ZASTOSOWANIA MATEMATYKI
VI (1961)

W. RUDZKI (Warszawa)

DOKŁADNOŚĆ OCENY WŁAŚCIWOŚCI ŚREDNIEJ PRODUKTÓW BEZKSZTAŁTNYCH

1. W pracach [1], [2] i [3] rozpatrzono zagadnienie szacowania właściwości średnich towarów bezkształtnych zawartych w opakowaniach o kształcie dowolnej bryły geometrycznej. Rozważania dotyczyły takich właściwości, których wartość w każdym punkcie Q wewnątrz opakowania zależy od położenia tego punktu w bryle. Wartość właściwości w punkcie Q jest funkcją odległości od pewnego punktu P wewnątrz bryły na dowolnej półprostej łączącej punkt P z powierzchnią bryły. Do właściwości takich należą między innymi: wilgotność, intensywność zapachu, zawartość pewnych składników itp.

Za miarę odległości przyjęto w pracy [2] stosunek r długości odcinka PQ do długości odcinka łączącego punkt P z powierzchnią bryły na tej samej prostej, a więc $0 \leqslant r \leqslant 1$.

Przyjęto dalej, że wartość właściwości w jest na każdej półprostej, wychodzącej z punktu P, określona przez tę samą funkcję $w = f(r)$. Przyjęto również upraszczające założenie (nie wypowiedziane zresztą wyraźnie w cytowanych pracach), że w jest zmienną rzeczywistą. W gruncie rzeczy mamy tu do czynienia ze zmienną losową W i przyjmujemy, że jej wartość oczekiwana w danym punkcie jest równa w. Faktycznie więc zależność $w = f(r)$ jest zależnością między wartością oczekiwaną zmiennej losowej W a odległością r od punktu P. Powierzchnie stałej wartości w są jednokładne z powierzchnią ograniczającą bryłę ze środkem jednokładności w P. Średnia wartość $\overline{w}$ badanej właściwości jest określona wzorem

$$\overline{w} = 3\int_0^1 r^2 f(r)\,dr. \tag{1}$$

Z zależności $w = f(r)$ wyznaczyć można taką odległość $\bar{r}$, w której $w = \overline{w}$. Dla oszacowania średniej $\overline{w}$ nasuwa się więc najprostszy przepis postępowania polegający na pobieraniu próbki w odległości $\bar{r}$.

Rozważania te można zastosować nie tylko do badania towarów bezkształtnych w opakowaniach, ale również towarów ułożonych, bądź zsypanych luzem w pewne bryły geometryczne, jak stożki, pryzmy, sterty, kopce itp.

W przypadku towaru w opakowaniu prostopadłościennym środek jednokładności P jest geometrycznym środkiem opakowania. W przypadku towarów zsypanych luzem, gdy podłoże podstawy nie wpływa na zmianę badanej właściwości, środek (lub środki) jednokładności może znajdować się w jednym lub więcej punktów podstawy. Ponieważ jednak towary zsypywane luzem tworzą zwykle bryły o symetrycznych przekrojach pionowych, środek jednokładności w tym przypadku pokrywa się ze środkiem geometrycznym podstawy.

Nie zmniejszając ogólności rozważań, można przyjąć, że $f(0) = 1$ i $f(1) = 0$. Przyjmujemy zależność między wartością właściwości w a odległością r typu

$$w = f(r) = 1 - r^m, \tag{2}$$

gdzie m może być dowolną wartością w przedziale $\langle 1, \infty)$.

Przyjęcie tego typu zależności wydaje się praktycznie w wielu przypadkach uzasadnione (patrz praca [2]).

