ZASTOSOWANIA MATEMATYKI
VI (1961)

F. ZÍTEK (Praga)

O MIERZENIU PRZEZ PORÓWNYWANIE

1. Mierzeniem przez porównywanie zajmowano się w Zastosowaniach Matematyki już dwukrotnie, w pracach [1] i [2]. Szczególnie w pracy [2] rozpatrywali autorzy niektóre ciekawe zagadnienia, z których trzy główne tutaj przytoczymy:

A. Jak należy podzielić odcinek $\langle 0,1\rangle$ na k części $S_1, S_2, \ldots, S_k$, by informacja $x \epsilon S_j$ miała możliwie największą wartość dla ustalenia nieznanej wielkości x, o której jednak z góry na pewno wiemy, że $0 \leqslant x \leqslant 1$?

B. Jak należy dobrać n liczb $t_1, \ldots, t_n$ zawartych w $\langle 0,1\rangle$, by kolejne porównania nieznanej wielkości $x \epsilon \langle 0,1\rangle$ z liczbami $t_1, \ldots, t_n$ prowadziły do możliwie najdokładniejszego określenia x?

C. Jak trzeba dobrać układ ciężarków, który pozwala określić nieznany ciężar przedmiotu, o którym wiemy, że nie waży więcej niż m jednostek wagi (np. dekagramów), z błędem mniejszym niż użyta jednostka wagi, i to w możliwie najmniejszej ilości ważeń?

Autorzy pracy [2] stosowali do rozwiązania tych zagadnień aparat teorii gier; w niniejszej pracy pokażemy, jakie wyniki daje zastosowana do tych zagadnień teoria informacji.

2. Przypomnimy najpierw kilka podstawowych pojęć teorii informacji, którymi będziemy się posługiwali. Najważniejsze będą dla nas pojęcia *informacji* i *entropii.*

Niech ξ będzie zmienną losową typu skokowego, przyjmującą wartości $x_1, x_2, \ldots$, z funkcją rozkładu P:

$$P(x_i) = \Pr\{\xi = x_i\}, \quad i = 1, 2, \ldots$$

Przez *entropię* zmiennej losowej ξ będziemy rozumieli liczbę

$$H(\xi) = -\sum_i P(x_i) \log P(x_i), \tag{1}$$

gdzie, tak samo jak wszędzie w dalszym ciągu, symbol log oznacza logarytm o dowolnej, lecz ustalonej podstawie większej od jedności. Przyjmujemy, że $0 \cdot \log 0 = 0$.

Jeżeli n ($\leqslant \infty$) oznacza liczbę dodatnich prawdopodobieństw $P(x_i)$, to $0 \leqslant H(\xi) \leqslant \log n$.

Niech ξ będzie dowolną zmienną losową o funkcji rozkładu P i niech Q oznacza inną miarę taką, że P jest bezwzględnie ciągła względem Q. *Uogólnioną entropią* zmiennej losowej ξ względem miary Q nazwiemy wtedy (por. [3])

$$H_Q(\xi) = -\int \log f dP = -\int f \log f dQ, \tag{2}$$

gdzie $f = dP/dQ$. Jeżeli Q jest prawdopodobieństwem, co w dalszym ciągu wszędzie zakładamy, to $H_Q(\xi) \leqslant 0$.

Jeżeli ξ jest zmienną losową typu skokowego ze skończoną liczbą n wartości $x_1, x_2, \ldots, x_n$, $P(x_j) > 0$, i Q jest funkcją rozkładu taką, iż $Q(x_j) = n^{-1}$, $j = 1, 2, \ldots, n$, to

$$H_Q(\xi) = H(\xi) - \log n.$$

Niech będą dane dwie zmienne losowe ξ i η z funkcjami rozkładu P_1 i P_2. Oznaczmy przez P funkcję rozkładu zmiennej losowej dwuwymiarowej $[\xi, \eta]$. Informację (ilość informacji) $I(\xi, \eta)$ o zmiennej ξ zawartą w η (lub też przeciwnie) określimy w sposób następujący (por. [3] i [5]):

(a) jeżeli miara P jest bezwzględnie ciągła względem miary iloczynowej $P_1 \times P_2$ i $a = dP/d(P_1 \times P_2)$, to

$$I(\xi, \eta) = \int \log a dP = \int a \log a d(P_1 \times P_2); \tag{3}$$

(b) jeżeli miara P nie jest bezwzględnie ciągła względem $P_1 \times P_2$, to przyjmujemy $I(\xi, \eta) = \infty$.

