Testarea numerelor aleatoare uniform distrbuite

A 1. Testul χ^2

Se considera variabila aleatoare S cu distributia

$$\begin{pmatrix} s_1 & s_2 & \cdots & s_k \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}$$ k = numarul de valori posibile ale lui S

Din n experimente rezulta ca valoarea s_1 a aparut de Y_1 ori, s_2 de Y_2 , s.a.m.d. Rezulta varibila aleatoare Y_s

$Y_s = Y_1 \quad Y_2 \; . . . \; Y_k$ cu probabilitatea de aparitie

$n*p_s = n*p_1 \; n*p_2 \qquad n*p_k$

Se calculeaza suma patratelor diferentelor:

$V = (Y_1 - n*p_1)^2 /(n*p_1) + (Y_2 - n*p_2)^2 /(n*p2) + \; . . + (Y_k -n*p_k)^2 /(n*p_k)$

care este statistica χ^2

sau pentru ca $Y_1 +Y_2+ . .+Y_k = n$ si $p_1 + p_2+ . .+p_k = 1$

$V = 1/n * (Y_1^2 /p_1 + Y_2^2 /p_2 + . . + Y_k^2 /p_k) - n \qquad (1)$

Valoarea lui V trebuie sa fie minima. Aceasta valoare se ia din tabele functie de k (KNUTH, Alogoritmi seminumerici, pag.58).

Tema 5 Sa se implementeze algoritmul de mai sus pentru unul din fisierele obtinute la temele anterioare.

A 2. Testul Kolmogorov-Smirmov(KS)

Se aplica asupra unor variabile aleatoare continue cu o functie de repartitie cunoscuta.

Functia de repartitie a unei variabile aleatoare X este:

$F(x)=p(X\le x)$ p = probabilitatea ca $X\le x$

Pentru numere aleatoare uniform distribuite $F(x)=x$.

Testul se desfasoara astfel:

Se fac n observatii asupra lui X obtinindu-se valorile $X_1, X_{2,} \ldots, X_n$

Rezulta functia de repartitie empirica:

$$F_n(x) = \frac{numarul \quad de \quad valori \quad X_1, X_2 \cdots X_n \quad \leq \quad X}{n}$$

si statisticile:

$K_n^+ = sqrt(n) * max(F_n(x) - F(x))$

$K_n^- = sqrt(n) * max(F(x) - F_n(x))$

Valorile obtinute K_n^+, K_n^- se compara cu cele din tabele (Knuth).

De obicei testul KS se aplica in cascada. Se considera portiuni diferite ale sirului X, se calculeaza K_{n_1}, K_{n_2} .... si se aplica testul KS asupra acestor valori.

A.3. Testul golurilor

Pentru U_j variabila aleatoare uniform distribuita pe (0,1) se ia un interval $0 \leq a < b \leq 1$. Marimea golului intre 2 aparitii succesive ale lui U_j apartine lui (a,b), este masurata de variabila aleatoare r care este egala cu numarul elementelor $U_j \ldots U_{j+r}$ pentru care $U_{j-1} \in (a,b)$ si $U_{j+r} \in (a,b)$ si in rest nu apartine lui (a,b).

Daca notam p = b-a variabila aleatoare r are distributie

$p_0 = p \quad p_1 = p(1-p) \quad \ldots \quad p_i = p(1-p)^{i-1} \quad \ldots \quad p_{t-1} = p(1-p)^{t-1} \quad p_t = (1-p)^t$

unde $p = P(r=i)$ si $0 \leq i \leq a-1$ si $p_t = P(r \geq t)$ (3)

Asupra variabilei r se aplica testul χ^2 pentru t (nr.grade libertate) notind *V* = numarul de goluri de lungime *i*; $0 \leq I \leq t-1$ dintre *n* goluri. Valoarea lui *V* se determina cu (1) cu p_s din (3)

Algoritmul pentru determinarea frecventelor V_i este :

Pas 1. Initializare j=1, s=0, V[r]=0 pentru $0 \leq r \leq t$; t=1000, n=1000, a=0, b=1/2

Pas 2. r=0

Pas 3. j=j+1, daca $a \leq U[j] < b$ atunci pas 5

Pas 4. r=r+1 , pas 3

Pas 5. daca r < t atunci V[r] =V[r]+1

altfel V[t]=V[t]+1

Pas 6. s=s+1 daca s < n atunci pas 2

Pas 7. Se calculeaza V

A.4 Testul permutarilor

Se imparte sirul U_j in *n* grupe a cite *t* elemente. Elementele din fiecare grupa admit t! ordonate posibile.

