TESTY LOSOWOŚCI

Badanie losowości próby - test serii.

W wielu zagadnieniach wnioskowania statystycznego istotnym założeniem jest losowość próby. Prostym testem do weryfikacji tej własności jest test serii.

Dla rozpatrywanego ciągu danych statystycznych obliczamy medianę m_e (wartość środkowa).

Jeśli $x_1 \le x_2 \le \ldots\ldots \le x_n$ dane uporządkowane to

$$m_e = \begin{cases} x_{\frac{n+1}{2}} & \text{dla n nieparzystych} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n+2}{2}} \right) & \text{dla n parzystych} \end{cases}$$

Przykład.

Dla danych (po uporządkowaniu)

2, 2, 3, 3, 4, 5, 5, 5, 5 medianą jest 4.

Dla danych (po uporządkowaniu)

2, 2, 2, 3, 3, 4, 5, 5, 5, 5

medianą jest 3,5.

Elementom próby przypisujemy symbol *a* lub *b*:

a - gdy $x_i > m_e$,

b - gdy $x_i < m_e$

(elementów $x_i = m_e$ nie rozpatrujemy).

Serie to podciągi złożone z jednakowych symboli.

Rozpatrujemy hipotezy
H_0(elementy próby mają charakter losowy),
H_1(elementy próby nie mają charakteru losowego),

Stosujemy statystykę:

$$U_n = \text{liczba serii}$$

Zbiór krytyczny:

$$K = (-\infty; k_1> \cup < k_{2;} \infty)$$

gdzie k_1 odczytujemy z tablicy dla poziomu istotności $\alpha/2$ i liczb n_1 oraz n_2 ,
gdzie k_2 odczytujemy z tablicy dla poziomu istotności $1 - \alpha/2$ i liczb n_1 oraz n_2 ,
gdzie n_1 - liczba symboli *a*, n_2 - liczba symboli *b*,
Decyzje:
Jeśli $U_n \in K$ to H_0 odrzucamy ,
Jeśli $U_n \notin K$ to nie ma podstaw do odrzucenia H_0 .

Uwaga.

Gdy n_1 lub n_2 jest większe od 20, to liczba serii ma w przybliżeniu rozkład

$$N\left(\frac{2n_1n_2}{n}+1;\sqrt{\frac{2n_1n_2\left(2n_1n_2-n\right)}{n^2(n-1)}}\right)$$

Dla rozkładu **równomiernego** i bardzo dużych n można stosować rozkład

$$N\left(\frac{n}{2};\frac{\sqrt{n}}{2}\right)$$

Tablica rozkładu serii

Tablica dla $\alpha = 0{,}025$: (tablica jest symetryczna)

n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5			2	2															
6		2	2	3	3														
7		2	2	3	3	3													
8		2	3	3	3	4	4												
9		2	3	3	4	4	5	5											
10		2	3	3	4	5	5	5	6										
11		2	3	4	4	5	5	6	6	7									
12	2	2	3	4	4	5	6	6	7	7	7								
13	2	2	3	4	5	5	6	6	7	7	8	8							
14	2	2	3	4	5	5	6	7	7	8	8	9	9						
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10					
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11				
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11			
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12		
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	
20	2	3	4	5	6	6	7	8	9	9	10	10	12	12	13	13	13	13	14

Tablica rozkładu serii

Tablica dla $\alpha = 0,975$: (tablica jest symetryczna)

n_2 \ n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	4																		
3	5	6																	
4	5	7	8																
5	5	7	8	9															
6	5	7	8	9	10														
7	5	7	9	10	11	12													
8	5	7	9	10	11	12	13												
9	5	7	9	11	12	13	13	14											
10	5	7	9	11	12	13	14	15	15										
11	5	7	9	11	12	13	14	15	16	16									
12	5	7	9	11	12	13	15	15	16	17	18								
13	5	7	9	11	13	14	15	16	17	18	18	19							
14	5	7	9	11	13	14	15	16	17	18	19	19	20						
15	5	7	9	11	13	14	15	17	17	18	19	20	21	21					
16	5	7	9	11	13	15	16	17	18	19	20	20	21	22	22				
17	5	7	9	11	13	15	16	17	18	19	20	21	22	22	23	24			
18	5	7	9	11	13	15	16	17	18	19	20	21	22	23	24	24	25		
19	5	7	9	11	13	15	16	17	19	20	21	22	22	23	24	25	25	26	
20	5	7	9	11	13	15	16	17	19	20	21	22	23	24	24	25	26	26	27