Jak wykazano w pracy [2], w przypadku zależności wyrażonej wzorem (2) zachodzi

$$\overline{w} = \frac{m}{m+3} \tag{3}$$

oraz

$$\overline{r} = \left(\frac{3}{m+3}\right)^{1/m}. \tag{4}$$

Ponieważ wartości $\overline{r}$ dla małych m niewiele różnią się między sobą i są bliskie 0,75, a więc bliskie wartości $\overline{r}$ odpowiadającej $m = 1$, w cytowanych pracach [1], [2] i [3] do oszacowania wartości $\overline{w}$ zalecono pobieranie próbek w odległości $\overline{r} = 0{,}75$. Teoretycznie do oszacowania $\overline{w}$ wystarczy jedna próbka z opakowania bądź towaru luzem.

2. Jest rzeczą oczywistą, że pobieranie próbki w odległości $\overline{r} = 0{,}75$ od środka bryły da bezbłędne teoretycznie oszacowanie średniej $\overline{u}$ tylko w przypadku, gdy naprawdę $m = 1$, dla pozostałych zaś m oszacowanie będzie obarczone błędem. Zajmijmy się obecnie znalezieniem wielkości tego błędu dla różnych m.

Za miarę błędu w ocenie średniej $\overline{w}$ przyjmijmy różnicę między wartością funkcji (2) w punkcie $\overline{r} = 0{,}75$ a wartością średniej $\overline{w}$ obliczoną z wzoru (3) dla danego m. Oznaczmy ten błąd przez Δw_m, a więc

$$\Delta w_m = f(0{,}75) - \overline{w};$$

korzystając z (2) i (3) po przekształceniach otrzymujemy

$$\Delta w_m = \frac{3}{m+3} - 0{,}75^m. \tag{5}$$

Zauważmy, że gdy $m = 1$, wówczas $\Delta w_1 = 0$, dla $m > 1$ zachodzi $\Delta w_m > 0$ oraz $\lim\limits_{m\to\infty} \Delta w_m = 0$. Stąd wynika, że istnieje przynajmniej jedno takie m, dla którego Δw_m przyjmuje maksimum. Z charakteru zaś funkcji (5) wynika, że ma ona tylko jedno ekstremum w przedziale $(1, \infty)$.

Ponieważ m jest ciągłe, Δw_m osiąga maksimum w punkcie $m = m_0$, w którym

$$\frac{d\Delta w_m}{dm} = 0.$$

Po zróżniczkowaniu otrzymujemy

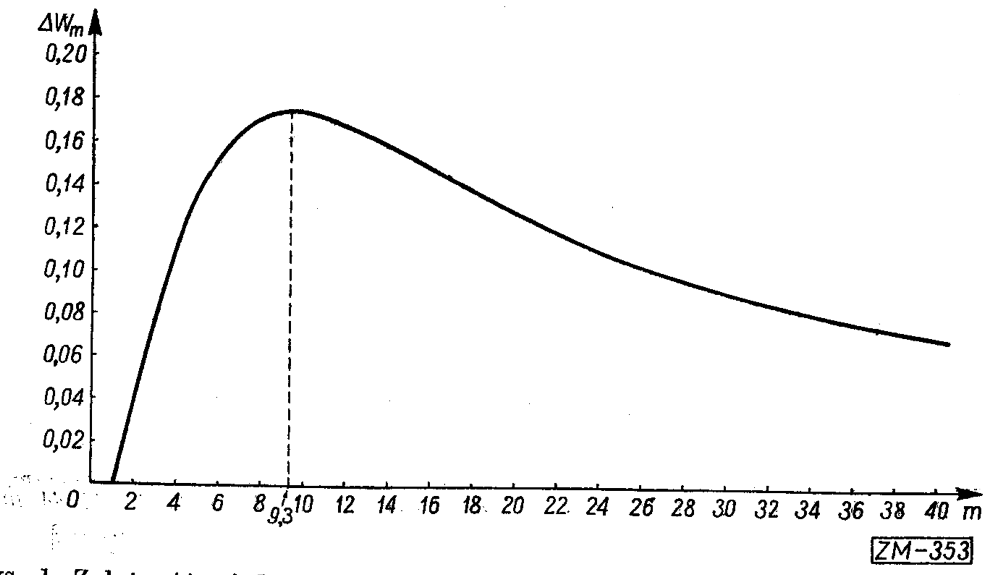
$$\frac{3}{(m_0+3)^2} + 0{,}75^{m_0} \ln 0{,}75 = 0. \tag{6}$$

Stąd

$$m_0 \approx 9{,}3.$$

W przypadku więc pobierania próbki w odległości $\bar{r} = 0{,}75$ od środka P, największy błąd w oszacowaniu średniej $\bar{w}$ będzie wówczas, gdy $m = 9{,}3$ i będzie wynosił $\Delta w_{m_0} = 0{,}175$, a więc $17{,}5\%$.

