ZASOTSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
VIII (1966)

S. ZUBRZYCKI (Wrocław)

O NIEKTÓRYCH PRACACH SEMINARIUM Z ZASTOSOWAŃ MATEMATYKI

Wstęp. Współpraca z przedstawicielami różnych dziedzin nauki i praktyki przy formułowaniu i rozwiązywaniu zadań jest głównym punktem programu pracy Działu Zastosowań Przyrodniczych, Gospodarczych i Technicznych Instytutu Matematycznego PAN (w skrócie: DZPGiT). Jednym z celów obecnej sesji (1) jest manifestacja takiej współpracy. Jeśli współpraca ma być skuteczna, nie można się obejść bez teoretycznej refleksji nad stosowanymi metodami, nad dostosowywaniem ich do napotykanych zadań, nad tworzeniem nowych metod, stwarzających możliwość nowych zastosowań. W związku z działalnością seminarium z zastosowań matematyki prowadzonym w DZPGiT powstała pewna liczba prac o bardziej teoretycznym charakterze, zawierających pewne pomysły nie będące li tylko zastosowaniem gotowej techniki. W mym przeglądzie na nie właśnie chciałbym zwrócić uwagę i o nich powiedzieć słów kilka.

Chciałbym jednak zacząć od przypomnienia, że istnieje już kilka przeglądowych publikacji poświęconych omówieniu prac DZPGiT w pewnych dziedzinach. Jest artykuł J. Perkala [1] z 1956 r. o zastosowaniach rolniczych, inny artykuł tegoż autora o współpracy z medykami [2], 1959, artykuł Bartkowiakowej i Zubrzyckiej o badaniach języka za pomocą metod matematycznych [3], 1963, wreszcie obszerny, liczący 62 stronice druku i ponad 50 cytowanych pozycji bibliograficznych, przegląd polskich prac z zastosowań matematyki pióra J. Łukaszewicza [4] będący streszczeniem (w języku chińskim) jego wykładów, wygłoszonych w Pekinie w 1958 roku.

Niniejszy przegląd będzie się w pewnym stopniu pokrywał z niektórymi poprzednimi, przynajmniej jeśli chodzi o listę prac cytowanych, i to bodaj najmocniej ze wspomnianym chińskim artykułem J. Łukaszewicza. Będą w nim także wspomniane pewne prace powstałe poza semi-

(1) Niniejszy przegląd jest nieco rozwiniętą postacią referatu przedstawionego 11 lutego 1964 na sesji naukowej z okazji 500 posiedzenia seminarium z zastosowań matematyki prowadzonego w DZPGiT.

narium, którego jubileusz obchodzimy. Wreszcie skorzystam z przywileju mówienia o pracach najlepiej mi znanych, skutkiem czego przegląd ten będzie na pewno niekompletny. W końcu będzie on pobieżny, stanowiąc jedynie rodzaj przewodnika po liście prac cytowanych.

§ 1. Wiarogodność i reguła Bayesa. Jednym z podstawowych kłopotów w zagadnieniach statystycznych jest nieznajomość tego, co się nazywa rozkładem a priori. Weźmy najprostszy przykład. Statystyk, wylosowawszy (ze zwracaniem) z badanej partii towaru próbkę liczącą n elementów i stwierdziwszy, że zawiera ona k elementów wadliwych, jest zapytywany, co może na tej podstawie orzec o wadliwości partii, czyli o frakcji p sztuk wadliwych. Statystyk wie, jaki jest warunkowy rozkład prawdopodobieństwa liczby K sztuk wadliwych w próbce, gdy jest ona losowana w podany wyżej sposób z partii o zadanej wadliwości p, mianowicie prawdopodobieństwo znalezienia w próbce $K = k$ sztuk wadliwych jest dane znanym wzorem binomialnym

$$f(k|p) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Gdyby statystyk znał ponadto rozkład a priori wadliwości p, czyli gdyby dla każdego a wiedział przed przystąpieniem do losowania próbki, jakie jest prawdopodobieństwo, że przedstawiona do badania partia ma wadliwość $p < a$, to mógłby na podstawie twierdzenia Bayesa obliczyć dla każdego a prawdopodobieństwo a posteriori, że wadliwość p jest mniejsza niż a, skoro w próbce zaobserwowano k sztuk wadliwych. Oznaczmy je przez $P(p < a | K = k)$. Zazwyczaj jednak za znane można uważać tylko prawdopodobieństwa warunkowe $f(k|p)$, natomiast rozkład a priori wadliwości p nie jest znany. W takiej sytuacji reguła Bayesa radzi obliczać rozkład a posteriori wadliwości p przy arbitralnie przyjętej hipotezie, że a priori wadliwość p ma rozkład jednostajny w przedziale $0 < p < 1$. Takie stawianie sprawy jako nieuzasadnione spotkało się z krytyką. Zaczęto poszukiwać sposobów orzekania o wadliwości p na podstawie obserwacji k w sposób niezależny od rozkładów a priori. Jednym z nich jest argument fiducjalny R. A. Fishera, który każe tym mocniej wierzyć w zachodzenie nierówności $p < a$ im większe jest prawdopodobieństwo tego, że przy losowaniu (ze zwracaniem) n-elementowej próbki z partii o wadliwości $p = a$ liczba K sztuk wadliwych w próbce okaże się większa od k. Nazywając to ostatnie prawdopodobieństwo *wiarogodnością* nierówności $p < a$ pod warunkiem, że $K = k$, możemy napisać definicję