Przykład

W celu zbadania rozkładu wydajności pracy zarejestrowano czas wykonania detalu przez 15 wylosowanych pracowników i otrzymano wyniki (min):

16, 20, 25, 34, 22, 33, 47, 30, 28, 19, 22, 40, 36, 31, 38.

Sprawdzimy na poziomie istotności 0,05 hipotezę, że wybór próby był losowy.

Rozwiązanie.
Wyznaczamy medianę (po uporządkowaniu danych niemalejąco) i otrzymujemy $m_e = 30$.
Kolejnym danym przyporządkowujemy symbole
a i b:

16	20	25	34	22	33	47	30
b	b	b	a	b	a	a	-

28	19	22	40	36	31	38
b	b	b	a	a	a	a

Liczba serii wynosi u = 6

Z tablic rozkładu serii odczytujemy

$$K = (-\infty; 3> \cup < 12; \infty)$$

Ponieważ $u \notin K$ to nie ma podstaw do odrzucenia hipotezy H_0 , zatem możemy sądzić, że próba ma charakter losowy.

Przykład
Wylosować 2000 liczb o rozkładzie jednostajnym w (0, 1).
Dla poziomu istotności 0,05 sprawdzić ich losowość testem serii.

Badanie losowości próby - ogólny test serii.

Rozpatrzmy rodzinę testów serii do badania losowości.
Ustalamy liczbę $0 < p < 1$.
Dla rozpatrywanego ciągu danych statystycznych obliczamy **kwantyl** x_p (wartość dzieląca uporządkowane dane na części p% i (1 - p)%).
Liczbę p nazywamy **rzędem kwantyla**.

Elementom próby przypisujemy symbol *a* lub *b*:

a - gdy $x_i > x_p$,

b - gdy $x_i < x_p$

(elementów $x_i = x_p$ nie rozpatrujemy).

Serie to podciągi złożone z jednakowych symboli.

Rozpatrujemy hipotezy
H_0(elementy próby mają charakter losowy),
H_1(elementy próby nie mają charakteru losowego),

Stosujemy statystykę:

$$R = \text{liczba serii}$$

Statystyka ta ma parametry:

$$E(R) = 2np(1-p) + p^2 + (1-p)^2,$$

$$D^2(R) = 4np(1-p)(1-3p(1-p)) - 2p(1-p)(3-10p(1-p)),$$

gdzie *n* to długość badanego ciągu, a *p* to rząd kwantyla.

Dla dużych n statystyka

$$U = \frac{R - E(R)}{\sqrt{D^2(R)}}$$

jest zmienną losową o asymptotycznym rozkładzie normalnym N(0, 1).

Zbiór krytyczny:

$$K = (-\infty; -k> \cup < k; \infty)$$

gdzie k odczytujemy z tablicy N(0, 1) dla poziomu istotności 1- $\alpha/2$.
Decyzje:
Jeśli $U_n \in K$ to H_0 odrzucamy ,
Jeśli $U_n \notin K$ to nie ma podstaw do odrzucenia H_0 .

Przykład.

Wyznaczyć wartości

$$E(R) = 2np(1-p) + p^2 + (1-p)^2,$$

gdzie n to długość badanego ciągu, a p to rząd kwantyla.

dla n = 500; n = 1000; n = 5000 i p = 0,1; …..0,9

Wykonać wykresy.