Wykres wartości Δw_m w zależności od m podaje rysunek 1.



Rys. 1. Zależność między błędem Δw_m w ocenie średniej a wartością wykładnika m

Jak wynika z wykresu, błąd w ocenie średniej początkowo szybko rośnie ze wzrostem m, a następnie po osiągnięciu maksimum maleje dość wolno i dla zakresu m mniej więcej od 3,5 do 27 przekracza 10%. Jest to więc błąd na ogół duży i w praktyce może w wielu przypadkach czynić omawianą metodę szacowania średniej za mało dokładną.

3. Zastanówmy się obecnie, w jaki sposób można by zmniejszyć maksymalny błąd oszacowania średniej, zachowując jednak przepis postępowania zalecający pobieranie próbki z jednego miejsca, określonego równością $r = R$. Zauważmy, że błąd maksymalny jest funkcją R i m.

Zdefiniujmy teraz błąd jako

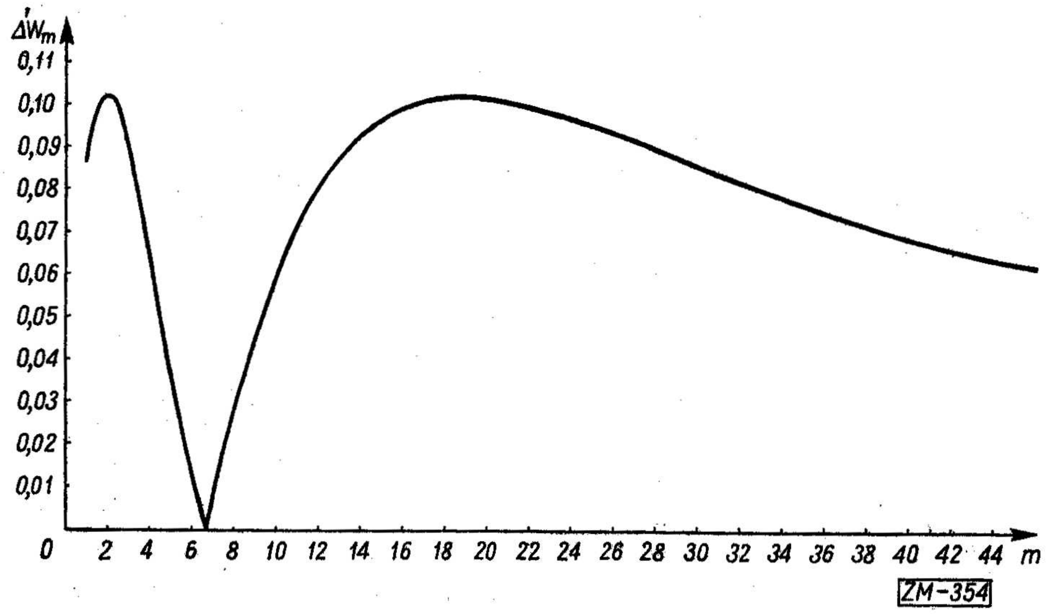
(7) $$\Delta = 1 - R^m - \frac{m}{m+3} = \frac{3}{m+3} - R^m.$$

Zadanie polega na znalezieniu takiej wartości R (oznaczmy ją R_0), żeby wyrażenie

(8) $$\delta^2 = \max_{m \geqslant 1} \Delta^2$$

osiągało minimum ze względu na R $(0 \leqslant R \leqslant 1)$.

