

A. PLUCIŃSKA (Warszawa)

O PEWNEJ UOGÓLNIONEJ POSTACI FUNKCJI GĘSTOŚCI I JEJ ZASTOSOWANIU DO BADANIA ROZKŁADU OPORU OPORNIKÓW

Funkcja, która jest tematem niniejszej pracy, została skonstruowana dla opisu pewnych zjawisk, nazwijmy je A oraz B, będących przedmiotem badań teorii jakości produkcji.

A. Elementy pochodzące z masowej produkcji są często dzielone na klasy o różnej dokładności ze względu na pewną określoną cechę X tych elementów. Załóżmy, że wyjściową populacją jest populacja normalna $N(m, w)$. Celem dokonywanej selekcji jest podział tej populacji na n części takich, że dla każdej z części jest, z prawdopodobieństwem bardzo bliskim jedności, spełniony warunek

$$|X - m| < \delta,$$

δ zaś jest to pewna z góry dana liczba.

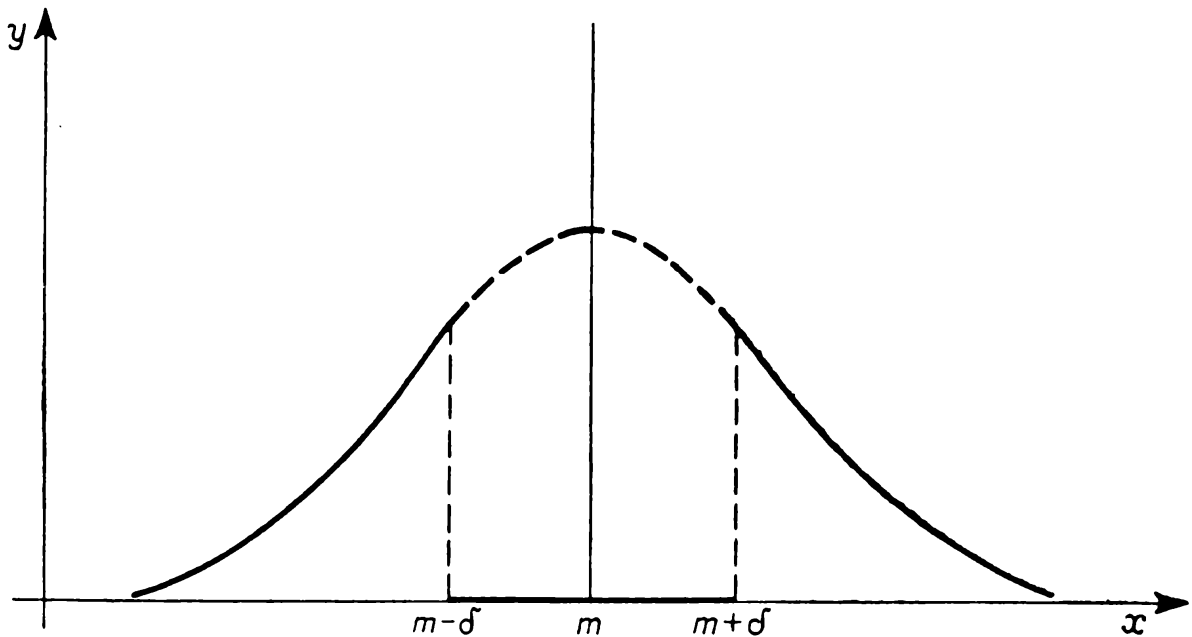
Podział populacji można scharakteryzować za pomocą rozkładów uciętych; ilustruje to rys. 1 dla przypadku $n = 2$. Operowanie rozkładami uciętymi jest analitycznie niewygodne przy badaniu własności funkcji zmiennych losowych, a ponadto postuluje bezbłędną pracę automatu segregującego. W pracach [1] i [2] została omówiona funkcja, za pomocą której wygodnie jest charakteryzować populacje wyselekcjonowane, czyli powstałe z podziału na części, wyjściowej populacji na ogół normalnej. Funkcja ta, napisana tu w nieco ogólniejszej postaci, dana jest wzorem

$$(1) \quad f(x; a, d, p) = \frac{p}{2\Gamma\left(\frac{d}{p}\right) a^d} |x|^{d-1} e^{-\left|\frac{x}{a}\right|^p}, \quad -\infty < x < +\infty,$$

przy czym parametry a, d, p spełniają warunki $a > 0, d \geq 1, p > 0$.