	500	1000	5000
p	E(R)	E(R)	E(R)
0,1			
0,2			
0,3			
0,4			
0,5			
0,6			
0,7			
0,8			
0,9			

Przykład

Wyznaczyć wartości

$$D^2(R) = 4np(1-p)(1-3p(1-p)) - 2p(1-p)(3-10p(1-p)),$$

gdzie *n* to długość badanego ciągu, a *p* to rząd kwantyla.

dla n = 500; n = 1000; n = 5000 i p = 0,1; …..0,9

Wykonać wykresy.

	500	1000	5000
p	D^2(R)	D^2(R)	D^2(R)
0,1			
0,2			
0,3			
0,4			
0,5			
0,6			
0,7			
0,8			
0,9			

Przykład

Wylosować 2000 liczb o rozkładzie jednostajnym w (0, 1).

Dla poziomu istotności 0,05 sprawdzić ich losowość testem serii dla p = 0,25.

$E(R) = 750$
$D^2(R) = 655$

Uwaga

Zgodność uzyskanej liczby serii poszczególnych długości z rozkładem teoretycznym można zbadać wykorzystując test zgodności chi-kwadrat. Po wyznaczeniu liczby wystąpień każdej z długości serii symboli a lub b w badanym ciągu obliczamy statystykę testową:

$$U_n = \sum_k \frac{\left(N_k - E(k)\right)^2}{E(k)},$$

gdzie N_k – uzyskana liczba serii długości *k*,

$E(k) = (n - k + 3)\left(p^k(1-p)^2 + p^2(1-p)^k\right)$ – wartość teoretyczna liczby serii długości *k* w ciągu o długości *n*, kwantyla rzędu *p*.

Statystyka U_n ma w przybliżeniu rozkład χ^2 o r–1 stopniach swobody, gdzie *r* jest liczbą długości serii, które zaobserwowano.

Hipoteza zerowa H_0 (Ciąg ma charakter losowy).

Hipoteza alternatywna H_1 (Ciąg nie ma charakteru losowego).

Poziom istotności α.

Obliczamy wartość u_n statystyki U_n.

Wyznaczamy zbiór krytyczny

$$K = \langle k \ ; \infty),$$

gdzie *k* wyznaczamy z tablicy rozkładu χ^2 z r - 1 stopniami swobody .

$$P\left(Y_{r-1} \geq k\right) = \alpha,$$

Podejmujemy decyzję:

odrzucamy hipotezę H_0, gdy $u_n \in K$

przyjmujemy hipotezę H_0, gdy $u_n \notin K$

Przykład

n = 200, ciąg o rozkładzie jednostajnym w (0, 1), p = 0,5 .

$$U_n = \sum_k \frac{\left(N_k - E(k)\right)^2}{E(k)},$$

gdzie N_k – uzyskana liczba serii długości *k*,

$$E(k) = (n-k+3)\left(p^k(1-p)^2 + p^2(1-p)^k\right)$$

	n	p		
	200	0,5		
dł serii	ni	E(i)	składniki	
1	38			
2	21			
3	11			
4	6			
5	5			
				U