$$W(p < a | K = k) = P(K > k | p = a). \tag{1}$$

W pracy [5] z 1948 roku Steinhaus po raz pierwszy wystąpił w obronie reguły Bayesa, a argumenty za regułą Bayesa przedstawił ponownie w pracy [7] z 1951 roku w bardziej rozwiniętej formie, opierającej się na odkryciu Oderfelda [6], 1951, nazwanym *regułą dualizmu*, a stwierdzającym, że wiarogodność obliczona według wzoru (1) jest równa prawdopodobieństwu a posteriori obliczonemu przy założeniu jednostajnego rozkładu a priori wadliwości p w przedziale $0 \leqslant p \leqslant 1$, z tym jednak, że odnosi się ono do próbki o tej samej liczbie sztuk wadliwych, ale o liczności o 1 mniejszej. Reguła dualizmu była przedmiotem badań kilku dalszych prac. I tak Sarkadi [8], 1953, odkrył między innymi analogiczną regułę dla rozkładów Poissona, Steinhaus [9], 1954, dla parametru przesunięcia, Romejko [10], 1957, dla pewnego przypadku porównywania dwóch procesów Poissona, a wyniki Romejki i poprzedników uogólnili Steinhaus i Zubrzycki w pracach [11], 1957, i [12], 1958, usuwając między innymi mankament argumentu Oderfelda polegający na niejednakowej liczności próbek, dla których w regule dualizmu oblicza się wiarogodność i prawdopodobieństwo a posteriori. We wszystkich przypadkach okazało się, że hipoteza co do rozkładu a priori kryjąca się w argumencie fiducjalnym jest dużo bardziej pesymistyczna, niż to, co podpowiada reguła (czy raczej postulat) Bayesa. Wnioski wyciągnięte przez Steinhausa w cytowanej już pracy [9], 1954, z odkrytych reguł dualizmu na temat stosowalności reguły Bayesa tak przypominają niektóre argumenty dotyczące podstaw wnioskowania statystycznego w dyskusji toczącej się na Zachodzie (zobacz [13], 1962, w szczególności wypowiedzi G. A. Barnarda), iż wydało się stosowne wznowić ją w języku angielskim [14], 1963.

§ 2. Zagadnienia estymacji. Zasada minimaks. Od czasu ukazania się książki Morgensterna i von Neumanna [15], 1944, zaczęto systematycznie eksploatować ideę traktowania zagadnień statystycznych jako gry między statystykiem i naturą w sensie nowoczesnej teorii gier. Książka Walda [16], 1950, jest pierwszym obszernym przedstawieniem zagadnień statystycznych w takim ujęciu. Podejście to wyróżnia się tym, iż próbuje się w jawny sposób uwzględnić skutki błędów przy podejmowaniu decyzji. I tak w przypadku szacowania nieznanego parametru p przyjmuje się, że dana jest funkcja $L(p, p')$, zwana *stratą*, która wyraża szkodę ponoszoną wtedy, gdy p jest prawdziwą wartością parametru, a my orzekamy, że jest ona p'. *Estymatorem* nazywamy funkcję, która wynikowi obserwacji przyporządkowuje przybliżoną wartość estymowanego parametru. Wartość estymatora, a wraz z nią strata L, staje się zmienną losową, gdy rozważamy go jako funkcję zmiennej losowej X reprezentującej wyniki obserwacji. Dla danej wartości parametru p, i danego estymatora $\hat{p}$, określona jest wówczas oczekiwana strata, zwana *ryzykiem*:

$$r(p, \hat{p}) = E_p L(p, \hat{p}). \tag{2}$$

Tak określone ryzyko r, traktowane przy danym estymatorze $\hat{p}$ jako funkcja parametru p, opisuje oczekiwane szkody wynikające z błędów oszacowania przy używaniu danego estymatora. Gdy więc chodzi o wybór estymatora do szacowania parametru na podstawie próbki o zadanej liczności, a więc gdy nie odgrywają roli koszty badania, charakteryzuje ono przydatność estymatora $\hat{p}$ do tego celu. Wymagania stawiane estymatorowi powinny być przeto wyrażone przy pomocy tego ryzyka. Gdy rozważa się zagadnienie sekwencyjne, trzeba uwzględnić jeszcze koszt badania zależny od liczby obserwacji. Przy wyborze estymatora trzeba wówczas w jakiś sposób skonfrontować straty płynące z błędów oszacowania, które maleją, gdy liczność próbki rośnie, i koszty badania, które rosną w miarę zwiększania liczby obserwacji.

Steinhaus w pracy [17], 1950, dyskutuje związek między kosztami badania i osiąganą dokładnością oszacowania dla przypadku, kiedy wartość partii towaru jest sumą wartości sztuk z jakich się ona składa, i proponuje pewien sekwencyjny sposób wyceny wartości partii.

W pracy [18], 1953, porównuje Steinhaus w terminach ryzyka różne postulaty prowadzące do ustalenia przepisu odbiorczego w statystycznej kontroli jakości, między innymi metodę prospektywną opartą na pojęciu wiarogodności i metodę retrospektywną opartą na postulacie Bayesa, podkreślając równorzędny charakter tej ostatniej w stosunku do innych propozycji. W pracy [19], 1956, na przykładzie prognozy pogody pokazuje Steinhaus niecelowość używania prognozy zrandomizowanej dla szerokiej klasy funkcji straty i wskazuje na istotną zależność optymalnej prognozy od przyjętej funkcji straty.

