ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 1986

Seria: MATEMATYKA-FIZYKA z. 51 Nr kol. 874

Jan KAŁUSKI

Instytut Automatyki

Politechnika Śląska

NIECENTRALNY ROZKŁAD χ^2 I JEGO ZASTOSOWANIE

Streszczenie. Przy opracowywaniu danych doświadczalnych podstawową rolę odgrywają statystyki otrzymane jako określone funkcje zmiennych losowych wyników badań odpowiednio dobranej próbki losowej prostej.

W pracy omówiono rzadko stosowany niecentralny rozkład χ^2 oraz jego zastosowanie do opisu statystyk, otrzymanych przy badaniach zmiennych losowych normalnych o niezerowej wartości oczekiwanej oraz wariancji będącej dowolną liczbą dodatnią lub jednością. Wychodząc z ogólnie znanego centralnego rozkładu χ^2_n o n stopniach swobody, pokazano powstawanie niecentralnego rozkładu χ^2 i jego powiązania z innymi rozkładami. Podano literaturę, dotyczącą tego rozkładu.

W dalszej części pracy przedstawiono, zdaniem autora, kilka istotnych zastosowań niecentralnego rozkładu χ^2 do opisu statystyk próbkowych jedno- i wielowymiarowych. Oprócz uogólnienia szeroko znanej statystyki t-Studenta przedstawiono sposób oceny niecentralnego momentu zwykłego rzędu drugiego zmiennej losowej normalnej za pomocą niecentralnego rozkładu χ^2.

To ostatnie zastosowanie jest rozwijane przez autora od szeregu lat i dotyczy zagadnień z dziedziny metrologii teoretycznej.

1. Wprowadzenie

Niech dane są zmienne losowe $X_1, X_2, \ldots X_n$ o jednakowych rozkładach normalnych $N(0,1)$. Wówczas zmienna losowa

$$Z = \sum_{i=1}^{n} X_i^2, \quad (i=\overline{1,n}) \tag{1}$$

będzie miała rozkład centralny χ^2_n o n-stopniach swobody. Gęstość tego rozkładu ma postać:

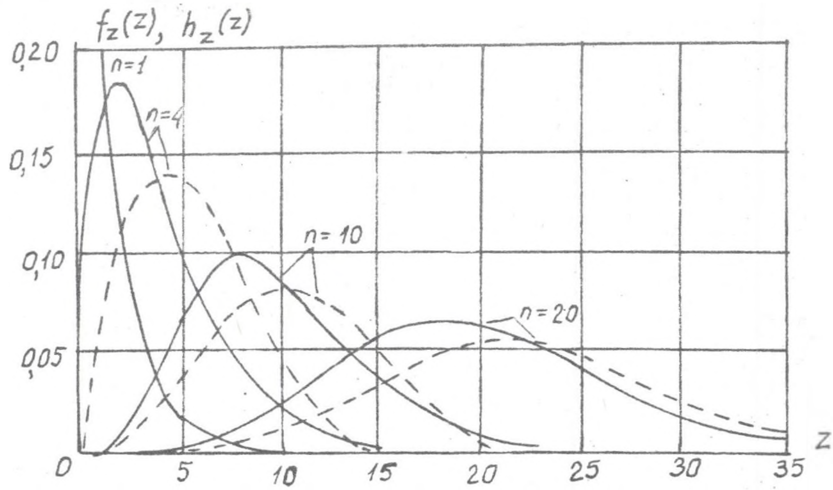
$$f_Z(z) = \begin{cases} 0 \quad \text{dla} \quad z \leqslant 0, \\ \dfrac{1}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot z^{\frac{n-2}{2}} \cdot e^{-\frac{z}{2}}, \quad z > 0 . \end{cases} \tag{2}$$

Jest ona szczególnym przypadkiem gęstości rozkładu gamma

$$f_z(z) = \begin{cases} 0 \quad \text{dla} \quad z \leqslant 0, \\ \frac{b^r}{\Gamma(r)} \cdot z^{r-1} \cdot e^{-bz}, \quad z > 0 \end{cases} \tag{3}$$

przy $b = \frac{1}{2}$ i $r = \frac{n}{2}$, gdzie n-liczba naturalna, zwana liczbą stopni swobody.

