 0,130$, urmează că măsurătorile au o repartiţie normală.

V.16. *Trei grupuri de oameni, aleşi din regiuni geografice deosebite au următoarea repartiţie a culorii părului:*

Tabelul V.10

	Blond	Brunet	Şaten	Total ($n_{i.}$)
Regiunea X	15(19,165)	12(7,291)	3(3,541)	30
Regiunea Y	53(51,111)	17(19,444)	10(9,444)	80
Regiunea Z	24(21,666)	6(8,262)	4(4,014)	34
Total ($n_{.j}$)	92	35	17	144(n)

Folosind testul X^2 *şi un prag de semnificaţie* $\alpha = 0,01$, *să se verifice ipoteza conform căreia culoarea părului nu este influenţată de regiunea geografică.*

Soluţie. În ipoteza de independenţă, statistica

$$X^2 = \sum_{1}^{3}\sum_{1}^{3}\frac{(n_{ij} - np_{i.}p_{.j})^2}{np_{i.}p_{.j}}$$

este repartizată $\chi^2(3.3-1)$.

Deoarece $p_{i.}$ şi $p_{.j}$ sînt necunoscute le vom estima cu ajutorul metodei verosimilităţii maxime şi a frecvenţelor observate. Găsim

$$\hat{p}_{i.} = \frac{n_{i.}}{n}, \qquad \hat{p}_{.j} = \frac{n_{.j}}{n},$$

*) Tabelul pentru testul lui Massey pentru $8 \leqslant n \leqslant 32$ şi $\alpha = 0,05;\ 0,10$ se găseşte în [25] pag. 120.

statistica

$$X^2=\sum_1^3\sum_1^3\frac{\left(n_{ij}-\frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}}$$

avînd însă o repartiție $\chi^2[(3.3-1)-(3-1)-(3-1)]$.

Frecvențele teoretice corespunzătoare cazului de independență sînt trecute în paranteze.

Găsim $X^2 = 5{,}289$.

Cum

$$X^2 = 5{,}289 < 13{,}3 = X^2_{4;0{,}01}$$

acceptăm ipoteza de independență.

V.17. *Apartenența la una din grupele de sînge O, A, B, AB a 355 de indivizi din colectivitatea C și a 364 de indivizi din colectivitatea D este dată în tabelul V.11.*

Tabelul V.11

Colectiv \ Grupa (i)	0	A	B	AB	Total ($n_{i.}$)
$C_{(n_{1i})}$	121	120	79	33	353
D	118	95	121	30	364
Total ($n_{.i}$)	239	215	200	63	717 (n)

Folosindu-se testul de omogenitate pentru selecții paralele și un prag de semnificație $\alpha = 0{,}05$, *să se verifice ipoteza egalității proporțiilor a claselor de grupe de sînge în cele două colectivități.*

S o l u ț i e. Avem

$$p_1 = \frac{n_{11}}{n_{.1}} = 0{,}5063; \quad p_2 = \frac{n_{12}}{n_{.2}} = 0{,}5581,$$

$$p_3 = \frac{n_{13}}{n_{.3}} = 0{,}3950; \quad p_4 = \frac{n_{14}}{n_{.4}} = 0{,}5238; \quad p = \frac{n_{1\cdot}}{n} = 0{,}4923.$$

Cu aceste date obţinem

$$\sum_{1}^{4} n_{1i}p_i - n_{1.}p = 121(0,5063) + 120(0,5581) + 79(0,3950) +$$

$$+ 33(0,5238) - 353(0,4923) = 2,9428$$

şi

$$p(1 - p) = 0,24994,$$

de unde

$$X^2 = \frac{1}{p(1-p)} \sum_{1}^{4} (n_{1i}p_i - n_{1.}p) = 11,77.$$

Statistica X^2 este repartizată $\chi^2(3)$.

Din tabele găsim $\chi^2_{3;0,05} = 7,815$ şi deoarece $X^2 > \chi^2_{3;0,05}$ respingem ipoteza egalităţii proporţiilor.

V.18. *Folosind testul medianei şi un prag de semnificaţie* $\alpha = 0,05$ *să se verifice ipoteza conform căreia elementele liniilor* 6 *şi* 7 *ale tabelului* V.7 *aparţin aceleiaşi populaţii.*

Soluţie. Deoarece într-o scriere ordonată mediana ocupă locul central, urmează că în cazul de faţă mediana va fi dată de media aritmetică a numerelor 1,81 şi 1,83, adică de 1,82.

Repartiţia rezistenţelor ohmice a elementelor liniilor 6 şi 7 ale tabelului V.7 faţă de mediană este dată în tabelul V.12.

Tabelul V.12

	Linia 6	Linia 7	Total
Valori $x_i < M_e$	4(5)	6(5)	10(10)
Valori $x_i > M_e$	6(5)	4(5)	10(10)
Total	10	10	20

Dacă ipoteza emisă este adevărată, repartiția teoretică a datelor, trebuie să fie în privința medianei variabile aleatoare X ce dă rezistența ohmică, cea trecută în parantezele din tabelul V.12.

Numărul elementelor celor două selecții fiind mic ne vom servi în analiză de testul probabilităților exacte.

Deoarece frecvența din prima celulă (4) este mai mică decît frecvența teoretică (5), va trebui să determinăm probabilitățile de a obține tabelele

$$\begin{vmatrix} 4 & 6 \\ 6 & 4 \end{vmatrix} \quad \begin{vmatrix} 3 & 7 \\ 7 & 3 \end{vmatrix} \quad \begin{vmatrix} 2 & 8 \\ 8 & 2 \end{vmatrix} \quad \begin{vmatrix} 1 & 9 \\ 9 & 1 \end{vmatrix} \quad \begin{vmatrix} 0 & 10 \\ 10 & 0 \end{vmatrix}.$$

Găsim

$$p(4) = \frac{10!\,10!\,10!\,10!}{20!\,4!\,6!\,4!\,6!} = 0{,}23868,$$

$$p(3) = \frac{10!\,10!\,10!\,10!}{20!\,3!\,7!\,3!\,7!} = 0{,}07794,$$

$$p(2) = \frac{10!\,10!\,10!\,10!}{20!\,2!\,8!\,2!\,8!} = 0{,}01093,$$

$$p(1) = \frac{10!\,10!\,10!\,10!}{20!\,1!\,9!\,1!\,9!} = 0{,}00054,$$

$$p(0) = \frac{10!\,10!\,10!\,10!}{20!\,0!\,10!\,10!\,0!} = 0{,}00000.$$

Deoarece

$$\sum_{0}^{4} p(i) = 0{,}32809 > 0{,}05$$

nu putem respinge ipoteza emisă.

V.19. *La o întrebare asupra necesității introducerii unei noi emisiuni în programul de televiziune, răspund* 100 *de persoane. Din acestea* 60 *de persoane răspund „da", este necesară" și* 40 *de persoane „ nu, nu este necesară". După ce aceste persoane au vizionat emisiunea, se re-*

petă din nou întrebarea. Din cele 60 *de persoane care au spus prima oară „da“ numai* 10 *si-au menţint părerea, iar din cele* 40 *de persoane care au spus „nu“* , 39 *îşi menţin părerea.*

Folosind testul lui McNemar şi un prag de semnificaţie $\alpha = 0{,}05$, *să se verifice ipoteza conform căreia vizionarea programului nu a influenţat părerea persoanelor în legătură cu noua emisiune.*

S o l u ţ i e. Rezultatele experimentului sînt trecute în tabelul V.13.

T a b e l u l V.13

		Prima probă	
		Da	Nu
A 2-a probă	Nu	50(a)	39
	Da	10	1(b)

În cazul în care ipoteza emisă este adevărată, statistica

$$X^2 = \frac{(|a - b| - 1)^2}{a + b}$$

urmează o repartiţie $\chi^2(1)$.

Avem

$$X^2 = \frac{(|50 - 1| - 1)^2}{50 + 1} = \frac{2304}{51} = 45{,}17,$$

iar din tabele găsim $\chi^2_{1;0,05} = 3{,}84$.

Deoarece $X^2 > \chi^2_{1;0,05}$, respingem ipoteza emisă şi deci există diferenţa semnificativă între răspunsuri date înainte şi după vizionarea emisiunii.

V.20. *Selecţii independente din populaţia în care caracteristica sub cercetare este un vector aleator bidimensional de volum* 21, 30, 39, 26, 35 *au condus la următoarele valori pentru coeficientul de corelaţie de selecţie:* 0,39; 0,61; 0,43; 0,54 *şi respectiv* 0,48.

a) *La un prag de semnificaţie* $\alpha = 0{,}05$, *să se verifice ipoteza de omogeneitate*;

b) *În ipoteza de omogeneitate, să se găsească estimaţia coeficientului de corelaţie al populaţiei.*

S o l u ţ i e. a) Ţinînd seama de transformarea lui Fisher

$$z_i = \frac{1}{2} \lg \frac{1 + r_i}{1 - r_i}, \qquad i = 1, 2, 3, 4, 5,$$

obţinem tabelul V.14.

Tabelul V.14

r_i	z_i	$n_i - 3$	$(n_i - 3)\, z_i$	$(n_i - 3)\, z_i^2$
0,39	0,412	18	7,416	3,055
0,61	0,709	27	19,143	13,572
0,43	0,460	36	16,560	7,618
0,54	0,604	23	13,892	8,391
0,48	0,524	32	16,768	8,786
		136	73,799	41,422

Urmează că statistica

$$X^2 = \sum_1^5 (n_i - 3)\; z_i^2 - \frac{\left[\sum\limits_1^5 (n_i - 3)\, z_i\right]^2}{\sum\limits_1^5 (n_i - 3)},$$

care este repartizată $\chi^2(4)$ are valoarea 1,4.

Din tabele găsim $\chi^2_{4;0,05} = 9{,}49$.

Deoarece $X^2 < \chi^2_{4;0,05}$, acceptăm ipoteza de omogeneitate.

b) Obţinem

$$\bar{z} = \frac{\sum\limits_1^5 (n_i - 3)\, z_i}{\sum\limits_1^5 (n_i - 3)} = \frac{73{,}799}{136} = 0{,}5425$$

şi din tabele, corespunzător la această valoare, găsim $\rho = 0{,}495$.

V.21. *Pentru caracteristica „alungire“ a firului de relon s-au cules date pe o perioadă de 11 luni.*

În tabelul V.15 sînt trecute numărul gradelor de libertate şi dispersiile de selecţie pe aceste luni.

Tabelul V.15

Nr.	$\nu_i = n_i - 1$	s_i^2	s_i
1	151	6,8306	2,61
2	159	7,3437	2,70
3	159	8,9359	2,98
4	159	7,4673	2,73
5	151	6,5592	2,56
6	159	7,9302	2,81
7	127	6,4025	2,53
8	103	8,5293	2,92
9	167	6,9790	2,64
10	119	6,1055	2,47
11	103	9,4061	3,06

Folosind testul lui Bartlett şi un prag de semnificaţie $\alpha = 0{,}05$, să se verifice ipoteza stabilităţii procesului tehnologic relativ la caracteristica alungire.

Soluţie. Dacă este adevărată ipoteza emisă ($\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$) atunci statistica

$$X^2 = \frac{-\sum\limits_1^k \nu_i \ln \frac{s_i^2}{s^2}}{1 + \frac{1}{k-1}\sum\limits_1^k \left(\frac{1}{\nu_i} - \frac{1}{\nu}\right)},$$

unde

$$s^2 = \sum_1^k \nu_i s_i^2 \Big/ \sum_1^k \nu_i,$$

este repartizată $\chi^2(k-1)$.

Din tabelul V.15, găsim

$$\nu = \sum_1^{11} \nu_i = 1557; \quad \sum_1^{11} \frac{1}{\nu_i} = 0{,}080081; \quad \nu \ln s^2 = 2{,}009193,$$

$$\sum_1^{11} \nu_i \ln s_i^2 = 3115{,}496553,$$

ceea ce conduce la valoarea 13,165788 pentru X^2.

Din tabele găsim $\chi^2_{10;0,05} = 18,3$ şi deoarece $X^2 < \chi^2_{10;0,05}$ se acceptă ipoteza stabilităţii procesului tehnologic pentru această caracteristică.

V.22. *Pentru caracteristica rezistenţă* *) *a betoanelor marca* B 200 *livrate de o staţie de betoane s-au cules* 6 *serii de date, fiecare serie conţinînd* $n = 37$ *de date. În tabelul* V.16 *sînt trecute dispersiile de selecţie ale acestor* 6 *serii de date.*

Tabelul V.16

Seria	1	2	3	4	5	6
Dispersia s_i^2	52,81	43,47	39,72	45,86	50,16	49,32

Folosind testul lui Cochran şi un prag de semnificaţie $\alpha = 0{,}05$, *să se verifice ipoteza stabilităţii procesului tehnologic relativ la caracteristica rezistenţă.*

Soluţie. Deoarece

$$\max_{1 \leqslant i \leqslant 6} s_i^2 = 52{,}81 \quad \text{şi} \quad \sum_1^6 s_i^2 = 281{,}34,$$

statistica

$$Q = \frac{\max\limits_{1 \leqslant i \leqslant 6} s_i^2}{\sum\limits_1^6 s_i^2} \text{ are valoarea } 0{,}183.$$

Din tabele găsim $Q_{\alpha,k,n-1} = Q_{0,05;6;36} = 0{,}2612$ şi deoarece $Q < Q_{\alpha,k,n-1}$, urmează că procesul tehnologic relativ la caracteristica rezistenţei este stabil.

Observaţie. Se mai putea aplica şi testul lui Bartlett.

V.23. *Masa medie a locuitorilor unui oraş poate fi considerată ca o variabilă aleatoare* X *repartizată* $N(70,\ 5^2)$. *O selecţie de* 100 *locuitori*

*) Rezistenţa de sfărîmare prin compresiune a cuburilor de latură 20 cm, la vîrsta de 28 zile, ţinute în condiţii standard.

ai oraşului, cu domiciliul în zona parcurilor, este găsită ca avînd o masă medie de 80 kg.

a) *Acest rezultat indică faptul că locuitorii avînd domiciliul în zona parcurilor au o masă mai mare decît a celorlalţi locuitori* ($\alpha = 0{,}05$);

b) *Care este puterea testului pentru* $\mu_1 = 71$;

c) *Pentru ce volum al selecţiei testul are puterea* 0,8849.

S o l u ţ i e. a) Trebuie să verificăm ipoteza

$$H_0 : \mu = \mu_0 = 70$$

faţă de ipoteza alternativă

$$H_1 : \mu = \mu_1 > 70.$$

Din tabele găsim $u_{1-\alpha} = 1{,}64$, ceea ce conduce la valoarea

$$x_c = \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha} = 70 + \frac{5}{10} 1{,}64 = 70{,}82$$

pentru marginea inferioară a regiunii critice.

Deoarece

$$\bar{x} = 80 > x_c = 70{,}82$$

respingem ipoteza H_0.

b) Avem

$$\pi(\mu_1) = \Phi\left(u_\alpha + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)$$

şi cum $u_\alpha = -1{,}64$

$$\pi(\mu_1) = \Phi\left(-1{,}64 + \frac{71 - 70}{5/10}\right) = \Phi(0{,}36).$$

Din tabele $\Phi(0{,}36) = 0{,}6406$; deci $\pi(71) = 0{,}6406$.

c) Din tabele găsim $u_{1-\beta} = u_{0,8849} = 1{,}2$; obţinem

$$n = (u_{1-\beta} + u_{1-\alpha})^2 \left(\frac{\sigma}{\mu_1 - \mu_0}\right)^2 \simeq 202.$$

V.24. *O selecţie de* 16 *loturi de caprolactamă* *) *cristalizată, tip A are conţinutul de baze volatile* $\bar{x} = 0{,}22$ *miliechivalenţi pe kg. Presupunem că conţinutul de baze volatile este o variabilă aleatoare avînd repartiţia* $N(\mu,(0{,}08)^2)$.

---

*) Produs intermediar utilizat în producerea relonului, precum şi în alte sinteze în industria chimică şi farmaceutică.

a) *Să se verifice la un prag de semnificație* $\alpha = 0{,}01$ *ipoteza*

$$H_0 : \mu = \mu_0 = 0{,}20$$

față de ipoteza alternativă

$$H_1 : \mu = \mu_1 > 0{,}20.$$

b) *Care este puterea testului pentru* $\mu_1 = 0{,}21$.

S o l u ț i e. *a)* Din tabele găsim $u_{1-\alpha} = 2{,}33$ ceea ce conduce la valoarea

$$x_c = \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha} = 0{,}20 + \frac{0{,}08}{4} . 2{,}33 = 0{,}2466$$

pentru marginea inferioară a regiunii critice.

Deoarece

$$\bar{x} = 0{,}22 < 0{,}2466 = x_c$$

nu putem respinge ipoteza H_0 la pragul de semnificație $\alpha = 0{,}01$.

b) Avem

$$\pi(\mu_1) = \Phi\left(u_\alpha; + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)$$

și cum $u_\alpha = -2{,}33$

$$\pi(\mu_1) = \Phi\left(-2{,}33 + \frac{0{,}21 - 0{,}20}{0{,}08/4}\right) = \Phi(-1{,}83)\cdot$$

Din tabele găsim $\Phi(-1{,}83) = 0{,}0338$; deci $\pi(0{,}21) = 0{,}0338$.