Wykres funkcji (1) jest symetryczny względem prostej $x = 0$. Za pomocą tej funkcji możemy aproksymować rozkład odpowiadający kon-

krétnej sytuacji praktycznej, a mianowicie polegającej na tym, że środek m wyjściowej populacji pokrywa się z środkiem x_0 nastawienia automatu segregującego. Środek nastawienia automatu segregującego jest zmienną losową i można się spodziewać, że wartość oczekiwana tej zmiennej losowej jest równa wartości oczekiwanej wyjściowej populacji. Jednakże dla poszczególnych serii produkcji obserwuje się nieraz duże różnice mię-



Rys. 1. Funkcja gęstości uciętego rozkładu normalnego

dzy wartościami tych zmiennych losowych. Po selekcji rozkład badanej cechy jest niesymetryczny. Wówczas możemy stosować aproksymację za pomocą funkcji

$$(2) \quad g(x; c, C, a, A, d, D, p, P) = \frac{2c}{c+C} f(x; A, D, P) \eta(-x) + \\ + \frac{2C}{c+C} f(x; a, d, p) \eta(x),$$

gdzie funkcja f jest określona wzorem (1), a $\eta(x)$ jest funkcją skoku jednostkowego daną wzorem

$$\eta(x) = \begin{cases} 1 & \text{dla } x > 0, \\ \frac{1}{2} & \text{dla } x = 0, \\ 0 & \text{dla } x < 0. \end{cases}$$

Funkcję (2) można traktować jako funkcję gęstości rozkładu prawdopodobieństwa, bowiem

$$1^\circ \quad g(x) \geq 0,$$

$$2^\circ \quad \int_{-\infty}^{+\infty} g(x) dx = \frac{2c}{c+C} \int_{-\infty}^0 f(x; A, D, P) dx + \frac{2C}{c+C} \int_0^{\infty} f(x; a, d, p) dx = \\ = \frac{2c}{c+C} \cdot \frac{1}{2} + \frac{2C}{c+C} \cdot \frac{1}{2} = 1.$$

B. Funkcja (1) lub (2) może być przydatna również przy badaniu zagadnienia mieszaniny rozkładów, czyli do opisu populacji powstałej np. przez połączenie dwu populacji o różnych środkach m_1, m_2 . Jeśli zostały połączone dwie populacje normalne $N(m_1, w_1), N(m_2, w_2)$, gdzie $m_1 \neq m_2$, to rozkład badanej cechy w tak powstałej populacji jest rozkładem dwumodalnym. Jeśli $w_1 = w_2$, to wydaje się celowe korzystanie ze wzoru (1); jeśli $w_1 \neq w_2$, to przypuszczalnie lepszą aproksymacją będzie wzór (2). Oczywiście stosowanie którejkolwiek z tych postaci wzorów wymaga sprawdzenia odpowiednim testem statystycznym zgodności rozkładu rozważanej populacji empirycznej z proponowanym rozkładem teoretycznym. Szczegółowy opis zagadnienia mieszaniny rozkładów, ujęty od strony praktycznej, podany jest w [3]

Zauważmy, że nadając pewne konkretne wartości parametrom we wzorze (2), otrzymujemy różne znane rozkłady prawdopodobieństwa.

Niech $c = 0$, wówczas $g(x)$ jest uogólnionym rozkładem gamma rozważanym w [4]. Szczególnymi przypadkami tego rozkładu są rozkłady gamma, chi, chi-kwadrat, wykładniczy, Weibulla.