Wróćmy na chwilę do ryzyka określonego wzorem (2). Teoria gier każe zadanie wyboru estymatora rozpatrywać jako grę między naturą, której strategiami są wartości parametru, a statystykiem, którego strategiami są estymatory. Ryzyko r jest wypłatą w tej grze płaconą przez statystyka naturze. Dążeniem statystyka jest uczynić ryzyko możliwie małym. Ideałem byłoby znaleźć taki estymator, dla którego ryzyko osiągałoby najmniejszą możliwą wartość jednocześnie dla wszystkich wartości parametru. Takie jednostajnie najlepsze estymatory rzadko istnieją. Teoria gier radzi wówczas stosować zasadę minimaks, to znaczy szukać takiego estymatora, dla którego maksymalna wartość ryzyka ze względu na różne wartości parametru byłaby możliwie mała. Formalnie mówiąc, estymator $\hat{p}_0$ nazywa się *minimaksowym* wtedy i tylko wtedy, gdy

$$\sup_p r(p, \hat{p}_0) = \inf_{\hat{p}} \sup_p r(p, \hat{p}). \tag{3}$$

Pomijając kwestię ustalenia najogólniejszych warunków istnienia i jednoznaczności minimaksowych strategii statystyka w grach statystycznych, interesujące jest zobaczyć, jakie są minimaksowe strategie statystyka w różnych zadaniach statystycznych spotykanych często w zastosowaniach. Odczyt Steinhausa [20], 1954, inauguruje serię takich prac. I tak, wyznaczono minimaksowe estymatory w zadaniu szacowania składu $p = (p_1, \ldots, p_k)$ urny zawierającej kule k kolorów ($p_i \geqslant 0$, $p_1 + \ldots + p_k = 1$) na podstawie składu $m = (m_1, \ldots, m_k)$ próbki n-elementowej losowanej ze zwracaniem przy kwadratowej funkcji straty $L(p, p') = \sum (p_i - p_i')^2$ (Steinhaus [21], 1957), a także podano kompletny opis minimaksowych estymatorów przy ogólniejszej funkcji straty $L(p, p') = \sum (p_i - p_i')^2 c_i$ i uogólnienie tego wyniku na przypadek przeliczalnie wielu kolorów kul (Trybuła [23], 1958). Osobnego potraktowania wymagało finitystyczne zadanie szacowania składu $M = (M_1, \ldots, M_k)$ urny zawierającej N kul k kolorów na podstawie n-elementowej próbki losowanej bez zwracania. Przy kwadratowej funkcji straty $L(M, M') = \sum c_i (M_i - M_i')^2$ rozpatrzył je Trybuła [24], 1958. Trybuła [25], 1960, odkrył dla zadania szacowania składu urny przy obu sposobach pobierania próbki przy jakiej funkcji straty znane estymatory największej wiarogodności okazują się minimaksowe. Tu należy praca Trybuły [22], 1957, o minimaksowej prognozie, praca Mikiewicza [26], 1961, w której podano minimaksowy estymator średniej w oparciu o serię pomiarów o ustalonej liczności, lecz niejednakowej dokładności, praca Dzan Dzo-i [27], 1961, o minimaksowych estymatorach frakcji sztuk wadliwych w przypadku, gdy zasób strategii natury jest ograniczony do pewnego podprzedziału przedziału $0 \leqslant p \leqslant 1$, praca Czen Pina [28], 1962, i Zubrzyckiego [29], 1963, o minimaksowych estymatorach liczności populacji na podstawie próbek odwrotnych, związane z pracą Czen Pina i Dzan Dzo-i [30], 1961, poświęconą dyskusji różnych wariantów zadania szacowania liczności populacji na podstawie obserwowania w próbce elementów uprzednio znaczonych i zwróconych do populacji.

Była wreszcie seria kilku prac o mierzeniu, w których stosowano zasadę minimaks przy nie zawsze probabilistycznym charakterze zadań: Łukaszewicz i Steinhaus [31], 1955, Steinhaus i Trybuła [32], 1959, oraz seria kilku prac zainicjowanych w Dziale Zastosowań Przemysłowych (Mikulski, Rudzki i Wiśniewski [33], 1959, Oderfeld [34], 1959, i kilka innych) poświęconych ogólnie mówiąc wyznaczaniu najlepszego systematycznego sposobu rozstawiania zadanej liczby obserwacji w bryle towaru bezkształtnego w celu możliwie dokładnego szacowania średniej wartości cechy tego towaru, gdy mamy pewne informacje o funkcji opisującej rozkład wartości badanej cechy wewnątrz geometrycznej figury wypełnionej przez towar. Do nich nawiązuje praca Zubrzyckiego [35], 1963. Sklasyfikowano w niej te zadania w pewien jednolity sposób i w przypadku gdy klasa funkcji opisującej wartości badanej cechy jest dość obszerna, obszerniejsza niż w przypadkach rozpatrywanych wcześniej przez Mikulskiego, Rudzkiego, Wiśniewskiego i Oderfelda, wyznaczono dla dowolnego n minimaksowe wzory przybliżonego całkowania, które wskazują sposób rozstawienia obserwacji i wagi, z jakimi należy te obserwacje wyśredniować. Wyniki te stanowią więc pewne rozwiązanie rozważanego zadania systematycznego rozmieszczania obserwacji. Dodajmy, że sformułowania dotyczące minimaksowych wzorów przybliżonego całkowania znalezione w [35] są bardzo bliskie wyników zawartych w [36], 1959, a także [37], 1957.

