J. ODERFELD (Warszawa)

O SKUPIENIU ROZKŁADU

§ 1. Skupienie a wartość użytkowa. Niechaj X będzie zmienną losową ciągłą, jednowymiarową, określoną w zbiorze S, który jest sumą przedziałów $\langle d_{i-1}, d_i)$, $i = 1, 2, \ldots, N$. Oznaczmy gęstość prawdopodobieństwa zmiennej losowej X przez $\varphi(x)$.

Utwórzmy wyrażenia

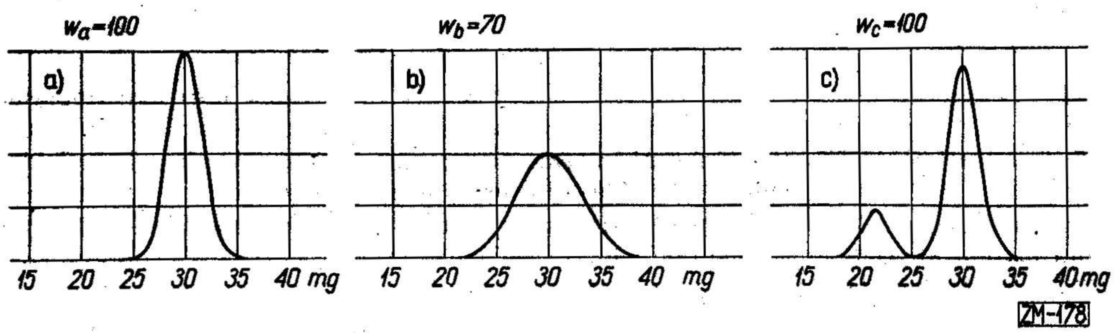
$$w = \Big(\int_{(S)} \varphi^2(x)\,dx\Big)^{1/2} \tag{1}$$

$$v_i = \Big(\int_{d_{i-1}}^{d_i} \varphi^2(x)\,dx\Big)^{1/2}. \tag{2}$$

Wyrażenie w będziemy nazywali *skupieniem względnym*, a wyrażenie v_i *skupieniem częściowym*.

Przedmiotem artykułu są własności wyrażeń w i v_i. Przed ich omówieniem damy pewną interpretację, która w gruncie rzeczy stała się punktem wyjścia do badania ścisłego.

Na rysunku 1 pokazano trzy warianty a, b i c rozkładów X i przy każdym napisano, ile wynosi skupienie względne wyrażone w umownych jednostkach tak dobranych, żeby $w_a = 100$. Mamy $w_b = 70$, $w_c = 100$.



Rys. 1

Liczby te pokazują, że w jest tym większe, im — *caeteris paribus* — rozkład jest bardziej skupiony wokół wartości modalnej, lub wartości modalnych, jeśli jest ich kilka.

Przypuśćmy[1] teraz, że rysunek 1 przedstawia rozkład ciężaru ziarn zboża przeznaczonego do siewu. Otóż jednym z kryteriów kwalifikujących ziarno do siewu jest mała dyspersja ciężaru. A więc ziarno o rozkładzie 1b ma mniejszą wartość użytkową niż ziarno o rozkładzie 1a.

Wiadomo również, że ziarno o rozkładzie 1c łatwiej można rozsortować niż 1b, a więc znowu ziarno 1b ma mniejszą wartość użytkową niż 1c.

Jest więc — przynajmniej w rozważanym przykładzie — dość wyraźna jednokierunkowa zależność między skupieniem względnym a wartością użytkową.

Na te związki zwrócił uwagę już w 1954 r. H. Steinhaus [1]. Ponadto w grudniu 1955 r. podał on (bez dowodu) tezę twierdzenia 1, które dowodzę w niniejszej pracy i postawił jako temat (w postaci szczególnej) zadanie, którego ogólnym rozwiązaniem jest twierdzenie 2.

§ 2. Skupienie względne a odchylenie średnie. Wzrokowe porównanie rysunków 1a i 1b nasuwa przypuszczenie, że istnieje jakaś relacja między skupieniem względnym w a odchyleniem średnim σ. Oczywiście nie ma żadnej zależności ogólnej, istnieje bowiem nieskończenie wiele rozkładów o różnych σ i tym samym w. Jednakże są pewne zależności, prawdziwe w przypadkach szczególnych, ale często spotykanych w praktyce.