Przykładowy wykres gęstości rozkładu χ^2_n dla niektórych n pokazuje rys. 1 (linia ciągła).



Rys. 1. Przykłady gęstości rozkładu chi-kwadrat oraz niecentralnego rozkładu chi-kwadrat dla $\tau = 2,5$

Fig. 2. Examples of densities of chi-square and noncentral chi-square distributions for $\tau = 2,5$

Wykorzystanie tego rozkładu jest szeroko znane i nie będzie tu omawiane.

Niech teraz dane są zmienne losowe $X_1, X_2, \ldots, X_n$ o rozkładzie normalnych $N(m_i, \sigma)$, $(i=\overline{1,n})$. Wówczas zmienna losowa

$$z = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i^2 \tag{4}$$

ma rozkład niecentralny $\chi^2_n(\tau)$ o n-stopniach swobody i parametrze niecentralności

$$\tau = \frac{1}{\sigma^2} \sum_{i=1}^{n} m_i^2.$$

Rozkład ten ma gęstość:

$$h_z(z) = \frac{1}{2^{\frac{n}{2}} 6^n \sqrt{n}} e^{-\frac{1}{2}(\tau + \frac{z}{6^2})} . z^{\frac{n}{2} - 1} \times \sum_{j=1}^{\infty} \frac{(\tau . z)^j}{(2j)!} \frac{\Gamma(\frac{1}{2} + j)}{\Gamma(\frac{n}{2} + j)} \tag{5}$$

lub kładąc

$$B(\tfrac{n-1}{2}, \tfrac{j+1}{2}) = \frac{\Gamma(\frac{n-1}{2}) . \Gamma(\frac{j+1}{2})}{\Gamma(\frac{h-1}{2} + \frac{j+1}{2})} \tag{6}$$

otrzymamy inną postać tej gęstości:

$$h_z(z) = \frac{1}{2^{\frac{n}{2}} . \Gamma(\frac{n-1}{2}) 6^n \sqrt{\pi}} . e^{-\frac{1}{2}(\tau + \frac{z}{6^2})} . z^{\frac{n}{2} - 1} \times$$

$$\times \sum_{j=0}^{\infty} \frac{\tau^{\frac{j}{2}} . z^{\frac{j}{2}}}{j!} . B(\tfrac{n-1}{2}, \tfrac{j+1}{2}). \tag{7}$$

We wzorach (2-7) $\Gamma(p)$ oraz $B(p,q)$ są to odpowiednio funkcje gamma Eulera i beta Eulera.

Wyprowadzenie wzorów (5) i (7) dla przypadku gdy $X_i : N(m_i, 1)$, $i = \overline{1,n}$ zostało podane u Andersona [1] oraz FOURGEAUD'a i FUCHS'ego [3]. Bez wyprowadzenia wzór ten podaje Fisz [2].

Jeżeli $X = (X_1, X_2, \ldots, X_N)^T$ jest wektorem losowym w R^N o niezdegenerowanym rozkładzie normalnym N-wymiarowym $N(m, \Sigma)$ o wektorze wartości oczekiwanych $m = (m_1, \ldots, m_N)$ i macierzy wariancyjno-kowariancyjnej

$$\Sigma = [6_{ij}], \quad (i,j = 1,2,\ldots,N),$$

to zmienna losowa

$$Z = X^T \Sigma^{-1} X \tag{8}$$

ma [3] niecentralny rozkład $\chi_n^2(\tau)$ o n-stopniach swobody i parametrze niecentralności $\tau = m^T \Sigma^{-1} m$.

Dalej w pracy omówiono właściwości i zastosowanie niecentralnego rozkładu $\chi_n^2(\tau)$.

2. Właściwości niecentralnego rozkładu $\chi_n^2(\tau)$

Niecentralny rozkład $\chi_n^2(\tau)$ jest uogólnieniem centralnego rozkładu χ_n^2 dla przypadku, gdy wartości oczekiwane zmiennych losowych składowych X_i, $(i=\overline{1,n})$ o rozkładzie normalnym są różne od zera.