Zawsze jest $0 \leqslant I(\xi, \eta) \leqslant \infty$ oraz (por. [3], str. 199)

$$I(\xi, \eta) = -H_{P_1 \times P_2}([\xi, \eta]). \tag{4}$$

3. Wrócimy teraz do naszych trzech zagadnień mierzenia przez porównywanie. Weźmy pierwsze z nich w nieco ogólniejszym ujęciu, nie ograniczając wartości x do przedziału $\langle 0, 1 \rangle$. Nieznaną wielkość x uważać będziemy za realizację pewnej zmiennej losowej ξ (przyjmującej wartości rzeczywiste) o funkcji rozkładu P oraz dystrybuancie F.

Niech będzie dany podział osi rzeczywistej $R = (-\infty, \infty)$ na k rozłącznych części $S_1, S_2, \ldots, S_k$, $\bigcup_{j=1}^{k} S_j = R$, mierzalnych względem miary P. Weźmy pod uwagę zmienną losową η przyjmującą wartości $1, 2, \ldots, k$ i określoną przez równości

$$\eta = j, \quad \text{gdy} \quad \xi \in S_j, \quad j = 1, 2, \ldots, k.$$

Obliczymy teraz informację o ξ zawartą w η; stosując wzór (3) otrzymujemy

$$I(\xi,\eta) = -\sum_{j=1}^{k} P(S_j)\log P(S_j) = H(\eta). \tag{5}$$

Informacja ta będzie jak wiadomo największa (przy zadanym k) wtedy, gdy wszystkie prawdopodobieństwa $P(S_j)$ będą jednakowe, a więc równe k^{-1}; będzie wówczas $I(\xi,\eta) = \log k$.

Zagadnienie A w ogólnym sensie zostało tym samym już właściwie rozwiązane: zbiory $S_j, j = 1, 2, \ldots, k$, należy dobrać tak, ażeby były one P-mierzalne i miary $P(S_j) = k^{-1}$. Kształt zbiorów S_j nie ma wpływu na wartość informacji I.

Z praktycznego punktu widzenia chodzi jednak o to, by zbiory S_j były możliwie prostej postaci; wygodne będzie np. jeżeli S_j będą odcinkami. Niech więc $S_1 = (-\infty, s_1)$, $S_j = \langle s_{j-1}, s_j)$, $j = 2, \ldots, k-1$, $S_k = \langle s_{k-1}, \infty)$. W tym szczególnym przypadku będziemy mieli, wyrażając informację (5) za pomocą dystrybuanty F,

$$I(\xi,\eta) = -\sum_{j=1}^{k} [F(s_j)-F(s_{j-1})]\log[F(s_j)-F(s_{j-1})],$$

gdzie $s_0 = -\infty$, $s_k = \infty$. Wartość informacji I nie zależy od kształtu dystrybuanty, lecz tylko od jej wartości w punktach $s_j, j = 0, 1, \ldots, k$. Warunek $P(S_j) = k^{-1}$ wyraża się wtedy w postaci

$$F(s_j) = jk^{-1}, \quad j = 0, 1, \ldots, k. \tag{6}$$

Trzeba jeszcze zauważyć, że taki podział optymalny, tzn. dający maksimum informacji, nie zawsze jest wykonalny, nawet jeżeli nie ograniczymy się do podziału na odcinki, gdyż nic nie zakładaliśmy o dystrybuancie F (lub o mierze P). Ciągłość dystrybuanty F jest warunkiem wystarczającym dla istnienia podziału optymalnego. Jeżeli zażądamy, żeby dla dowolnego naturalnego k istniał podział optymalny, to ciągłość dystrybuanty F będzie nawet konieczna. Rzeczywiście, jeśli istnieje wartość x taka, iż $\Pr\{\xi = x\} = a > 0$, to dla żadnego $k > a^{-1}$ nie można zrealizować podziału optymalnego, gdyż w tym przypadku zawsze choć jeden ze zbiorów S_j będzie miał miarę P co najmniej równą a. W związku z tym można przypomnieć pracę [4], w której rozwiązuje się zadanie maksymizacji entropii $H(\eta)$ w tym przypadku, gdy niektóre z prawdopodobieństw $P(S_j)$ są z góry zadane.