V.25. *Experiența anterioară arată că durabilitatea unei anvelope auto poate fi considerată o variabilă aleatoare* $N(30\,000 \text{ km}, (800 \text{ km})^2)$. *Se face o schimbare în procesul de producție. O selecție de* 100 *anvelope are* $\bar{x} = 29\,000$ km. *Pe baza acestei selecții și la un prag de semnificație* $\alpha = 0{,}05$ *putem spune că noua metodă conduce la scăderea durabilității anvelopelor?*

S o l u ț i e. Trebuie să verificăm ipoteza

$$H_0 : \mu = \mu_0 = 30\,000,$$

față de ipoteza alternativă

$$H_1 : \mu = \mu_1 < 30\,000.$$

Din tabele găsim $u_\alpha = u_{0,05} = -1{,}64$, ceea ce conduce la valoarea

$$x_c = \mu_0 + \frac{\sigma}{\sqrt{n}} u_\alpha = 30\,000 - \frac{800}{10} 1{,}64 = 29\,864{,}8.$$

Deoarece

$$\bar{x} = 29\,000 < 29\,864{,}8 = x_c$$

respingem ipoteza H_0, deci noua metodă duce la scăderea durabilității anvelopelor.

V.26. *Durata de funcționare a unui tip oarecare de bec electric de* 100 *watti poate fi considerată ca o variabilă aleatoare* X *repartizată* N (1 500, 200²). *O selecție de* 25 *de astfel de becuri dă o durată medie de funcționare de* 1 380 *de ore.*

a) *La un prag de semnificație* $\alpha = 0{,}01$, *să se verifice ipoteza*

$$H_0 : \mu = \mu_0 = 1\,500$$

față de ipoteza alternativă

$$H_1 : \mu = \mu_1 < 1\,500.$$

b) *Care este puterea testului pentru* $\mu_1 = 1\,400$.

Soluție. *a)* Din tabele găsim $u_\alpha = u_{0,01} = -2{,}33$, ceea ce conduce la valoarea

$$x_c = \mu_0 + \frac{\sigma}{\sqrt{n}} u_\alpha = 1\,500 - \frac{200}{5} 2{,}33 = 1\,406{,}80$$

pentru marginea superioară a regiunii critice.

Deoarece

$$\bar{x} = 1\,380 < x_c = 1\,406{,}80$$

respingem ipoteza H_0.

b) Avem

$$\pi(\mu_1) = \Phi\left(u_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right) =$$

$$= \varnothing\left(-2{,}33 + \frac{1\,500 - 1\,400}{200/5}\right) = \Phi(0{,}17).$$

Din tabele găsim $\Phi(0{,}17) = 0{,}5675$; deci $\pi(1\,400) = 0{,}5675$.

V.27. *O selecție de volum* $n = 28$, *dintr-o populație în care caracteristica sub cercetare este un vector aleator bidimensional conduce la va-*

loarea 0,65. *O asemenea valoare poate proveni dintr-o populație în care coeficientul* ρ *are valoarea* 0,72?

S o l u ţ i e. Avem

$$z = \frac{1}{2}\lg\frac{1+r}{1-r} = 0,775.$$

$$M(z) = \frac{1}{2}\lg\frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)} = 0,908 + 0,013 = 0,921,$$

de unde

$$\frac{z - M(z)}{D(z)} = (n-3)^{1/2}[z - M(z)] = 5(-0,146) = -0,73.$$

Din tabele rezultă că

$$P\left(\frac{|z - M(z)|}{D(z)} > \frac{|0,775 - 0,921|}{1/5}\right) = P(|U| > 0,73) =$$

$$= 2[1 - \Phi(0,73)] = 0,4652 > 0,05.$$

Astfel coeficientul de corelație de selecție $r = 0,65$, *nu diferă semnificativ de* 0,72 *la pragul de semnificație* $\alpha = 0,05$.

V.28. *Două caracteristici sînt notate prin* (X, Y) *la bărbați și prin* (U, V) *la femei. În urma unei selecții de volum* $m = 28$, *s-a obținut pentru coeficientul de corelație de selecție* r_{xy} *valoarea* 0,64, *iar în urma unei selecții de volum* $n = 32$, *s-a obținut pentru coeficientul de selecție* r_{uv} *valoarea* 0,75.

În presupunerea că vectorii (X, Y) *și* (U, V) *sînt vectori normali, să se verifice ipoteza conform căreia coeficienții de corelație nu diferă pentru sexe, adică* $\rho_{xy} = \rho_{uv}$.

S o l u ţ i e. Obţinem

$$z_{xy} = \frac{1}{2}\lg\frac{1+r_{xy}}{1-r_{xy}} = 0,758 \quad \text{şi} \quad z_{uv} = \frac{1}{2}\lg\frac{1+r_{uv}}{1-r_{uv}} = 0,973.$$

Deoarece

$$\frac{|\rho|}{2}\left|\frac{1}{27} - \frac{1}{31}\right| < \frac{2}{837},$$

urmează că $z_{xy} - z_{uv}$ are o repartiţie normală de valoare medie zero şi dispersie

$$\frac{1}{m-3} + \frac{1}{n-3} = \frac{1}{25} + \frac{1}{29} = \frac{54}{725}.$$

Putem scrie

$$P(|Z_{xy} - Z_{uv}| \geqslant |0,758 - 0,973|) = P(|Z_{xy} - Z_{uv}| \geqslant 0,215) =$$

$$= P\left(\frac{|Z_{xy} - Z_{uv}|}{\sqrt{54/725}} > \frac{0,215}{\sqrt{54/725}}\right) = P(|U| \geqslant 0,78).$$

Din tabele găsim

$$P(|U| \geqslant 0,78) = 1 - P(|U| < 0,78) = 0,4354;$$

ceea ce ne permite să afirmăm că ipoteza emisă este adevărată.

V.29. *Presupunem că două maşini M_1 şi M_2 sînt folosite pentru ambalarea cărnii în pachete de 1 000 g. Din experienţa trecută se ştie că cantităţile de carne ambalată pot fi considerate variabile aleatoare repartizate $N(\mu_1, (3\text{ g})^2)$, respectiv $N(\mu_2\ (4\text{ g})^2)$.*

Presupunem că în urma cîntăririi a 100 pachete de carne din producţia fiecărei maşini s-au obţinut mediile:

$$\bar{x}_1 = 1\,007\text{ g} \quad \text{şi} \quad \bar{x}_2 = 1\,002\text{ g}.$$

La un prag de semnificaţie $\alpha = 0,01$ să se verifice:

a) *ipoteza*

$$H_0 : \mu_1 = \mu_2$$

faţă de ipoteza alternativă

$$H_1 : \mu_1 > \mu_2;$$

b) *ipoteza*

$$H_0 : \mu_1 = \mu_2$$

faţă de ipoteza alternativă

$$H_1 : \mu_1 \neq \mu_2.$$

Soluţie. *a)* Din tabele găsim $u_{1-\alpha} = u_{0,99} = 2,33$, ceea ce conduce la

$$x_c = u_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 2,33\sqrt{\frac{9}{100} + \frac{16}{100}} = 1,165$$

pentru marginea inferioară a regiunii critice.

Deoarece

$$\bar{x}_1 - \bar{x}_2 = 5 > 1,165 = x_c$$

respingem ipoteza H_0.

b) Din tabele găsim $u_{1-\frac{\alpha}{2}} = u_{0,995} = 2,57$. Deoarece

$$\left|\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right| = \left|\frac{5}{\frac{1}{2}}\right| = 10 > 2,57$$

respingem ipoteza H_0.

V.30. *Experiența acumulată arată că rezistența la rupere a cablurilor fabricate la o anumită întreprindere poate fi considerată o variabilă aleatoare X, repartizată normal de abatere medie pătratică* 15 kg.

O selecție de volum 9, *conduce la o rezistență medie de rupere* $\bar{x} = 400$ **kg.** *Dacă* μ *este media rezistenței la rupere:*

a) *Să se verifice, la un prag de semnificație* $\alpha = 0,01$, *ipoteza*

$$H_0 : \mu_0 = \mu = 450$$

față de ipoteza alternativă

$$H_1 : \mu = \mu_1 \neq \mu_0.$$

b) *Să se determine puterea testului pentru* $\mu_1 = 460$.

c) *Să se determine volumul n al selecției astfel încît testul să aibe puterea* 0,99.

Soluție. *a)* Din tabele găsim $u_{1-\frac{\alpha}{2}} = u_{0,995} = 2,57$.

Deoarece

$$\left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| = \left|\frac{400 - 450}{15/3}\right| = 10 > 2,57 = u_{1-\frac{\alpha}{2}}$$

respingem ipoteza H_0.

b) Avem

$$\pi(\mu_1) \simeq \Phi\left(u_{\frac{\alpha}{2}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)$$

și cum $u_{\frac{\alpha}{2}} = -2,57$, găsim

$$\pi(460) \simeq \Phi\left(-2,57 + \frac{460 - 450}{15/3}\right) = \Phi(-0,57).$$

Din tabele găsim $\pi(460) = 0,2843$.

c) Avem

$$n \simeq (u_{1-\beta} + u_{1-\frac{\alpha}{2}})^2 \left(\frac{\sigma}{\mu_1 - \mu_0}\right)^2$$

şi cum $u_{1-\beta} = u_{0,99} = 2{,}33$, găsim

$$n \simeq (2{,}33 + 2{,}57)^2 \left(\frac{15}{460 - 450}\right)^2 \simeq 54.$$

V.31. 10 *determinări ale procentului de apă într-o anumită soluţie au condus la* $\bar{x} = 0{,}832\%$ *şi* $s = 0{,}02\%$. *Dacă* μ *este procentul „adevărat" de apă în soluţia respectivă, la un prag de semnificaţie* $\alpha = 0{,}05$, *să se verifice:*

a) *Ipoteza*

$$H_0 : \mu = \mu_0 = 0{,}9\%$$

faţă de ipoteza alternativă

$$H_1 : \mu = \mu_1 < 0{,}9\%.$$

b) *Ipoteza*

$$H_0' : \sigma = \sigma_0 = 0{,}03\%$$

faţă de alternativa

$$H_1' : \sigma_1 < 0{,}03\%; \quad H_1'' : \sigma_1 > 0{,}03\%; \quad H_1''' : \sigma_1 \neq 0{,}03\%.$$

Soluţie. *a)* din tabele găsim $t_{9;\,0,05} = 2{,}262$, ceea ce conduce la valoarea

$$x_c = \mu_0 + t_{n-1;\,\alpha}\frac{s}{\sqrt{n}} = 0{,}9 + 2{,}262\,\frac{0{,}02}{\sqrt{10}} = 0{,}91$$

pentru marginea superioară a regiunii critice.

Deoarece $\bar{x} = 0{,}832\% < x_c = 0{,}91$ respingem ipoteza H_0.

b) Din tabele găsim $\chi^2_{9,0,05} = 16{,}9$; avem

$$\frac{\sigma_0^2\chi^2_{n-1,\alpha}}{n-1} = \frac{(0{,}0^39)\ (16{,}9)}{9} = 0{,}0^2169$$

şi cum

$$s^2 = 0{,}0^34 < \frac{\sigma_0^2\chi^2_{n-1,\alpha}}{n-1} = 0{,}0^2169$$

respingem ipoteza H_0.

Din tabele găsim $\chi^2_{n-1;\,1-\alpha} = 3{,}3$; avem

$$\frac{\sigma_0^2 \chi^2_{n-1,\,1-\alpha}}{n-1} = \frac{(0{,}0^3 9)\,(3{,}3)}{9} = 0{,}0^3 147$$

şi cum

$$s^2 = 0{,}0^3 4 > 0{,}0^3 33 = \frac{\sigma_0^2 \chi^2_{n-1,\,1-\alpha}}{n-1}$$

respingem ipoteza H_0.

Din tabele găsim

$$\chi^2_{n-1;\,\frac{\alpha}{2}} \simeq 19 \quad \text{şi} \quad \chi^2_{n-1;\,1-\frac{\alpha}{2}} = \chi^2_{9;\,0{,}975} = 2{,}7;$$

avem

$$\frac{\sigma_0^2 \chi^2_{n-1;\,\frac{\alpha}{2}}}{n-1} = \frac{(0{,}0^3 9)\cdot 19}{9} = 0{,}0^2 19,$$

$$\frac{\sigma_0^2 \chi^2_{n-1;\,1-\frac{\alpha}{2}}}{n-1} = \frac{(0{,}0^3 9)\cdot 2{,}7}{9} = 0{,}0^3 27.$$

V.32. *În urma a 8 determinări a viscozităţii a două tipuri A şi B de ulei pentru maşini auto s-au obţinut rezultatele din tabelul 17.*

Tabelul V.17

Marca A Viscozitatea (x_i)	Marca B Viscozitatea (y_j)
10,28	10,31
10,27	10,31
10,30	10,26
10,32	10,30
10,27	10,27
10,27	10,31
10,28	10,29
10,29	10,26

Presupunînd că viscozităţile pentru cele două tipuri de ulei auto sînt variabile aleatoare repartizate normal să se verifice, la un prag de semnificaţie $\alpha = 0{,}05$:

a) *Ipoteza*

$$H_0 : \mu_A = \mu_B$$

față de ipoteza alternativă

$$H_1 : \mu_A \neq \mu_B.$$

b) *Ipoteza* $\sigma_A^2 = \sigma_B^2$.

Soluție. *a)* Avem

$$\bar{x} = 10,285, \quad s_A^2 = \frac{1}{7}\sum_1^8 (x_i - \bar{x})^2 = 0,0^3314,$$

$$\bar{y} = 10,289, \quad s_B^2 = \frac{1}{7}\sum_1^8 (y_i - \bar{y})^2 = 0,0^3498,$$

$$c = \frac{\frac{s_A^2}{8}}{\frac{s_A^2}{8} + \frac{s_B^2}{8}} = 0,3867,$$

de unde

$$\frac{1}{n^*} = \frac{c^2}{7} + \frac{(1-c)^2}{7} = 0,0750962.$$

Urmează că în ipoteza H_0, statistica

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{s_A^2}{8} + \frac{s_B^2}{8}}}$$

este repatizată student cu $n^* = 13$ grade de libertate.
Din tabele găsim $t_{n^*,\alpha} = t_{13;\,0,05} = 2,160$.

Deoarece $t = 0,398 < t_{n^*,\alpha} = 2,160$ acceptăm ipoteza H_0.

b) În ipoteza $\sigma_A^2 = \sigma_B^2$, statistica

$$F = \frac{s_A^2}{s_B^2}$$

este repartizată $F_{7,7}$ și are valoarea 0,630522.

Din tabele găsim $F_{1-\alpha;\,7,7} = F_{0,95;\,7,7} = 3,79$.

Deoarece $F = 0,630522 < F_{0,95;\,7,7} = 3,79$ acceptăm ipoteza H_0

V.33. *Următoarele date dau efectele coroziunii în diferite terenuri asupra conductelor de oţel acoperite şi neacoperite.*

Tabelul V.18

Acoperite (x_{i1})	42	47	51	52	59	54	32	47	66
Neacoperite (x_{i2})	40	43	50	44	57	56	44	37	70

Aceste rezultate indică faptul că acoperirea conductelor de oţel micşorează efectul coroziunii? ($\alpha = 0{,}01$).

Soluţie. Avem $\bar{d} = 1$,

Tabelul V.19

$d_i = x_{i1} - x_{i2}$	2	4	1	8	2	-2	-12	10	-4
$d_i - \bar{d}$	1	3	0	7	1	-3	-13	9	-5
$(d_i - \bar{d})^2$	1	9	0	49	1	9	169	81	25

Trebuie să verificăm ipoteza

$$H_0 : \frac{1}{N} \sum_{1}^{N} (X_{i1} - X_{i2}) = 0.$$

În ipoteza H_0:

$$t = \frac{\bar{d} - \frac{1}{N} \sum_{1}^{N} (X_{i1} - X_{i2})}{\frac{1}{\sqrt{n}} \sqrt{\frac{\sum_{1}^{n} (d_i - \bar{d})^2}{n-1}}} = \frac{1 - 0}{\frac{1}{\sqrt{9}} \sqrt{\frac{312}{8}}} = 0{,}457.$$

Din tabele pentru 8 grade de libertate şi un prag de semnificaţie $\alpha = 0{,}01$, găsim $t_{8;\,0{,}01} = 3{,}355$.

Deoarece $t = 0{,}457 < 3{,}355 = t_{8;\,0{,}01}$, nu putem respinge ipoteza H_0.

V.34. *Să se determine valoarea cea mai mică a coeficientului de corelație r dintr-o selecție de volum $n = 18$, dintr-o populație normală bidimensională, care să fie semnificativă la un prag de semnificație $\alpha = 0,01$.*

S o l u ț i e. Din tabele găsim $t_{n-2,\alpha} = t_{16;\,0,01} = 2,921$.

Pentru ca r să fie semnificativ la un prag de semnificație $\alpha = 0,01$ trebuie ca

$$t = \sqrt{n-2}\,\frac{r}{\sqrt{1-r^2}} = 4\,\frac{r}{\sqrt{1-r^2}} > 2,91 = t_{16;\,0,01}.$$

Rezultă $r^2 > 0,36$ sau $|r| > 0,6$.