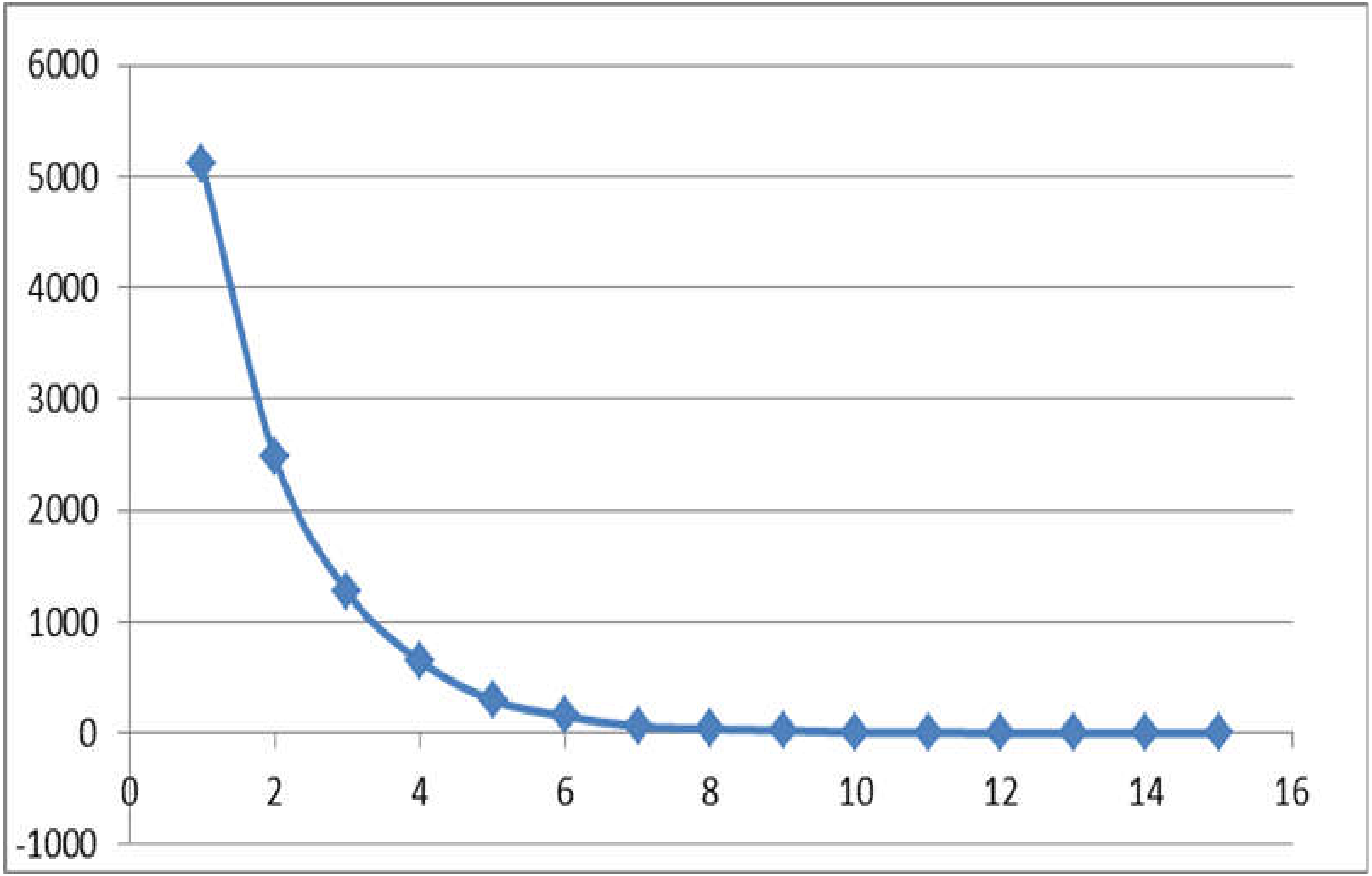
	n	p	
	200	0,5	
dł serii	ni	E(i)	składniki
1	38	50,5	
2	21	25,125	
3	11	12,5	
4	6	6,21875	
5	5	3,09375	
			U

alfa	0,01
r	5
k	
alfa^	

Przykład

n = 20000, ciąg binarny, p = 0,5

dł serii	ni
1	5109
2	2479
3	1267
4	647
5	298
6	154
7	61
8	37
9	24
10	9
11	7
12	2
13	0
14	0
15	0
	10094



Liczba serii danej długości

dł serii	ni	E(i)	składniki	
1	5109			
2	2479			
3	1267			
4	647			
5	298			
6	154			
7	61			
8	37			
9	24			
10	9			
11	7			
12	2			
13	0			U
14	0			
15	0			
	10094			

dł serii	ni	E(i)	składniki
1	5109	5000,5	
2	2479	2500,125	
3	1267	1250	
4	647	624,9688	
5	298	312,4688	
6	154	156,2266	
7	61	78,10938	
8	37	39,05273	
9	24	19,52539	
10	9	9,762207	
11	7	4,880859	
12	2	2,440308	
13	0		
14	0		
15	0		
	10094		

alfa	0,01
r	12
k	
alfa^	

Decyzja?