Rozważanie numeryczno-graficzne daje jako wynik przybliżony $R_0 = 0{,}838$ i $\delta = 0{,}1025$. Oznacza to, że jeżeli do oszacowania średniej $\overline{w}$ będziemy pobierali próbkę w odległości $R_0 = 0{,}838$ od środka P i jeżeli spełniona jest zależność (2), to błąd w oszacowaniu $\overline{w}$ nie przekroczy $10{,}25\%$.



Rys. 2. Zależność między błędem Δ' w ocenie średniej a wartością wykładnika m

Rysunek 2 przedstawia zależność między błędem w oszacowaniu średniej $\overline{w}$, wyrażającym się wzorem

(9) $$\Delta' = \frac{3}{m+3} - 0{,}838^m$$

(analogicznie do wzoru (5)), a wartością wykładnika m, w przypadku pobierania próbki w odległości $R_0 = 0{,}838$.

Jak wynika z porównania rysunków 1 i 2, błąd w przypadku pobierania próbki w odległości R_0 jest z wyjątkiem początkowych wartości m (1-3,5) dla całego pozostałego zakresu m mniejszy niż w przypadku pobierania próbki w odległości $r = 0{,}75$.

4. Rozważania dotyczące błędu oszacowania i miejsca pobierania próbki przeprowadziliśmy przy założeniu, że m może przyjąć dowolną wartość w przedziale $\langle 1, \infty)$. W praktyce może okazać się, że w pewnych przypadkach można ustalić zakres wartości m. Poniższa tabelka podaje, dla niektórych zakresów m, odległość R_0 w jakich należy pobierać próbkę, aby błąd maksymalny oszacowania był najmniejszy, oraz wielkości błędów maksymalnych.

Zakres m	1-3	1-5	1-10	3-8
odległość R_0	0,778	0,800	0,825	0,824
błąd maksymalny	2,85%	5,0%	8,5%	6,0%

5. Zajmijmy się obecnie innym zagadnieniem, a mianowicie zbadaniem rozrzutu właściwości w dla wartości r w przedziale $\langle 0, 1\rangle$. Za miarę rozrzutu przyjmijmy wariancję σ_w^2 badanej właściwości. Wariancja σ_w^2 zmiennej w, będąca zgodnie z definicją drugim momentem centralnym, wyraża się wzorem

$$\sigma_w^2 = \int w^2 \varphi(w)\, dw - \overline{w}^2, \tag{10}$$

gdzie $\varphi(w)$ jest gęstością prawdopodobieństwa właściwości w.

Przyjmujemy tu taki schemat losowania, że prawdopodobieństwo wylosowania każdego punktu wewnątrz bryły i na jej powierzchni jest takie samo.

Niech teraz r będzie zmienną losową (oznaczmy ją przez ϱ) o gęstości $g(r)$. Zajmijmy się jej znalezieniem. Przyjęcie stosunku r za miarę odległości powoduje, że każdy punkt powierzchni bryły jest oddalony od środka jednokładności o $r = 1$. Każdą więc bryłę w opakowaniu można w jednostkach miary odległości r traktować jako kulę o promieniu 1. Bryły natomiast utworzone przez towary luzem — jako półkule o promieniu również równym jedności.

Objętość kuli o promieniu r, która jest zbiorem wszystkich punktów bryły oddalonych od punktu P co najwyżej o r, jest $\frac{4}{3}\pi r^3$. Objętość natomiast całej bryły, będącej zbiorem wszystkich punktów bryły, jest $\frac{4}{3}\pi$.

Stosunek tych objętości jest

$$\frac{\frac{4}{3}\pi r^3}{\frac{4}{3}\pi} = r^3. \tag{11}$$

Wyraża on prawdopodobieństwo wylosowania z bryły punktu oddalonego od środka jednokładności co najwyżej o r; jest więc dystrybuantą $G(r)$ zmiennej losowej ϱ, a zatem

$$G(r) = \int_0^r g(r)\,dr = r^3. \tag{12}$$

Stąd

$$g(r) = 3r^2. \tag{13}$$

Jak łatwo sprawdzić, (13) zachodzi też w przypadku towaru zsypanego luzem.