Jeśli zmienna losowa X ma gęstość prawdopodobieństwa $\varphi(x)$, to zmienna losowa $Y = ax+b$ ma gęstość

$$\psi(y) = \frac{1}{|a|}\varphi\left(\frac{y-b}{a}\right).$$

Wstawiając gęstość $\varphi(x)$ i $\psi(y)$ po prawej stronie (1) znajdujemy, że między skupieniami w_x a w_y jest zależność

$$w_y = \frac{1}{\sqrt{|a|}}\, w_x.$$

Jak wiadomo, $\sigma_y = |a|\sigma_x$, więc

$$w_y\sqrt{\sigma_y} = w_x\sqrt{\sigma_x} = K. \tag{3}$$

A więc w przekształceniu liniowym zmiennej losowej wyrażenie $w\sqrt{\sigma}$ pozostaje stałe.

---

[1] Aby uniknąć ewentualnego nieporozumienia, zastrzegamy się, że we wszystkich przykładach podanych w tej pracy staraliśmy się o poprawność techniczną tylko w takim zakresie, jaki był każdorazowo potrzebny do zilustrowania pewnej koncepcji matematycznej.

Łatwy rachunek daje następujące wartości stałej K:

w rozkładzie prostokątnym $K_1 = 2^{-1/2}3^{-1/4} \sim 0{,}537$,

„ trójkątnym $K_2 = \frac{1}{3}\cdot 2^{3/4} \sim 0{,}561$,

„ normalnym $K_3 = 2^{-1/2}\pi^{-1/4} \sim 0{,}531$.

Weźmy jeszcze pod uwagę symetryczny rozkład beta o gęstości

$$\varphi(x) = \frac{\Gamma(2p)}{\Gamma^2(p)}[x(1-x)]^{p-1} \qquad (0 < x < 1).$$

Rozkład ten gra dość dużą rolę w badaniu automatycznej produkcji metalowej. Znajdujemy zależność

$$w\sqrt{\sigma} = \frac{\Gamma(2p)}{\Gamma^2(p)} \cdot \frac{\Gamma(2p-1)}{\sqrt{\Gamma(4p-2)}} \cdot \frac{1}{\sqrt{2}\sqrt[4]{2p+1}} = K_4(p). \tag{4}$$

W poniższej tabliczce spisano kilka wartości (z dokładnością do trzeciego znaku) funkcji $K_4(p)$:

p	2	3	5	10	20	50	∞
$K_4(p)$	0,518	0,520	0,523	0,527	0,529	0,530	0,531

§ 3. Wpływ podziału na skupienie względne.

TWIERDZENIE 1. *Niechaj $\varphi(x)$, $\varphi_1(x)$ i $\varphi_2(x)$ będą gęstościami prawdopodobieństwa i niechaj*

$$\varphi(x) = c\varphi_1(x) + (1-c)\varphi_2(x), \tag{5}$$

gdzie c jest ustalone i spełnia podwójną nierówność $0 \leqslant c \leqslant 1$. Niechaj skupienie względne w będzie określone wzorem (1).

Wtedy

$$w \leqslant \max(w_1, w_2). \tag{6}$$

Znak równości w (6) *zachodzi tylko wtedy, gdy*

$$\varphi(x) = \varphi_1(x) = \varphi_2(x). \tag{7}$$

Dowód([2]).

$$\int_{(S)} \varphi^2(x)dx = \int_{(S)} [c\varphi_1(x)+(1-c)\varphi_2(x)]^2 dx = \tag{8}$$

$$= c^2\int_{(S)} \varphi_1^2(x)dx + (1-c)^2\int_{(S)}\varphi_2^2(x)dx + 2c(1-c)\int_{(S)}\varphi_1(x)\varphi_2(x)dx.$$

([2]) Mój dowód w pierwszej wersji tej pracy opierał się, jak obecnie, na nierówności Schwarza, ale był dłuższy; skrócenie zawdzięczam S. Zubrzyckiemu, który ponadto zauważył, że twierdzenie 1 stosuje się również do rozkładów skokowych, ze stosownymi zmianami formalnymi.

