

I N S T Y T U T Ł A C Z N O S C I

Na prawach rękopisu

R E F E R A T Y P R O B L E M O W E

Zeszyt 1

Romuald Biało-brzeski, Stanisław Sońta

ZASTOSOWANIE TESTU CHI KWADRAT PEARSONA
DO WERYFIKACJI HIPOTEZY STATYSTYCZNEJ
NA PODSTAWIE
EMPIRYCZNEJ GĘSTOŚCI PRAWDOPODOBIEŃSTWA

Warszawa, grudzień 1977

S-8198

Dr inż. Romuald Biało-brzeski /Z-12/

Mgr inż. Stanisław Sońta /Z-2/

Nr pracy 5/2-14

Instytut Łączności

04-894 Warszawa, ul. Szachowa 1, tel. 128

Maszynopis dostarczono dnia 7.12.1977 r.

BIBLIOTEKA
Instytutu Łączności
Nr 5-8198

W zagadnieniach technicznych występuje często potrzeba sprawdzenia hipotezy statystycznej, że badana cecha populacji generalnej ma określony rozkład statystyczny. Zagadnienie polega na przeprowadzeniu testu weryfikującego przyjętą hipotezę, że badana zmienna losowa ma rozkład zbliżony do określonego rozkładu teoretycznego. W tego typu testach nie precyzuje się wartości parametrów rozkładu, stąd testy takie zwane są nieparametrycznymi testami zgodności.

Szerzej znaną statystyką służącą za miarę rozbieżności między rozkładem empirycznym a teoretycznym jest statystyka χ^2 Pearsona [1, 2]:

$$\chi^2 = \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i} \quad /1/$$

gdzie: r - liczba rozłącznych przedziałów możliwych wartości badanej cechy, np_i ,

n - całkowita liczebność badanej próby zmiennej losowej,

n_i - liczebność zaobserwowanych wartości cechy U w przedziale i -tym,

p_i - częstość teoretyczna zaobserwowania cechy U w i -tym przedziale^{1/}.

^{1/}Wartości p_1 i p_r wyznacza się zakładając, że zmienna losowa przyjmuje wartości odpowiednio mniejsze i większe od granic przedziałów [1].

W zastosowaniach praktycznych często wykorzystuje się zależność przekształconą do postaci:

$$\chi^2 = \sum_{i=1}^r \frac{n_i^2}{np_i} - n = n \left[\sum_{i=1}^r \left(\frac{n_i}{n} \right)^2 \frac{1}{p_i} - 1 \right] \quad /2/$$

Przy czym obowiązują oczywiście zależności:

$$\sum_{i=1}^r p_i = 1 \quad \text{oraz} \quad \sum_{i=1}^r n_i = n \quad /3/$$

Stosowany do badań właściwości sygnałów losowych analizator statystyczny /korelator/ typ HP 3721 A umożliwia uzyskanie w sposób bezpośredni dystrybuanty, gęstości prawdopodobieństwa oraz funkcji korelacyjnych sygnałów losowych, będących realizacjami stacjonarnych procesów losowych [4]. Empiryczne prawdopodobieństwo uzyskiwane za pomocą analizatora jest podawane w postaci wykresu oscyloskopowego dla 100 punktów i w postaci wartości cyfrowych dla 100 wartości zmiennej losowej - napięcia u badanego sygnału. Wartości prawdopodobieństwa /ściślej - liczebności n_i / przedstawiane są w postaci cyfrowej /wartości liczbowych/, przy czym linia podstawy wykresu /zerowe prawdopodobieństwo/ reprezentowane jest przez 0, zaś górny kraniec wykresu przez liczbę 2048. Prawdopodobieństwo empiryczne najwyższego punktu wykresu wynosi [4]:

$$p_{\max}/u/ = \frac{2048}{r-100} \sum_{i=1} y_i \quad /4/$$

gdzie: y_i - wartość liczbową i -tego punktu.