Ponieważ zmienna w jest związana funkcyjnie ze zmienną ϱ, więc

$$\varphi(w) = g(r)\left|\frac{dr}{dw}\right|. \tag{14}$$

W przypadku zależności (2), korzystając z (10) i (14) znajdujemy że

$$\sigma_w^2 = 3\int_0^1 r^2(1-r^m)^2 dr - \bar{u}^2, \tag{15}$$

a stąd, po całkowaniu i podstawieniu (3), otrzymujemy ostatecznie

$$\sigma_w^2 = \frac{3m^2}{(2m+3)(m+3)^2}. \tag{16}$$

A więc odchylenie średnie właściwości w jest

$$\sigma_w = \frac{m}{m+3}\sqrt{\frac{3}{2m+3}}. \tag{17}$$

Zauważmy, co jest zresztą zgodne z intuicją, że $\lim\limits_{m\to\infty}\sigma_w = 0$. Na rysunku 3 podano zależność wyrażoną wzorem (17) między odchyleniem średnim σ_w a wartością wykładnika m.

Jak wynika z tej zależności, rozrzut właściwości w jest duży i wolno maleje z wzrostem m.

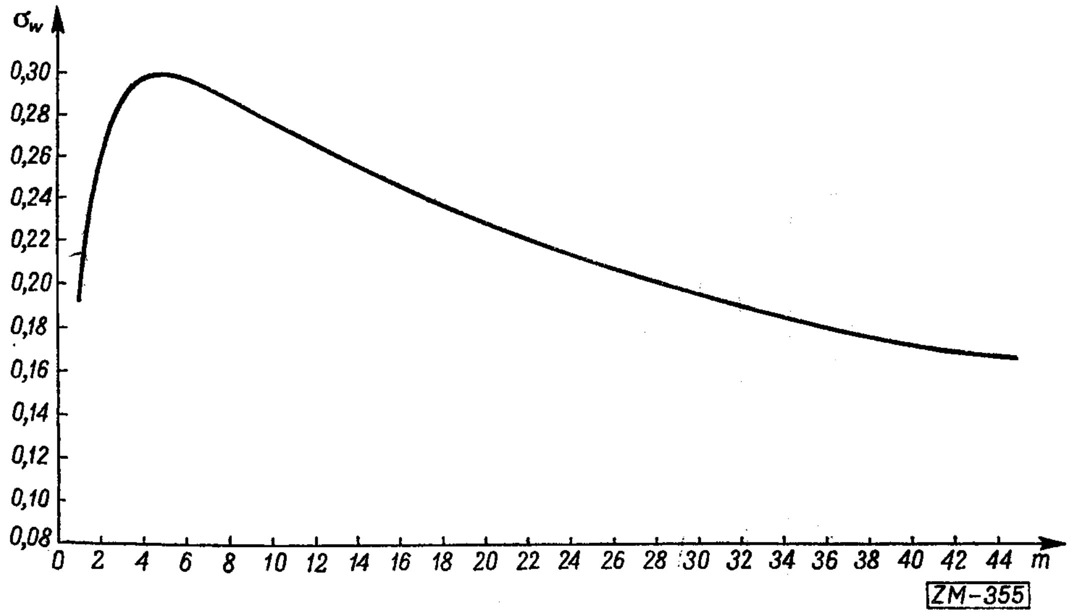
6. Mając odchylenie średnie σ_w badanej właściwości w w bryle, możemy określić błąd, jakim będzie obarczone oszacowanie wartości średniej $\bar{w}$ w bryle w przypadku szacowania jej na podstawie wyników próbki o liczności n pobranej w sposób losowy. Zakładamy tu, jak poprzednio, że właściwość w w miejscu r jest zmienną rzeczywistą. Ponieważ średnia z próbki, oznaczmy ją przez $\hat{w}$, ma rozkład asymptotycznie normalny, możemy przyjąć, że błąd δ_w w oszacowaniu $\overline{w}$ za pomocą $\hat{w}$ praktycznie będzie niewiększy, co do wartości bezwzględnej, od $3\sigma_{\hat{w}}$, gdzie $\sigma_{\hat{w}}$ jest odchyleniem średnim średniej $\hat{w}$.