V.35. *O selecție de volum $n = 102$ dintr-o populație normală bidimensională conduce la valoarea 0,8 pentru coeficientul de corelație.*

Este această valoare a coeficientului de corelație, la un prag de semnificație $\alpha = 0,05$, semnificativă?

S o l u ț i e. Din tabele găsim $t_{n-2,\alpha} = t_{100;\,0,05} = 1,984$.
Avem

$$t = \sqrt{n-2}\,\frac{r}{\sqrt{1-r^2}} = 10\,\frac{0,8}{0,6} = 13,33.$$

Deoarece $t > t_{100;\,0,05}$ urmează că $r = 0,8$ este semnificativă.

V.36. *Folosind un prag de semnificație $\alpha = 0,05$ și coeficienții:*

a) *Spearman* (R_S).
b) *Kendall* (R_K).

să se verifice ipoteza conform căreia punctajele X și Y obținute de 10 studenți la muzică și matematică, date în tabelul V.20, sînt variabile aleatoare independente.

T a b e l u l V.20

		1	2	3	4	5	6	7	8	9	10
Punctajul	Muzică (X)	10	3	24	14	25	26	36	29	40	45
	Matematică (Y)	30	89	56	23	81	50	17	70	75	40

S o l u ţ i e. Rangurile celor 10 studenţi ţinînd seama de punctajele la muzică şi matematică sînt trecute în tabelul V.21

Tabelul V.21

		1	2	3	4	5	6	7	8	9	10
Rangul	Muzică	9	10	7	8	6	5	3	4	2	1
	Matematică	8	1	5	9	2	6	10	4	3	7

a) Avem $R_S = -0,1515$ şi $n = 10$.

Statistica

$$T = \sqrt{n-2}\,\frac{R_S}{\sqrt{1 - R_S^2}}$$

care este repartizată t_{n-2} are valoarea $-0,4335$.

Deoarece $|T| < t_{n-2,\frac{\alpha}{2}} = t_{10;\,0,025}$, acceptăm ipoteza conform căreia variabilele X şi Y sînt independente.

b) Dacă X şi Y sînt variabile aleatoare independente atunci

$$M(R_K) = 0, \quad D^2(R_K) = \frac{2n+5}{18n(n-1)}.$$

Găsim $R_K = -0,111$ şi $D^2(R_K) = 0,015432$.

Urmează că

$$T = \frac{R_K}{\sqrt{D^2(R_K)}}$$

care este repartizată $N(0,1)$, are valoarea $-0,896$, valoare care nu este semnificativă la un prag de semnificaţie $\alpha = 0,05$.

V.37. *Măsurătorile X şi Y a două oase diferite dintr-o specie zoologică S în viaţă au un coeficient de regresie* $A = \rho\,\frac{\sigma_2}{\sigma_1} = 0,60$.

Pentru o colecţie de 18 schelete măsurătorile au condus la rezultatele din primele două coloane ale tabelului V.22.

Tabelul V.22

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_j - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
31,4	51,8	0,8	−5,2	0,64	27,04	−4,16
30,6	52,0	0,0	−5,0	0,00	25,00	0,00
30,8	52,1	0,2	−4,9	0,04	24,01	−0,98
30,4	52,3	−0,2	−4,7	0,04	22,09	0,94
30,9	52,5	0,3	−4,5	0,09	20,25	−1,35
29,8	52,4	−0,8	−4,6	0,64	21,16	3,68
30,3	52,3	−0,3	−4,7	0,09	22,09	1,41
28,8	53,4	−1,8	−3,6	3,24	12,96	6,48
32,4	59,4	1,8	2,4	3,24	5,76	4,32
32,6	54,6	2,0	−2,4	4,00	5,76	−4,80
32,8	61,6	2,2	4,6	4,84	21,16	10,12
28,4	60,6	−2,2	3,6	4,84	12,96	−7,92
28,6	61,7	−2,0	4,7	4,00	22,09	−9,40
28,5	61,9	−2,1	4,9	4,41	24,01	10,29
32,7	61,5	2,1	4,5	4,41	20,25	9,45
27,7	61,7	−2,9	4,7	8,41	22,09	−13,63
33,5	62,0	2,9	5,0	8,41	25,00	14,50
30,6	62,2	0	5,2	0,00	27,04	0,00

La un prag de semnificaţie $\alpha = 0{,}05$, *să se verifice ipoteza conform căreia scheletele ar aparţine aceleiaşi specii ca şi animalele din specia* S.

Soluţie. Avem

$$\bar{x} = 30{,}6 \qquad \bar{y} = 57$$

$$\sum_1^{18} (x_i - \bar{x})^2 = 51{,}34; \qquad \sum_1^{18} (y_i - \bar{y})^2 = 360{,}72,$$

$$\sum_1^{18} (x_i - \bar{x})(y_i - \bar{y}) = 61{,}19 - 42{,}24 = 18{,}95,$$

de unde

$$W = \frac{1}{n} \sum_1^n (x_i - \bar{x})^2 = 2{,}85; \qquad Z = \frac{1}{n} \sum_1^n (y_i - \bar{y})^2 = 20{,}04,$$

$$R = \frac{1}{n} \sum_1^n (x_i - \bar{x})(y_i - \bar{y}) = 1{,}052,$$

$$r = \frac{R}{\sqrt{WZ}} = \frac{1{,}052}{\sqrt{57{,}114}} = 0{,}139, \quad a = \frac{R}{W} = 0{,}369,$$

şi

$$t = \sqrt{\frac{W(n-2)}{Z(1-r^2)}}\,(a - A) = \sqrt{\frac{2{,}85 \cdot 16}{20{,}04(1 - 0{,}19321)}}\,(0{,}369 - 0{,}60) = -0{,}3927.$$

Din tabele găsim $t_{0,05;\,16} = 2{,}120$.

Deoarece $|t| < t_{0,05;\,16}$ acceptăm ipoteza conform căreia oasele au aparţinut la animalele din specia S.

V.38. *În tabelul* V.23

Tabelul V.23

i	Îngrăşăminte	x_{ij}				$\bar{x}_i$
1	Fără	67	67	55	42	57,75
2	$K_2O + N$	98	96	91	66	87,75
3	$K_2O + P_2O_5$	60	69	50	35	53,50
4	$N + P_2O_5$	79	64	81	70	73,50
5	$K_2O + P_2O_5 + N$	90	70	79	88	81,75

sînt date 20 *observaţii independente despre recolta produsă de o varietate anumită cultivată într-o oală de pămînt. Observaţiile sînt împărţite în* 5 *grupe a cîte* 4 *observaţii fiecare, în funcţie de îngrăşămîntul folosit, la o grupă nefolosindu-se nici un îngrăşămînt.*

Presupunem că variabilele x_{ij} *au repartiţii* $N(\mu_i, \sigma)$.

Să se verifice ipoteza $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$, *conform căreia nici un fel de îngrăşămînt nu influenţează recolta medie.*

Soluţie. Avem $n = 20$, $r = 5$, $n_i = 4$ $(i = 1, 2, 3, 4, 5, 6)$. Dacă ipoteza este adevărată, abaterile observate de la valorile x_i, vor fi date numai pentru fluctuaţii variabile. Avem

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} x_{ij} = \frac{1}{n}\sum_{i=1}^{r} n_i\bar{x}_i = 70{,}85,$$

$$s_1^2 = \frac{1}{r-1}\sum_{i=1}^{r} n_i(\bar{x}_i - \bar{x})^2 = \sum_{i=1}^{5}(\bar{x}_i - x)^2 = 884{,}075,$$

$$s_2^2 = \frac{1}{n-r}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 = \frac{1}{15}\sum_{i=1}^{5}\sum_{j=1}^{4}(x_{ij} - \bar{x}_i)^2 = 144{,}150,$$

de unde

$$F = \frac{s_1^2}{s_2^2} = \frac{884{,}075}{144{,}150} = 6{,}13.$$

Deoarece $s_1^2[s_2^2]$ este repartizată ca $4\ \chi^2(4)\ [15\chi^2(15)]$, $\frac{s_1^2}{s_2^2}$ este repartzată $F(4,15)$. Din tabele găsim $F_{0,01}\ (4,15) = 5,04$. Deoarece

$$F = 6,13 > 5,04 = F_{0,01}(4,15)$$

respingem ipoteza H_0.

După respingerea ipotezei conform căreia nici unul dintre îngrășăminte nu afectează recolta medie, putem verifica ipoteza H_1: $\mu_2 = \mu_3 = \mu_4 = \mu_5$, adică ipoteza că aplicarea unui fel de îngrășămînt nu afectează recolta medie.

Vom folosi liniile 2, 3, 4, 5. Aici avem $n = 16$, $r = 4$ și

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{r} n_i\bar{x}_i = 76,125,$$

$$s_1^2 = \frac{1}{3}\sum_{i=2}^{5} 4(\bar{x}_i - \bar{x})^2 = 892,55,$$

$$s_2^2 = \frac{1}{12}\sum_{i=2}^{5}\sum_{j=1}^{4} (x_{ij} - \bar{x}_i) = 144,625.$$

Prin urmare

$$F = \frac{892,55}{141,625} = 6,30.$$

Din tabele găsim $F_{0.01}\ (3,12) = 5,95$. Deoarece

$$F = 6,30 > 5,95 = F_{0,01}\ (3,12),$$

respingem ipoteza H_1, la un prag de semnificație de 0,01.

Verificăm ipoteza H_2 conform căreia adăugarea de îngrășămînt P_2O_5 nu afectează recolta medie. Prin urmare trebuie să verificăm ipoteza $\mu_2 = \mu_5$.

În acest caz avem

$$\bar{x} = 84,75,$$

$$s_1^2 = 4 \cdot 3^2 + 4 \cdot 3^2 = 72,$$

$$s_2^2 = \frac{1}{6}\sum_{j=1}^{4} (x_{2j} - \bar{x}_2)^2 + \frac{1}{6}\sum_{j=1}^{4} (x_{5j} - \bar{x}_5)^2 = 151,58.$$

Cum $s_1^2 < s_2^2$, vom schimba rolurile lui s_1^2 și s_2^2. Obținem

$$F = \frac{151,58}{72} = 2,10.$$

Din tabele găsim $F_{0,01}(6,1) > 2,10$. Deci ipoteza H_2 nu poate fi respinsă.

Ipoteza de mai sus poate fi tratată ca ipoteza: că două populaţii care au aceeaşi abatere medie pătratică σ necunoscută, din care au fost extrase două selecţii independente, să aibe aceeaşi valoare medie; prin urmare repartiţiile lor sînt identice. O astfel de ipoteză poate fi verificată şi cu testul t.

V.39. *Se măsoară în grame, greutatea a* 100 *bucăţi de fire de lînă a* 95 m *fiecare. Aceste bucăţi de fire sînt împărţite în* 4 *grupe a* 25 *fiecare şi fiecare grupă este luată dintr-un ghem diferit.*

Tabelul V.24

Bucata	Ghemul				Greutatea	Greutatea medie
	1	2	3	4		
1	7,50	7,23	7,50	7,53	29,76	7,44
2	7,52	7,81	7,77	8,05	31,15	7,79
3	7,70	7,94	7,83	8,16	31,63	7,91
4	7,93	7,94	7,96	7,76	31,59	7,90
5	7,78	7,89	8,02	7,85	31,54	7,89
6	7,73	8,23	7,99	8,14	32,09	8,02
7	8,07	8,27	8,25	8,26	32,85	8,21
8	8,01	8,54	8,24	8,54	33,33	8,33
9	8,22	8,24	8,37	8,10	32,93	8,23
10	8,24	8,35	8,43	8,15	33,17	8,29
11	8,17	8,29	8,46	8,38	33,30	8,33
12	8,09	8,54	8,33	8,47	33,43	8,36
13	8,11	8,45	8,27	8,38	33,21	8,30
14	7,96	8,43	8,24	8,60	33,23	8,31
15	8,09	8,47	8,12	8,45	33,13	8,28
16	8,04	8,33	8,14	8,43	32,94	8,24
17	7,78	8,47	8,19	8,57	33,01	8,25
18	8,11	8,63	8,36	8,38	33,48	8,37
19	8,17	8,31	8,31	8,16	32,95	8,24
20	8,12	8,31	8,47	8,41	33,31	8,33
21	8,13	8,10	8,19	8,27	32,69	8,17
22	8,01	8,01	8,37	7,96	32,35	8,09
23	8,17	7,92	8,27	8,08	32,44	8,11
24	8,05	8,27	8,07	8,16	32,55	8,14
25	7,91	7,92	8,28	8,52	32,63	8,16
Greutatea	199,61	204,89	204,43	205,76	816,69	
Greutatea medie	7,98	8,20	8,18	8,23		8,15

Presupunînd că fiecare din aceste 100 observații este valoarea observată a unei variabile cu repartiția $N(\mu_{k,l}, \sigma)$, să se verifice ipoteza conform căreia toate mediile $\mu_{k,l}$ sînt identice.

Soluție. Pentru a verifica această ipoteză aplicăm analiză dispersională. Rezultatele acestei analize sînt sintetizate în tabelul V.25

Tabelul V.25

	Suma pătratelor	Nr. gr. de libertate	Raportul
Între segmente	4,637814	24	$s_1^2 = 0{,}1932$
Între gheme	0,916707	3	$s_2^2 = 0{,}3056$
Reziduală	1,670418	72	$s_3^2 = 0{,}0232$
Total	7,224939	99	$s^2 = 0{,}0730$

Am obținut acest tabel folosindu-ne de formulele:

— suma pătratelor diferențelor dintre segmente

$$(r-1)s_1^2 = v\sum_{k=1}^{r}(\bar{x}_{k.} - \bar{x})^2;$$

— suma pătratelor diferențelor dintre gheme

$$(v-1)\, s_2^2 = r\sum_{l=1}^{v}(\bar{x}_{.l} - \bar{x})^2;$$

— suma pătratelor reziduală

$$(r-1)\,(v-1)\, s_3^2 = \sum_{k=1}^{r}\sum_{l=1}^{v}(x_{kl} - \bar{x}_{k.} - \bar{x}_{.l} + \bar{x})^2;$$

— totalul sumei pătratelor

$$(rv-1)\, s^2 = \sum_{k=1}^{r}\sum_{l=1}^{v}(x_{kl} - \bar{x})^2.$$

În cazul de față $v = 25$, $r = 4$.

Ţinînd seama de aceste date, obţinem

$$z_1 = \frac{1}{2}\lg\frac{0{,}1932}{0{,}0232} = \frac{1}{2}\lg 8{,}33 = 1{,}0599,$$

$$z_2 = \frac{1}{2}\lg\frac{0{,}3056}{0{,}0232} = \frac{1}{2}\lg 1{,}31 = 1{,}2863.$$

Deoarece $s_1^2[s_3^2]$ este repartizat ca $24\chi^2(24)$ $[72\chi^2(72)]$ din tabele găsim că valoarea z_0 care satisface relaţia $P(z \geqslant z_0) = 0{,}01$ este cuprinsă între 0,3339 şi 0,3746. După cum vedem valoarea z_1 observată în exemplul nostru este mult mai mare ca z_0. Deoarece $s_2^2[s_3^2]$ este repartizată $3\ \chi^3(3)$ $[72\chi^2(72)]$, din tabele găsim că valoarea z_0 pentru care relaţia $P(z \geqslant z_0) = 0{,}01$ are loc, este cuprinsă între 0,6867 şi 0,7086. Prin urmare z_2 este mult mai mare ca această valoare.

Tragem concluzia că valorile observate z_1 şi z_2 nu pot fi explicate ca rezultat al fluctuaţiilor variabile — cu un prag de semnificaţie de 0,01 — dacă ipoteza H_0 este adevărată. Astfel respingem H_0.

V.40. *Fie* $(x_1, x_2, ..., x_n)$ *o selecţie de volum n dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X repartizată* $N(\mu, \sigma^2)$.

Să se găsească puterea testului raportului de verosimilitate pentru verificarea ipotezei H_0: $\sigma^2 \in \omega$, ω *submulţimea lui* Ω *pentru care* $\sigma^2 = \sigma_0^2$.

Soluţie. Pentru funcţia de verosimilitate

$$P(x, \mu, \sigma^2) = \frac{1}{(\sigma\sqrt{2\pi})^n}\exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_i - \mu)^2\right].$$

a) Parametrii fiind în Ω, găsim estimaţiile de maximă verosimilitate

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n}\Sigma(x - \bar{x})^2 = s^2,$$

astfel că

$$(P)_\Omega = (2\pi s^2)^{-n/2}\exp\left(-\frac{1}{2}n\right).$$

b) Parametrii fiind în ω, $\sigma^2 = \sigma_0^2$, iar estimaţia de maximă verosimilitate pentru μ este $\bar{x}$, astfel încît

$$(P)_\omega = (2\pi\sigma_0^2)^{-n/2}\exp\left(-\frac{\Sigma(x - \bar{x})}{2\sigma_0^2}\right].$$

Urmează că raportul de verosimilitate este

$$l = \left(\frac{s^2}{\sigma_0^2}\right)^{n/2} \exp\left[-\frac{1}{2}\, n\left(\frac{s^2}{\sigma_0^2} - 1\right)\right],$$

astfel că

$$z = e^{-1} l^{n/2} = \frac{t}{n} e^{-t/n} \tag{V.1}$$

unde $t = ns^2/\sigma_0^2$. z este o funcţie monolonă în l, însă nu este o funcţie monolonă în t/n. Avem

$$\frac{dz}{dt} = \frac{1}{n}\left(1 - \frac{t}{n}\right) e^{-t/n},$$

astfel că z creşte pentru $t < n$, devenind maxim pentru $t = n$ şi atunci descreşte. Punînd $l \leqslant C_\alpha$ este echivalent cu

$$t \leqslant a_\alpha, \quad t \geqslant b_\alpha,$$

unde a_α, b_α sînt determinate din (V.1), de

$$P(t \leqslant a_\alpha) + P(t \geqslant b_\beta) = \alpha,$$

$$a_\alpha\, e^{-a_\alpha/n} = b_\alpha\, e^{-b_\alpha/n}. \tag{V.2}$$

Deoarece atunci cînd este adevărată ipoteza H_0, statistica t este repartizată $\chi^2(n-1)$, putem folosi tabela repartiţiei, ce satisface (V.2).