Niech $a = A, p = P = 2, d = D = 1, c = C$, wówczas $g(x)$ jest funkcją gęstości rozkładu normalnego.

Niech $a = A, p = P = 2, d = D = 3, c = C$, wówczas $g(x)$ jest funkcją gęstości rozkładu Maxwella.

Nasuwa się pytanie, czy celowe jest konstruowanie funkcji zależnej od wielu parametrów, ale jednocześnie zawierającej jako szczególne przypadki różne znane rozkłady prawdopodobieństwa. Odpowiedź na to pytanie składa się z dwu części:

1) pewne występujące w praktyce rozkłady są bardzo nieregularne i do ich opisu trzeba wykorzystywać funkcje zależne od wielu parametrów, czego dowodem jest przykład liczbowy rozważany w dalszym ciągu niniejszej pracy;

2) jeśli okaże się możliwe podanie pewnych własności wspólnych dla wielu rozkładów, to tym samym można zrezygnować z badania tychże własności dla każdego rozkładu oddzielnie, co niewątpliwie przemawia za celowością konstruowania takiego ogólnego rozkładu.

Podajmy prosty przykład takiego ogólnego ujęcia zagadnienia.

Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi, z których każda ma rozkład określony funkcją rodziny (2). Oznaczmy

funkcję gęstości zmiennej losowej $X_i (i = 1, 2, \dots, n)$ przez $g_i(x)$, gdzie

$$g_i(x) = g(x; c_i, C_i, a_i, A_i, d_i, D_i, p_i, P_i).$$

Wiadomo, że wygodnym aparatem do badania rozkładu zmiennej losowej

$$Y = X_1^{a_1} \cdot X_2^{a_2} \cdot \dots \cdot X_n^{a_n},$$

gdzie a_1, a_2, \dots, a_n pewne stałe rzeczywiste takie, by Y miało sens liczbowy w dziedzinie rzeczywistej, jest funkcja Mellina. Oznaczmy przez $h_{|Y|}(s)$ funkcje Mellina odpowiednio zmiennych losowych $|Y|$ i $|X_i|^{a_i}$, wówczas

$$\begin{aligned} (3) \quad h_{|Y|}(s) &= E[|Y|^s] = h_{|X_1|^{a_1}} \cdot \dots \cdot h_{|X_n|^{a_n}} = \\ &= \prod_{i=1}^n \int_0^{\infty} |x|^{sa_i} \left[\frac{2c_i}{c_i + C_i} f(x; A_i, D_i, P_i) + \right. \\ &\quad \left. + \frac{2C_i}{c_i + C_i} f(x; a_i, d_i, p_i) \right] dx = \prod_{i=1}^n \left[\frac{c_i}{c_i + C_i} \cdot \frac{A_i^{sa_i} \Gamma\left(\frac{a_i s + D_i}{P_i}\right)}{\Gamma\left(\frac{D_i}{P_i}\right)} + \right. \\ &\quad \left. + \frac{C_i}{c_i + C_i} \frac{a_i^{sa_i} \Gamma\left(\frac{a_i s + d_i}{p_i}\right)}{\Gamma\left(\frac{d_i}{p_i}\right)} \right], \end{aligned}$$

gdzie

$$\Gamma(s) = \lim_{m \rightarrow \infty} \frac{m! m^s}{s(s+1) \cdot \dots \cdot (s+m)} \quad (1).$$

Ze wzoru (3) otrzymujemy wzory na momenty m_{2r} rzędu parzystego $2r$ zmiennej losowej Y , a mianowicie

$$\begin{aligned} (4) \quad m_{2r} = h_{|Y|}(2r) &= \prod_{i=1}^n \left[\frac{c_i}{c_i + C_i} \frac{A_i^{2r+a_i} \Gamma\left(\frac{2ra_i + D_i}{P_i}\right)}{\Gamma\left(\frac{D_i}{P_i}\right)} + \right. \\ &\quad \left. + \frac{C_i}{c_i + C_i} \frac{a_i^{2ra_i} \Gamma\left(\frac{2ra_i + d_i}{p_i}\right)}{\Gamma\left(\frac{d_i}{p_i}\right)} \right]. \end{aligned}$$