Test częstości dla ciągów binarnych (NIST)

Niech n – długość ciągu ($n \geq 100$), n_1 – liczba jedynek, n_0 – liczba zer,
Stosujemy statystykę (rozkład asymptotyczny N(0, 1))

$$u = \frac{|n_1 - n_0|}{\sqrt{n}}$$

Wynik testu jest pozytywny gdy $2(1 - \Phi(u)) > 0{,}01$.

Przykład

Wygenerować ciąg binarny n = 400 i sprawdzić go testem częstości.

$$u = \frac{|n_1 - n_0|}{\sqrt{n}}$$

$$2(1 - \Phi(u)) > 0{,}01\ \textbf{?}.$$

Test częstości dla bloków ciągów binarnych (NIST)

Niech n – długość ciągu ($n \geq 100$),
M – długość bloku ($M \geq 20$, $M > 0,01n$),
N liczba bloków $N = \left\lfloor \frac{n}{M} \right\rfloor$, ($N < 100$).

π_i = (suma jedynek w i-tym bloku)/M,
Stosujemy statystykę (rozkład asymptotyczny Y_N)

$$u = \frac{\sum_{i=1}^{N} \left(M\pi_i - M\,0,5\right)^2}{0,5M} = 2M \sum_{i=1}^{N} \left(\pi_i - 0,5\right)^2$$

Wynik testu jest pozytywny gdy $Y_N(u) > 0,01$.

Przykład

Wygenerować 10 bloków po 100 elementów binarnych i sprawdzić testem częstości dla bloków.

Test jednostajności dla ciągów binarnych (NIST)

Test spektralny (test transformaty Fouriera)

Niech n – długość ciągu, (n ≥ 1000),

Wygenerowany ciąg zerojedynkowy x_i przekształcamy na ciąg y_i złożony z 1 i -1 przez funkcję $y_i = 2x_i - 1$.

Otrzymany ciąg y_i przekształcamy przez dyskretne przekształcenie Fouriera.

$$f_j = \sum_{k=1}^{n} y_k \exp\big(2\pi i(k-1)j/n\big) = \sum_{k=1}^{n} y_k [\cos\big(2\pi(k-1)j/n\big) + i \sin\big(2\pi(k-1)j/n\big)]$$

gdzie i-jedynka urojona,

Test opiera się na własności, że średnio 95% wartości ciągu f_j ma moduł

mniejszy od $T = \sqrt{\ln\left(\frac{1}{0{,}05}\right) n} = \sqrt{2{,}995732274n} \approx \sqrt{3n}$

Wartości dyskretnego przekształcenia Fouriera są symetryczne; wystarczy, zatem dalej rozpatrywać tylko elementy o numerach od 1 do n/2.

Niech $N_0 = 0{,}95n/2$.

Niech N_1 = liczba elementów ciągu f_j o numerach od 1 do n/2, których moduł jest mniejszy od T.

Stosujemy statystykę (rozkład asymptotyczny N(0, 1))

$$u = \frac{N_1 - N_0}{\sqrt{0{,}95 \cdot 0{,}05 \cdot \frac{n}{2}}}$$

Wynik testu jest pozytywny, gdy $2(1 - \Phi(|u|)) > 0{,}01$.

Przykład

Niech (x_i) = 1,0,0,1,0,1,0,0,1,1,1,0,1,1,1,0

Sprawdzić jednostajność tego ciągu.