Ale z nierówności Schwarza [2] wynika, że

$$\int_{(S)} \varphi_1(x)\varphi_2(x)\,dx \leqslant \Big(\int_{(S)} \varphi_1^2(x)\,dx\Big)^{1/2}\Big(\int_{(S)} \varphi_2^2(x)\,dx\Big)^{1/2}$$

Wskutek tego

$$\int_{(S)} \varphi^2(x)\,dx \leqslant c^2 \int_{(S)} \varphi_1^2(x)\,dx + (1-c)^2 \int_{(S)} \varphi_2^2(x)\,dx + \\ + 2c(1-c)\Big(\int_{(S)} \varphi_1^2(x)\,dx\Big)^{1/2}\Big(\int_{(S)} \varphi_2^2(x)\,dx\Big)^{1/2},$$

czyli

$$\Big(\int_{(S)} \varphi^2(x)\,dx\Big)^{1/2} \leqslant c\Big(\int_{(S)} \varphi_1^2(x)\,dx\Big)^{1/2} + (1-c)\Big(\int_{(S)} \varphi_2^2(x)\,dx\Big)^{1/2},$$

a więc — wobec (1) —

(9) $$w \leqslant cw_1 + (1-c)w_2.$$

Przy założeniach przyjętych co do c prawa strona (9) jest niewiększa niż $\max(w_1, w_2)$, co stanowi dowód pierwszej części twierdzenia 1.

Ponieważ w nierówności Schwarza równość zachodzi wtedy i tylko wtedy, gdy $w_1 = w_2 = w$, więc prawdziwa jest druga część twierdzenia 1.

Uwaga. Z warunków, w których zachodzi nierówność Schwarza, wynika, że twierdzenie 1 jest prawdziwe nie tylko dla rozkładów ciągłych, ale również skokowych, ze stosownymi zmianami formalnymi.

Twierdzenie 1 ma następującą interpretację:

Przypuśćmy, że każdemu przedmiotowi pewnej partii można przyporządkować pewną realizację cechy X. Jeśli tę partię podzielimy w dowolny sposób, to na ogół otrzymamy, przynajmniej w jednej z części, skupienie względne *większe* niż przed podziałem. W szczególnym przypadku ($w_1 = w_2$) może ono być *niemniejsze* niż przed podziałem; taki przypadek może zajść wtedy, gdy partię podzielono przez losowanie i gdy zarówno partia, jak jej obie części, są duże.

Nasuwa się pytanie, jakie warunki należałoby dołączyć do (5), aby otrzymać tezę mocniejszą od (6), mianowicie tezę, że

(10) $$w < \min(w_1, w_2).$$

Przypuśćmy, że obszar S całkowania jest przedziałem $\langle d_0, d_2)$ zawierającym punkt d_1 i że

(11) $$\varphi_1(x) = \begin{cases} \varphi(x)/c & \text{dla} \quad d_0 < x < d_1, \\ 0 & \text{dla} \quad x < d_0 \text{ oraz } x \geqslant d_1; \end{cases}$$
$$\varphi_2(x) = \begin{cases} \varphi(x)/(1-c) & \text{dla} \quad d_1 < x < d_2, \\ 0 & \text{dla} \quad x < d_1 \text{ oraz } x \geqslant d_2. \end{cases}$$

W naszej interpretacji posługującej się partią przedmiotów znaczy to, że podzielono partię na dwie części rozłączne według cechy X.

Wobec (11) jest

$$\int_{(S)} \varphi_1(x)\varphi_2(x)\,dx = 0 \tag{12}$$

i wzór (8) przyjmuje postać

$$\int_{(S)} \varphi^2(x)\,dx = c^2\int_{(S)} \varphi_1^2(x)\,dx+(1-c)^2\int_{(S)} \varphi_2^2(x)\,dx,$$

czyli

$$w^2 = c^2w_1^2+(1-c)^2w_2^2. \tag{13}$$

Jeśli zaś w jest określone przez (13), to nierówność (10) jest spełniona, gdy

$$\frac{1-c}{1+c} < \left(\frac{w_1}{w_2}\right)^2 < \frac{2-c}{c}. \tag{14}$$

Ostatecznie więc, warunkami *dostatecznymi* mocniejszej tezy (10) są łącznie warunki (11) i (14).