(1) Dla $\text{Re } s > 0$ funkcję $\Gamma(s)$ możemy również określać wzorem

$$\Gamma(s) = \int_0^{\infty} e^{-t} t^{s-1} dt;$$

występująca we wzorze całka jest zbieżna w półpłaszczyźnie $\text{Re } s > 0$.

Momenty m_{2r+1} rzędu nieparzystego $2r+1$ określone są wzorami

$$(5) \quad m_{2r+1} = \prod_{i=1}^n \left[(-1)^{(2r+1)\alpha_i} \frac{c_i}{c_i + C_i} \cdot \frac{A_i^{(2r+1)\alpha_i} \Gamma\left(\frac{(2r+1)\alpha_i + D_i}{P_i}\right)}{\Gamma\left(\frac{D_i}{P_i}\right)} + \frac{C_i}{c_i + C_i} \cdot \frac{a_i^{(2r+1)\alpha_i} \Gamma\left(\frac{(2r+1)\alpha_i + d_i}{p_i}\right)}{\Gamma\left(\frac{d_i}{p_i}\right)} \right].$$

Znaleziono zostały wzory na momenty dowolnego rzędu zmiennej losowej Y . Podaliśmy więc przykład polegający na określeniu pewnej własności łącznie dla wszystkich rozkładów określonych funkcją gęstości (2).

Przykład liczbowy. Podamy teraz przykład liczbowy aproksymacji za pomocą funkcji rodziny (2) pewnego zaobserwowanego rozkładu empirycznego.

Zbadano próbkę składającą się z $N = 100$ oporników pyrolitycznych OWS $1k\Omega, \pm 10\%$, $2W^{(2)}$. Wyniki pomiarów podane są w tabelicy 1, gdzie przez n_k oznaczono liczby oporników, dla których wartość liczbową oporu zawarta jest w przedziale o numerze k ($k = 1, 2, \dots, 20$).

TABLICA 1

Numer k przedziału	Przedział	Liczebność n_k
1	$[-100, -90)$	0
2	$[-90, -80)$	5
3	$[-80, -70)$	4
4	$[-70, -60)$	12
5	$[-60, -50)$	28
6	$[-50, -40)$	29
7	$[-40, -30)$	11
8-14	$[-30, +40]$	0
15	$(40, 50]$	1
16	$(50, 60]$	1
17	$(60, 70]$	2
18	$(70, 80]$	5
19	$(80, 90]$	2
20	$(90, 100]$	0

(²) Pomiarzy były wykonane w Katedrze Konstrukcji Telekomunikacyjnych Wydz. Łączności Politechniki Warszawskiej.

Przesuwamy układ współrzędnych równoległe w prawo tak, by początek znalazł się w punkcie $R = 1000$ (rys. 2). Wysuwamy hipotezę, że rozkład jest rozkładem rodziny (2), gdzie dla uproszczenia przyjmujemy $p = P = 2$. Wyznaczamy najpierw wartości parametrów c, C . Oznaczmy przez N_1 liczbę elementów, dla których wartość liczbowa badanej cechy jest nie większa od zera, a przez N_2 liczbę elementów, dla których wartość liczbowa badanej cechy jest większa od zera. Zatem $N_1 = 89, N_2 = 11$, oczywiście $N_1 + N_2 = N = 100$. Przyjmujemy $c = N_1, C = N_2$.