Oś odciętych uzyskiwanego wykresu jest podzielona na 100 punktów ponumerowanych od 1 do 100, przy czym numer $i=1$ jest

przyporządkowany lewemu skrajnemu punktowi wykresu, a numer $i=100$ - prawemu. Odstęp między dwoma kolejnymi punktami tej osi jest szerokością przedziału napięciowego i wynosi:

$$\Delta u = \frac{\text{pozycja przełącznika HORIZ V/CM}}{10} \text{ [V]} \quad /5/$$

Kalibracja osi odciętych pozwala na przyporządkowanie punktom osi odciętych wartości napięcia /woltach/ zgodnie z zależnością

$$u_i = \frac{i-50}{10} \cdot \text{/pozycja przełącznika HORIZ V/CM/ [V]} \quad /6/$$

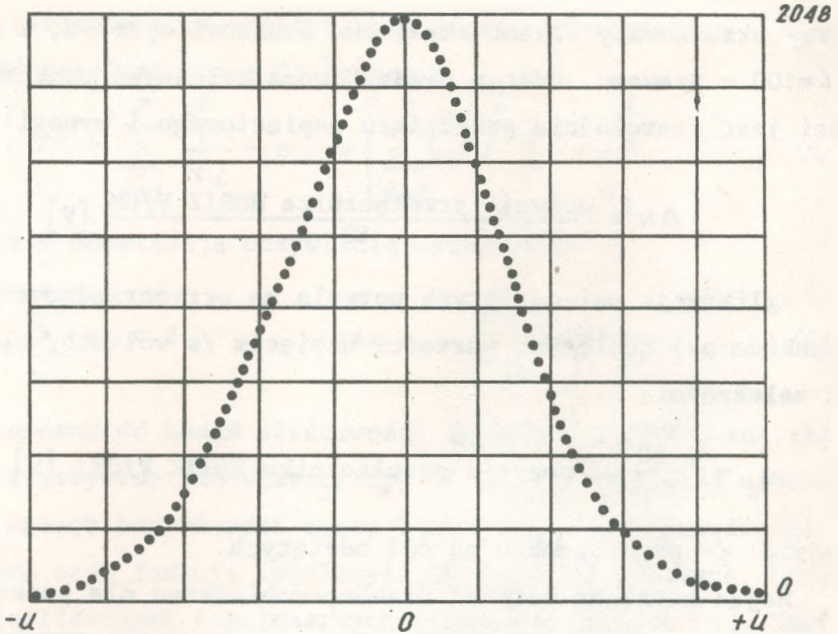
gdzie: i - numer punktu na osi odciętych.

Mając ustaloną wartość prawdopodobieństwa dla punktu o największym prawdopodobieństwie $p_{max}/u/$, można uzyskać proporcjonalną kalibrację wszystkich pozostałych wartości prawdopodobieństwa empirycznego p_{ei} zgodnie z zależnością:

$$p_{ei} = \frac{y_i}{2048} p_{max}/u/ = \frac{y_i}{\sum_{i=1}^{r=100} y_i} \quad /7/$$

Przy wyznaczaniu gęstości prawdopodobieństwa za pomocą korelatora HP 3721 A analiza sygnału zostaje zatrzymana, gdy zostanie pobranych N^{∞} próbek dla punktu o największym prawdopodobieństwie. Przykład wykresu gęstości prawdopodobieństwa uzyskanego za pomocą tego analizatora jest przedstawiony na str. 4.

Uzyskany wykres gęstości prawdopodobieństwa może służyć do przeprowadzenia testu zgodności rozkładu empirycznego z



Obraz oscyloskopowy gęstości prawdopodobieństwa dla sygnału o rozkładzie normalnym wartości napięć chwilowych

założonym rozkładem teoretycznym. Z uwagi na stosunkowo dużą liczebność próby uzyskiwaną z analizatora, znakomitym testem może być w tym przypadku test chi kwadrat Pearsona. Jednakże bezpośrednio przeprowadzenie testu na podstawie gęstości empirycznej jest niemożliwe. Należy dokonać pewnych przekształceń otrzymywanych danych i wyliczeń niezbędnych parametrów. Przede wszystkim należy wyznaczyć wartość odchylenia standardowego na podstawie wykresu gęstości prawdopodobieństwa.

Przyjmując zatem, że wartość napięcia odpowiadająca danemu punktowi o numerze " i " na osi odciętych jest wartością odpowiadającą środkowi przedziału o szerokości Δu można dokonać standaryzacji zmiennej losowej empirycznej.

Dla sygnałów, których wartość oczekiwana $E/U/ = 0$, estymator centralnego momentu rzędu drugiego wynosi [3]:

$$\hat{\mu}_2 = \sum_{i=1}^r u_i^2 p_{ei} \quad /8/$$

gdzie: u_i - środek i -tego przedziału,

p_{ei} - empiryczna częstość względna w i -tym przedziale.