Nie zająłem się poszukiwaniem warunków koniecznych i dostatecznych, w których zachodzi (10), pragnę jednak zwrócić uwagę, że już sama klasa przypadków spełniających warunki (11) i (14) jest obszerna i ważna dla praktyki. Mianowicie dzieląc partię o rozkładzie cechy X prostokątnym, trójkątnym lub normalnym na dwie podpartie rozłączne według cechy X, mamy $w_1 > w$ oraz $w_2 > w$. Ponieważ dowody tych własności są elementarne, poprzestanę na zanotowaniu, że dzieląc partię o rozkładzie normalnym i gęstości prawdopodobieństwa

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \qquad (-\infty < X < \infty),$$

na część I, taką że $X \leqslant d$, i część II, taką że $X > d$, mamy

$$\frac{w_1}{w} = \sqrt{\frac{1}{2}+\Theta\left(\frac{d-\mu}{\sigma/\sqrt{2}}\right)} : \left[\frac{1}{2}+\Theta\left(\frac{d-\mu}{\sigma}\right)\right],$$

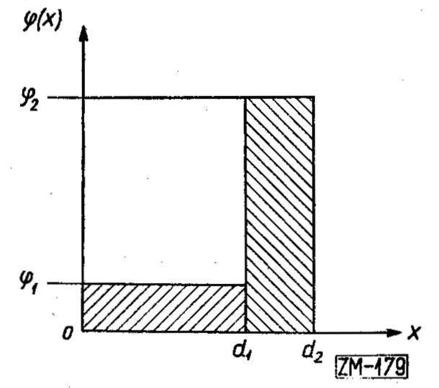
$$\frac{w_2}{w} = \sqrt{\frac{1}{2}-\Theta\left(\frac{d-\mu}{\sigma/\sqrt{2}}\right)} : \left[\frac{1}{2}-\Theta\left(\frac{d-\mu}{\sigma}\right)\right],$$

gdzie $\Theta(\ldots)$ oznacza funkcję Laplace'a.

Łatwo zauważyć, że wtedy zawsze $w_1/w > 1$ oraz $w_2/w > 1$. Gdy d rośnie od μ do ∞, to w_1/w maleje monotonicznie od $\sqrt{2}$ do 1, a w_2/w rośnie monotonicznie od $\sqrt{2}$ do ∞. Dla $d \to \infty$, asymptotyczna wartość w_2/w jest równa $\sqrt{\pi}\sqrt{(d-\mu)/\sigma}$.

Pospolitość przykładów, w których zachodzi mocniejsze twierdzenie (10), może nasunąć przypuszczenie, że zachodzi ono zawsze. Tak jednak nie jest, a nawet bywa tak, że prawdziwość twierdzenia (10) zależy nie tylko od typu rozkładu i sposobu podziału, ale również od relacji metrycznych.

Dla ilustracji rozważmy następujący przykład.



Rys. 2

Niechaj zmienna losowa X w partii ma gęstość jak na rysunku 2. Podzielmy partię na dwie części, tak żeby

$$\text{w części I} \quad X \in \langle 0, d_1),$$
$$\text{w części II} \quad X \in \langle d_1, d_2).$$

Mamy

$$w = \frac{\sqrt{d_1\varphi_1^2+(d_2-d_1)\varphi_2^2}}{d_1\varphi_1+(d_2-d_1)\varphi_2}, \quad w_1 = \frac{1}{\sqrt{d_1}}, \quad w_2 = \frac{1}{\sqrt{d_2-d_1}}.$$

Żeby zachodziły jednocześnie nierówności $w < w_1$ i $w < w_2$, potrzeba i wystarcza, żeby

$$\frac{d_2-d_1}{d_1}+2\frac{\varphi_1}{\varphi_2} > 1 \quad \text{oraz} \quad \frac{d_1}{d_2-d_1}+2\frac{\varphi_2}{\varphi_1} > 1.$$

A więc na przykład, jeśli $d_2 = 2d_1$, to twierdzenie (10) jest prawdziwe dla wszelkich wartości φ_1 i φ_2. Jeśli natomiast $(d_2-d_1)/d_1 = \varphi_1/\varphi_2 = 0{,}1$, to twierdzenie (10) nie jest prawdziwe.