§ 3. Długość empiryczna i kształt. Paradoksowi długości, związanemu z generalizacją map, a polegającemu na tym, że np. długość wybrzeża morskiego rośnie w miarę mierzenia go na coraz dokładniejszych mapach, poświęcił Steinhaus serię prac datujących się od 1931 r. Zaproponował on wolne od takiego paradoksu pojęcie długości rzędu n, związane z miarą Croftona zbioru prostych na płaszczyźnie. Jego prace o długości, kształcie i polu [38], 1954, i [39], 1956, są przeglądem tej problematyki ze wskazaniem nierozwiązanych problemów.

Praca Perkala o geometrycznych wskaźnikach łąk [40], 1955, w której zdefiniowano pewne liczbowe charakterystyki odnoszące się do kształtu, wielkości i sposobu rozmieszczenia kępek trawy na łąkach uprawnych i badano różne sposoby estymowania tych charakterystyk, stała się powodem do podjęcia na nowo dyskusji kształtu i długości obiektów empirycznych takich jak kępka trawy, tym razem z nieco innego punktu widzenia. Mianowicie, Perkal zaproponował oprzeć pojęcie zgeneralizowanego kształtu, pola i długości (obwód kępki trawy) na pojęciu zbiorów ε-wypukłych. Idee te opracował teoretycznie i praktycznie, opisując między innymi longimetr do wyznaczania zgeneralizowanej długości, w serii prac (zobacz [41], 1956, [42], 1956, [43], 1958).

§ 4. Szacowanie parametrów złóż geologicznych. Systematyczne rozmieszczanie obserwacji na płaszczyźnie. Oprócz wspomnianej już pracy Perkala o geometrycznych wskaźnikach łąk na problematykę szacowania średniej wartości procesów rozciągających się na płaszczyźnie natknęliśmy się w związku z szacowaniem złóż geologicznych. Rozpatrywane zadanie można sformułować tak: na danym obszarze D, na którym, jak się spodziewamy na podstawie uprzednich badań, zalega złoże, należy rozmieścić daną liczbę obserwacji tak, żeby na ich podstawie jak najdokładniej oszacować wielkość zasobów znajdujących się na tym obszarze. Potoczne metody statystyczne służące do oceny błędu oszacowania opierają się na założeniu statystycznej niezależności obserwacji. Założenia te można zrealizować, dla danego obszaru D, przez losowe rozstawianie obserwacji. Nie daje to jednak możliwości uwzględniania wpływu sposobu rozmieszczania obserwacji na dokładność oszacowania. Zubrzycki [44], 1957, proponuje do opisu złoża użyć pojęcia procesu stochastycznego stacjonarnego. Funkcja korelacyjna takiego procesu jest tym narzędziem, które umożliwia ocenę wpływu sposobu rozmieszczenia obserwacji na dokładność oszacowania. W wymienionej pracy omówiono sposób wyznaczania funkcji korelacyjnej złoża na podstawie obserwacji obciążonych błędem losowym, obliczono empiryczne funkcje korelacyjne na podstawie dwu serii obserwacji z górnośląskich złóż cynkowych, wyrównano je przy pomocy kilku teoretycznych funkcji korelacyjnych, z których najlepiej pasowała do obserwacji funkcja wykładnicza, przedstawiono sposób obliczania błędu oszacowania przy danym rozstawieniu obserwacji, gdy znana jest funkcja korelacyjna, podano przykłady numeryczne ilustrujące różnice w wielkości błędu oszacowań przy różnych rozstawieniach tej samej liczby obserwacji.

Ponieważ jednak funkcja korelacyjna złoża nie zawsze znana jest dość dokładnie, interesujące wydało się pytanie, czy można wskazać jakiś ustalony sposób rozmieszczania obserwacji, który byłby optymalny dla każdej funkcji korelacyjnej z możliwie szerokiej klasy takich funkcji. Zubrzycki [45], 1958, próbował przenieść na płaszczyznę znane dla prostej porównanie trzech sposobów rozmieszczania obserwacji: losowego, warstwowego i systematycznego. Są to wszystko sposoby o tej własności, że oczekiwana liczba obserwacji przypadających na dowolną część obszaru D jest proporcjonalna do pola tej części. Na prostej wiadomo, że dla wypukłych funkcji korelacyjnych wśród sposobów rozmieszczania obserwacji o wymienionej wyżej własności najlepszy jest sposób systematyczny. Na płaszczyźnie okazało się jedynie, że warstwowe rozmieszczanie obserwacji jest lepsze od losowego, ale nie ma generalnej wskazówki co do lepszości systematycznego rozmieszczania obserwacji w porównaniu z warstwowym, nawet przy daleko posuniętych ograniczeniach co do funkcji korelacyjnej i kształtu warstw (zobacz [45] wraz z [47], 1961).