Pozostałe parametry wyznaczymy, badając oddzielnie gałęzie wykresu odpowiadające ujemnym i dodatnim wartościom argumentu. Zadanie sprowadza się do wyznaczenia na podstawie zbioru danych empirycznych parametrów funkcji

$$(6) \quad y = f(x; a, d) = \frac{2}{\Gamma\left(\frac{d}{2}\right) a^d} |x|^{d-1} e^{-\frac{x^2}{a^2}}, \quad x < 0$$

(badamy oddzielnie próbkę o liczebności N_1 i oddzielnie próbkę o liczebności N_2).

Nakładamy warunki, aby moment zwykły rzędu drugiego i moda rozkładu empirycznego były odpowiednio równe momentowi rzędu drugiego i modzie rozkładu teoretycznego. Ze względów rachunkowych wydaje się, że są to najprostsze warunki, jakie możemy wykorzystać do wyznaczenia parametrów. Warto zauważyć, że warunek równości momentów rzędu drugiego jest warunkiem wynikającym z zastosowania do rozważanej funkcji metody największej wiarogodności. Mamy bowiem

$$\begin{aligned} L &= f(x_1; a, d) \cdot f(x_2; a, d) \cdot \dots \cdot f(x_n; a, d) = \\ &= \left(\frac{2}{\Gamma\left(\frac{d}{2}\right) a^d} \right)^n |x_1 \cdot x_2 \cdot \dots \cdot x_n|^{d-1} e^{-\frac{1}{a^2} (x_1^2 + x_2^2 + \dots + x_n^2)}. \end{aligned}$$

Stąd

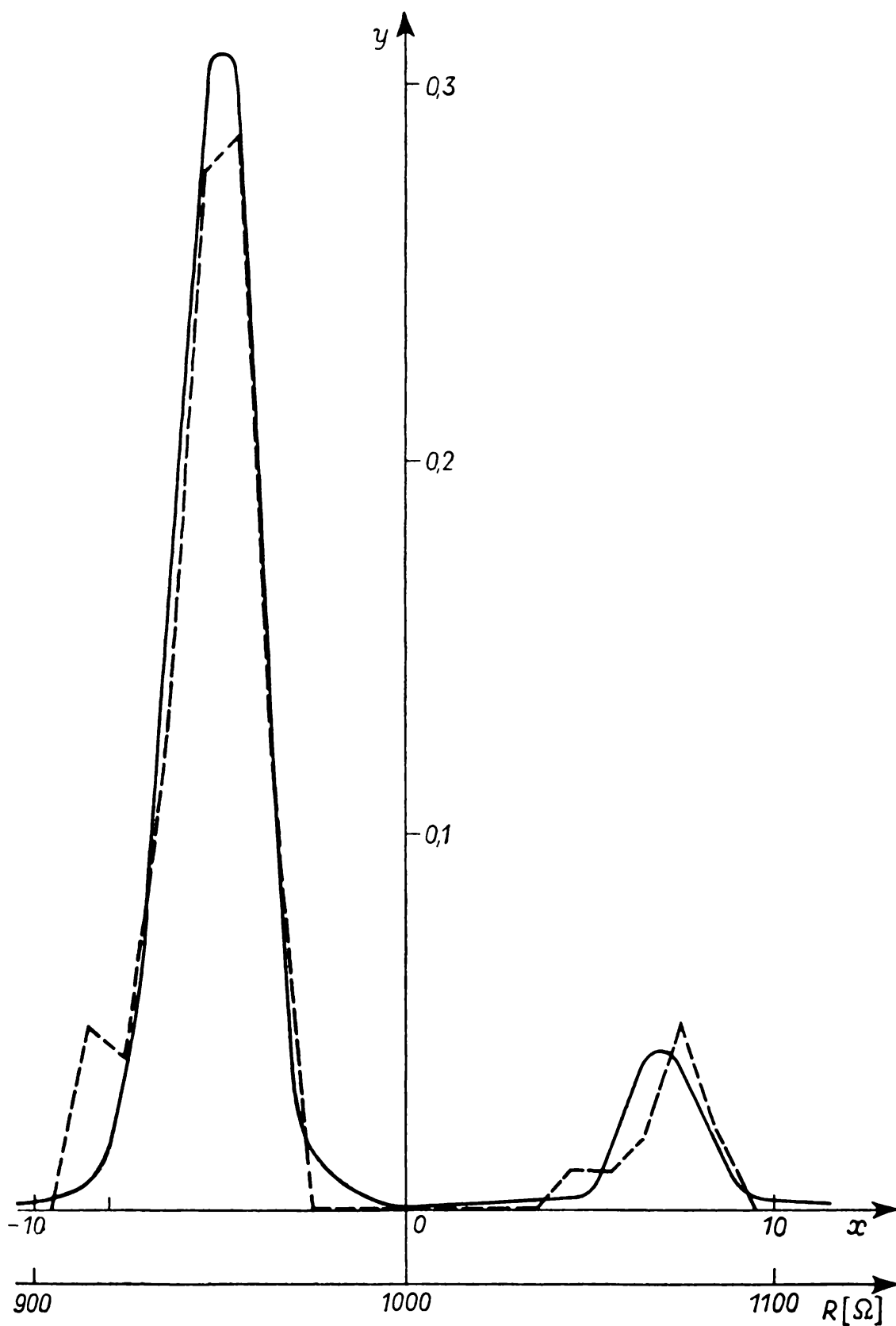
$$\frac{\partial \log L}{\partial a} = -\frac{nd}{a} + \frac{2}{a^3} (x_1^2 + x_2^2 + \dots + x_n^2).$$