n=16

wtedy (y_i) = 1,-1, -1,1, -1,1, -1, -1,1,1, 1, -1, 1,1,1, -1

xi	yi	f_i	moduły
1	1	2	2
0	-1	-1,08239220029239+3,74603492445379i	3,899276
0	-1	-0,828427124746192	0,828427
1	1	-2,61312592975275+3,44155305449895i	4,321194
0	-1	2-4i	4,472136
1	1	2,61312592975275+1,78469880500656i	3,164424
0	-1	4,82842712474619	4,828427
0	-1	1,08239220029238-5,91081932503859i	6,009106
1	1	2	2
1	1	1,0823922002924+5,91081932503858i	6,009106
1	1	4,82842712474619	4,828427
0	-1	2,61312592975275-1,78469880500657i	3,164424
1	1	2+4i	4,472136
1	1	-2,61312592975276-3,44155305449894i	4,321194
1	1	-0,828427124746189	0,828427
0	-1	-1,0823922002924-3,7460349244538i	3,899276
9			

T	N0	N1	u	p

Wniosek:

Przykład

Wygenerować 1024 elementów binarnych i sprawdzić testem jednostajności.

Wyniki orientacyjne

T	N_0	N_1	u	p

Wniosek:

Test losowości dla ciągów binarnych (NIST)

Test sum skumulowanych (analizuje się maksymalne odchylenie sumy elementów od 0).

Niech n – długość ciągu, (n ≥ 100),

Wygenerowany ciąg zerojedynkowy x_i przekształcamy na ciąg y_i złożony z 1 i -1 przez funkcję $y_i = 2x_i - 1$.

Niech $z_1 = \max_{k=1,..,n}\left(\sum_{i=1}^{k} y_i\right)$; $z_2 = \max_{k=1,..,n}\left(\sum_{i=n-k+1}^{k} y_i\right)$ to sumy częściowe wprost i wstecz.

Dla $z = z_1$ i $z = z_2$ obliczamy

$$p = 1 - \sum_{k=k_1}^{k_2}\left[\Phi\left(\frac{(4k+1)z}{\sqrt{n}}\right) - \Phi\left(\frac{(4k-1)z}{\sqrt{n}}\right)\right] + \sum_{k=k_3}^{k_2}\left[\Phi\left(\frac{(4k+3)z}{\sqrt{n}}\right) - \Phi\left(\frac{(4k+1)z}{\sqrt{n}}\right)\right]$$

gdzie

$$k_1 = \left\lceil \frac{-n/z+1}{4} \right\rceil$$

$$k_2 = \left\lfloor \frac{n/z-1}{4} \right\rfloor$$

$$k_3 = \left\lceil \frac{-n/z-3}{4} \right\rceil$$

Wynik testu jest pozytywny, gdy $p > 0{,}01$ dla $z = z_1$ i $z = z_2$.

Przykład

Niech $(x_i) = 1,0,1,1,0,1,0,1,1,1$

Sprawdzić losowość tego ciągu.

n=10

wtedy $(y_i) = 1,-1, 1,1, -1,1, -1, 1,1,1$

$z_1 = ?$

$z_2 = ?$

1
0
1
2
1
2
1
2
3
4

4
3
4
3
2
3
2
3
2
1

z1
4

z2
4

$k_1 = ?$

$k_2 = ?$

$k_3 = ?$

k1
-0,375
0

k2
0,375
0

k3
-1,375
-1

p=?

p

Wniosek:

Test losowości χ^2
Testowanie losowości generatora ciągu binarnego.

Hipoteza zerowa H_0(Ciąg ma charakter losowy).

Hipoteza alternatywna H_1(Ciąg nie ma charakteru losowego).

poziom istotności α.

Weryfikacja powyższych hipotez za pomocą testu χ^2 przebiega następująco:

1. Generujemy długi binarny ciąg losowy. Dzielimy go na bloki np. 4-bitowe.
n – liczba bloków, n > 80,
k – liczba możliwych wartości w bloku, (dla bloków 2-bitowych k = 4 liczb dwubitowych, dla bloków 4-bitowych k = 16 liczb czterobitowych)
2. Przyjmujemy, że $p_i = \frac{1}{k}$.
3. Wyznaczamy liczbę n_i wystąpień i – tej wartości we wszystkich blokach, i = 1, 2, , k. $\sum_{i=1}^{k} n_i = n$

4. Obliczamy

$$u_n = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = \frac{1}{n} \sum_{i=1}^{k} \frac{(n_i)^2}{p_i} - n$$