Korzystając z zależności między odchyleniem średnim w populacji (w naszym przypadku w bryle) a odchyleniem średnim średniej z próbki o liczności n, wyrażającej się wzorem

$$\sigma_{\hat{w}} = \frac{\sigma_w}{\sqrt{n}}, \tag{18}$$

możemy uwzględniając (17) napisać, że praktycznie największy błąd w oszacowaniu w dla jednej bryły za pomocą $\hat{w}$ przy danym m jest

$$\delta_w = \frac{3}{\sqrt{n}} \cdot \frac{m}{m+3} \sqrt{\frac{3}{2m+3}}. \tag{19}$$



Rys. 3. Odchylenie średnie właściwości w w zależności od wykładnika m

Z drugiej strony, jak wykazano poprzednio, jeżeli będziemy pobierali jedną próbkę z bryły w odległości R_0, to w przypadku istnienia zależności (2), błąd oszacowania średniej $\overline{w}$ w bryle przy danym m nie przekroczy wartości Δ' określonej wzorem (9).

Oba kresy błędów zależą od m, a ponadto (19) zależy od n. Porównując je dla wszelkich m, możemy określić, jaka powinna być liczność n próbki, aby błąd oszacowania średniej za pomocą wyników próbki losowej nie przekraczał błędu Δ' dla danego m.

Korzystając z (9) i (19) możemy więc napisać

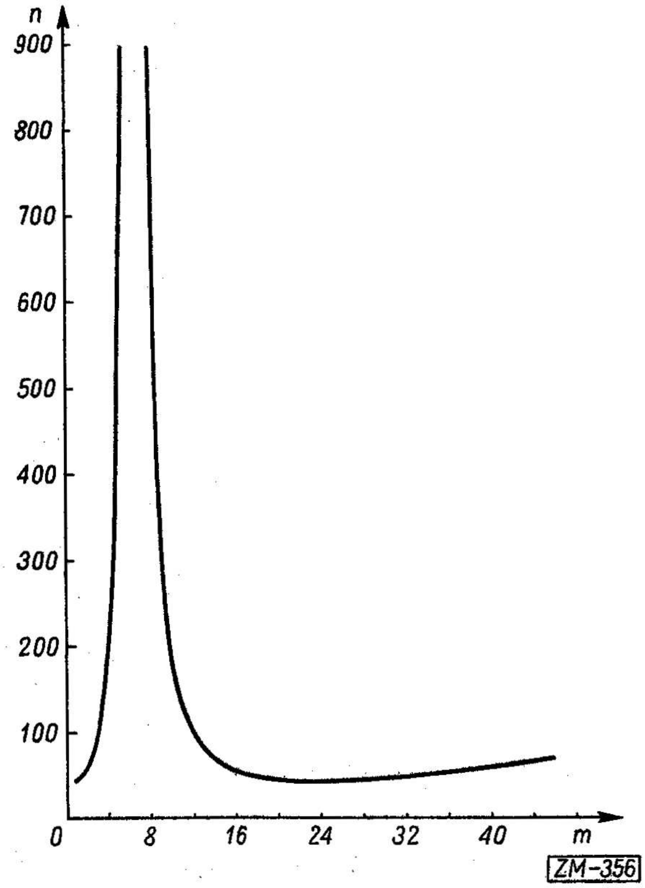
$$\frac{3}{\sqrt{n}}\cdot\frac{m}{m+3}\sqrt{\frac{3}{2m+3}}=\left|\frac{3}{m+3}-0{,}838^m\right|. \tag{20}$$

Stąd możemy wyznaczyć liczność n próbki, a mianowicie

$$n=\frac{27m^2}{(2m+3)(m+3)^2\left[\frac{3}{m+3}-0{,}838^m\right]^2}. \tag{21}$$

Powyższą zależność przedstawia rysunek 4.