Dalenius, Hájek i Zubrzycki [48], 1961, zajęli się pytaniem, jaki jest najlepszy kształt regularnej sieci punktów na płaszczyźnie z punktu widzenia dokładności szacowania średniej procesu stochastycznego (stacjonarnego, izotropowego) na danym obszarze przez średnią arytmetyczną obserwacji. Tu również w przypadku procesów określonych na prostej odpowiedź jest jasna i dość generalna: dla wypukłych funkcji korelacyjnych najlepsze jest rozstawianie obserwacji w równych odstępach. Dość obiecujące było przypuszczenie, że dla procesów izotropowych, tzn. takich, że korelacja między wartościami procesu w różnych punktach zależy tylko od odległości tych punktów, a nie zależy od kierunku odcinka łączącego te punkty, najlepszą wśród siatek o ustalonej gęstości okaże się siatka trójkątów równobocznych. Niestety, okazało się, że tak nie jest, i że optymalny kształt sieci zależy od funkcji korelacyjnej i od gęstości sieci.

Kontrprzykłady uzyskano w cytowanej pracy [48] dla pewnej specjalnej funkcji korelacyjnej. W górnośląskich złożach cynkowych do empirycznych funkcji korelacyjnych najlepiej pasowała funkcja wykładnicza (zobacz [44]). Pozostało nie rozstrzygnięte przypuszczenie, któremu nie przeczą przykłady numeryczne u Matérna [46], 1960, że sieć trójkątów równobocznych jest optymalna wśród sieci o zadanej gęstości dla wykładniczych funkcji korelacyjnych. Przypuszczenie to zostało podtrzymane przez wyniki Kopocińskiego ([49], 1965, dalsza część w przygotowaniu). Pokazał on mianowicie, że gdy funkcja korelacyjna izotropowego stacjonarnego procesu stochastycznego na płaszczyźnie jest wykładnicza, to dokładności oszacowania uzyskiwanej przy użyciu sieci trójkątów równobocznych nie da się poprawić ani przez małe affiniczne deformacje takiej sieci (zachowujące gęstość), ani nawet przez małe (infinitesymalne) deformacje ogólniejsze od affinicznych. Ponieważ funkcja wykładnicza najlepiej wyrównywała empiryczną funkcję korelacyjną obliczoną dla złóż cynkowych (zobacz [44], 1957), można w tym widzieć wskazówkę, że korzystnie jest rozmieszczać wiercenia geologiczne w sieć trójkątów równobocznych.

§ 5. Tablice. Tablice liczb losowych są pomyślane jako narzędzie pomocne przy wybieraniu próbki z partii liczącej wiele sztuk. Jednym z mankamentów takich tablic jest to, że dopuszczalne są powtórzenia, a skutkiem tego zdarza się również, że pewne liczby w ogóle nie występują. Ponadto przy używaniu takich tablic, co odpowiada losowaniu ze zwracaniem, występuje tendencja do powstawania losowych nierównomierności w rozstawianiu elementów próbkowych. Jest to niepożądane szczególnie wtedy, gdy wchodzi w rachubę jakaś zależność między jakością sztuki, a jej numerem, np. gdy sztuki są numerowane w kolejności produkcji. Tym mankamentom można by zaradzić przez systematyczne losowanie, np. biorąc do próbki co dziesiąty element. Lecz i w takim sposobie pobierania próbki można widzieć pewne niedogodności: jest pewne niebezpieczeństwo natrafienia na okresowy rytm zmiany jakości sztuk w zależności od numeru, a ponadto równomierność rozmieszczenia sztuk próbkowych ulegałaby drastycznemu pogorszeniu, gdybyśmy chcieli próbkę zwiększyć o kilka sztuk. H. Steinhaus starał się zaradzić wszystkim tym mankamentom przez zaprojektowanie tablicy zawierającej wszystkie liczby czterocyfrowe od 0000 do 9999, każdą dokładnie jeden raz, powstałej przez wymieszanie tych liczb przy pomocy odpowiedniego ściśle określonego algorytmu. Jeden taki dość zawiły algorytm został użyty do konstrukcji tablicy [51], 1954, liczb przetasowanych, opisanej w artykule [50], 1954. Innego sposobu mieszania użyto przy budowaniu tablicy liczb żelaznych [52], 1956. Wykorzystano do tego celu reszty modulo 1 wielokrotności liczby złotej $\alpha = (\sqrt{5}-1)/2 = 0{,}618\ldots$, określającej tzw. złoty podział odcinka.

Oznaczmy przez $\{n\alpha\}$ resztę modulo 1 z liczby $n\alpha$. Ustawmy następnie reszty $\{\alpha\}$, $\{2\alpha\}, \ldots, \{10000\alpha\}$ w porządku rosnącym, a następnie w tak powstałym ciągu zastąpmy resztę $\{n\alpha\}$ przez n, dla $n = 1, 2, \ldots$

$\ldots, 10000$. Otrzymamy wówczas pewne uporządkowanie liczb $1, 2, \ldots$ $\ldots, 10000$. Jeśli liczby te napiszemy w tak otrzymanym porządku jako liczby czterocyfrowe, identyfikując 1 z 0001, 2 z 0002, ..., 10000 z 0000, otrzymamy tablicę liczb żelaznych. Przy jej pomocy możemy otrzymywać próbki rozstawione dużo równomierniej, niż średnio biorąc próbki otrzymywane w oparciu o tablice liczb losowych, bez niebezpieczeństw związanych z systematycznym pobieraniem próbek i z możliwością powiększania próbki sztuka po sztuce z zachowaniem owej dość dużej równomierności rozstawienia przy każdej liczności próbki.