Twierdzenie 1 można łatwo uogólnić na podział partii na N części. Dzieląc najpierw partię na 2 części stwierdzamy, że zgodnie z (6), $w \leqslant \max(w_1, w_2)$. Dzieląc teraz część I na części I′ i I″, mamy $w_1 \leqslant \max(w_1', w_1'')$, a więc $w \leqslant \max(w_1', w_1'', w_2)$ itd. Ogólnie:

$$w \leqslant \max(w_i) \quad (i = 1, 2, \ldots, N).$$

A więc dzieląc partię w *dowolny* sposób na N części, mamy przynajmniej w jednej z części skupienie względne nie mniejsze niż przed podziałem.

§ 4. Podział według największej sumy skupień częściowych. Niechaj zmienna losowa X określona w obszarze S ma gęstość prawdopodobieństwa $\varphi(x)$ i niechaj obszar S będzie sumą przedziałów $\langle d_{i-1}, d_i)$, $i = 1, 2, \ldots, N$.

TWIERDZENIE 2. *Jeśli* v_i *jest określone przez* (2), *to* $\sum_1^N v_i = \max$, *gdy*

$$v_1 = v_2 = \ldots = v_N. \tag{15}$$

Dowód. Z warunku koniecznego

$$\frac{\partial}{\partial d_i} \sum_1^N v_i = 0 \quad (i = 1, 2, \ldots, N-1)$$

wynika, że

$$\frac{1}{2v_i^2}\varphi^2(d_i) - \frac{1}{2v_{i+1}^2}\varphi^2(d_i) = 0 \quad (i = 1, 2, \ldots, N-1),$$

czyli

$$v_i^2 = v_{i+1}^2 \quad (i = 1, 2, \ldots, N-1),$$

co jest równoznaczne z (16).

Pomijamy banalny dowód, że ekstremum tak znalezione jest maksimum.

WNIOSEK. *Wielkości* d_i *są kwantylami rzędu* N *w rozkładzie o gęstości prawdopodobieństwa* $\propto \varphi^2(x)$. *W przypadku szczególnym, gdy* $N = 2$, d_i *jest medianą.*

Twierdzenie 2 ma następującą interpretację:

Jeśli partię należy podzielić na $N-1$ części i jeśli wartość i-tej części jest proporcjonalna do i-tego skupienia częściowego, to największą wartość łączną uzyskuje się przy podziale według warunków (15).

Dla ilustracji przypuśćmy, że dana jest partia złożona z 1000 klocków; rozkład ich ciężarów jest normalny ze średnią m gramów i odchyleniem średnim $\sigma = 10$ gramów. Klocki są przeznaczone do montażu sprzęgieł odśrodkowych. Z warunków pracy wynika, że komplety 5-cio sztukowe przeznaczone do jednego sprzęgła powinny mieć możliwie zbliżone ciężary. Z warunków technicznych wynika, że nie opłaca się dobierać indywidualnie klocków do takich kompletów, ale że wystarczy całą partię podzielić na 4 części i wybierać na chybił trafił po 5 klocków z jednej części. Jak należy podzielić partię?

Przyjmujemy, że wartość części partii jest proporcjonalna do skupień częściowych. Ma więc zastosowanie twierdzenie 2. Z wniosku z tego twierdzenia wynika, że punkty podziału powinny być kwartylami w rozkładzie normalnym $N(m, \sigma/\sqrt{2})$, czyli że

$$a_1 = m - \frac{0{,}674}{\sqrt{2}}\sigma, \quad a_2 = m, \quad a_3 = m + \frac{0{,}674}{\sqrt{2}}\sigma.$$

A więc powinno być:

w części I około 315 klocków o ciężarach do $m-4{,}8$ g,

w części II „ 185 klocków o ciężarach od $m-4{,}8$ do m g,

w części III „ 185 klocków o ciężarach od m do $m+4{,}8$ g,

w części IV „ 315 klocków o ciężarach powyżej $m+4{,}8$ g.