Z warunku

$$\frac{\partial \log L}{\partial a} = 0$$

wynika, że

$$\frac{a^2 d}{2} = \frac{1}{n} \sum_{i=1}^n x_i^2,$$



Rys. 2. Wielobok częstości rozkładu empirycznego i odpowiadająca mu funkcja gęstości rozkładu teoretycznego określonego wzorem (9)

a $\frac{1}{2}a^2d$ jest właśnie wartością momentu rzędu drugiego, określoną przez wzór (4).

Moda rozkładu teoretycznego m_0 jest to wartość odciętej, dla której funkcja (6) przyjmuje maksimum. Obliczając pochodną tej funkcji i przyrównując ją do zera, znajdujemy, że

$$m_0 = -a\sqrt{\frac{1}{2}(d-1)}.$$

Z danych liczbowych rozważanego przykładu wynika, że

$$\frac{1}{2}a^2d = 27,66, \quad -a\sqrt{\frac{1}{2}(d-1)} = -5,$$

czyli przyjmujemy $a^2 = 5,32$, $d = 10,4$.

Wysuwamy hipotezę, że rozkład odpowiadający ujemnej części osi ma postać

$$(7) \quad y = 2f(x; \sqrt{5,32}; 10,4; 2)\eta(-x).$$

(funkcja (7) ma ekstremum w punkcie o odciętej $m_0 \approx -5$).

Sprawdzamy tę hipotezę testem statystycznym lambda dla prób o małej liczebności, dla którego odpowiednie tablice podane są w [5].

Oznaczmy przez $F(x)$ dystrybuantę teoretyczną, przez $F_n(x)$ dystrybuantę empiryczną; niech

$$D_n = \sup_x |F(x) - F_n(x)|.$$

Wartości dystrybuanty teoretycznej $F(x)$ obliczamy, korzystając z tablic niekompletnej funkcji gamma [6]. Mamy bowiem

$$F(x) = 2 \int_{-\infty}^x f(x; a, d, 2) dx = 1 - \frac{1}{\Gamma\left(\frac{d}{2}\right)} \int_{x^2/a^2}^{+\infty} y^{\frac{d}{2}-1} e^{-y} dy.$$

Wartości $F(x)$ i $F_{N_1}(x)$ zestawione są w tabelicy 2.