Same reszty $\{n\alpha\}$ stanowią ciąg liczb z przedziału jednostkowego, które mogą służyć za argumenty dla przybliżonego całkowania wtedy, gdy nie jest z góry ustalona liczba punktów, w których chcemy obliczać czy też obserwować wartości całkowanej funkcji, natomiast chcemy móc dodawać po jednym punkcie tak, aby mieć *za każdym razem* układ punktów możliwie równomiernie rozstawiony po całym przedziale całkowania. Tablicę reszt $\{n\alpha\}$ nazwał H. Steinhaus tablicą liczb złotych [53], 1956.

Analogiczne własności, jak liczby złote na odcinku jednostkowym, mają tzw. punkty złote w kwadracie jednostkowym. Są one zdefiniowane jako obrazy reszt $\{n\alpha\}$ przez dziewiątkową krzywą Peano, odwzorowującą odcinek jednostkowy na kwadrat jednostkowy w sposób ciągły i z zachowaniem miary. Tablicę 248 takich punktów złotych z objaśnieniem ich użycia do przybliżonego całkowania funkcji określonych na obszarach płaskich podaje L. Zubrzycka [54], 1960.

Niektóre z opisanych tu tablic zostały spopularyzowane w zbiorze [55], 1957, tablic statystycznych, a także w podręczniku [56], 1958–1963.

§ 6. Badanie zespołu cech. W badaniach przyrodniczych, a w szczególności w psychologii, bardzo istotną rolę gra statystyczna analiza zespołu cech. Zwłaszcza w psychologii popularne są badania zwane analizą czynnikową. Hotelling, Spearman, Thurstone są tu najbardziej znanymi nazwiskami. Jednak ze znanymi metodami Spearmana, czy Thurstone'a związane są duże trudności zarówno rachunkowe jak i interpretacyjne. J. Perkal ([57], 1953, [58], 1960) zaproponował pewien rachunkowo prostszy sposób określania faktorów. Metodę tę można z grubsza streścić tak: niech x_{ij}, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, n$, oznacza wartość i-tej cechy u j-tego indywiduum i niech przy tym cechy będą unormowane na średnią zero i wariancję jeden. Jeśli wszystkie cechy są skorelowane nieujemnie, to $m_j = \frac{1}{k}\sum_{i}^{n} x_{ij}$ jest z definicji wartością wspólnego czynnika u j-tego indywiduum a reszty $w_{ij} = x_{ij} - m_j$, zwane wskaźnikami, zawierają resztę informacji o j-tym indywiduum. Jeśli cechy nie są skorelowane nieujemnie, dzieli się je uprzednio na podgrupy cech nieujemnie skorelowanych, a sumaryczny wspólny faktor określa się osobno dla tak wyróżnionych grup cech. Wektor k-wymiarowy, którego wszystkie składowe są m_j w pierwszym przypadku, a który ma jednakowe składowe dla wyróżnionych grup cech w drugim przypadku, reprezentuje z definicji główny wspólny czynnik. To postępowanie nazywa autor metodą pojedynczego wektora. Przez iterowanie powyższego postępowania w stosunku do wskaźników, które po unormowaniu traktuje się jako nowe cechy, uzyskuje się możliwość określania dalszych faktorów. Metoda ta, choć nie prowadzi do wyróżnienia faktorów nieskorelowanych ze sobą i z pojawiającymi się resztami, jak to ma miejsce u metod wcześniej znanych, znalazła uznanie u przyrodników, którzy cenią ją za prostotę rachunków i łatwość interpretacji.

Pewne kwestie związane z wydzielaniem zgodnych, to znaczy nieujemnie skorelowanych, zespołów cech, były rozpatrywane w pracach Szwarca [60], 1960, oraz Kopocińskiego i Zubrzyckiej [61], 1964.

Z analizą zespołów cech związana jest blisko zaproponowana przez J. Perkala definicja podobieństwa indywiduów na tle populacji (zobacz [57], 1953, i [59], 1960), według której dwa indywidua uważa się za podobne wtedy i tylko wtedy, gdy wektory różnic między wartościami cech tych indywiduów i średnimi wartościami cech w populacji są proporcjonalne.

§ 7. Taksonomia wrocławska, podział zbioru na części. Nawiązując do badań przyrodniczych związanych z porządkowaniem indywiduów ogłoszono w 1951 roku (zobacz [62], 1951, i [63], 1951) metodę przedstawiania podobieństw między indywiduami scharakteryzowanymi wieloma cechami przy pomocy dendrytu. Przede wszystkim na podstawie wartości cech u indywiduów oblicza się w jeden z wielu możliwych sposobów odległości między indywiduami. Wybór sposobu obliczania odległości jest ważny i determinuje w znacznej mierze wynik. Proponowana metoda dotyczy sytuacji, kiedy odległości między indywiduami są już ustalone. Pokazuje ona, jak znaleźć najkrótszy spójny zbiór odcinków łączący wszystkie indywidua. Spójny znaczy tu tyle, że między dowolnymi dwoma indywiduami istnieje droga złożona z odcinków należących do zbioru i łącząca te indywidua; najkrótszy znaczy tyle, że wyznacza się spójny zbiór odcinków o najmniejszej sumie długości należących do niego odcinków. Taki najkrótszy spójny zbiór odcinków nie może zawierać łamanej zamkniętej i dlatego jest dendrytem.