Dla porównania przykładowe gęstości prawdopodobieństwa zmiennej losowej Z, mającej niecentralny rozkład $\chi_n^2(\tau)$, pokazano na rys. 1 linię przerywaną, dla $n > 2$ i dużych τ. Wartość oczekiwana i wariancja zmiennej losowej o niecentralnym rozkładzie $\chi_n^2(\tau)$ wynoszą:

$$\left.\begin{aligned} E(Z) &\cong n + \tau \\ V(Z) &\cong 2(n+2\tau) \end{aligned}\right\} \tag{9}$$

Jeżeli $\tau \to 0$, to zmienna losowa $Z = \chi_n^2(\tau)$ dąży do rozkładu centralnego χ_n^2. Dla dużych τ rozkład $\chi_n^2(\tau)$ ma w przybliżeniu rozkład normalny $N((n+\tau), \sqrt{2(n+2\tau)})$. Jeżeli $\chi_n^2(\tau)$ ma niecentralny rozkład χ_n^2 o n stopniach swobody i parametrze niecentralności τ, to zmienna losowa

$$\chi^2 = \frac{n+2\tau}{n+3\tau}\left[\chi_n^2(\tau) + \frac{\tau^2}{n+3\tau}\right] \tag{10}$$

ma w przybliżeniu centralny rozkład χ_r^2 o

$$r = \frac{(n+2\tau)^3}{(n+3\tau)^2} \tag{11}$$

stopniach swobody.

Mniej dokładne przekształcenie daje, że zmienna losowa $\frac{n+\tau}{n+2}\chi_n^2(\tau)$ ma w przybliżeniu centralny rozkład χ_l^2 o

$$l = (n+\tau)^2/(n+2\tau) \tag{12}$$

stopniach swobody [6].

3. Zastosowanie niecentralnego rozkładu $\chi_n^2(\tau)$

Podamy teraz kilka istotnych zastosowań tego rozkładu.

1. Wykorzystując właściwości niecentralnego rozkładu $\chi_n^2(\tau)$ o n stopniach swobody i parametrze niecentralności τ, buduje się statystykę

$$\chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2, \tag{13}$$

która służy do weryfikacji hipotezy. $H : \tau = 0$ dla przypadku, gdy zmienne losowe $X_1, X_2, \ldots, X_n$ mają rozkład normalny $X_i : N(m_i, \sigma)$, $(i=\overline{1,n})$ o jednakowym znanym odchyleniu standardowym σ. Jeżeli hipoteza jest prawdziwa, to statystyka (13) ma centralny (zwykły) rozkład χ^2_n o n stopniach swobody.
Hipotezę H odrzuca się, gdy zaobserwowana wartość statystyki χ^2 jest większa lub równa pewnej wartości krytycznej $\chi^2(\alpha, n)$. Wówczas prawdopodobieństwo

$$P\left\{\chi^2_n(\tau) \geq \chi^2(\alpha, n)\right\}, \quad (0 < \alpha < 1),$$

jest funkcją mocy $\mu(\vec{\tau}(n,\alpha))$ testu χ^2.

W tablicach (patrz np. [6]) podaje się wartości funkcji $\tau/\mu(n,\alpha) = \tau$, dla których funkcja mocy $\mu/\tau(n,\alpha) = \mu$ $(0 < \mu < 1)$.

2. Jedno z najważniejszych zadań jednowymiarowej statystyki polega na znalezieniu oceny wartości oczekiwanej określonej zmiennej losowej, której odchylenie standardowe nie jest znane.
W tym przypadku wykorzystuje się znany rozkład t-Studenta i zbudowaną na jego podstawie statystykę t.