4. Zanim przejdziemy do zagadnienia B, zobaczymy, w jaki sposób można powiększyć informację kombinując dwa różne podziały. Niech więc będzie znowu dana zmienna losowa ξ oraz dwa podziały osi R na zbiory rozłączne: $S_1, \ldots, S_k$ i $T_1, \ldots, T_l$. Wprowadzimy odpowiednie zmienne losowe η i ϑ:

$$\eta = i, \quad \text{gdy} \quad \xi \epsilon S_i, \quad i = 1, 2, \ldots, k;$$

$$\vartheta = j, \quad \text{gdy} \quad \xi \epsilon T_j, \quad j = 1, 2, \ldots, l.$$

Oznaczając przez p_{ij} prawdopodobieństwa

$$p_{ij} = \Pr\{\xi \epsilon (S_i \cap T_j)\}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, l,$$

otrzymujemy wyrażenie dla informacji o ξ zawartej w parze zmiennych $[\eta, \vartheta]$:

$$I(\xi, [\eta, \vartheta]) = -\sum_{i=1}^{k} \sum_{j=1}^{l} p_{ij} \log p_{ij} = H([\eta, \vartheta]).$$

Zgodnie z poprzednimi wynikami informacja $I(\xi, [\eta, \vartheta])$ będzie największa, jeżeli wszystkie prawdopodobieństwa p_{ij} będą jednakowe, $p_{ij} = (kl)^{-1}$, $i = 1, \ldots, k$, $j = 1, \ldots, l$. Dla naszych celów ważne będzie także zadanie maksymizacji I, przy zadanych z góry wartościach $P(S_i)$ oraz $P(T_j)$; proste rachunki prowadzą wówczas do wyniku, że zbiory S_i i T_j należy dobierać tak, by one (lub, co na jedno wychodzi, zmienne η i ϑ) były stochastycznie niezależne, tzn. żeby

$$P(S_i \cap T_j) = P(S_i) P(T_j).$$

Analogiczny wynik otrzymujemy dla ogólnej kombinacji n różnych podziałów. Weźmy dla przykładu — który zresztą przyda nam się jeszcze później — n podziałów na dwie części: $R = S_j \cup S'_j$, $j = 1, 2, \ldots, n$. Określając zmienne losowe η_j tym razem równościami

$$\eta_j = \begin{cases} 1, & \text{jeśli} \quad \xi \epsilon S_j, \\ 0, & \text{jeśli} \quad \xi \epsilon S'_j, \quad j = 1, 2, \ldots, n, \end{cases}$$

możemy znowu postawić zadanie maksymizacji wartości informacji $I(\xi, [\eta_1, \ldots, \eta_n])$. Analogicznie jak poprzednio, otrzymamy wtedy odpowiedź: zbiory S_j należy dobierać tak, żeby zmienne losowe η_j były stochastycznie niezależne o jednakowym rozkładzie

$$\Pr\{\eta = 0\} = \tfrac{1}{2} = \Pr\{\eta = 1\}, \quad j = 1, 2, \ldots, n.$$

5. Wyniki poprzedniego rozdziału pozwalają nam teraz rozwiązać zagadnienie B; ξ zachowuje sens taki jak w zagadnieniu A (uogólnionym). Dla $n = 1$ zagadnienie B sprowadza się do najprostszego przypadku zagadnienia A dla $k = 2$. Porównaniu ξ z liczbą porównawczą t_1 odpowiada podział R na dwa podzbiory: $S_1^{(1)} = \langle t_1, \infty)$ i $S_1^{(0)} = (-\infty, t_1)$ oraz odpowiednia zmienna losowa η_1:

$$\eta_1 = 0, \quad \text{gdy} \quad \xi < t_1, \qquad \eta_1 = 1, \quad \text{gdy} \quad \xi \geqslant t_1.$$

Żeby otrzymać maksymalną informację $I(\xi, \eta)$, należy dobrać t_1 tak, by $F(t_1) = \frac{1}{2}$. Zakładamy najpierw dla uproszczenia, iż F jest ciągła, tak że takie t_1 rzeczywiście istnieje.