TABLICA 2

k	-8	-7	-6	-5	-4	-3
$F_{N_1}(k)$	0,056	0,101	0,236	0,550	0,876	1
$F(k)$	0,008	0,058	0,217	0,531	0,841	0,982

Z tabelicy 2 widać, że największą wartością zmiennej losowej D_{N_1} jest

$$F_{N_1}(-7) - F(-7) = 0,043.$$

Przyjmijmy poziom istotności $\alpha = 0,1$. Z tablic w cytowanej pracy [5] odczytujemy, że

$$P(D_{N_1} \geq 0,043) \approx 0,99 \gg 0,1.$$

Test ten nie prowadzi więc do odrzucenia hipotezy (7).

Rozważając gałąź wykresu odpowiadającą dodatnim wartościom argumentu i odpowiadającą jej próbkę o liczebności $N_2 = 11$ znajdujemy analogiczną metodą, że $d \approx 25,6$ i $a^2 \approx 4$. Wysuwamy hipotezę, że rozkład odpowiadający dodatniej części osi ma postać

$$(8) \quad y = 2f(x; 2; 25,6; 2)\eta(x).$$

(funkcja (8) przyjmuje ekstremum w punkcie $m_0 \approx 7$).

Wartości dystrybuanty empirycznej i teoretycznej dla tego rozkładu podane są w tabelicy 3.

TABLICA 3

k	5	6	7	8	9
$F_{N_2}(k)$	0,091	0,182	0,364	0,818	1
$F(k)$	0,013	0,137	0,470	0,821	0,968

Z tabelicy tej wynika, że największą wartością zmiennej losowej D_{N_2} jest

$$|F_{N_2}(7) - F(7)| = 0,106,$$

natomiast

$$P(D_{N_2} \geq 0,106) \approx 0,9 \gg 0,1.$$

Test nie prowadzi więc do odrzucenia hipotezy.

W wyniku przeprowadzonego testu istotności możemy uzyskać dwa orzeczenia: 1) należy odrzucić wysuwaną hipotezę, 2) test nie prowadzi do odrzucenia wysuwanej hipotezy. W przypadku gdy test nie prowadzi do odrzucenia wysuwanej hipotezy, stosuje się nieraz dodatkowo inne testy dla potwierdzenia słuszności uzyskanych wyników. W rozważanym przykładzie wydaje się to tym bardziej celowe, że na podstawie próbki były wyznaczane parametry funkcji, a dopiero potem była sprawdzana hipoteza postaci rozkładu.

Aby dodatkowo sprawdzić uzyskane wyniki, zastosujemy znów test lambda, ale tym razem w odniesieniu do całej próbki, tzn. w odniesieniu do próbki o liczebności $N = N_1 + N_2 = 100$ i hipotezy

$$(9) \quad y = g(x; 89; 11; \sqrt{5,32}; 2; 10,4; 25,6; 2; 2),$$

czyli łącznie badamy obie części wykresu odpowiadające ujemnej i dodatniej części osi.

Dane liczbowe dotyczące wartości odpowiednich dystrybuant zestawione są w tabelicy 4.

TABLICA 4

k	-8	-7	-6	-5	-4	-3	0	5	6	7	8	9
$F_{100}(k)$	0,05	0,09	0,21	0,49	0,78	0,89	0,890	0,90	0,91	0,93	0,98	1
$F(k)$	0,007	0,052	0,193	0,473	0,748	0,874	0,890	0,891	0,905	0,942	0,98	0,991

Z tabelicy 4 wynika, że największą wartością zmiennej losowej D_{100} jest

$$F(-7) - F_{100}(-7) = 0,038.$$

Z tablic odczytujemy, że

$$P(D_{100} \geq 0,038) \approx 0,99 \gg 0,1.$$

Test nie prowadzi więc do odrzucenia hipotezy (9).

Na rys. 2 podane są wykresy funkcji gęstości rozkładu teoretycznego określonego wzorem (9) i odpowiedniego wieloboku częstości rozkładu empirycznego.