Jeżeli pobrano próbkę losową prostą z populacji o rozkładzie normalnym $N(m, \sigma)$, to wielkość

$$t = \sqrt{n}\,\frac{\bar{x} - m}{s} \tag{14}$$

ma rozkład t-Studenta o n-1 stopniach swobody. Na podstawie tej statystyki buduje się test, służący do weryfikacji hipotezy $H : m = m_0$, gdzie m_0 - wartość zadana lub buduje się przedział ufności dla nieznanej wartości m.

Wielowymiarowym analoganem kwadratu wielkości t według (14) jest wielkość:

$$T^2 = N(\bar{x}-m)^T S^{-1}(\bar{x}-m) \tag{15}$$

gdzie:

$\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N)^T$ jest wektorem wartości średnich otrzymanym z badaniem wektora losowego $X = (X_1, \ldots, X_N)^T$,

S - nieobciążona ocena macierzy wariancyjno-kowariancyjnej Σ.

Statystykę T^2 wykorzystuje się do sprawdzenia hipotezy o wektorze wartości oczekiwanych $m = (m_1, m_2, \ldots, m_N)$ w populacji o rozkładzie normalnym N-wymiarowym, a także w celu otrzymania przedziałów ufności dla nieznanej wartości wektora m.

Rozkład statystyki T^2 związany jest z niecentralnym rozkładem $\chi_n^2(\tau)$ w następujący sposób:

Jeżeli zmienna losowa Z ma niecentralny rozkład $\chi_n^2(\tau)$ oraz niezależna od niej zmienna losowa Y ma centralny rozkład χ_m^2 o m stopniach swobody, to zmienna losowa

$$F = \frac{Z/n}{Y/m} \tag{16}$$

ma niecentralny rozkład F-Fishera o gęstości [1]:

$$g_F(f) = \frac{n \cdot e^{-\frac{1}{2}\tau}}{m \cdot T(\frac{1}{2}m)} \sum_{j=0}^{\infty} \frac{(\tau/2)^j (n.f/m)^{\frac{1}{2}n+j-1} \Gamma\left[\frac{1}{2}(n+m)+j\right]}{j! \cdot \Gamma(\frac{1}{2}n+j)(1+n.f/m)^{\frac{1}{2}(n+m)+j}} \tag{17}$$

gdzie $f = \frac{z/n}{y/m}$ jest realizacją statystyki F.

Jeżeli $T^2 = N(\bar{x}-m_0)^T \cdot S^{-1}(\bar{x}-m_0)$, otrzymana na podstawie badania N-wymiarowej próbki, każda o liczności n, pochodzącej z populacji $N(m,\Sigma)$, to wielkość

$$(T^2/N-1)(N-n)/n \tag{18}$$

ma niecentralny rozkład F o parze stopni swobody (n, N-n) i parametrze niecentralności $\tau = N(m_0-m_0)^T \Sigma^{-1} \cdot (m-m_0)$.

Na podstawie gęstości statystyki F można otrzymać gęstość statystyki T^2 w postaci:

$$\varphi_{T^2}(t^2) = \frac{e^{-\frac{1}{2}\tau}}{(N-1)\Gamma\left[\frac{1}{2}(N-n)\right]} \sum_{j=0}^{\infty} \frac{(\tau/2)^j \left[(t^2)/(N-1)\right]^{\frac{1}{2}n+j-1} \Gamma(\frac{1}{2}N+j)}{j!\Gamma(\frac{1}{2}n+j)\left[1+(t^2)/(N-1)\right]^{\frac{1}{2}N+j}} \tag{19}$$

Wyprowadzenie tego wzoru można znaleźć u Andersona [1]. Tablice funkcji mocy testu F podaje Zieliński [6]. Prawdopodobieństwa przyjęcia hipotezy H : $\tau = 0$ (tzn. prawdopodobieństwo popełnienia błędu II rodzaju) dla różnych wartości τ i poziomów istotności 0,05 i 0,01 podał Tang [5].