Considerăm repartiţia aproximativă:

$$-2\lg l = (t - n) - n \lg\left(\frac{t}{n}\right)\cdot$$

Deoarece

$$M(t) = n - 1, \quad D^2(t) = 2(n-1),$$

putem scrie

$$-2\lg l = (t-n) - n\lg\left(1 - \frac{t-n}{n}\right) =$$

$$= (t-n) - n\sum_{r=1}^{\infty} (-1)^{r-1}\left(\frac{t-n}{n}\right)^r \Big/ r =$$

$$= (t-n) - n\left[\frac{t-n}{n} - \frac{(t-n)^2}{2n^2} + 0(n^{-2})\right] =$$

$$= \frac{(t-n)^2}{2n} + 0(n^{-1}).$$

Se ştie că pentru $n \to \infty$, repartiţia $\chi^2(n-1)$ tinde către repartiţia $N(n-1, 2(n-1))$; prin urmare $(t-n)/\sqrt{2n}$ tinde către o variabilă redusă, adică $(t-n)^2/2n$ este o variabilă $\chi^2(1)$. Cînd H_0 este adevărată, $-2 \lg l$ are această repartiţie. Acest rezultat se păstrează şi cînd H_0 este falsă, $-2\lg l$ avînd repartiţia χ^2 necentrată cu un grad de libertate şi parametru de necentricitate

$$\lambda = -M\left[\frac{\partial^2 \ln P}{\partial(\sigma^2)^2}\right](\sigma^2-\sigma_0^2)^2 = \frac{n}{2\sigma^4}(\sigma^2 - b_0^2)^2$$

$$= \frac{n}{2}\left(1-\frac{\sigma_0^2}{\sigma^2}\right)^2.$$

Puterea testului raportului de verosimilitate este:

$$P = \int_{\left(\frac{r+\lambda}{r+2\lambda}\right)\chi_\alpha^2(r)}^{\infty} d\chi^2\left(r+\frac{\lambda^2}{r+2\lambda}\right) = \int_{\left(\frac{1+\lambda}{1+2\lambda}\right)\chi_\alpha^2(1)} d\chi^2\left(1+\frac{\lambda^2}{1+2\lambda}\right), \quad \text{(V.3)}$$

unde $1 = \nu = r$ este numărul gradelor de libertate, iar $\chi_\alpha^2(r)$ cuantila de ordin $1-\alpha$.

Punînd $\lambda = 0$, în (V.3) se obţine volumul testului.

V.41. *Fie $(x_{ij})_{1\leqslant j\leqslant n}$ $(i = 1, 2)$ o selecţie dintr-o populaţie în care caracteristica sub cercetare este o variabilă aleatoare X_i repartizată*

$$N(\mu_i, \sigma^2)\ (i = 1, 2) \quad \text{şi} \quad \Delta = \frac{\mu_1-\mu_2}{\sigma}.$$

a) *Să se verifice ipoteza*

$$H_0\colon\ |\Delta| = K$$

faţă de alternativa

$$H_1\colon\ -K < \Delta < K$$

şi să se calculeze puterea testului $\pi(\Delta, n)$;

b) *Să se determine volumul de selecţie astfel încît testul să aibă o putere dată.*

S o l u ţ i e. a) Considerăm statistica

$$Y = \frac{\bar{x}_1 - \bar{x}_2}{\sigma}\sqrt{\frac{n}{2}},$$

unde $\bar{x}_j$, $j = 1, 2$ este media de selecţie a selecţiei ce provine din populaţia $N(\mu_j\,;\,\sigma^2)$. Respingem H_0 dacă şi numai dacă $-a < Y < a$ $(a > 0)$. Atunci, deoarece $Y - \Delta\sqrt{\frac{n}{2}}$ are o repartiţie normală $N(0,1)$, putem calcula probabilitatea de a respinge ipoteza H_0 condiţionată de faptul că Δ şi n au valori determinate.

$$P(\text{resping } H_0 \mid \Delta, n) = P(-a < Y < a \mid \Delta, n) =$$

$$= \Phi\left(a - \Delta\sqrt{\frac{n}{2}}\right) - \Phi\left(-a - \Delta\sqrt{\frac{n}{2}}\right).$$

Cum
$$\alpha = P(\text{Erori de genul I}),$$
$$\pi = 1 - P\ (\text{Erori de genul II}),$$

în continuare va fi suficient să considerăm numai valorile nenegative ale lui Δ, deoarece

$$P(-a < Y < a \mid \Delta, n) = P(-a < Y < a \mid -\Delta, n).$$

Dacă în plus $K' > K$, atunci

$$P(\text{resping } H_0 \mid K', n) < P(\text{resping } H_0 \mid K, n).$$

Rezultă atunci că dacă $\Delta = K$, probabilitatea unei erori de genul I este maximă pentru familia ipotezei nule. Punem acum (după cum rezultă de mai sus)

$$\alpha = \Phi\left(a - K\sqrt{\frac{n}{2}}\right) - \Phi\left(-a - K\sqrt{\frac{n}{2}}\right).$$

Dacă ipoteza H_1 este adevărată, adică în cazul nostru dacă $\Delta \in (-K, K)$, putem scrie

$$\pi(\Delta; n) = \Phi\left(a - \Delta\sqrt{\frac{n}{2}}\right) - \Phi\left(-a - \Delta\sqrt{\frac{n}{2}}\right)$$

şi în particular pentru $\Delta = 0$ obţinem

$$\pi(0; n) = \Phi(a) - \Phi(-a).$$

Observaţie. Putem schimba regula de decizie în funcţie de statistica pe care o adoptăm şi simplificările care dorim să le obţinem atunci cînd calculăm nivelul α sau puterea π a testului. Aşa de exemplu dacă considerăm statistica

$$X = \frac{\bar{x}_1 - \bar{x}_2}{\sigma}\sqrt{\frac{n}{2}} - K\sqrt{\frac{n}{2}}$$

şi regula de decizie: respingem H_0 dacă şi numai dacă $X < b$ cu $b > -K\sqrt{\frac{n}{2}}$, atunci observăm că deoarece $X = |Y| - K\sqrt{\frac{n}{2}}$ putem scrie

$$P(X < b) = P\left(-b - K\sqrt{\frac{n}{2}} < Y < b + K\sqrt{\frac{n}{2}}\right).$$

Pentru Δ şi n daţi obţinem

$$P(X < b \mid \Delta, n) = P(\text{respingem } H_0 \mid \Delta, n) =$$

$$= \Phi\left(\quad + [K - \Delta]\sqrt{\frac{n}{2}}\right) - \Phi\left(-b - [k + \Delta]\sqrt{\frac{n}{2}}\right).$$

Ca şi mai înainte

$$P(X < b \mid \Delta, n) = P(X < b \mid -\Delta, n),$$

iar $P(X < b \mid K, n)$ este maximă pentru familia din ipoteza nulă.

Dacă Δ face parte din familia H_1 atunci

$$P(X < b \mid \Delta, n) = \Phi\left(b + [K - \Delta]\sqrt{\frac{n}{2}}\right) - \Phi\left(-b - [K + \Delta]\sqrt{\frac{n}{2}}\right)$$

devine puterea $\pi(\Delta, n)$.

b) Vom presupune că $[-\Delta^*, \Delta^*] \subset (-K, K)$ şi că dorim ca puterea testului să nu fie mai mică decît P^* dacă $\Delta \in [-\Delta^*, \Delta^*]$.

Deoarece $\pi(\Delta, n) \geqslant \pi(\Delta^*, n)$ dacă $|\Delta| \leqslant \Delta^*$ este suficient să găsim o valoare n astfel încît

$$\pi(\Delta^*, n) \geqslant P^*.$$

Punînd

$$z = b + [K - \Delta^*]\sqrt{\frac{n}{2}}$$

avem

$$\pi(\Delta^*, n) = \Phi(z) - \Phi\left(-z - 2\Delta^*\sqrt{\frac{n}{2}}\right) =$$

$$= \Phi(z) - \Phi\left(-b - \frac{K + \Delta^*}{K - \Delta^*}\left[z - b\right]\right).$$

Definim acum z' prin $\Phi(z') = P^*$ şi notînd n' valoarea lui n care satisface relaţia

$$z' = b + [K - \Delta^*]\sqrt{\frac{n}{2}},$$

atunci rezultă că

$$P^* - \pi(\Delta^*, n') = \Phi\left(-b - \frac{K+\Delta^*}{K-\Delta^*}(z'-b)\right) > 0$$

şi deci $\pi(\Delta^*, n') < P^*$.

Fie z^* definit de

$$\Phi(z^*) = P^* + \Phi\left(-b - \frac{K+\Delta^*}{K-\Delta^*}(z'-b)\right)$$

şi n^* soluţia ecuaţiei

$$z^* = b + [K - \Delta^*]\sqrt{\frac{n}{2}}.$$

Atunci

$$\pi(\Delta^*, n^*) = \Phi(z^*) - \Phi\left(-b - \frac{K+\Delta^*}{K-\Delta^*}(z^*-b)\right) =$$

$$= P^* + \Phi\left(-b - \frac{K+\Delta^*}{K-\Delta^*}(z'-b)\right) -$$

$$- \Phi\left(-b - \frac{K+\Delta^*}{K-\Delta^*}(z^*-b)\right) > P^*$$

cu $n^* = 2\left(\frac{z^* - b}{K - \Delta^*}\right)^2$ volumul de selecţie pentru care puterea testului este mai mare decît P^* dacă $\Delta \in [-\Delta^*, \Delta^*]$.

V.42. *Pe baza unei singure observaţii asupra variabilei X avînd densitatea de repartiţie*

$$f(x) = \frac{1}{\pi[1 + (x-\theta)^2]}, \quad -\infty < x < \infty,$$

să se verifice ipoteza $H_0: \theta = 0$, faţă de ipoteza alternativă $H_1: \theta = 1$.

Soluţie. Cea mai bună regiune critică pentru verificarea ipotezei H_0 faţă de H_1 este dată de acea munţime W pentru care

$$\frac{P(x \mid H_0)}{P(x \mid H_1)} \leqslant c_\alpha, \tag{V.4}$$

c_α fiind ales astfel ca $\int_W P(x \mid H_0)\, dx = \alpha$.

În cazul unei singure observaţii x, (V.4) se scrie

$$\frac{P(x \mid H_0)}{P(x \mid H_1)} = \frac{1 + (x - 1)^2}{1 + x^2} \leqslant c_\alpha,$$

de unde

$$x^2(c_\alpha - 1) + 2x + (c_\alpha - 2) \geqslant 0. \tag{V.5}$$

Forma C.M.B.R.C. astfel definită depinde de valoarea α aleasă. Dacă luăm $c_\alpha = 1$, (V.5) se reduce la $x \geqslant 1/2$, astfel că vom respinge ipoteza H_0 în favoarea ipotezei H_1, ori de cîte ori valoarea observată x este mai apropiată de 1 decît de 0.

Dacă luăm $c_\alpha = 0,5$, (V.5) devine

$$x^2 - 4x + 3 \geqslant 0 \quad \text{sau} \quad (x - 2)^2 \leqslant 1,$$

care are loc cînd $1 \leqslant x \leqslant 3$. Aceasta este regiunea critică.

Deoarece repartiţia Cauchy este o repartiţie $t(1)$ şi ţinînd seama că

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x,$$

putem calcula volumul testului în fiecare din cele două cazuri considerate.

Pentru $c_\alpha = 1$, volumul este

$$P\left(t \geqslant \frac{1}{2}\right) = 0,352,$$

în timp ce pentru $c_\alpha = 0,5$, volumul este

$$P(1 \leqslant t \leqslant 3) = 0,148.$$

V.43. *Fie* $(X_1, X_2, \ldots, X_n)$ *un vector aleator repartizat normal cu*

$$M(X_1) = n\theta, \quad \theta > 0,$$

$$M(X_r) = 0, \quad r > 1$$

şi matricea de covarianţă:

$$\Sigma = \begin{pmatrix} n - 1 + \theta^2 \ldots - 1 \ldots -1 \\ -1 \quad\quad 1 \ldots \;\; 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ -1 \quad \ldots \quad 0 \ldots \;\; 1 \end{pmatrix}.$$

Pe baza unei singure observații $x = (x_1, x_2, \dots, x_n)$ să se verifice ipoteza $H_0: \theta = \theta_0 = 0$ față de ipoteza alternativă $H_1: \theta = \theta_1 > 0$.

Soluție. Cum $|\Sigma| = \theta^2$, obținem

$$\Sigma^{-1} = \frac{1}{\theta^2}\begin{pmatrix} 1 & 1 & \dots 1 \\ 1 & 1+\theta^2 & 1 \\ \vdots & 1 & \\ 1 & & 1+\theta^2 \end{pmatrix}.$$

Densitatea de repartiție a vectorului x este

$$f(x) = (2\pi)^{-n/2}|\Sigma^{-1}|^{1/2} \exp\left[-\frac{1}{2}(x-\mu)'\,\Sigma^{-1}(x-\mu)\right] =$$

$$= \frac{1}{\theta(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\left[\frac{n^2}{\theta^2}(\bar{x}-\theta)^2 + \sum_{2}^{n} x_i^2\right]\right\}.$$

C.M.B.R.C. pentru verificarea ipotezei $H_0: \theta = \theta_0 > 0$ față de $H_1: \theta = \theta_1 > 0$ pe baza unei singure observații este dată de

$$\frac{P(x \mid H_0)}{P(x \mid H_1)} = \frac{\theta_1}{\theta_0} \exp\left\{-\frac{n^2}{2}\left[\frac{(\bar{x}-\theta_0)^2}{\theta_0^2} - \frac{(\bar{x}-\theta_1)^2}{\theta_1^2}\right]\right\} \leqslant c_\alpha,$$

care se reduce la

$$\frac{(\bar{x}-\theta_1)^2}{\theta_1^2} - \frac{(\bar{x}-\theta_0)^2}{\theta_0^2} \leqslant \frac{2}{n^2}\ln(c_\alpha\theta_0/\theta_1)$$

sau

$$\bar{x}^2(\theta_0^2 - \theta_1^2) - 2\bar{x}(\theta_0 - \theta_1) \leqslant \frac{2\theta_0^2\theta_1^2}{n^2}\ln(c_\alpha\theta_0/\theta_1).$$

Dacă $\theta_0 > \theta_1$ aceasta este de forma

$$\bar{x}^2(\theta_0 + \theta_1) - 2\bar{x} \leqslant a_\alpha, \tag{V.6}$$

care implică

$$b_\alpha \leqslant \bar{x} \leqslant c_\alpha. \tag{V.7}$$

Dacă $\theta_0 < \theta_1$, C.M.B.R.C. este de forma

$$\bar{x}^2(\theta_0 + \theta_1) - 2\bar{x} \geqslant d_\alpha \tag{V.8}$$

implicînd

$$\bar{x} \leqslant e_\alpha \quad \text{sau} \quad \bar{x} \geqslant f_\alpha. \tag{V.9}$$

Observații. Atît în (V.7) cît și în (V.9) limitele între care (în afara cărora) variază $\bar{x}$ sînt funcții de valoare exactă θ_1. Această dificultate, care apare din faptul că θ_1 apare în coeficientul lui $\bar{x}^2$ în formele pătratice (V.6) și (V.8), arată că nu există C.M.B.R.C. chiar și pentru o mulțime de ipoteze alternative unilaterale; prin urmare nu există un test uniform cel mai puternic.

$\bar{x}$ este singura statistică suficientă pentru θ; deci, suficiența nu implică existența testului uniform cel mai puternic.

V.44. *Fie* $(x_1, ..., x_n)$ $[y_1, ..., y_m]$ *o selecție dintr-o populație în care caracteristica sub cercetare este o variabilă* $N(0, \sigma_1^2)$, $[N(0, \sigma_1^2 + \sigma_2^2)]$. *Să se construiască un interval de încredere pentru* σ_2^2, *avînd coeficientul de încredere* $1 - \alpha$.

Soluție. Problema pusă este echivalentă cu a construi un test de volum α pentru verificarea ipotezei

$$H_0\colon c_2^2 = \delta^2$$

față de

$$H_1\colon \sigma_2^2 \neq \delta^2,$$

adică, testul pentru care $P(\text{resping } H_0 \mid H_0) = \alpha$, ori care ar fi σ_1^2. Un astfel de test va conduce la un interval de încredere pentru σ_2^2.