Określenie drugiej liczby porównawczej t_2 będzie oczywiście zależało od wyniku pierwszego porównania (a więc od wartości zmiennej η_1), gdyż na tym właśnie polega zaleta kolejnego postępowania. Będziemy więc mieli do czynienia nie ze stałą, ale ze zmienną losową $t_2 = t_2(\eta_1)$, która przyjmuje dwie różne wartości $t_2(0)$ i $t_2(1)$. Widać natychmiast, że jeśli okazało się że $\xi < t_1$, to nie ma już sensu porównywać ξ w drugim kroku z $t_2(0) \geqslant t_1$; jeśli natomiast było $\xi \geqslant t_1$, to porównanie z $t_2(1) \leqslant t_1$ także nie dałoby dodatkowej informacji. Można więc od razu przyjąć ograniczenie

$$t_2(0) < t_1 < t_2(1).$$

Porównaniu ξ z $t_2(\eta_1)$ odpowiadać będzie zmienna losowa η_2 określona równościami

$$\eta_2 = \begin{cases} 1, & \text{jeśli} \quad \xi \geqslant t_2(\eta_1), \\ 0, & \text{jeśli} \quad \xi < t_2(\eta_1). \end{cases}$$

Informacja $I(\xi, [\eta_1, \eta_2])$ będzie maksymalna, jeżeli

$$F[t_2(0)] = \tfrac{1}{4}, \qquad F(t_1) = \tfrac{1}{2}, \qquad F[t_2(1)] = \tfrac{3}{4};$$

będzie wówczas $I = \log 4 = 2 \log 2$.

Ogólne rozwiązanie zagadnienia B dla dowolnego n otrzymamy w sposób analogiczny: ξ porównujemy kolejno z $t_1, t_2(\eta_1), t_3(\eta_1, \eta_2), \ldots, t_n(\eta_1, \ldots, \eta_{n-1})$. Zmienne losowe η_j, $j = 1, 2, \ldots, n$, odpowiadają wynikom porównań:

$$\eta_j = \begin{cases} 1, & \text{jeśli} \quad \xi \geqslant t_j(\eta_1, \ldots, \eta_{j-1}), \\ 0, & \text{jeśli} \quad \xi < t_j(\eta_1, \ldots, \eta_{j-1}), \qquad j = 1, 2, \ldots, n. \end{cases}$$

Warunek maksymalnej informacji wyraża się wtedy ogólnie przez

$$F[t_k(i_1, \ldots, i_{k-1})] = 2^{-k} + \sum_{j=1}^{k-1} i_j 2^{-j}, \qquad k = 1, \ldots, n, \; i_j = 0 \text{ lub } 1. \tag{7}$$

(por. wzór (6) w pracy [2]).

Jest to zgodne z wynikami rozdziału 4; każdemu porównaniu odpowiada bowiem podział R na dwa podzbiory rozłączne $S_j^{(1)}$ i $S_j^{(0)} = R - S_j^{(1)}$.

Mamy rzeczywiście $S_1^{(1)} = \langle t_1, \infty)$, $S_2^{(1)} = \langle t_1(0), t_1) \cup \langle t_2(1), \infty)$ itd.; ogólnie $S_k^{(1)}$ będzie sumą 2^{k-1} przedziałów półotwartych postaci

$$(8)\qquad \langle t_k(i_1, \ldots, i_{k-1}), t_l(i_1, \ldots, i_{l-1})), \quad i_j = 0 \text{ lub } 1, \\ \text{dla } j = 1, 2, \ldots, k-1; \quad l = \max_{i_j=0} j;$$

ostatni przedział będzie zawsze $\langle t_k(1, \ldots, 1), \infty)$.

Wobec tego relacje $\xi \in S_k^{(i)}$ i $\eta_k = i$ są równoważne. Jednocześnie każdy przedział typu (8) jest częścią wspólną zbiorów $S_1^{(i_1)} \cap \ldots \cap S_{k-1}^{(i_{k-1})} \cap S_k^{(1)}$.