Jak wynika z wykresu, w rozpatrywanym na rysunku 4 przedziale dla m, liczność próbki jest większa od 40.



Rys. 4. Liczność próbki potrzebna do oszacowania średniej właściwości, aby błąd δ_w był równy błędowi Δ'.

Tak więc, aby uzyskać drogą losowego pobierania próbki taką samą co do dokładności ocenę wartości średniej badanej właściwości jak przy pobieraniu tylko jednej próbki w odległości R_0, próbka losowa powinna się składać co najmniej z około 50-ciu próbek pierwotnych.

Wynik powyższy dobitnie świadczy na korzyść zalecanej metody pobierania próbki z jednego ściśle określonego miejsca w przypadku istnienia zależności (2), mimo popełnianych przy tej metodzie błędów oszacowania średniej.

7. Omawianą metodę badania właściwości średniej można z powodzeniem zastosować również do badania właściwości średnich roztworów i ciał mazistych w naczyniach, jak ropa, smoła itp., oraz produktów rozwarstwiających się. W tym przypadku zamiast punktu P wystąpi powierzchnia P^* wewnątrz badanego produktu, równoległa do jego górnej powierzchni. Położenie P^* będzie zależało od rodzaju badanego produktu i od badanej właściwości; może ona w pewnych przypadkach pokrywać się z górną lub dolną powierzchnią produktu. W zależności od położenia powierzchni P^* odległość r będzie mierzona w górę lub w dół od niej, bądź też w obie strony. Przyjmujemy, że naczynia mają stały przekrój poziomy, równoległy do płaszczyzny P^*.

W tym przypadku, rozumując analogicznie jak dla towarów w opakowaniu, otrzymujemy funkcję gęstości $h(r)$ zmiennej ϱ w postaci

$$h(r) = 1. \tag{22}$$

Jeżeli znów przyjmiemy, że zachodzi zależność (2), to otrzymamy

$$\begin{cases} \overline{w} = \dfrac{m}{m+1}, \\ \sigma_w^2 = \dfrac{m^2}{(2m+1)(m+1)^2}, \end{cases} \tag{23}$$

oraz odległość $\bar{r}$, w której $w = \overline{w}$

$$\bar{r} = \left(\frac{1}{m+1}\right)^{1/m}. \tag{24}$$

8. W praktyce mogą istnieć inne przepisy postępowania przy badaniu produktów w opakowaniu, zalecające inny schemat losowania niż omówiony w części 5 niniejszego studium. Jednym z nich może być przepis polecający losowanie nie punktów wewnątrz bryły, lecz warstw i pobranie z każdej z nich jednej próbki pierwotnej. Takie losowanie jest więc losowaniem odległości r.

Zajmijmy się porównaniem losowania punktów z losowaniem odległości, z punktu widzenia dokładności oszacowania średniej $\overline{w}$ badanej właściwości. W przypadku losowania warstwowego, a więc losowania odległości r, każda odległość ma to samo prawdopodobieństwo wylosowania. Rozkład zmiennej ϱ jest więc równomierny na odcinku $\langle 0, 1\rangle$ i jej funkcja gęstości $g'(r)$ jest

$$g'(r) = 1. \tag{25}$$

W przypadku zależności (2) wartość oczekiwana zmiennej w przy takim schemacie losowania jest

$$E(w) = \frac{m}{m+1}, \tag{26}$$

a wariancja

$$\sigma_w'^2 = \frac{m^2}{(2m+1)(m+1)^2}. \tag{27}$$

Jak wynika z porównania (26) z (3), to jest z prawdziwą średnią $\overline{w}$ właściwości w, w przypadku losowania warstw oszacowanie średniej $\overline{w}$ zawsze będzie obarczone błędem. Błąd ten definiujemy jako różnicę między wartością oczekiwaną oszacowania średniej a prawdziwą średnią:

$$E(w) - \overline{w} = \frac{2m}{m^2+4m+3}. \tag{28}$$

Łatwo zauważyć, że wielkość tego błędu maleje ze wzrostem m i w granicy, przy $m \to \infty$, jest równa zeru. Praktycznie oznaczałoby to, że błąd jest równy zeru, gdy wartość właściwości w jest taka sama w każdym punkcie bryły. Oczywiście w takim przypadku każdy schemat losowania jest równie dobry.