Este comod să se construiască un astfel de test prin adăugarea unei selecții $(z_1, z_2, ..., z_n)$ din $N(0,1)$; statistica

$$W = \frac{\sum_1^m y^2/m}{\sum_1^n (x + \delta z)^{2/n}},$$

atunci cînd ipoteza H este adevărată, are repartiție $F(m, n)$. De exemplu, regiunea de acceptare pentru H este definită de

$$F_1 = F_{\frac{\alpha}{2}}(m, n) \leqslant W \leqslant F_{1-\frac{\alpha}{2}}(m, n) = F_2. \tag{V.10}$$

Vom defini W numai pentru $\delta \geqslant 0$, corespunzătoare la rădăcina pozitivă a lui σ_2^2; semnul lui δ nu trebuie să influiențeze valabilitatea semnificației testului, dar un semn potrivit pentru δ este necesar în obținerea limitelor de încredere.

La obţinerea limitelor de încredere pentru σ_2^2 plecînd cu regiunea de acceptare bazată pe W, avem

$$\{F_1 \leqslant W \leqslant F_2\} = \{F_1^{-1} \leqslant W \leqslant F_1^{-1}\} = \tag{V.11}$$

$$= \left\{\frac{n\Sigma y^2}{mF_2} \leqslant \Sigma(x + \delta z)^2 \leqslant \frac{n\Sigma y^2}{mF_1}\right\} =$$

$$= \left\{\frac{n\Sigma y^2}{mF_2} - \Sigma x^2 \leqslant 2\delta \sum xz + \delta^2\Sigma z^2 \leqslant \frac{n\Sigma y^2}{mF_1} - \Sigma x^2\right\}. \tag{V.12}$$

Se observă că în (V.12) s-a strecurat o particularitate nedorită; dacă continuăm ca în (V.12) limitele rezultate vor implica termeni în Σxz. Aceşti termeni dau inconveniente în aplicaţii la analiza componenţei dispersiei, deoarece variabilele independente $x_1, x_2, ..., x_n$ avînd repartiţia $N(0, \sigma_1^2)$ nu sînt observate aici, şi vor deveni accesibile numai după o transformare convenabilă a datelor iniţiale. E mai bine ca limitele de încredere să depindă numai de datele iniţiale.

Pentru a obţine limite mai convenabile procedăm în felul următor: Împărţind inegalităţile (V.12) prin $2\delta(\Sigma x^2)^{1/2}$ obţinem

$$\left\{\frac{1}{2\delta(\Sigma x^2)^{1/2}}\left[\frac{n\Sigma y^2}{mF_2} - \Sigma x^2\right] \leqslant \frac{\Sigma xz}{(\Sigma x^2)^{1/2}} + \frac{\delta\Sigma z^2}{2(\Sigma x^2)^{1/2}} \leqslant \right.$$

$$\left. \leqslant \frac{1}{2\delta(\Sigma x^2)^{1/2}}\left[\frac{n\Sigma y^2}{mF_1} - \Sigma x^2\right]\right\}, \tag{V.13}$$

şi fie

$$t = \frac{\Sigma xz}{(\Sigma x^2)^{1/2}} + \frac{\delta\Sigma z^2}{2(\Sigma x^2)^{1/2}}.$$

Acum pentru $(x_1, x_2, ..., x_n)$ fixat, o transformare ortogonală de la $(z_1, z_2, ..., z_n)$ la $(z_1', z_2', ..., z_n')$ cu $z_1' = (\Sigma xz)/(\Sigma x^2)^{1/2}$ dă

$$t' = z_1' + \frac{\delta}{2(\Sigma x^2)^{1/2}} \sum_1^n z'^2, \tag{V.14}$$

unde $z_1', z_2', ..., z_n'$ sînt fiecare repartizate $N(0, 1)$ mutual independente şi independente de $(x_1, x_2, ..., x_n)$. Statisticile t şi t' au aceeaşi repartiţie şi deoarece t' este calculabilă direct din Σx^2 şi $(z_1', z_2', ..., z_n')$, va putea înlocui pe t. O a treia substituţie posibilă pentru t este

$$t'' = z + \frac{\delta}{2(\Sigma x^2)^{1/2}} \cdot z^2 + \frac{\delta}{2(\Sigma x^2)^{1/2}} \chi^2(n - 1),$$

unde z este repartizată $N(0,1)$; z şi χ_{n-1}^2 sînt independente.

Înlocuind t prin t' în (V.13), avem

$$\left\{\frac{1}{2\delta(\Sigma x^2)^{1/2}}\left[\frac{n\Sigma y^2}{mF_2}-\Sigma x^2\right]\leqslant z_1+\frac{\delta}{2(\Sigma x^2)^{1/2}}\sum_1^n z^2\leqslant\right.$$

$$\left.\leqslant\frac{1}{2\delta(\Sigma x^2)^{1/2}}\left[\frac{n\Sigma y^2}{mF_1}-\Sigma x^2\right]\right\},$$

sau

$$\left\{\frac{1}{\Sigma z^2}\left[\frac{n\Sigma y^2}{mF_2}-\Sigma x^2\right]+\left[\frac{z_1(\Sigma x^2)^{1/2}}{\Sigma z^2}\right]^2\leqslant\left[\delta+\frac{z_1(\Sigma x^2)^{1/2}}{\Sigma z^2}\right]^2\leqslant\right.$$

$$\left.\leqslant\frac{1}{\Sigma z^2}\left[\frac{n\Sigma y^2}{mF_1}-\Sigma x^2\right]+\left[\frac{z_1(\Sigma x^2)^{1/2}}{\Sigma z^2}\right]^2\right\}. \tag{V.15}$$

Un interval de încredere pentru σ_2 este, atunci, definit ca mulţimea valorilor lui δ, satisfăcînd (V.15) şi $\delta \geqslant 0$. Acestea sînt valorile lui σ_2 pentru care ipoteza H_0 va fi acceptată şi deoarece atunci cînd este adevărată ipoteza H_0 probabilitatea de cădere în regiunea de acceptare (V.11) este $1-\alpha$, regiunea de încredere definită de (V.15) şi $\delta \geqslant 0$ are coeficientul de încredere $1-\alpha$.

Rămîne să determinăm limitele definind această regiune. Punînd

$$k=\frac{1}{\Sigma z^2}\left[\frac{n\Sigma y^2}{mF_2}-\Sigma x^2\right],$$

$$l=\frac{1}{\Sigma z^2}\left[\frac{n\Sigma y^2}{mF_1}-\Sigma x^2\right]$$

şi

$$b=-\frac{z_1(\Sigma x^2)^{1/2}}{\Sigma z^2},$$

(V.15) se scrie

$$k+b^2\leqslant(\delta-b)^2\leqslant l+b^2.$$

De aici

$$b-(l+b^2)^{1/2}\leqslant\delta\leqslant b-[\max(0, k+b^2)]^{1/2} \tag{V.16}$$

sau

$$b+[\max(0, k+b^2)]^{1/2}\leqslant\delta\leqslant b+[l+b^2]^{1/2}. \tag{V.17}$$

(Am presupus $l+b^2\geqslant 0$; astfel regiunea de încredere este vidă).

Considerarea lui (V.16) şi (V.17), împreună cu $\delta\geqslant 0$, arată că intervalul de încredere va fi definit de (V.17) singur, afară de cazul

$-b^2 \leqslant k < 0$ şi $b > 0$. În ultimul caz, regiunea de încredere constă din (V.17) şi intervalul

$$\max\left[b - (l + b^2)^{1/2}\right] \leqslant \delta \leqslant b - [\max(0, k + b^2)^2]^{1/2}.$$

Astfel există posibilitatea ca limitele de încredere să poată defini o regiune constînd în două intervale separate. Este interesant să examinăm şansa obţinerii a două intervale, şi notăm în primul rînd că b^2 converge către zero în probabilitate, astfel că asimptotic şansa a două intervale este neglijabilă. De asemenea $P(b > 0) = \frac{1}{2}$. În plus $k < 0$ este condiţia ca raportul $(\Sigma y^2/m)$ $(\Sigma x^2/n)$ să nu fie semnificativ la pragul $\frac{1}{2}$ α pentru a verifica ipoteza $\sigma_2 = 0$. Calculele pot da o margine grosolană a probabilităţilor ca să rezulte două intervale. Folosind definiţia lui k, condiţia $-b^2 \leqslant k < 0$, este echivalentă cu condiţia

$$\left(F_2 - \frac{z_1^2}{\Sigma z^2}\right)\left(\frac{\sigma_1^2 + \delta^2}{\sigma_1^2 + \sigma_2^2}\right) \leqslant F \leqslant F_2\left(\frac{\sigma_1^2 + \delta^2}{\sigma_1^2 + \sigma_2^2}\right).$$

Probabilitatea acestui eveniment va fi maximă cînd F_2 $[(\sigma_1^2 + \delta^2)/(\sigma_1^2 + \sigma_2^2)$ este apropiată de modulul $(mn - 2n)/(mn + 2m)$ corespunzătoare repartiţiei F, sau cînd $F_2 \sim (\sigma_1^2 + \sigma_2^2)/(\sigma_1^2 + \sigma^2)$.

V.45. *În aceleaşi condiţii tehnologice ca şi în exemplul* V.14 *se determină rezistenţa ohmică a 18 tronsoane de ceramică acoperite cu carbon, obţinîndu-se următoarele valori*

1,65	1,72	1,70	1,84	1,96	1,58	1,85	1,78	1,68
1,71	1,92	1,60	1,75	2,05	1,90	1,76	1,82	1,88

Folosindu-ne de testul Shapiro-Wilk şi de un prag de semnificaţie $\alpha = 0{,}05$, *să se verifice ipoteza conform căreia rezistenţa ohmică este o variabilă repartizată normal.*

Soluţie. Găsim

$$\bar{x} = \frac{1}{18}\, 32{,}18 = 1{,}786, \quad \Sigma(x_i - \bar{x})^2 = 0{,}2743.$$

Modul de lucru cu testul Shapiro-Wilk este sintetizat în tabelul V.26.

Tabelul V.26

i	$x_{(n-i+1)}$	$x_{(i)}$	$W_{(i)}$	a_{n-i+1}	$a_{n-i+1} \cdot W_{(i)}$
1	2,05	1,58	0,47	0,4886	0,229642
2	1,96	1,60	0,36	0,3253	0,117108
3	1,92	1,65	0,27	0,2553	0,068931
4	1,90	1,68	0,22	0,2027	0,044594
5	1,88	1,70	0,18	0,1587	0,028536
6	1,85	1,71	0,14	0,1197	0,016758
7	1,84	1,72	0,12	0,0837	0,010044
8	1,82	1,75	0,07	0,0496	0,003472
9	1,78	1,76	0,02	0,0163	0,000326

Primele 9 observații în ordinea mărimii au fost trecute în coloana a treia a tabelului V.26, celelalte 9 observații au fost trecute în coloana a doua a aceluiași tabel începînd de jos în sus. În coloana a patra s-au trecut i — amplitudinile

$$W_{(i)} = x_{(n-i+1)} - x_{(i)}.$$

Elementele coloanei a cincea sînt extrase din tabele.

Din ultima coloană obținem

$$b = \sum_{i=1}^{9} a_{n-i+1}[x_{(n-i+1)} - x_{(i)}] = \sum_{i=1}^{9} a_{n-i+1} W_{(i)} = 0,51941 \simeq 0,52$$

Deci statistica

$$W = \frac{b^2}{\Sigma(x_i - \bar{x})^2} \text{ are valoarea } 0,985.$$

Din tabele găsim $W_{18;\,0,95} = 0,982.$

Deoarece $W > W_{18;\,0,95}$, acceptăm ipoteza de normalitate.

V.46. *În aceleași condiții tehnologice ca și în exemplul* V.14 *să determinăm rezistența ohmică a 15 tronsoane de ceramică acoperite cu carbon, obținîndu-se următoarele valori:*

1,68	1,95	1,83	1,56	1,94	1,84	1,78	
1,92	1,86	1,73	1,70	1,76	1,95	1,90	1,85

Folosindu-ne de testul Shapiro-Wilk şi de un prag de semnificaţie $\alpha = 0{,}05$, să se verifice ipoteza conform căreia rezistenţa ohmică este o variabilă repartizată normal.

S o l u ţ i e. Găsim

$$\bar{x} = 1{,}816; \quad \sum_{1}^{15} (x_i - \bar{x})^2 = 0{,}18340.$$

Deoarece în acest caz n este impar, mediana 1,84 se ia zero. Avem

Tabelul V. 27

i	$x_{(n-i+1)}$	$x_{(i)}$	$W_{(i)}$	a_{n-i+1}	$a_{n-i+1} W_{(i)}$
1	1,95	1,56	0,39	0,5150	0,200850
2	1,95	1,68	0,27	0,3306	0,089262
3	1,94	1,70	0,24	0,2495	0,059880
4	1,92	1,73	0,19	0,1878	0,035682
5	1,90	1,76	0,14	0,1353	0,018942
6	1,86	1,78	0,08	0,0880	0,007004
7	1,85	1,83	0,02	0,0433	0,000866

Din ultima coloană, obţinem

$$b = a_n[x_{(n)} - x_{(1)}] + \dots + a_{l+2}[x_{(l+2)} - x_{(l)}] = 0{,}41522.$$

Deci statistica

$$W = \frac{b^2}{\Sigma(x_i - \bar{x})^2} \text{ are valoarea } 0{,}94.$$

Din tabele găsim $W_{15;\,0,95} = 0{,}980$.

Deoarece $W < W_{15;\,0,95}$ respingem ipoteza de normalitate.

V.47. *Doi chimişti, A şi B fac $m = 9$ şi respectiv $n = 12$ determinări ale greutăţii unui anumit compus. Rezultatele acestor determinări sînt trecute în tabelul V.28. (Numerele din paranteze sînt rangurile determinărilor în selecţia combinată).*

Tabelul V.28

Chimistul A x	Chimistul$B$ x'
174,18(5)	174,19(6)
174,29(13)	174,40(20)
174,23(9)	174,21(8)
174,30(14)	174,35(18)
174,36(19)	174,32(16)
174,16(3)	174,14(1)
174,41(21)	174,27(11)
174,20(7)	174,31(15)
174,25(10)	174,17(4)
	174,34(17)
	174,28(12)
	174,15(2)

Folosind testul Mann-Whitney (Wilcoxon) să se verifice ipoteza conform căreia rezultatele celor doi chimişti nu diferă semnificativ.

Soluţie. Combinînd cele două grupe de două măsurători într-o singură grupă şi ordonînd observaţiile găsim că valoarea statisticii T a lui Wilcoxon, suma rangurilor determinărilor x, este 101.

Statistica lui Mann-Whitney

$$W = mn + \frac{m(m+1)}{2} - T$$

ce reprezintă numărul celor mn perechi posibile (x_i, x'_1) cu $x_i < x'_1$ are valoarea 52.

Dacă este adevărată ipoteza emisă $(H_0: F_1(x) \equiv F_2(x))$, adică cele două selecţii provin din populaţii cu funcţii de repartiţie identice, atunci pentru valori ale lui m şi n mai mari decît 8, W este aproximativ repartizată normal de medie $\frac{mn}{2}$ şi dispersie $\frac{mn(m+n+1)}{12}$.

Deci statistica

$$\frac{W - \frac{mn}{2}}{\sqrt{mn(m+n+1)/12}} \tag{V.18}$$

este repartizată $N(0,1)$.

În acest caz $M(W) = 54$, $D^2(W) = 198$; deci statistica (V.38) are valoarea $-0{,}142$ şi astfel ipoteza conform căreia rezultatele celor doi chimişti nu diferă semnificativ.

V.48. *Fie populaţiile normale p-dimensionale* $N(\mu_i, \Sigma)$, $1 \leqslant i \leqslant k$, *unde* Σ *este o matrice cunoscută şi nesingulară.*

Se consideră o selecţie de volum N *din populaţiile* $N(\mu_i, \Sigma)$, $1 \leqslant i \leqslant k$ *şi pe baza acestei selecţii se cere să se verifice ipoteza*

$$H_0 \colon H\mu_i = \zeta \leqslant i \leqslant k,$$

unde H *este o matrice* (s, p) *de rang* $s < p$ *iar* $\zeta' = (\zeta_1, \zeta_2, \ldots, \zeta_s)$ *şi unde componentele matricei* H *şi ale vectorului* ζ *sînt constante cunoscute date prin ipoteza de verificat.*

S o l u ţ i e. Cum Σ este cunoscută şi nesingulară rezultă că densitatea de repartiţie a valorilor medii observate este

$$P(\mu_1, \ldots, \mu_k) = K \, e^{\frac{1}{2}\sum\limits_{i=1}^{k} N(\overline{U}-\mu_i)' \Sigma^{-1}(\overline{U}_i-\mu_i)}.$$

Raportul de verosimilităţi pentru verificarea ipotezei H_0 este

$$\chi^2 = -2 \lg \frac{\mathrm{Max}_{H_0} P(\mu_1, \ldots, \mu_k)}{\mathrm{Max}\; P(\mu_1, \ldots, \mu_k)}.$$

Dar $\mathrm{Max}_{H_0} P(\mu_1, \ldots, \mu_k)$ corespunde în cazul nostru la

$$\mathrm{Min}_{H_0} \sum_{i=1}^{k} N(\overline{U}_i - \mu_i)' \Sigma^{-1}(\overline{U}_i - \mu_i) = \lambda_{r+1} + \ldots + \lambda_p,$$

unde $\lambda_1 > \lambda_2 > \ldots > \lambda_p$ sînt rădăcinile ecuaţiei

$$\left| \sum_1^k \overline{U}_i \overline{U}_1' - k\overline{U}\,\overline{U}' - \lambda N^{-1}\Sigma \right| = 0 \quad \text{cu} \quad \overline{U} = \frac{1}{k} \sum_1^k \overline{U}_i.$$

Pentru a arăta acest lucru să considerăm vectorul coloană x cu N componente, o matrice B de ordinul $q \times N$ şi de rang $q < N$, o matrice A pozitiv definită, simetrică de ordinul $N \times N$ şi μ un vector coloană cu N componente.