Moment ten przy przyjętych założeniach można utożsamiać z wariancją w próbie. Stąd estymator odchylenia standardowego w próbie można wyrazić jako:

$$\hat{s} = \sqrt{\hat{\mu}_2} = \sqrt{\sum_{i=1}^r u_i^2 p_{ei}} = \sqrt{\sum_{i=1}^r u_i^2 \frac{y_i}{r=100}} = \sqrt{\frac{\sum_{i=1}^r u_i^2 y_i}{\sum_{i=1}^r y_i}} \quad /9/$$

Zatem standaryzowaną wartość zmiennej losowej - napięcia - można wyrazić za pomocą zależności:

$$z_i = \frac{U_i}{\hat{s}} \quad /10/$$

a standaryzowaną szerokość przedziału przez:

$$\Delta z = \frac{\Delta u}{\hat{s}} \quad /11/$$

Wykorzystując fakt, że częstość empiryczna $\frac{n_i}{n}$ jest równa prawdopodobieństwu empirycznemu p_{ei} , można podstawiając równanie /7/ do równania /2/ otrzymać wartość statystyki χ^2 zgodnie z zależnością:

$$\chi^2 = \sum_{i=1}^{r=100} y_i \left\{ \sum_{i=1}^{r=100} \left(\frac{y_i}{\sum_{i=1}^{r=100} y_i} \right)^2 \frac{1}{p_i} - 1 \right\} \quad /12/$$

przy czym:

$$n = \sum_{i=1}^{r=100} y_i \quad /13/$$

Występującą w równaniu /12/ częstość teoretyczną p_i cechy U w i -tym przedziale można wyznaczyć w oparciu o zależności definicyjne [3]. Mamy bowiem:

$$p_i = F/z_i + \frac{\Delta z}{2} - F/z_i - \frac{\Delta z}{2} \quad /14/$$

gdzie: $F/x/$ - dystrybuanta teoretyczna rozkładu normalnego w punkcie x .

Zatem ostatecznie wartość statystyki χ^2 wynosi:

$$\chi^2 = \sum_{i=1}^{r=100} y_i \left\{ \sum_{i=1}^{r=100} \left(\frac{y_i}{\sum_{i=1}^{r=100} y_i} \right)^2 \frac{1}{F/z_i + \frac{\Delta z}{2} - F/z_i - \frac{\Delta z}{2}} - 1 \right\} \quad /15/$$

Po wyliczeniu statystyki χ^2 według równania /15/ pozostaje tylko porównać ją z wartością statystyki $\chi^2_{m, \alpha}$ wziętej z tablic dla określonego poziomu istotności α i dla określonej liczby stopni swobody - m .

Dla prób o dostatecznie dużych liczebnościach, jak to ma miejsce w przypadku analizatora HP 3721 A, statystyka χ^2 ma

$m = r - c - 1$ stopni swobody, gdzie c - jest liczbą ocenianych parametrów [2]. Przy stosowaniu analizatora HP 3721 A założono E/u oraz wyznaczono \hat{S} , wobec tego liczba stopni swobody wynosi $m = 100 - 2 - 1 = 97$.

Przedstawiona modyfikacja testu χ^2 Pearsona umożliwia zastosowanie tego nieparametrycznego testu do weryfikacji hipotezy o rozkładzie wartości chwilowych napięć badanych sygnałów przy wykorzystaniu wyników pomiarów uzyskanych z analizatora HP 3721 A w postaci gęstości prawdopodobieństwa. Analizator ten jest wyposażony w przystawkę do wyprowadzania wyników na taśmie dziurkowanej, tak że test zgodności może być przeprowadzony całkowicie przy wykorzystaniu techniki komputerowej.

BIBLIOTEKA
Instytutu Łączności
Nr 5-8198

WYKAZ LITERATURY

1. Fisz M.: Rachunek prawdopodobieństwa i statystyka matematyczna PWN. Warszawa 1969.
2. Smirnow W.N., Dunin I.W., Borkowski A.M.: Kurs rachunku prawdopodobieństwa i statystyki matematycznej dla zastosowań technicznych. PWN Warszawa 1969.
3. PN-58/N-01051 Rachunek prawdopodobieństwa i statystyka matematyczna. WN Warszawa 1973.
4. Instrukcja eksploatacji analizatora HP 3721 A. IŁ Warszawa 1972.

BIBLIOTEKA
Instytutu Łączności
Nr _____

Biblioteka

Iz

S-8198