5. Wyznaczamy zbiór krytyczny obustronny

$$K =< 0\ ;\ k_1 > \cup < k_2\ ;\ \infty)\ ,$$

gdzie k_1, k_2 wyznaczamy z tablicy rozkładu χ^2 z k - 1 stopniami swobody .

$$P(Y_{k-1} \geq k_1) = 1 - \frac{\alpha}{2}, \qquad P(Y_{k-1} \geq k_2) = \frac{\alpha}{2}$$

Podejmujemy decyzję:

odrzucamy hipotezę H_0, gdy $u_n \in K$

przyjmujemy hipotezę H_0, gdy $u_n \notin K$

Test pokerowy 4 bitowy.

Hipoteza zerowa H_0(Ciąg ma charakter losowy).

poziom istotności α.

Generujemy długi binarny ciąg losowy. Dzielimy go na bloki 4-bitowe.

n – liczba bloków, n > 80,

k – liczba możliwych wartości w bloku, dla bloków 4-bitowych k = 16 liczb czterobitowych)

Przyjmujemy, że $p_i = \dfrac{1}{16}$.

Wyznaczamy liczbę n_i wystąpień i – tej wartości we wszystkich blokach, i = 1, 2, , 16. $\sum_{i=1}^{16} n_i = n$

Obliczamy

$$u_n = \frac{16}{n} \sum_{i=1}^{16} (n_i)^2 - n$$

6. Wyznaczamy zbiór krytyczny obustronny

$$K =< 0 \; ; \; k_1 > \cup < k_2 \; ; \; \infty) \; ,$$

gdzie k_1, k_2 wyznaczamy z tablicy rozkładu χ^2 z k - 1 stopniami swobody .

$$P\left(Y_{k-1} \geq k_1\right) = 1 - \frac{\alpha}{2} \; ,$$

$$P\left(Y_{k-1} \geq k_2\right) = \frac{\alpha}{2}$$

Podejmujemy decyzję:
Nie ma podstaw do odrzucenia ciągu gdy

$$k_1 < u_n < k_2$$

Wyznaczanie liczby jedynek w ciągu binarnym

Przykład. (*test monobitowy*)

Wyznaczanie liczby jedynek w równomiernym ciągu binarnym.

Niech n = 10000, poziom istotności α = 0,01.

$$H_0(p = 0{,}5), \qquad H_1(p \neq 0{,}5)$$

$$K = (-\infty\,; -k > \cup < k\,; \infty), \quad \Phi(k) = 1 - \frac{\alpha}{2}$$

$$U_n = \frac{W - 0{,}5}{\sqrt{0{,}5(1 - 0{,}5)}}\sqrt{n}$$

gdzie W – średnia częstość jedynek w ciągu.

Wtedy k = 2,58,

$$K = (-\infty\ ;\ -k > \cup < k\ ;\ \infty) = (-\infty\ ;\ -2{,}58 > \cup < 2{,}58\ ;\ \infty)$$

Aby test był pozytywny dla generatora liczba jedynek *k* powinna być w granicach

$$-2{,}58 < \frac{\frac{k}{10000} - 0{,}5}{\sqrt{0{,}5(1-0{,}5)}}\sqrt{10000} < 2{,}58$$

Czyli

$$\text{od } 4872 \text{ do } 5128$$

Zadanie:
wykonać powyższy przykład dla n = 10000 poziomu istotności $\alpha = 0{,}001$.

Zadanie (NIST):
wykonać powyższy przykład dla n = 20000 i poziomu istotności $\alpha = 0{,}0001$.

Test entropii (Maurera)

Generujemy długi binarny ciąg losowy. Dzielimy go na bloki L-bitowe

(zwykle L = 8, ... , 16).

Q – sekwencja inicjująca,
K – sekwencja testowa,
Długość próbki N = (Q + K)L.