Warunek (7) wyraża się wtedy w postaci

$$(9)\qquad P(S_1^{(i_1)} \cap \ldots \cap S_k^{(i_k)}) = 2^{-k}, \quad i_j = 0 \text{ lub } 1 \text{ dla } j = 1, 2, \ldots, k.$$

Maksymalna informacja uzyskana przez n kolejnych porównań jest $n\log 2$. Żeby otrzymać tę samą informację przez porównania z k z góry zadanymi liczbami — co odpowiada zagadnieniu A — trzeba, żeby $k = 2^n - 1$. Widać stąd, w jakim stopniu powiększa się informację przez uzależnienie liczb porównawczych od wyników poprzednich porównań.

6. Dotychczas zakładaliśmy, iż F jest funkcją ciągłą. Jeżeli opuścić to założenie, to sprawa staje się bardziej skomplikowana, gdyż — jak widzieliśmy już w przypadku zagadnienia A — maksimum informacji może nie być osiągalne: nieciągłości funkcji F mogą uniemożliwić dokładne spełnienie warunku (7). W takim przypadku trzeba dobierać liczby $t_k(i_1, \ldots, i_{k-1})$ tak, żeby warunek (9) był jak najdokładniej spełniony dla $k = n$, bez względu na dokładność spełnienia (9) dla $k = 1, \ldots, n-1$.

W razie nieciągłości funkcji F trzeba więc określać łącznie wszystkie liczby porównawcze $t_1, t_2(0), t_2(1), \ldots, t_n(1, \ldots, 1)$. Podamy teraz bardzo prosty przykład, który pokazuje, iż kolejne ustalanie liczb porównawczych (tzn. tak, że liczby t_k określamy, gdy liczby $t_j, j = 1, 2, \ldots, k-1$, są już dane) może dać rzeczywiście gorsze wyniki.

Niech będzie

$$F(x) = \begin{cases} 0 & \text{dla} \quad -\infty < x \leqslant 0, \\ \frac{1}{2} & \text{dla} \quad 0 < x \leqslant 1, \\ \frac{1}{2}x & \text{dla} \quad 1 \leqslant x \leqslant 2, \\ 1 & \text{dla} \quad x \geqslant 2. \end{cases}$$

Jeżeli wybieramy najpierw t_1, to możemy wziąć np. $t_1 = \frac{1}{2}$. Określając następnie (przy zadanym już t_1) liczby $t_2(0)$ i $t_2(1)$, otrzymamy $t_2(1) = \frac{3}{2}$ oraz np. $t_2(0) = 0$. Odpowiadająca wartość informacji $I(\xi, [\eta_1, \eta_2])$ będzie wówczas $\frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4 = \frac{1}{2}\log 8$. Jeżeli natomiast wziąć od razu $t_2(0) = \frac{1}{2}$, $t_1 = \frac{4}{3}$, $t_2(1) = \frac{5}{3}$, to będziemy mieli

$$I(\xi, [\eta_1, \eta_2]) = \tfrac{1}{2}\log 2 + 3\cdot\tfrac{1}{6}\log 6 = \tfrac{1}{2}\log 12 > \tfrac{1}{2}\log 8.$$

7. Przejdźmy w końcu do zagadnienia C. Nieznany ciężar x uważamy znów za realizację zmiennej losowej ξ, która tym razem może przyjąć tylko jedną z wartości $1, 2, \ldots, m$. Ponieważ ważenie to właściwie porównywanie wagi przedmiotu z wagą ciężarków leżących na drugiej szali, zagadnienie wyboru tych ciężarków sprowadza się do zagadnienia B dla ξ typu skokowego. Układ ciężarków powinien więc być taki, żeby pozwalał realizować liczby mt_k, przy czym t_k spełniają warunki (7). Układ będzie wobec tego zależał w znacznym stopniu od rozkładu zmiennej losowej ξ.

Podane rozwiązanie zagadnienia C pozwala minimizować średnią ilość ważeń potrzebnych do ustalenia nieznanego ciężaru. Jeżeli jednak chcemy minimizować maksymalną ilość ważeń, to należy stosować układ $m/2$, $m/2^2, \ldots$, określony w pracy [2].