Atunci

$$\mathrm{Min}_x (x - \mu)' A(x - \mu)$$

supus la restricţia

$$Bx = U$$

este dat de

$$(U - B\mu)' (BA^{-1}B')^{-1} (U - B\mu)$$

şi este atins pentru o anumită valoare a vectorului x.

Se ştie că

$$\text{Min}_{y_1, \dots, y_k} \sum_{1}^{k} \frac{y_i' A y_i}{y_i' y} = \text{Min}_Y \text{Ur}\, YAY' = \lambda_{p-k+1} + \dots + \lambda_p,$$

unde $\lambda_1 > \lambda_2 > \dots > \lambda_p$ sînt rădăcinile caracteristice ale matricei A, Y este o matrice $k \times p$ de rang k astfel încît $YY' = I$.

În acest caz minimul este atins pentru y_i proporţional cu o funcţie liniară de vectorii proprii ai matricei A.

Rezultatul de mai sus se stabileşte ţinînd cont de faptul că

$$\text{Min}_y\, y'Ay$$

cu restricţiile $y'y = 1$, $Hy = 0$ este λ_{p-k} unde H este o matrice $k \times p$ de rang k, minimul fiind luat pentru toate variaţiile posibile ale matricei H.

În felul acesta obţinem că

$$\text{Min}_{y_1} \frac{y_1' A y}{y_1' y_1}$$

cu restricţiile $y_1'y_1 = 1$, $y_1'y_2 = 0, \dots, y_1'y_k = 0$, este λ_{p-k+1}.

Considerăm acum

$$\text{Min}_{y_2} \frac{y_2' A y_2}{y_2' y_2}$$

cu restricţiile $y_2'y_2 = 1$, $y_2'y_3 = 0, \dots, y_2'y_k = 0$, iar minimul este λ_{p-k+2} şi tot aşa mai departe în final considerăm

$\text{Min}_{y_k} \frac{y_k' A y_k}{y_k' y_k}$ cu restricţia $y_k'y_k = 1$, care ştim că este λ_p. Adunînd toate aceste valori obţinem

$$\text{Min}_{y_1, y_2, \dots, y_k} \sum_{1}^{k} \frac{y_i' A y_i}{y_i' y_i} = \lambda_{p-k+1} + \dots + \lambda_p.$$

Deoarece fiecare termen al expresiei

$$\text{Min}_{H_0} \sum_{1}^{k} N(\bar{U}_i - \mu_i)' \Sigma^{-1} (\bar{U}_i - \mu_i)$$

este pozitiv, vom avea minim cînd fiecare termen al sumei este minim cu restricţie

$$H\mu_i = \zeta, \quad 1 \leqslant i \leqslant k,$$

unde H este o matrice $s\times p$ de rang $s < p$, iar ζ este o constantă astfel încît ecuația $H\mu_i = \zeta$ este consistentă.

Minimul fiecărui termen al expresiei de mai sus se poate calcula acum utilizînd relațiile obținute mai înainte.

Pentru aceasta observăm că $\zeta - H\overline{U}_i$ este o matrice $s \times 1$ iar $(H\Sigma H')^{-1}$ este o matrice $s \times s$. Deci

$$(\zeta - H\overline{U}_i)'(H\Sigma H')^{-1}(\zeta - H\overline{U}_i) =$$
$$= \mathrm{Ur}\,(H\Sigma H')^{-1}(\zeta - H\overline{U}_i)(\zeta - H\overline{U}_i).$$

Folosind rezultatele obținute putem scrie acum că

$$\mathrm{Min}_H \sum_1^k N(\zeta - H\overline{U}_i)'(H\Sigma H')^{-1}(\zeta - H\overline{U}_i) =$$
$$= \mathrm{Min}_H\,\mathrm{Ur}\,N(H\Sigma H')^{-1}\sum_1^k(\zeta - H\overline{U}_i)(\zeta - H\overline{U}_i)' =$$
$$= \mathrm{Min}_H\,\mathrm{Ur}\,N(H\Sigma H')^{-1}\Big[k(\zeta - H\overline{U})(\zeta - H\overline{U})' +$$
$$+ H\left(\sum_1^k \overline{U}_i\overline{U}_i' - kU\overline{U}'\right)H'\Big] =$$
$$= \mathrm{Min}_H\,\mathrm{Ur}\,N(H\Sigma H')^{-1}\left[H\left(\sum_1^k \overline{U}_i\overline{U}_i' - k\overline{U}\,\overline{U}'\right)H'\right].$$

Însă

$$\mathrm{Min}_H\,\mathrm{Ur}\,N(H\Sigma H')^{-1}\left[H\left(\sum_1^k \overline{U}_i\overline{U}_i' - k\overline{U}\,\overline{U}'\right)H'\right] =$$
$$= \mathrm{Min}_T\,\mathrm{Ur}\,N(TT')^{-1}\left[T\Sigma^{-\frac{1}{2}}\left(\sum_1^k \overline{U}_i\overline{U}_i' - k\overline{U}\,\overline{U}'\right)\Sigma^{-\frac{1}{2}}T'\right] =$$
$$= \mathrm{Min}_T\,\mathrm{Ur}\,NT\,\Sigma^{-\frac{1}{2}}\left(\sum_1^k \overline{U}_i\overline{U}_i' - k\overline{U}\,\overline{U}'\right)\Sigma^{-\frac{1}{2}}T' =$$
$$= \lambda_{r+1} + \ldots + \lambda_p$$

cu restricția $TT' = I$,
unde $\lambda_1 > \lambda_2 > \ldots > \lambda_p$ sînt rădăcinile ecuației

$$\left|\sum_1^k \overline{U}_i\overline{U}_i' - k\overline{U}\,\overline{U}' - \lambda N^{-1}\Sigma\right| = 0.$$

V.49. *La un examen fiecare student răspunde la trei întrebări, răspunsurile fiind notate în mod independent de doi examinatori A și B. Diferența dintre punctajele date de cei doi examinatori (A — B) pentru întrebarea i se notează prin x_i, $i = 1, 2, 3$. Vectorul valoare medie și matricea de covarianță pentru x_1, x_2, x_3 estimate dintr-o selecție de volum 50, sînt trecute în tabelul V.29.*

Tabelul V.29

Media	Matricea de covarianțe		
	x_1	x_2	x_3
$\bar{x}_1 = 2,54$	16,90	22,01	12,53
$\bar{x}_2 = -1,72$		51,72	19,46
$\bar{x}_3 = 0,94$			28,73

La un prag de semnificație $\alpha = 0,01$, în presupunerea că (x_1, x_2, x_3) este un vector normal să se verifice ipoteza conform căreia media diferenței dintre cei doi examinatori pentru fiecare întrebare este zero ($\mu = 0$).

Soluție. În acest caz $p = 3$ și $\nu = n - 1 = 49$. Estimația $\hat{\Sigma}$ a fost determinată după formula

$$\hat{\Sigma} = \frac{S}{n-1}, \quad \text{unde} \quad S_{ij} = \sum_{\lambda=1}^{n} x_{i\lambda} x_{j\lambda} - n\bar{x}_i \bar{x}_j, \quad \bar{x}_i = \frac{1}{n} \sum_{\lambda=1}^{n} x_{i\lambda}.$$

Mai întîi calculăm forma pătratică $\bar{x}\hat{\Sigma}^{-1}\bar{x}'$, unde $\bar{x} = (\bar{x}_1, \bar{x}_2, \bar{x}_3)$ folosindu-ne de metoda condensării pivotale, cu un zero adițional ca ultim element al diagonalei.

Tabelul V.30

$\hat{\Sigma}$			$\bar{x}'$	
16,90	22,01	12,35	2,54	53,80
	51,72	19,46	−1,72	91,47
		28,73	0,94	61,48
			0	1,76
16,90	1,302367	0,730767	0,150296	3,183432
22,01	23,054902	0,146423	−0,091163	1,055260
12,35	3,375774	19,210722	−0,031670	0,968330
2,54	−2,101752	−0,608408	−0,592622	

Ultimul element cu semnul schimbat dă $\bar{x}\hat{\Sigma}^{-1}\bar{x}' = 0,592622$.
Statistica

$$T = \frac{n(n-p)}{\nu p} (\bar{x} - \mu_0)\, S^{-1}(\bar{x} - \mu_0)' \qquad \text{(V.19)}$$

în ipoteza emisă $\mu_0 = 0$ este repartizată $F_{p,\, n-p}$.

În cazul de față (V.19) are valoarea

$$T = \frac{n(n-p)}{\nu p}\, \bar{x}\hat{\Sigma}^{-1}\bar{x}' = \frac{50(50-3)}{49 \cdot 3} \cdot 0,592622 = 9,473.$$

Deoarece $T > F_{3,47;\,0,01}$ respingem ipoteza emisă, sau cu alte cuvinte există diferență semnificativă între cei doi examinatori.

V.50. *În tabelul V.31 sînt date estimațiile mediei și matricei de covarianțe corespunzătoare la trei caracteristici x_1, x_2, x_3 pentru două grupe de lăcuste A și B.*

Tabelul V.31

Caracteristică	Mediile: A $n_1 = 20$	Mediile: B $n_2 = 72$	Mediile: Diferențele d'	Matricea de covarianțe Σ: x_1	x_2	x_3
x_1	25,80	28,35	−2,55	4,7350	0,5622	1,4685
x_2	7,81	7,41	0,40		0,1413	0,2174
x_3	10,77	10,75	−0,02			0,5702

La un prag de semnificație $\alpha = 0,01$, în presupunerea că (x_1, x_2, x_3) este un vector normal, să se verifice ipoteza conform căreia cele două grupe au mediile celor trei caractere egale.

Soluție. Numărul de caractere este $p = 3$, iar numărul gradelor de libertate este $\nu = n_1 + n_2 - 2 = 90$. Notînd prin d diferența dintre vectorii medie $(\bar{x}^{(1)} - \bar{x}^{(2)})$ calculăm mai întîi $d\Sigma^{-1}d'$. Aceste calcule sînt trecute în tabelul V.32

Tabelul V.32

$\hat{\Sigma}$			d'	Suma	
4,7350	0,5622	1,4685	−2,5500	4,2157	
	0,1413	0,2174	0,4000	1,3227	
	0	0,5702	0,0200	2,2761	
			0	−2,1300	
4,7350	0,1187328	0,3101373	−0,5385428	0,8903273	0,8903273
0,5622	0,0763484	0,5637433	9,2047587	10,7685020	10,7685046
1,4685	0,0430409	0,0904994	4,5820083	5,5820083	5,5820045
−2,5500	0,7027686	0,4146690	−9,7421163	1	1

Găsim $d\hat{\Sigma}^{-1}d' = 9{,}742116$.
Statistica

$$T = \frac{n_1 + n_2 - p - 1}{p} \cdot \frac{n_1 n_2}{n_1 + n_2} dS^{-1}d', \qquad \text{(V.20)}$$

unde $d = \bar{x}^{(1)} - \bar{x}^{(2)}$, $S = S^{(1)} + S^{(2)}$, $\bar{x}^{(i)}$ şi $S^{(i)}$ fiind vectorul medie şi matricea de covarianţe de selecţie pentru selecţia de volum n_i din populaţia $N(\mu^{(i)}, \Sigma)$, $i = 1, 2$, este repartizată $F_{p,\, n_1+n_2-p-1}$.

Din(V.20) obţinem

$$T = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} \cdot \frac{n_1 n_2}{n_1 + n_2} d\hat{\Sigma}^{-1}d' =$$

$$= \frac{20 + 72 - 3 - 1}{3(20 + 72 - 2)} \cdot \frac{20 \cdot 72}{20 + 72} \cdot 9{,}742116 = 49{,}699.$$

Deoarece $T > F_{3,90;\, 0,01}$, respingem ipoteza emisă.

V.51. *În tabelul V.33 sînt date estimaţiile matricelor de covarianţe a trei caractere antropometrice: x_1, x_2, x_3 pentru locuitorii din trei regiuni A, B şi C.*

Tabelul V.33

	A			B			C		
	Volumul selecției								
	$n_1 = 170$			$n_2 = 337$			$n_3 = 299$		
	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
x_1	43,38	4,62	9,90	41,09	7,05	8,66	40,37	4,16	4,97
x_2		27,85	11,60		27,72	10,97		25,37	9,75
x_3			23,05			18,75			13,79

La un prag de semnificație $\alpha = 0{,}05$, *în presupunerea că* (x_1, x_2, x_3) *pentru fiecare din regiunile* A, B *și* C *este un vector normal, să se verifice ipoteza conform căreia matricele de covarianțe sînt egale.*

Soluție. Dacă notăm $\hat{\Sigma}_1, \hat{\Sigma}_2, \hat{\Sigma}_3 (\hat{\Sigma}_i = S_i/\nu_i)$ estimațiile matricelor de covarianțe pentru regiunile A, B și C, avem

$$|\hat{\Sigma}_1| = 19849{,}81; \quad |\hat{\Sigma}_2| = 14\,740{,}42; \quad |\hat{\Sigma}_3| = 9\,853{,}69.$$

Gradele de libertate sînt $\nu_1 = n_1 - 1 = 169$, $\nu_2 = n_2 - 1 = 336$, $\nu_3 = n_3 - 1 = 298$ și $\nu = \nu_1 + \nu_2 + \nu_3 = 803$. Avem

$$\hat{\Sigma} = (\nu_1\Sigma_1 + \nu_2\Sigma_2 + \nu_3\Sigma_3/\nu = \begin{pmatrix} 41{,}30 & 5{,}47 & 7{,}48 \\ & 26{,}89 & 10{,}65 \\ & & 17{,}81 \end{pmatrix}$$

de unde $|\hat{\Sigma}| = 13\,928{,}57$.

Pentru a vedea dacă matricele de covarianțe pentru cele 3 regiuni sînt semnificativ diferite, calculăm statistica

$$T = -\rho \ln B,$$

unde

$$B = \frac{\prod_1^k |\hat{\Sigma}_i|^{\nu_i}}{|\hat{\Sigma}|^{\nu}}, \quad k = 3$$

și

$$\rho = 1 - \left(\frac{1}{\nu_1} + \dots + \frac{1}{\nu_k} - \frac{1}{\nu}\right) \frac{2p^2 + 3p - 1}{6(p+1)(k-1)}, \quad p = 3$$

care, în ipoteza emisă este repartizată $\chi^2(f)$ unde

$$f = \frac{1}{2}(k-1)\,p(p+1).$$

Găsim

$$\rho = 0{,}9940396, \quad f = 12$$

şi

$$T = \rho(\nu \ln |\hat{\Sigma}| - \nu_1 \ln|\hat{\Sigma}_1| - \nu_2 \ln |\hat{\Sigma}_2| - \nu_3 \ln |\hat{\Sigma}_3|) = 24{,}091.$$

Din tabele $\chi^2_{12;\,0{,}05} = 21$ şi cum $T > \chi^2_{12;\,0{,}05}$ respingem ipoteza egalităţii matricelor de covarianţe.

V.52. *În tabelul V.34 sînt date matricele de covarianţe ale punctajelor x_1, x_2, x_3, x_4 obţinute la 4 teste, primele două de literatură şi următoarele două de matematică, bazate pe o selecţie de volum $n = 300$.*

Tabelul V.34

	x_1	x_2	x_3	x_4
x_1	15,129	23,860	1,793	0,998
x_2		54,756	3,633	3,511
x_3			18,225	21,122
x_4				60,516

La un prag de semnificaţie $\alpha = 0{,}05$, să se verifice ipoteza conform căreia punctajele la literatură (x_1, x_2) sînt independente de punctajele la matematică (x_3, x_4).

Soluţie. Pentru a verifica această ipoteză folosim statistica

$$L = \frac{|S|}{\Pi\,|S_i|}.$$

Această statistică poate fi exprimată în forma echivalentă

$$L = |\hat{\Sigma}|/|\hat{\Sigma}_{11}|\,|\hat{\Sigma}_{22}|,$$

unde $\hat{\Sigma}$ este matricea de covarianţe pentru toate cele 4 punctaje $\hat{\Sigma}_{11}$ este matricea de covarianţe corespunzătoare numai punctajelor (x_1, x_2) şi $\hat{\Sigma}_{22}$ matricea de covarianţe corespunzătoare numai punctajelor (x_3, x_4).