Uwaga,

$$Q > 5 \cdot 2^L, \quad K >> Q$$

zatem dla np. L = 8 $\quad Q > 5 \cdot 2^8 = 1280$

Statystyka testowa

$$U = \frac{1}{K} \sum_{n=Q}^{Q+K-1} \log_2 A_n$$

A_n – odległość n-tego bloku od jego ostatniego wystąpienia (lub *n* gdy blok występuje pierwszy raz),

Stawiamy hipotezę: H_0(entropia istotna),
Pożądana jest istotna entropia.

Zbiór krytyczny

$$K_r = < 0 ; k_1 > \cup < k_2 ; \infty) =$$

Wyznaczanie $k_1 < k_2$:

$k_1 = E - k\sigma$,

$k_2 = E + k\sigma$

gdzie $\Phi(k) = 1 - \frac{\alpha}{2}$, dla poziomu istotności α.

$\sigma = \sqrt{\frac{V}{K}}$ K - długość sekwencji testowej

Zestawienie wartości E, V dla przykładowych L.

L	E	V
8	7,1836655	3,238
9	8,17642476	3,311
10	9,17232431	3,356
11	10,17003223	3,384
12	11,1687649	3,401
13	12,1680703	3,410
14	13,1676926	3,416
15	14,1674884	3,419
16	15,1673788	3,421

Nie ma podstaw do odrzucenia ciągu gdy

$$k_1 < u < k_2$$

Testy losowości - testy kombinatoryczne.

Test permutacji.

Hipoteza zerowa H_0(Ciąg ma charakter losowy).

poziom istotności α.

Generujemy ciąg losowy o rozkładzie jednostajnym w [0, 1).

Liczba elementów $n = k \cdot m$.

Dzielimy go na bloki m-elementowe.

Przyporządkowujemy każdemu blokowi permutację m elementową wg kolejności liczb w bloku (wg kolejności rosnącej). Bloki w których są powtarzające się elementy odrzucamy.

Wyznaczamy prawdopodobieństwo teoretyczne permutacji ($p_i = \frac{1}{m!}$) przy założeniu H_0,

Badamy testem chi kwadrat zgodność zaobserwowanej liczby wystąpień poszczególnych permutacji n_i z rozkładem teoretycznym (rozkład jednostajny)

$$u = \sum_{i=1}^{m!} \frac{\left(n_i - k/m!\right)^2}{k/m!}$$

Wyznaczamy zbiór krytyczny

$$K = < k \quad ; \infty)\ ,$$

gdzie *k* wyznaczamy z tablicy rozkładu χ^2 z m! - 1 stopniami swobody .

$$P(Y_{m!-1} \geq k) = \alpha\ ,$$

Podejmujemy decyzję:

odrzucamy hipotezę H_0, gdy $u_n \in K$

przyjmujemy hipotezę H_0, gdy $u_n \notin K$

Przykład.
Wygenerować ciąg 20 elementowy o rozkładzie jednostajnym w [0, 1).

Dla m = 2 i standardowego poziomu istotności sprawdzić jego losowość testem permutacji.

0,382	0,245033	1	0
0,100681	0,045473	1	0
0,596484	0,03238	1	0
0,899106	0,164129	1	0
0,88461	0,219611	1	0
0,958464	0,01709	1	0
0,014496	0,285043	0	1
0,407422	0,343089	1	0
0,863247	0,553636	1	0
0,138585	0,357372	0	1

	ni	pi	kpi		
n1					
n2					
					u
					k

Dane:
69, 49, 19, 73, 68, 71, 97, 22 ,04, 24, 40, 73, 82, 36, 96, 10, 63,
29, 41, 37, 32, 12, 95, 16, 33, 40, 50, 32, 92, 31, 56, 59, 29, 13,
15, 60, 63, 67, 18, 10, 83, 67, 18, 08, 57, 21, 28, 54, 01, 37, 92,
81, 33, 42, 34, 28, 67, 70, 02, 85, 90, 35, 99, 71, 50, 94, 73, 64,
76, 19, 84, 54, 01, 27, 22, 40, 30, 53, 14, 41, 70