Porównajmy jeszcze wariancje zmiennej w dla obu schematów losowania. Jak łatwo sprawdzić, wariancja w przypadku losowania odległości jest dla m większych od 2,85 mniejsza niż w przypadku losowania punktów.

Z powyższych rozważań wynika, że losowanie warstw, w przypadku badania towarów w opakowaniu bądź zsypanych luzem, jest przy zależności między r i w niekorzystne dla dokładności oszacowania średniej.

Prace cytowane

[1] P. Mikulski, W. Rudzki, K. Wiśniewski, *Badanie właściwości średnich towarów bezkształtnych pakowanych w prostopadłościenne bele*, Zastosowania Matematyki 4 (1959), str. 332-340.

[2] J. Oderfeld, *Powierzchnie o wilgotności średniej*, Zastosowania Matematyki 4 (1959), str. 341-349.

[3] W. Rudzki, K. Wiśniewski, *Metoda pobierania próbek i szacowania właściwości średnich produktów bezkształtnych pakowanych w prostopadłościenne bele*, Wiadomości PKN 5 (1959).

INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK

Praca wpłynęła 17. 12. 1959

В. РУДЗКИЙ (Варшава)

ТОЧНОСТЬ ОЦЕНКИ СРЕДНЕЙ ПРИЗНАКА ДЛЯ АМОРФНЫХ ИЗДЕЛИЙ

РЕЗЮМЕ

Многочисленные товары в упаковке, или без, исследуются на среднюю признаков. В работах [1], [2], [3] обсуждён метод изъятия выборки в случае, когда имеет место зависимость между значением исследуемого признака и положением точки с данным значением признака.

В настоящей статье исследована погрешность оценки средней указанным методом и определено такое место изъятия выборки, чтобы для зависимости данной формулой (2), погрешность оценки оказалась наименьшей.

Для того же типа зависимости найдены функция плотности рассматриваемого признака и его дисперсия.

Наконец, исследован вопрос, насколько большую случайную выборку надо взять для оценки средней признака, чтобы погрешность оценки при помощи среднего значения этой выборки совпала с погрешностью для случая изъятия выборки из определенного места.

Затем обобщается метод исследования средней признака при помощи изъятия выборки из определенного места на случай растворов, мазеобразных веществ и изделий, расслаивающихся в сосудах. Кроме того сравнены — с точки зрения точности оценки средней рассматриваемого признака — схема выбора точек со схемой выбора расстояний.

W. RUDZKI (Warszawa)

EFFECTIVENESS OF A CERTAIN ESTIMATE OF THE AVERAGE PROPERTY OF SHAPELESS PRODUCTS

SUMMARY

There are many kinds of merchandise which are shapeless and either packed or kept loose, and a certain average property of which has to be investigated. In papers [1], [2] and [3] a method of sampling is discussed for estimating the average property in the case of existence of a relation between the value of the property investigated and the position of point with given value of this property.

In the present paper the error of the discussed estimate of the average property has been investigated; also such sampling points have been determined, that, under relation (2), the error of the estimate is as small as possible.

For the same type of relation the density function and the variance of the property investigated have been found.

Finally, it has been found what should be the sample size for estimating the average property in order that the error of the estimate by sample mean be equal to the error in the case of sampling from a fixed point.

Next, the method of estimating the average property by a single sample from a given point has been generalized to the case of investigating liquid solutions, greasy substances and layer-forming products.

Also, the scheme of points sampling and the scheme of sampling the distances have been compared from the point of view of efficiency of estimating the average property.