3. Kolejny przykład zastosowania niecentralnego rozkładu $\chi_n^2(\tau)$ będzie dotyczył oceny momentu zwykłego (niecentralnego) rzędu drugiego zmiennej losowej normalnej. Z potrzebą takiej oceny spotykamy się zawsze tam, gdzie badana normalna zmienna losowa ma wartość oczekiwaną różną od zera. Na przykład, przy określaniu błędu pomiaru przyrządu, przy pomiarze konkretnej wartości wielkości mierzonej wykorzystujemy fakt, że wskazania przyrządu są na ogół obarczone losowym błędem wskazań, a otrzymane wartości

wskazań przy kilkukrotnym pomiarze są właśnie realizacjami pierwiastka z momentu rzędu drugiego zmiennej losowej wskazań przyrządu. Moment ten jest sumą kwadratu błędu przypadkowego (w postaci wariancji) i kwadratu wartości oczekiwanej zmiennej losowej wskazań. W tym przypadku interesuje nas zarówno ocena punktowa, jak i przedziałowa tego momentu.

Ocenę punktową momentu zwykłego rzędu drugiego otrzymuje się w prosty sposób, wykorzystując na przykład metodę momentów do budowy tego estymatora.

Jeżeli zmienna losowa Z ma rozkład normalny $N(m,\sigma)$, to interesujący nas moment definiuje się w następujący sposób:

$$m_2 = E(Z^2) = m^2 + \sigma^2 \tag{20}$$

Zbudowane oszacowanie punktowe tego momentu ma postać:

$$m_2^* = \frac{1}{n}\sum_{i=1}^{n} z_i, \quad (i=\overline{1,n}) \tag{21}$$

$$\hat{m}_2 = \frac{1}{n}\sum_{i=1}^{n} Z_i, \tag{22}$$

gdzie zmienna losowa Z_i, $(i=\overline{1,n})$ ma taki sam rozkład jak Z.

Wiadomo, że oszacowanie m_2^* jest zaobserwowaną w danym doswiadczeniu realizacją statystyki i jest oszacowaniem zgodnym i nieobciążonym momentu m_2 w populacji.

W celu oszacowania przedziału ufności dla nieznanej wartości m_2 wykorzystuje się powiązania statystyki $\hat{m}_2$ z niecentralnym rozkładem $\chi_n^2(\tau)$.

Jeżeli zmienna losowa Z_i, $(i=\overline{1,n})$ ma rozkład normalny $N(m,\sigma)$ to można wykazać, że [4] statystyka $\frac{1}{\sigma^2} n \cdot \hat{m}_2$ ma niecentralny rozkład $\chi_n^2(\tau)$ o n stopniach swobody i parametrze niecentralności $\tau = \frac{nm^2}{\sigma^2}$.

Dokonując standaryzacji Z_i statystykę $\hat{m}_2$ można zapisać w postaci:

$$\hat{m}_2 = \frac{1}{n}\sum_{i=1}^{n} (\sigma \cdot \eta_i + m)^2, \tag{23}$$

gdzie zmienna losowa η_i, $(i=\overline{1,n})$ ma rozkład $N(0,1)$.

Po podniesieniu do kwadratu i uporządkowaniu wyrazów oraz uwzględnieniu, że zmienna losowa χ_n^2 o rozkładzie centralnym χ^2 i zmienna losowa normalna standaryzowana, są niezależne, można otrzymać następnie rozkład statystyki $\hat{m}_2/m_2$ w postaci:

$$\hat{m}_2/m_2 = \frac{1}{n} a \cdot \chi^2_n + \frac{2}{\sqrt{n}} b \cdot U + c, \tag{24}$$

gdzie

$$a = \frac{\sigma^2}{\sigma^2 + m^2},$$

$$b = \frac{m \cdot \sigma}{\sigma^2 + m^2}, \tag{25}$$

$$c = \frac{m^2}{\sigma^2 + m^2}$$

a U jest zmienną losową $N(0,1)$.

Mając statystykę $\hat{m}_2/m_2$, można dla każdego poziomu ufności β, $(0<\beta<1)$ przy χ^2_n i U napisać następującą równość:

$$P\left\{\hat{m}_2 \leqslant m_2 \left(\frac{1}{n} a \cdot \chi^2_{n,\beta} + \frac{2}{\sqrt{n}} b \cdot U_\beta + c\right)\right\} = \beta^2. \tag{26}$$

Dowód tej nierówności został naszkicowany w pracy [4].