Fiecare ordonare are probabilitatea $p_i = 1$. Se aplica testul χ^2 sirului V_i ce numara frecventa unei anumite ordonari din cele t! permutari pentru t! grade de libertate .

Algoritmul de analiza a unei permutari ce asociaza unei permutari $U_1 \ldots U_t$ o valoare $f(U_1 \ldots U_t)$ $0 \le f < t!$ este:

Se foloseste un vector auxiliar T[r], r=t

Pas 1. Initializari r=t

Pas 2. Se cauta maxim din sir (U_1 U_2 ... U_r) => U_s ; T[r]=s-1

Pas 3. Se schimba elementele U_s si U_r intre ele

Pas 4. r=r-1 daca r>0 se trece la pas 2

Pas 5. Se calculeaza f dupa formula

f = T[t] +t * T[t-1] + t * (t-1) * T[t-2] + ... + t!*T[1]

Pas 6. V[f+1] = V[f+1]+1

Aplicind algoritmul tuturor celor *t* grupe se obtine vectorul

F[1] ...F[L] unde L=t! Se calculeaza *V* din (1) cu $p_s = p_i$

Ex. pentru t=7

A.5 Testul iteratilor

Se desparte sirul U_i in subsiruri ascendente

Se masoara numarul de elemente consecutive ascendente din fiecare subsir , rezulta ca lungimea iteratiei care este ≤ *n*. Frecventele lungimilor iteratiilor formeaza un sir aleator ce se verifica cu testul χ^2 pentru t=6 cu

$$V = \frac{1}{n} \sum_{i,j=1}^{6} a_{ij} \left(F_i - nb_i\right) * \left(F_j - nb_j\right)$$

cu a_{ij} b_{ij} din Vaduva pag. 89

Algoritmul de date a frecventelor F_i este:

P 1. $F[1] = F[2] = ... = F[6] = 0$, $j=0$ conditia $U_{n-1} > U_n$

P 2. $r = 0$

P 3. $r = r + 1$, $j = j + 1$ daca $U_j < U_{j+1}$ repetam 3

P 4. daca $r \geq 6$ atunci $F[6] = F[6] + 1$

altfel $F[r] = F[r] + 1$

P 5. daca $j < n - 1$ atunci pas 2

P 6. Calcul V

Teste referitoare la primele d cifre semnificative

Sirul U_i apartinind (0,1) devine $Yi = [k * Ui]$ $0<=Yi <= k-1$

A6. Testul frecvential

Se iau primele d cifre semnificative ale numerelor $U_i = Y_i$

Se aplica testul χ^2 cu $l=k$, $p_i = 1/k$ $1 \leq i \leq k$, $k=10^d$

k este astfel ales ca $n \geq 5 * k$, $(n*p_i \geq 5)$, V_i sint frecventele valorilor $Y_j = i-1$ pentru $1 \leq j \leq n$

A7. Testul serial

Se imparte sirul Y_i in perechi succesive $(Y_{2j}, Y_{2j+1})=(q,r)$

Se considera V_i numarul de perechi (q,r)

Se presupune ca Y_i are 2*n numere, rezulta n perechi

Se aplica χ^2 pentru $l=k^2$, cu $p_i = 1/k^2$, $1 \leq i \leq k^2$

Conditia $n > 5 * k^2$ (Perechile trebuie sa fie uniform distrbuite)

A8. Testul corelatiei seriale

Se calculeaza coeficientul :

$$C = \frac{n\,(U_0 U_1 + U_1 U_2 + \ldots + U_{n-2} U_{n-1} + U_{n-1} U_0) - (U_0 + U_1 + \ldots + U_{n-1})^2}{n(U_0^2 + U_1 + \ldots + U_{n-1}^2) - (U_0 + U_1 + \ldots + U_{n-1})^2}$$

$$m_n - s_n \le c < m_n + 2s_n$$

m_n = - 1/n-1   T = (1/n-1)* SQRT(n (n-3) / (n+1))   n>2

Concluzii

Testul serial si frecvential sint slabe, toate nr.trec

Testul iteratiilor este recomandat