Okazało się potem, że zadanie wyznaczania najkrótszego dendrytu było rozpatrywane w literaturze zarówno przed rokiem 1951, jak później, w związku z wyznaczaniem najkrótszej sieci połączeń elektroenergetycznych lub komunikacyjnych (zobacz [64], 1926, [65], 1956, i [66], 1957). Metoda taksonomii wrocławskiej była stosowana wielokrotnie. Prace od [67], 1953, do [71], 1961, są kilkoma przykładami. Zagadnieniami statystycznymi, związanymi z identyfikacją najkrótszego dendrytu między populacjami statystycznymi, gdy odległości między nimi wyznacza się na podstawie próbek, zajął się ostatnio J. Mikiewicz w pracy [74], 1963.

Podstawą dla taksonomii wrocławskiej jest definicja odległości między indywiduami. Pewne kwestie związane z definicją odległości między zespołami ekologicznymi, używaną przez przyrodników, były rozpatrywane przez Marczewskiego i Steinhausa w pracach [72], 1958 i [73], 1959.

Poza poglądowym przedstawieniem podobieństw między indywiduami przy pomocy dendrytu rozpatrywane było również zadanie wyróżniania grup indywiduów na podstawie ich wzajemnych odległości. Blisko z tym związane jest zadanie podziału na części obszaru płaskiego lub bryły przestrzennej, obłożonych masą o pewnej gęstości, tak, aby spełniony był pewien warunek optymalności, np. aby suma momentów bezwładności części względem ich środków ciężkości osiągnęła minimum. Zagadnienie to rozpatrzył z punktu widzenia istnienia rozwiązania i jego jednoznaczności H. Steinhaus [75], 1956, a podział Polski na części w oparciu o gęstość zaludnienia zademonstrował B. Kopociński [76], 1960.

Prace cytowane

[1] J. Perkal, *Zastosowania metematyki do rolnictwa w środowisku wrocławskim*, Zast. Matem. 3 (1956), str. 90-99.

[2] J. Perkal, *O współpracy matematyków i medyków w środowisku wrocławskim*, Zast. Matem. 4 (1959), str. 265-278.

[3] A. Bartkowiakowa i L. Zubrzycka, *O polskich badaniach języka przy użyciu metod matematycznych*, Wiadomości Matem. 7 (1963), str. 87-97.

[4] J. Łukaszewicz, *Przegląd pewnych wyników z dziedziny zastosowań matematyki w Polsce* (po chińsku), Postępy Nauk Matematycznych (czasopismo chińskie), Pekin, 6 (1962), str. 1-62.

Ad § 1

[5] H. Steinhaus, *Sur l'interprétation des resultats statistiques*, Coll. Math. 1 (1948), str. 232-238.

[6] J. Oderfeld, *On the dual aspect of sampling plans*, Coll. Math. 2 (1951), str. 89-97.

[7] H. Steinhaus, *Quality control by sampling* (*A plea for Bayes rule*), Coll. Math. 2 (1951), str. 98-108.

[8] K. Sarkadi, *On the rule of dualism concerning the Bayes probability limits of the fraction defective* (po węg., ze streszcz. ang.), Alkalmazott Matematikai Intezetenek Kozlememenyei II, Budapest 1953, str. 275-285.

[9] H. Steinhaus, *Prawdopodobieństwo, wiarogodność, możliwość*, Zast. Matem. 1 (1954), str. 149-171.

[10] A. Romejko, *Porównywanie dwóch partii towaru*, Zast. Matem. 3 (1957), str. 204-215.

[11] H. Steinhaus and S. Zubrzycki, *On the comparison of two production processes and the rule of dualism*, Coll. Math. 5 (1957), str. 103-115.

[12] H. Steinhaus i S. Zubrzycki, *O porównywaniu dwóch procesów produkcyjnych i zasadzie dualizmu*, Zast. Matem. 3 (1958), str. 229-257.

[13] L. J. Savage i inni, *The foundations of statistical inference*, London-New York 1962, 112 str.

[14] H. Steinhaus, *Probability, credibility, possibility*, Zast. Matem. 6 (1963), str. 341-361.

Ad § 2

[15] O. Morgenstern and J. von Neumann, *Theory of Games and Economic Behavior*, Princeton 1944.

[16] A. Wald, *Statistical decision functions*, New York-London 1950.

[17] H. Steinhaus, *Wycena statystyczna jako metoda odbioru towarów produkcji masowej*, Studia i Prace Statystyczne 1950, str. 1-16.

[18] H. Steinhaus, *Podstawy kontroli statystycznej*, Zast. Matem. 1 (1953), str. 4-25.

[19] H. Steinhaus, *O prognozie*, Zast. Matem. 3 (1956), str. 1-7.

[20] H. Steinhaus, *Ueber einige prinzipielle Fragen der Mathematischen Statistik*, Tagung über Wahrscheinlichkeitsrechnung und mathematische Statistik, Berlin 19-21. X. 1954.

[21] H. Steinhaus, *The problem of estimation*, Annals of Mathematical Statistics 28 (1957), str. 633-648.