Należy zaznaczyć, że w rozważonym przykładzie weryfikowaliśmy hipotezę nieparametryczną, przy czym rozkład hipotetyczny zależał od 8 parametrów, które były estymowane w oparciu o dane empiryczne.

Oczywiście im większa jest próbka, tym mocniejszy jest test. Często jednak praktycznie nie można zebrać liczniejszego materiału empirycznego i trzeba wyciągać wnioski na podstawie dostępnego zbioru obserwacji.

Próbka o liczebności 100 w rozważanym zagadnieniu jest próbką o małej liczebności. Dlatego też wnioskowanie statystyczne zostało oparte na teście statystycznym λ dla prób o małej liczebności.

Gdybyśmy z tej samej populacji, z której została pobrana nasza 100-elementowa próbka, pobrali próbkę o większej liczebności, to pokazałyby się pewne różnice w oszacowaniach wartości parametrów, ale metoda ich szacowania byłaby ta sama, test nieparametryczny natomiast byłby oparty na rozkładzie granicznym zmiennej losowej D_n .

Prace cytowane

- [1] A. Plucińska, *O pewnych zagadnieniach związanych z podziałem populacji normalnej na części*, Zastosow. Mat. 8 (1965), str. 117-125.
 [2] A. Plucińska, *Probabilistyczna analiza błędów przy połączeniach typu szeregowego i równoległego*, Arch. Elektrotechn. 14 (1965), str. 583-595.

[3] П. А. Бородачев, *Основные вопросы теории точности производства*, Москва 1950, rozdz. I i II.

[4] E. W. Stacy, *A generalization of the gamma distribution*, Ann. Math. Stat. 33 (1962), str. 1187-1192.

[5] Z. W. Birnbaum, *Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size*, J. Am. Stat. Assoc. 47 (1952), str. 425-441.

[6] В. И. Пагурова, *Таблицы неполной гамма-функции*, Москва 1963.

Praca wpłynęła 22. 4. 1965

А. ПЛЮЦИНЬСКА (Варшава)

**ОБ ОДНОЙ ОБОБЩЁННОЙ ФОРМЕ ФУНКЦИИ ПЛОТНОСТИ
И ЕЁ ПРИМЕНЕНИЮ К ИССЛЕДОВАНИЮ РАСПРЕДЕЛЕНИЯ
СОПРОТИВЛЕНИЯ РЕОСТАТОВ**

РЕЗЮМЕ

Функцию (2), о которой говорится в работе, сконструировано с целью описания генеральной совокупности, возникшей из разделения нормальной генеральной совокупности на части.

Функция (2) является обобщением некоторых известных функций распределения вероятности. Придавая некоторые конкретные значения параметрам, получаем в качестве частных случаев распределения: гамма, хи, хи-квадрат экспоненциальное, Вейбулла, нормальное, Максвелла.

Дается метод — простой при вычислениях — с помощью которого можно на основании множества эмпирических данных найти параметры функции (2). Дается пример исследования согласия распределения сопротивления реостатов с кривой семейства (2) на основании пробы численностью $N = 100$.

А. PLUCIŃSKA (Warszawa)

**ON A GENERAL FORM OF THE PROBABILITY DENSITY FUNCTION AND ITS
APPLICATION TO THE INVESTIGATION OF THE DISTRIBUTION OF RHEOSTAT
RESISTANCE**

SUMMARY

The subject of this paper, the function (2), has been constructed for purposes of describing populations obtained from a division of normal populations into parts. The function (2) is a generalization of some known probability density functions. Certain parameter values yield known distributions, as e.g. the Gamma, Chi, Chi-square, exponential, Weibull, normal, and Maxwell distributions.

A non-complicated method of estimating the parameters of function (2) on the basis of empirical data is indicated. Given is also an example of testing the goodness of fit of the distribution of rheostat resistance and a suitable function of type (2).