Avem $|\hat{\Sigma}| = 167\,198{,}37$

$$|\hat{\Sigma}_{11}| = \begin{vmatrix} 15{,}129 & 23{,}860 \\ 23{,}860 & 54{,}756 \end{vmatrix} = 259{,}1039$$

şi

$$|\hat{\Sigma}_{22}| = \begin{vmatrix} 18{,}225 & 21{,}122 \\ 21{,}122 & 60{,}516 \end{vmatrix} = 656{,}7652.$$

Urmează că $L = 0{,}98253$.
Avem

$$p = 4, \quad p_1 = 2, \quad p_2 = 2.$$

Statistica

$$T = -(n - \lambda)\ln L$$

care este repartizată $\chi^2(f)$, unde

$$f = \frac{1}{2}p^2 - \frac{1}{2}(p_1^2 + p_2^2) = 4,$$

$$\lambda = \frac{3}{2} + \frac{p^3 - (p_1^3 + p_2^3)}{6f} = 3{,}5$$

are valoarea 5,226.

Din tabele găsim $\chi^2_{4;\,0,05} = 9{,}42$.

Deoarece $T < \chi^2_{4;\,0,05}$ acceptăm ipoteza de independenţă.

V.53. *Într-un proces de producţie de serie mare 100 p % (p necunoscut) din produse prezintă defecte. Cu ajutorul testului secvenţial să se verifice ipoteza* H_0: $p = p_0$ *faţă de alternativa* H_1: $p = p_1 > p_0$ *şi să se determine funcţia caracteristică operatoare.*

S o l u ţ i e. Considerăm variabila aleatoare x_i care poate lua valoarea 1 sau 0, după cum produsul controlat prezintă sau nu defect. Funcţia de verosimilitate corespunzătoare variabilelor $x_1, x_2, \ldots, x_n$ este

$$\prod_1^n p^{x_i}(1-p)^{1-x_i}$$

iar raportul de verosimilitate este

$$\frac{p_{1n}}{p_{0n}} = \prod_1^n \left[\left(\frac{p_1}{p_0}\right)^{x_i}\left(\frac{1-p_1}{1-p_0}\right)^{1-x_i}\right] = \left(\frac{p_1}{p_0}\right)^{s_n}\left(\frac{1-p_1}{1-p_0}\right)^{n-s_n},$$

unde $s_n = \sum_1^n = x_i$ = numărul total de articole cu defecte găsite în cele n produse controlate.

Dacă

$$\frac{p_{1n}}{p_{0n}} \geqslant A, \quad \text{acceptăm ipoteza } H_1;$$

$$\frac{p_{1n}}{p_{0n}} \leqslant B, \quad \text{acceptăm ipoteza } H_0;$$

$B < \frac{p_{1n}}{p_{0n}} < A$, experimentul se continuă şi se face o observaţie suplimentară ;

A şi B satisfac inegalităţile

$$A \leqslant \frac{1-\beta}{\alpha}, \quad B \geqslant \frac{\beta}{1-\alpha}.$$

În cazul de faţă avem următorul procedeu.

Dacă

$$s_n \ln\left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right] \geqslant \ln\frac{1-\beta}{\alpha} + n\ln\frac{1-p_0}{1-p_1} \quad \text{acceptăm } H_1;$$

$$s_n \ln\left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right] \leqslant \ln\frac{\beta}{1-\alpha} + n\ln\frac{1-p_0}{1-p_1} \quad \text{acceptăm } H_0;$$

$$\ln\frac{\beta}{1-\alpha} < s_n \ln\frac{p_1(1-p_0)}{p_0(1-p_1)} - n\ln\frac{1-p_0}{1-p_1} < \ln\frac{1-\beta}{\alpha}.$$

Folosind acest procedeu, avem

P(resp. H_0 | adev H_0) $= \alpha$, P(accept H_0 | adev. H_1) $= \beta$.

Grafic avem situaţia din fig. V. 1.

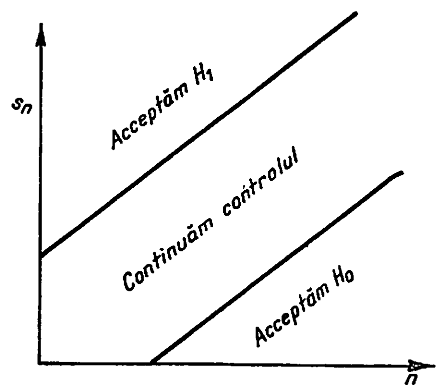

Fig. V.1

Pe abscisă am trecut numărul de produse controlate n, iar pe ordonată numărul total de produse defecte s_n găsite printre cele n produse controlate.

Atîta timp cît punctul (n, s_n) variază între dreptele paralele

$$s_n = \frac{\ln \frac{\beta}{1-\alpha} + n \ln \frac{1-p_0}{1-p_1}}{\ln \frac{p_1(1-p_0)}{p_0(1-p_1)}}$$

şi

$$s_n^* = \frac{\ln \frac{1-\beta}{\alpha} + n \ln \frac{1-p_0}{1-p_1}}{\ln \frac{p_1(1-p_0)}{p_0(1-p_1)}}$$

controlul se continuă. Cînd punctul (n, s_n) cade în afara acestor limite controlul se termină şi decizia este luată în concordanţă cu regulile de procedură.

Funcţia caracteristică operatoare, probabilitatea de acceptare a ipotezei H_0 este

$$L(p) \simeq \frac{1 - A^{h(p)}}{B^{h(p)} - A^{h(p)}} = \frac{1 - \left(\frac{1-\beta}{\alpha}\right)^{h(p)}}{\left(\frac{\beta}{1-\alpha}\right)^{h(p)} - \left(\frac{1-\beta}{\alpha}\right)^{h(p)}},$$

relaţia dintre h şi p

$$p = \frac{1 - \left(\frac{1-p_1}{1-p_0}\right)^h}{\left(\frac{p_1}{p_0}\right)^h - \left(\frac{1-p_1}{1-p_0}\right)^h}$$

obţinîndu-se din ecuaţia

$$\sum_{x=0}^{1} \left[\left(\frac{p_1}{p_0}\right)^x \left(\frac{1-p_1}{1-p_0}\right)^{1-x}\right]^h p^x(1-p)^{1-x} = 1.$$

Se observă că $h(p_0) = 1$ şi $(p_1) = -1$.

V.54. *Fie $x_1, x_2, \ldots$ un şir de variabile aleatoare independente toate avînd aceeaşi repartiţie $P_0(\lambda)$.*

Cu ajutorul testului secvenţial al raportului probabilităţilor să se determine regiunea critică pentru verificarea ipotezei

$$H_0: \lambda = \lambda_0$$

față de alternativa

$$H_1 : \lambda = \lambda_1 \, (> \lambda_0).$$

S o l u ț i e. Raportul de verosimilitate are valoarea

$$\frac{p_{1n}}{p_{0n}} = e^{-n\lambda_1} \frac{\lambda_1^{\sum\limits_1^n x_i}}{\prod\limits_1^n x_i!} \cdot \frac{1}{e^{-n\lambda_0} \dfrac{\lambda_0^{\sum\limits_1^n x_i}}{\prod\limits_1^n x_i!}} =$$

$$= e^{-n(\lambda_1 - \lambda_0)} \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum\limits_1^n x_i}.$$

Dacă α, β sînt probabilitățile erorilor de genul unu și respectiv doi atunci regiunea de continuare a experimentului este dată de

$$\frac{\beta}{1-\alpha} < e^{-n(\lambda_1 - \lambda_0)} \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum\limits_1 x_i} < \frac{1-\beta}{\alpha}$$

sau

$$\frac{\lambda_1 - \lambda_0}{\ln \dfrac{\lambda_1}{\lambda_0}} n + \frac{\ln \dfrac{\beta}{1-\alpha}}{\ln \dfrac{\lambda_1}{\lambda_0}} < \sum_1^n x_i < \frac{\lambda_1 - \lambda_0}{\ln \dfrac{\lambda_1}{\lambda_0}} n + \frac{\ln \dfrac{1-\beta}{\alpha}}{\ln \dfrac{\lambda_1}{\lambda_0}}.$$

De aici rezultă că regiunea critică pentru verificarea ipotezei H_0: $\lambda = \lambda_0$ față de alternativa H_1: $\lambda = \lambda_1 (\lambda_1 > \lambda_0)$ este dată de

$$U = \left\{ (x_1, \ldots, x_n) : \sum_1^n x_i \geqslant \frac{\lambda_1 - \lambda_0}{\ln \dfrac{\lambda_1}{\lambda_0}} n + \frac{\ln \dfrac{1-\beta}{\alpha}}{\ln \dfrac{\lambda_1}{\lambda_0}} \right\},$$

n fiind variabila aleatoare, numărul de probe necesar pentru acceptarea ipotezei H_0 sau H_1.

V.55. *Fie $x_1, x_2, \ldots$ un şir de variabile aleatoare independente toate avînd aceeaşi repartiţie $N(\mu, 1)$. Cu ajutorul testului secvenţial să se verifice ipoteza H_0: $\mu = \mu_0$ faţă de ipoteza alternativă H_1: $\mu = \mu_1 > \mu_0$, şi să se determine funcţia caracteristică operatoare.*

Soluţie. Raportul de verosimilităţi are valoarea

$$\frac{p_{1n}}{p_{0n}} = \frac{\exp\left[-\frac{1}{2}\sum_1^n (x_i - \mu_1)^2\right]}{\exp\left[-\frac{1}{2}\sum_1^n (x_i - \mu_0)^2\right]} = \exp\left[\Sigma x_i(\mu_1 - \mu_0) - \frac{n}{2}(\mu_1^2 - \mu_0^2)\right].$$

Dacă

$$\sum_1^n x_i \leqslant \frac{1}{\mu_1 - \mu_0} \ln \frac{\beta}{1 - \alpha} + \frac{n}{2}(\mu_1 + \mu_0) \quad \text{acceptăm } H_0,$$

$$\sum_1^n x_i \geqslant \frac{1}{\mu_1 - \mu_2} \ln \frac{1 - \beta}{\alpha} + \frac{n}{2}(\mu_1 + \mu_0) \quad \text{respingem } H_0,$$

iar dacă $\sum_1^n x_i$ variază între valorile indicate anterior se face a $(n+1)$ experienţă.

Funcţia caracteristică operatoare este

$$L(\mu) = \frac{\left(\frac{1 - \beta}{\alpha}\right)^h - 1}{\left(\frac{1 - \beta}{\alpha}\right)^h - \left(\frac{\beta}{1 - \alpha}\right)^h},$$

unde h determinat din condiţia

$$\int_{-\infty}^{\infty} \left(\frac{\varphi(x, \mu_1)}{\varphi(x, \mu_0)}\right)^h \varphi(x, \mu)\, dx = 1$$

are valoarea

$$h = \frac{\mu_1 + \mu_0 - 2\mu}{\mu_1 - \mu_0}.$$

V.56. *Fie $x_1, x_2, \ldots$ un şir de variabile aleatoare independente toate avînd aceeaşi repartiţie $N(0, \sigma^2)$. Cu ajutorul testului secvenţial să se veri-*

fice ipoteza H_0: $\sigma^2 = \sigma_0^2$ *față de ipoteza alternativă* H_1: $\sigma^2 = \sigma_1^2 > \sigma_0^2$ *și să se calculeze funcția caracteristică operatoare.*

Soluția. Raportul de verosimilități are valoarea

$$\frac{p_{1n}}{p_{0n}} = \frac{\exp\left[-\frac{1}{2\sigma_1^2}\Sigma x_i^2\right]}{\exp\left[-\frac{1}{2\sigma_0^2}\Sigma x_i^2\right]} = \exp\left[\frac{1}{2}\,\frac{\sigma_1^2 - \sigma_0^2}{\sigma_0^2 \cdot \sigma_1^2}\,\Sigma x_i^2\right].$$

Dacă

$$\Sigma x_i^2 \leqslant \frac{\sigma_0^2\sigma_1^2\left[2\lg\frac{\beta}{1-\alpha} + n\lg\frac{\sigma_1^2}{\sigma_0^2}\right]}{\sigma_1^2 - \sigma_0^2} \text{ acceptăm ipoteza } H_0,$$

$$\Sigma x_i^2 \geqslant \frac{\sigma_0^2\sigma_1^2\left[2\lg\frac{1-\beta}{\alpha} + n\lg\frac{\sigma_1^2}{\sigma_0^2}\right]}{\sigma_1^2 - \sigma_0^2} \text{ acceptăm ipoteza } H_1$$

facem o observație suplimentară [*a* $(n+1)$ — a] dacă Σx_i^2 variază între valorile date anterior.

Funcția caracteristică operatoare este

$$L(\mu) = \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h},$$

unde legătura dintre h și σ^2 determinată din condiția

$$\int_{-\infty}^{\infty}\left(\frac{\varphi(x, \sigma_1^2)}{\varphi(x, \sigma_0^2)}\right)^h \varphi(x, \sigma^2)\, dx = 1$$

are forma

$$\sigma^2 = \frac{\sigma_0^2\sigma_1^2\left[1 - \left(\frac{\sigma_0}{\sigma_1}\right)^{2h}\right]}{h(\sigma_1^2 - \sigma_0^2)}.$$

Graficul funcției $L(\sigma^2)$ trece prin punctele $(\sigma_0^2, 1-\alpha)$ și (σ_1^2, β). Avem

$$M(n \mid \sigma^2) \simeq \frac{2\sigma_0^2\sigma_1^2\left[L(\sigma^2)\lg\left(\frac{\beta}{1-\alpha}\right) + (1 - L(\sigma^2))\lg\left(\frac{1-\beta}{\alpha}\right)\right]}{(\sigma_1^2 - \sigma_0^2)\,\sigma^2 + \sigma_0^2\sigma_1^2\lg(\sigma_0^2/\sigma_1^2)}.$$

BIBLIOGRAFIE

1. Alan, J. Gross. *A note on the Convolution of Poisson Distributions with the zero class Missing.* Amer. Statist. 24,5, 1970.
2. Alan Stuart. *Reduced Mean-Square-Error Estimation of σ^p in Normal Samples.* Amer. Statist. 23,4, 1969.
3. Anderson T. W. *Introduction to multivariate statistical analysis.* New York-London, John Wiley, 1958.
4. Andreas, N. Philippon, Ram C. Dahiya. *Some Instructive Examples where the Maximum Likelihood Estimator of the Population Mean is not Sample Mean.* Amer. Statist. 24,3, 1970.
5. Bahadur, R. R. *Examples of inconsistency of maximum likelihood estimates.* Sankhya, 20, 207, 1958.
6. Barankin, E. W., Katz, Jr. *Sufficient statistics of minimal dimension* Sankhya, 21, 217 1959.
7. Bartlett, M.S. *Aproximate confidence intervals. Biometrika,* 40, 12, 306 şi 42, 201, 1953— 1955.
8. Barton, D. E. *A class of distributions for which the maximum likelihood estimate is unbiased and of minimum variance for all sample sizes. Biometrika,* 43, 200. 1956.
9. Bass, J. *Exercices de mathematiques.* Masson et Cie, 1965.
10. Basu, A. P. *Effect of trucation on a test for the scale parameter of the exponential distribution.* Ann. Math. Statist. 35,1, 1964.
11. Basu, D. *The family of ancillary statistics-Sankhya* 21,3, 1959.
12. Birnbaum Z. W. *Introduction to probability and mathematical Statististics.* New-York, Harper & Brother, 1962.
13. Brookner, R. J. 41. *A note on the mean as a poor estimate of central tendency.* J. Amer. Statist. Ass. 36, 1941.
14. Brownlee, K. A., *Statistical Theory and Methodology in science and Engineering,* John Wiley, New York London 1960.
15. Chakravarti, J. M., Laha, R. G., Roy, J. *Handbook of Methods of Applied Statistics,* vol. I, John Wiley, New York London, Sydney, 1967.
16. Chatterji, S. D., *Some elementary characterisation of the Poisson distribution.* Amer. Math. Monthly 70, 1963.
17. Chernoff. H., Lehmann, E. L. *The use of the maximum likelihood estimates in χ^2 test for goodness of fit.* Ann. Math. Statist., 25, 579, 1954.
18. Chernoff, H. and Svage I. R., *Asymptotic normality and efficieny o certain nonparametric test statistics.* Ann. Math. Statist, 29, 972, 1958.
19. Ciucu, G. *Elemente de teoria probabilităţilor şi statistică matematică,* Bucureşti Editura didactică şi pedagogică, 1963.