Twierdzenie 2 pozostaje w mocy, jeśli zamiast (15) rozważymy wyrażenie $\left(\int\limits_{d_{i-1}}^{d_i}\varphi^k(x)\,dx\right)^{1/k}$. Okazuje się, że optymalne punkty podziału są kwartylami w rozkładzie o gęstości prawdopodobieństwa $\propto\varphi^k(x)$. Może to być użyteczne w praktyce, gdyż swoboda w doborze wykładnika k pozwala na dobre powiązanie wartości ze skupieniem.

Można pójść jeszcze dalej w uogólnieniach, ale nie jest to ani ciekawe matematycznie, ani nie wydaje się użyteczne w praktyce.

Prace cytowane

[1] H. Steinhaus, *Über einige prinzipielle Fragen...*, Berlin, 1956, Deutscher Verlag d. Wissenschaften.

[2] H. Cramér, *Mathematical methods of statistics*, Princeton, 1946, str. 88.

INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK

Praca wpłynęła 21. 2. 1956

Я. ОДЕРФЕЛЬД (Варшава)

О СОСРЕДОТОЧЕНИИ РАСПРЕДЕЛЕНИЯ

РЕЗЮМЕ

Пусть случайная переменная величина X непрерывная, одномерна и определена в области S, которая является суммой интервалов $\langle d_{i-1}, d_i)$ $(i = 1, 2, \ldots, N)$. Обозначим плотность вероятности через $\varphi(x)$. Составим формулы

$$w = \left(\int\limits_{(S)}\varphi^2(x)\,dx\right)^{1/2}, \qquad v_i = \left(\int\limits_{d_{i-1}}^{d_i}\varphi^2(x)\,dx\right)^{1/2}.$$

Первая формула называется *сосредоточением односительным*, вторая — *сосредоточением частичным*.

Показано связь между w и средним отклонением. Доказано две теоремы.

I. ***Разделяя произвольно совокупность получаем по крайней мере в одной её части относительное сосредоточение не меньшее чем до разделения.***

Рассмотрено несколько примеров разделения, среди которых одно такое, что в каждой части сосредоточение относительное больше чем до разделения.

II. *Разделяя партию на части не пересекающиеся относительно признака X получаем наибольшую сумму частичных сосредоточений, когда точки разделения являются квантилями в распределении $\propto f^2(x)$.*

Приводятся некоторые обобщения этих теорем.

Выяснено связь между параметрами w и v и значением партии товара и приведено несколько примеров применения.

J. ODERFELD (Warszawa)

ON THE CONCENTRATION OF DISTRIBUTION

SUMMARY

Let X be a continuous one-dimensional random variable, defined in region S which is the sum of the intervals $\langle d_{i-1}, d_i)$, $i = 1, 2, \ldots, N$. Let us denote the density of probability by $\varphi(x)$. Let us form the expressions

$$w = \Big(\int\limits_{(S)} \varphi^2(x)\,dx\Big)^{1/2} \quad \text{and} \quad v_i = \Big(\int\limits_{d_{i-1}}^{d_i} \varphi^2(x)\,dx\Big)^{1/2}.$$

The first of them will be called relative concentration, and the second partial concentration.

The relation between w and the standard deviation is shown.

Two theorems are proved. Theorem 1 states that *by an arbitrary division of a population we obtain in at least one part of the population a relative concentration that is not smaller than before the division.* Several particular cases of division are considered, among them a case in which the relative concentration in each part is greater than before the division.

Theorem 2 states *that by dividing a lot into parts that are disjoint with respect to the characteristic X we obtain the greatest sum of partial concentrations when the points of division are quantiles in the distribution $\propto f^2(x)$.*

Some generalizations of both theorems are given.

In the part containing applications the author explains the relation between parameters w and v and the value of a lot of the product and gives a few examples of application.