Zagadnienia równoważne z zagadnieniami ważenia zostały poruszone także w zadaniach 19 i 20 rozdziału III książki [6].

8. Modyfikacją „metody maksymalnej informacji", którą stosowaliśmy do rozwiązywania naszych zagadnień, jest metoda oparta na pojęciu entropii interpretowanej jako miara niepewności. Pokażemy ją na przykładzie zagadnienia A.

Niech będą dane zmienne losowe ξ i η o tym samym znaczeniu co w rozdziale 3. Niech Q oznacza miarę taką, że $H_Q(\xi) < \infty$; wybór miary Q nie odgrywa przy tym istotnej roli. Dla danego podziału $S_1, S_2, \ldots, S_k$, $P(S_j) > 0$, $j = 1, 2, \ldots, k$, określimy k zmiennych losowych $\xi_1, \ldots, \xi_k$, z funkcjami rozkładu $P_1, \ldots, P_k$, odpowiadającymi rozkładom warunkowym zmiennej ξ pod warunkiem $\eta = j$, $j = 1, 2, \ldots, k$. Mamy więc dla $Z \subset S_j$

$$P_j(Z) = \Pr\{\xi_j \epsilon Z\} = P(Z)/P(S_j)$$

i stąd

$$\frac{dP_j}{dQ} = \frac{1}{P(S_j)} \cdot \frac{dP}{dQ}.$$

Uogólniona entropia zmiennej ξ_j względem Q będzie wówczas równa

$$H_Q(\xi_j) = -\int \left[\log\left(\frac{dP_j}{dQ}\right) - \log P(S_j)\right] dP,$$

wskutek czego zachodzi równość (por. [6])

$$H_Q(\xi) - I(\xi, \eta) = \sum_{j=1}^{k} P(S_j) H_Q(\xi_j). \tag{10}$$

Warunek maksymalnej informacji $I(\xi, \eta)$ jest więc równoważny z warunkiem minimalnej średniej entropii, która pozostaje po doświadczeniu określającym wartość zmiennej η

Prace cytowane

[1] J. Łukaszewicz, H. Steinhaus, *O mierzeniu przez kalibrowanie*, Zastosowania Matematyki 2 (1956), str. 225-231.

[2] H. Steinhaus, S. Trybuła, *Pomiar przez kolejne porównywanie*, Zastosowania Matematyki 4 (1959), str. 204-212.

[3] A. Perez, *Notions généralisées d'incertitude, d'entropie et d'information du point de vue de la théorie des martingales*, Transactions of the First Prague Conference on Information Theory, Praha 1957, str. 183-208.

[4] L. Votavová, *Ein Satz von Extremen der Entropie*, ibidem, str. 293-295.

[5] Р. Л. Добрушин, *Общая формулировка основной теоремы Шеннона в теории информации*, Успехи Мат. Наук 14 (1959), вып. 6, str. 3-104.

[6] А. М. Яглом, И. М. Яглом, *Вероятность и информация*, Москва 1957.

INSTYTUT MATEMATYCZNY CZECHOSŁOWACKIEJ AKADEMII NAUK

Praca wpłynęła 15. 2. 1960

Ф. ЗИТЕК (Прага)

ОБ ИЗМЕРЕНИИ ПУТЕМ СРАВНЕНИЯ

РЕЗЮМЕ

В статье [2] рассматривались с точки зрения теории игр интересные проблемы об измерении путем сравнения. В настоящей статье дано решение этих проблем при помощи понятий теории информации.

Характеристической задачей является здесь определение неизвестной величины x путем последовательного сравнения с некоторыми числами, данными для сравнения. Вместо того, чтобы минимизировать погрешность, максимизируется полученная информация.

F. ZÍTEK (Prague)

ON MEASURING BY COMPARISON

SUMMARY

Three interesting problems concerning measuring by comparison have been considered in paper [2] from the Game-Theoretical point of view. This paper gives the solution of these problems by using notions from Information Theory.

The typical problem may be stated as follows: how to determine the unknown value x by successive comparing it with some numbers; instead of minimizing the estimation error, we maximize the information gained.