20. Ciucu, G., Sîmboan G. *Teoria probabilităţilor şi statistică matematică*, Bucureşti, Editura tehnică, 1962.
21. Ciucu, G., Craiu, V. *Probleme de statistică Matematică*. Editura didactică şi pedagogică, Bucureşti, 1968.
22. *Ciucu, G., Craiu V. Teoria estimaţiei şi verificarea ipotezelor statistice.* Editura didactică şi pedagogică, Bucureşti, 1969.
23. Colombo, B. *Contributo statistico ad un tentativo di discriminazione biometrica di populazioni o razze geografiche di teleostei del genere Bregmaceros. Rendic.* Ac. Naz. Lincei. vol. 14, fasc. 1—2, 1955.
24. Craig A. T. *On the mathematics of the reprezentative method of sampling.* Ann. Math. Statist. 10, 26, 1939.
25. Craiu, V. *Verificarea ipotezelor statistice.* Editura didactică şi pedagogică, Bucureşti, 1972.
26. Cramér, H. *Mathematical Methods of Statistics.* Princeton, Princeton University Press, 1950.
27. Crow, E. L. *Confidence intervals for a proportion.* Biometrika, 43, 423, 1956.
28. Cusimano, G. *La metodologia statistica condizionata dall'analisi di piu variabili.* Delf. Palermo, 1955.
29. D'Agostino, Ralph, B. *Linear Estimation of the Normal Distribution Standard Deviation.* Amer. Statist. 24,3, 1970.
30. Dahiya, R. C., Gurland John. *Functions of the sample Mean and Sample Variance of a Poisson Variate.* Biometrics, vol. 25, No. 1, 1969.
31. Dàll'Aglio G. *Analisi delle medie, della varianza e della covarianza.* Inst. Stat. Roma, 1960.
32. Daniels, H. E. *Some problems of statistical interest in wool research.* Suppl. J. R. Statisti. Soc. , 5, 89, 1938.
33. Daniels, H. E. *The theory of position finding.* J. R. Statist. Soc. (B), 13, 186 şi 14, 246, 1951—1952.
34. Dantzig, G. B. Wald, A. *On the fudamental lemma of Neyman and Pearson.* Ann, Math. Statist., 22, 87 1951.
35. Darling D. A. *The Cramer Smirnow test in the parametric case.* Ann., Math, Statist. 28 1, 1955.
36. Darmois, G. *Lecons sur l'estimation statistique*, Paris, 1947.
37. David A. Pierce, Richard L. Dykstra. *Independence and the Normal Distribution.* Amer. Statist. vol. 23 nr. 4, 1969.
38. Donald J. Wheeler. *Letters to the Edition* 24,5, 1970.
39. D. Dumas de Rauly. *L'estimation statistique.* Gauthier-Villars. Paris, 1966.
40. Dugué, D. *Traité de Statistique Théoritique et Appliquée.* Paris, Masson et Cie, 1958.
41. Fischer, R. A. *Contributions to mathematical statistics.* New York. Wiley and Sons, 1950.
42. Fisz, M. *Probability Theory and Mathematical Statistics.* New York, John Wiley and Sons, 1963.
43. Fraser, D. A. S. and Guttman, I. *Tolerance regions.* Ann. Math. Statist. 27, 162, 1956.
44. Fraser, D. A. S. *Nonparametric Methods in Statistics.* New. York, J. Wiley 1957.
45. Freeman, H. *Introduction to statistical inference.* London, Addison-Wesley, 1963.
46. Gérard Letac. *Problèmes de probabilité.* Presses Universitaires de France Paris, 1970.
47. Ghosh, B. K. *Asymptotic expansions for the moments of the distribution of correlation coefficient.* Biometrica 53, 1966.
48. Girault, M. *Calcul des Probabilités en vue des applications*, 1964.

49. G l a s s e r G. J., *An Unbiased Estimator for Powers of the Aritmetic* Mean. J. Royal Statist. Soc. 23,1, 1961.
50. G l a s s e r, G. J. *Estimators for the Product of Aritmetic* Means. J. Royal Statist. Soc. B. 24, 1962.
51. G o o d m a n, L, A. *On the exact variance of products* J. Amer. Statist. Ass. 55, 1960.
52. G n e d e n k o, B. V. *Curs teorii veroiatnostei*. Gosud. Izdatelstvo. Fiziko-Mate-maticeskoi literaturi, Moskva 1961.
53. G r a y b i l l, F. A. and M a r s a g l i a, G. *Idempotent matrices and quadratic forms in the general hypothesis*. Ann. Math. Statist. 28, 678, 1957.
54. G u r l a n d, J, *Inversion formulae for the distribution of ratios*. Ann. Math. Statist. 19, 1948.
55. G u r l a n d, J. *Note on a Paper by Rayand Pitman J.R.* Statist. Soc. B. 24,2., 1962.
56. H a l p e r i n, M. *Extension of the Wilcoxon-Mann-Whitney test to samples censored at the fixed point*. J. Amer. Statist. Ass. 55, 125, 1960.
57. H a m d a n, M.A., A l-B a y y a t i, H. A. *A note on the bivariate Poisson Distribution*. Amer. Statist. 23,4, 1969.
58. H a n n a n, E. J. *The asymptotic powers of certain tests based on multiple correlations*. J. R. Statist. Soc. (B), 18, 227, 1956.
59. H a r t l e y, H. O. *Maximum likelihood estimations from incomplete data*. Biometrics, 14, 174, 1958.
60. H e a l y, W. C. *Limits for a variance component with an exact confidence coefficient*. Ann. Math. Statist, 32, 467, 1961.
61. H o d g e s, J. L. J r. and L e h m a n n, E. L. *The efficiency of some nonparametric competitors of the t-test*. Ann. Math. Statist. 27, 324, 1956.
62. H o e l, P. *Introduction to Mathematical Statistics*. Third Edition, New York, J. Wiley & Sons.
63. H o g g, R. V., C r a i g, A. T. *Sufficient statistics in elementary distribution theory* Sankhya 17, 209, 1965.
64. H o y t, P. J o h n. *Two Instructive Exemples of Maximum Likelihood Estimates*. Amer. Statist. 23, 2, 1969.
65. H u z u r b a z a r, V. S. *Confidence intervals for the parameters of a distribution admitting a sufficient statistic when the range depends on the parameter*, J.R. Statist. Soc. (B), 17, 86, 1955.
66. I r w i n, J. O. *On the estimation of the mean of a Poisson distribution from a sample with the zero class missing*, Biometrics, 15, 329, 1959.
67. J o s h i, S. W. *Construction of Certain Bivariate Distributions* Amer. Statist. 24, 2, 1970.
68. K a c, M., K i e f e r, J. a n d W o l f o w i t z, J. *On tests of normality and other tests of goodness of fit bosed on distance methods*. Ann. Math. Statist,. 26, 189, 1955.
69. K e m p e r m a n, J. H. B. *Generalised tolerance limits*. Ann. Math. Statist. 27, 180, 1956.
70. K e n d a l l, M. G. *Exercices in Theoretical Statistics*. New York, Hafner Publishing 1954.
71. K e n d a l l M. G. and S t u a r t A. *The Advance theory of Statistics Distribution theory*. Vol. 1. London, Charles Griffin & Company limites, 1958.
72. K e n d a l l M. G. a n d S t u a r t A. *The Advanced theory of Statistics. Inference and relationship*. Vol. II, London, Charles Griffin & Company Limited 1961.
73. K i e f e r J., a n d W o l f w i t z, J. *Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters*. Ann. Math. Statist,. 27, 887, 1965.

74. K o n i j n, H. S. *On the power of certain tests for independence in bivariate populations*. Ann. Math. Statist. 27, 300 şi 29, 935, 1956, 1958.
75. K r u s k a l, W. *When are Gauss-Markov and least squares estimators identical? A coordinate-free approach*. Ann. Math. Statist. 39,1, 1968.
76. L a n c a s t e r, H. O. *On tests of independence in several dimensions*. J. Aust. Math. Soc. 1, 241, 1969.
77. L a n d e n n a, G. *I tests non parametric basati sulle frequenze*. Inst. Stat., 1960
78. L a w l e y, D. H. *A general method for approximating to the distribution of likelihood ratio criteria*. Biometrika, 43, 295, 1956.
79. L e h m a n n, E. L. a n d S c h e f f é. H. *Completeness, similar regions and unbiased estimation*, Sankhya, 10, 305 şi 15, 219, 1955.
80. L e h m a n n, E. L. *Consitency and unbiassedness of certain non-parametric tests*, Ann. Math. Statist. 22, 165, 1951.
81. L e h m a n n, E. L. *Testing Statistical Hypotheses*. London, John Wiley A. Sons, 1959.
82. L e o n a r d W. D e a t o n *An Example on Maximum Likelihood Estimates'* 24, 1 1970.
83. M a n n, H. B., W a l d, A. *On the choice of the number of intervals in the application of chi-square test*. Ann. Math. Statist. 13, 306, 1942.
84. M a s s e y, F. J. *A note on the estimation of a distribution function by confidence limits*. Ann. Math. Statist. 21, 116, 1950.
85. M a r t i n S a n d e l i u s *A note on the variance of a Discrete Uniform Distribution* Amer. Statist. 21, 5, 1967.
86. M a s s e y F. J. *A note on the power of a non-parametric test*. Ann. Math. Statist. 21, 440 şi 23, 637, 1950, 1952.
87. M a s s e y F. J. *The Kolmogorov-Smirnov test of goodness of fit*. J. Amer. Statist. Ass. 46, 68, 1951.
88. M a x L. E a t o n *Letters to the Editor*. Amer. Statist, 23, 3, 1969.
89. M e s a l k i n, L. D. *Sbornic zadaci po teorii veroiatnostei*. Izdatelstvo Moskovskovo, Universiteta, 1963.
90. M i h o c, G., U r s e a n u, V. *Matematici aplicate în statistică*. Bucureşti, Editura Acad. R.P.R., 1962.
91. M o n j a l l o n, A. *Elements de statistique mathematique*. Paris, Vuibert, 1962.
92. M o o d, A. M., *On the asymptotic efficiency of certain nonparametric twosample tests*. Ann. Math. Statist. 25, 514, 1945.
93. M o o r e, P. G. *The estimation of the mean of a censored normal distribution by ordered variables*, Biometrika, 43, 482, 1956.
84. M o s t e l l e r, F. *Fifty challenging problems in probability with solutions*. Reading, Addison Wesley, 1965.
95. N e v e u, J. *Bases mathématiques du Calcul des Probabilités*. Masson et Cie, 1964.
96. N e w m a n, D. J. *Problème* 5367. Amer. Math. Monthly, 73, 1966.
97. N e y m a n J. *Existence of consistent estimates of the directional parameter in a linear structural relation between two variables*. Ann. Math. Statist. 22, 297, 1951.
98. N o e t h e r, G. E. *Two confidence intervals for the ratio of two probabilities*, and some measures of effectiveness, J. Amer. Statist. Ass. 52, 36, 1957.
99. O n i c e s c u, O., M i h o c G. *Lecţii de statistică. matematică*, Bucureşti, Editura tehnică, 1958.
100. P a r z e n, E. *Modern Probability Theory and its applications*. New York John Wiley and Sons, 1960.
101. P a t i l, G. P. *A caracterization of the exponential type distribution*, Biometrika, 51, 2, 205, 1963.
102. P l a c k e t t, R. L., *A historical note on the method of least squares*. Biometrika 36, 458, 1949.

103. Plackett, R. L. *Some theorems in least squares.* Biometrika, 37, 149, 1950.
104. Plackett, R. L. *Linear estimation from censored data.* Ann. Math. Statist. 29, 131, 1958.
105. Putter, J. *The treatment of ties in non-parametric tests.* Ann. Math. Statist. 26, 368, 1955.
106. Quenouille, M. H. *Notes on bias in estimation* Biometrika, 43, 353, 1956.
107. Rangarajan, G., Chatterjeé *A note on Comparison Between Correlation Coefficients of Original and Transformed Variables* Amer. Statist. 23, 4, 1969.
108. Ramachandran, K. V. *A test of variances.* J. Amer. Statist. Ass. 53, 741, 1958.
109. Rao C. R. *Advanced Statistical Methods in Biometric Research.* New York, John Wiley and Sons, 1952.
110. Rao, C. R. *Estimation of relative potency from multiple response data* Biometrics, 10, 208, 1954.
111. Rao, C. R. *Analysis of dispersion for multiplity clasified data with unequal numbers in cells,* Aankhya 15, 253, 1955.
112. Rao, C. R. *Analysis of dispersion with incomplete observations on one of the characters.* Journ. Roy. Stat, Soc. B., 18, 259, 1965.
113. Rao, C. R. *Linear Statistical Inference and its applications.* London and New York. J. Wiley, 1956.
114. Rao C. R., Chakravarte, I. M. *Some small samples tests of significance for a Poisson distribution.* Biometrika 12, 264, 1956.
115. Raymond H. Burros *A Multivariate Generalization of Chebychev's Inequality.* Amer. Statist. vol. 24 nr. 5, 1970.
116. Renyi, A. *Neues Kriterion zum Vergleich zweier Stichproben,* Magyar Tud. Akad. Mat. Inst. 2, 243, 1953.
117. Richard M. Mayer *The* k^{th} *largest coordinate of an ordered n-tuple.* The Amer. Statist. 23, 5, 1969.
118. Romanovski, V. I. *Matematiceskaia Statistika.* Pervain kniga, Izdatelstvo Akademii Nauk Uzbekskoi S. S. R. 1961.
119. Romanovski, V. I. *Matematiceskaia Statistica.* Vtaraia kniga. Izdatelstvo Akademii Nauk Uzbekskoi S.S.R., 1961.
120. Roy, S. N. and Bose R. C. *Simultaneous confidence interval estimation.* Ann. Math. Statist., 24, 513, 1953.
121. Roy, K. P. *A note on the asymptotic distribution of likelihood ratio.* Bull. Calcutta Statist. Ass. 7, 73, 1957.
122. Roy, S. N. *Some further results in simultaneous confidence interval estimation.* Ann. Math. Statist. 25, 752, 1954.
123. Sankaran, M. *On the non-central* X^2 *distribution.* Biometrika 46, 235, 1959.
124. Scheffé, H. *The Analysis of Variance.* New York, John Wiley, 1959.
125. Shah, B. K. Khatri, C. G. *Distribution of definite quadratic form for non--central normal variates.* Ann. Math. Statist, 32, 3, 883, 1961.
126. Shenton, L. R. *Maximum likelihood and the efficiency of the method of moments,* Biometrika, 37, 111, 1950.
127. Strivastava, A. B. L. *Effect of non-normality on the power function of t-test.* Biometrika, 45, 221, 1958.
128. Stein, C. *Unibased estimates with minimum Variance.* Ann. Math. Statist, 21, 406, 1950.
129. Stuart, A. *Equally correlated variates and the multinormal integral,* J. R. Statist, Soc. (B), 20, 1958.
130. Sukhatme, B. V. *On certain two-sample nonparametric tests for variances,* Ann. Math, Statist. 28, 188, 1957.

131. Suresh C. Rastogi *Note on the Distribution of a Test Statistic.* Amer. Statist 23,5, 1969.
132. Tate, R. F., and Klett, G. W, *Optimal confidence intervals for the variance of a normal distribution.* J. Amer. Statist. Ass. 54, 674, 1959.
133. Teicher, H. *Maximum likelihood caracterization of distribution.* Ann. Math. Statist., 32, 1214, 1961.
134. Toke Jayachandran, D. R. Barr *On the distribution of a Difference of two scaled chi-square Random variables.* Amer. Statist. vol. 24 no. 5, 1970.
135. Tortrat, A. *Calcul des Probabilités,* Paris, 1963.
136. Tortrat, A. Hennequin, P. L. *Théorie des probabilités et quelques applications.* Masson et Cie, 1965.
137. Tucker, H. G. *An introduction to probability and mathematical statistics.* New-York, Academic Press Textbooks in Mathematics, 1962.
138. Udny Yule G. and Kendall, M. G. *An Introduction to the theony of statistics.* London-Griffin, 1965.
139. Ventell, S. S. *Teoria veroiatnostei-Gosudarstvennoe,* Moskva, Izdatelstvo Fiziko-Matematiceskoi literaturi, 1958.
140. Vassereau, A. *Sur les conditions d'applications de criterion* χ^2 *de Pearson,* Rev. Stat. Appl. 6, nr. 2, 83, 1958.
141. Waerden van Der. *Mathematiche Statistik,* Berlin, Springer Verlag, 1957.
142. Wald, A. *Notes on the Theory of Statistical Estimation and Testing Hypothese.* New York, Columbia University, 1941.
143. Wald, A. *Note on the consistency of the maximum likelihood estimate,* Ann, Math. Statist, 20, 595, 1949.
144. Walsh, J. E. *Some non-parametric test of whether the largest observations of a set are too large or too small.* Ann. Math. Statist. 21, 583, 1950.
145. Warren H. Bondy. *Note on the Distribution of a Test Statistic.* Amer. Statist 23, 5, 1969.
146. Watson G. S. *The* χ^2 *goodness-of-fit test for normal distribution.* Biometrika, 44, 336, 1957.
147. Wayne Nelson *Estimation for a Population Depleted by Sampling.* Amer. Statist. 21, 4, 1967.
148. Wilks, S. S. *Mathematical Statistics,* New-York, John Wiley and Sons 1962.
149. Williams, E, J. *Some exact tests in multivariate analysis.* Biometrika 39, 17, 1952.
150. Wiliams, E. J. *Significance tests for discriminant functions and linear functional relationships.* Biometrika, 42, 360, 1955.
151. Williams, E. J., *Exact fiducial in non-linear estimation.* J. R. Statist. Soc. B, 24 ,125, 1962.
152. Wolfowitz, J. *Estimation by the minimum distance method in nonparametric statistic difference equation,* Ann. Math. Statist., 25. 203, 1954.
153. Wolfowitz, J. *The minimum distance method,* Ann. Math. Statist, 28, 75, 1957.
154. Yates. F. *The analysis of contingency tables with groupings based on quantitative caracters.* Biometrika, 35, 176, 1948.
155. Yates, F. *A note on the applications of the combination of probabilities test to a set of* 2×2 *tables.* Biometrika, 42, 404, 1955.