We wzorze (26) $\chi^2_{n,\beta}$ i U_β są kwantylami rzędu β zmiennych losowych χ^2_n i U.

Ze wzorów (21) i (26) można już bezpośrednio otrzymać wzory na granicy dwustronnego przedziału ufności dla nieznanej wartości m_2, na poziomie ufności β^2, w postaci:

$$m_{2,dd} = \frac{m_2^*}{\frac{1}{n} a \chi^2_{n,(1+\beta)/2} + \frac{2}{\sqrt{n}} b \cdot U_{(1+\beta)/2} + c} \tag{27}$$

dla granicy $m_{2,dd}$-dwustronnej dolnej oraz

$$m_{2,dg} = \frac{m_2^*}{\frac{1}{n} a \chi^2_{n,(1-\beta)/2} + \frac{2}{\sqrt{n}} b \cdot U_{(1-\beta)/2}} \tag{28}$$

dla granicy $m_{2,dg}$-dwustronnej górnej.

LITERATURA

[1] Anderson T.: An introduction to multivariete statistical analysis (tłumaczenie ros.). Moskwa 1963.

[2] Fisz M.: Rachunek prawdopodobieństwa i statystyka matematyczna. Wydanie IV, PWN, Warszawa 1969.

[3] Fourgraud C., Fucsh A.: Statistique. Wydanie II, Dunod, Paris 1972.

[4] Kałuski J.: Integralny wskaźnik dokładności wybranej klasy przyrządów pomiarowych. Praca doktorska. Gliwice 1977.

[5] Tang P.C.: The power function of the analysis of variance test with tables and illustrations of their use. Stat. Res. Mem. 2/1938.

[6] Zieliński R.: Tablice statystyczne. PWN, Warszawa 1972.

Recenzent: Doc. dr hab. inż. Janusz Karpiński

Wpłynęło do Redakcji w kwietniu 1985

НЕЦЕНТРАЛЬНОЕ χ^2 РАСПРЕДЕЛЕНИЕ И ЕГО ПРИМЕНЕНИЕ

Р е з ю м е

Статистики полученные как определённые функции случайных величин играют основную роль при обработке экспериментальных данных.

В работе оговорено редко употребляемое нецентральное χ^2 распределение в качестве описания статистик полученных при исследованиях нормальных случайных величин, имеющих отличное от нуля математическое ожидание и любую положительную дисспересию (в частности - равную единицы). Выходя из общеизвестного центрального χ^2 распределения с n-степенями свободы, показано центральное χ^2 распределение и его связи с другими распределениями. Дана литература касающаяся этого распределения.

В дальнейшем в работе приводится несколько существенных, по мнению автора, практических применений нецентрального χ^2 распределения для описания одно и многомерных статистик. Кроме обобщения широко известной статистики t-Студента, даныметод оценки нецентрального момента второго порядка нормальной случайной величины, выходя из нецэнтрального χ^2 распределения. Это последнее применение развивается автором уже в течении нескольких лет для метрологических приложений.

NONCENTRAL CHI-SQUARE DISTRIBUTION AND ITS APPLICATION

S u m m a r y

Statistics, i.e. some specific functions of random variables play important role in the processing of given data, obtained from experiments.

In the paper the noncentral chi-square distribution is discussed, that is of statistics not often use, as well as its applications for description. The statistics are obtained in an analysis of Gaussian random variables having non-zero expected values and having any positive number (for example var = 1) as their variances. Starting from the well - known

central chi-square distribution of n degrees of greedom, it is derived the noncentral distribution and it is shown its correlation with another known distributions. The relevant bibliography concerning the distribution, is given.

Then a few substantial (according to the author's opinion) applications of the noncentral distribution for the sample statistic, one - and many - dimensional, are presented. Besides the generalization of the wide known t-Student statistics, a way of an estimation of the noncentral second - order moment is given for the Gausian random variable, using the noncentral chi-square distribution. The last problem is being developed by the author for some years, for metrological applications.