[22] S. Trybuła, *On a problem of prognosis*, Bull. Ac. Pol. Sc., Cl. III, 5 (1957), str. 859-862.

[23] S. Trybuła, *O minimaksowej estymacji parametrów rozkładu wielomianowego*, Zast. Matem. 3 (1958), str. 307-322.

[24] S. Trybuła, *Some problems in simultaneous minimax estimation*, Annals of Mathematical Statistics 29 (1958), str. 245-253.

[25] S. Trybuła, *On some loss functions*, Coll. Math. 7 (1960), str. 297-305.

[26] J. Mikiewicz, *Ocena wspólnej przeciętnej na podstawie prób z populacyj normalnych o różnych wariancjach*, Zast. Matem. 6 (1961), str. 119-126.

[27] Dzan Dzo-i, *O minimaksowej estymacji parametru rozkładu dwumianowego*, Zast. Matem. 6 (1961), str. 31-42.

[28] Czen Pin, *O minimaksowym estymatorze liczności populacji*, Zast. Matem. 6 (1962), str. 137-148.

[29] S. Zubrzycki, *O minimaksowym szacowaniu liczności populacji*, Zast. Matem. 7 (1963), str. 183-194.

[30] Czen Pin i Dzan Dzo-i, *O estymowaniu liczności populacji metodą łowienia i znakowania*, Zast. Matem. 6 (1961), str. 51-63.

[31] J. Łukaszewicz i H. Steinhaus, *O mierzeniu przez porównywanie*, Zast. Matem. 2 (1955), str. 225-230.

[32] H. Steinhaus i S. Trybuła, *Pomiar przez kolejne porównywanie*, Zast. Matem. 4 (1959), str. 204-211.

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

[34] J. Oderfeld, *Powierzchnie o wilgotności średniej*, Zast. Matem. 4 (1959), str. 341-349.

[35] S. Zubrzycki, *Some approximate integration formulas of statistical interest*, Coll. Math. 11 (1963), str. 123-136.

[36] Т. А. Шайдаева, *Квадратурные формулы с наименьшей оценкой остатка для некоторых классов функций*, Труды Математического Института АН СССР 53 (1959), str. 313-341.

[37] J. Kiefer, *Optimum sequential search and approximation methods under minimum regularity assumptions*, J. Soc. Industr. Appl. Math. 5 (1957), str. 105-136.

Ad § 3

[38] H. Steinhaus, *Length, shape and area*, Coll. Math. 3 (1954), str. 1-13.

[39] H. Steinhaus, *Długość, kształt i pole*, Prace Matem. 2 (1956), str. 65-78.

[40] J. Perkal, *Geometryczne wskaźniki łąk*, Zast. Matem. 2 (1955), str. 133-149.

[41] J. Perkal, *Sur les ensembles ε-convexes*, Coll. Math. 4 (1956), str. 1-10.

[42] J. Perkal, *On the ε-length*, Bull. Ac. Polon. Sc., Cl. III, 4 (1956), str. 399-403.

[43] J. Perkal, *O długości krzywych empirycznych*, Zast. Matem. 3 (1958), str. 258-284.

Ad § 4

[44] S. Zubrzycki, *O szacowaniu parametrów złóż geologicznych*, Zast. Matem. 3 (1957), str. 105-153.

[45] S. Zubrzycki, *Remarks on random, stratified and systematic sampling in a plane*, Coll. Math. 6 (1958), str. 251-264.

[46] B. Matérn, *Spatial variation*, Meddelanden fran Statens Skogforskningsinstitut 49 (1960), No 5, 144 stronic,

[47] J. Hájek, *Concerning relative accuracy of stratified and systematic sampling in a plane*, Coll. Math. 8 (1961), str. 133-134.

[48] T. Dalenius, J. Hájek and S. Zubrzycki, *On plane sampling and related geometrical problems*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, Berkeley and Los Angeles 1961, str. 125-150.

[49] B. Kopociński, *On the local optimality of some regular sampling patterns in the plane*, Zast. Matem. 8 (1965), str. 1-12.

Ad § 5

[50] H. Steinhaus, *Liczby przetasowane*, Zast. Matem. 2 (1954), str. 34-45.

[51] H. Steinhaus, *Tablica liczb przetasowanych czterocyfrowych*, Rozprawy Matematyczne 6 (1954), 46 stronic.

[52] H. Steinhaus, *Liczby żelazne*, Warszawa GUS, 1956.

[53] H. Steinhaus, *Liczby złote i żelazne*, Zast. Matem. 3 (1956), str. 51-65.

[54] L. Zubrzycka, *O rozmieszczaniu punktów próbkowych na płaszczyźnie*, Zast. Matem. 5 (1960), str. 161-171.

[55] W. Sadowski, *Tablice statystyczne*, Warszawa 1957.

[56] J. Perkal, *Matematyka dla rolników, I-III*, Warszawa, 1958-1963.

Ad § 6

[57] J. Perkal, *O wskaźnikach antropologicznych*, Przegląd Antropologiczny 19 (1953), str. 209-221.

[58] J. Perkal, *On the analysis of a set of characteristics*, Zast. Matem. 5 (1960), str. 35-45.

[59] J. Perkal, *Sur la notion de la ressemblence relativiseé à la population*, Zast. Matem. 5 (1960), str. 281-288.

[60] W. Szwarc, *O pewnym macierzowym zagadnieniu Perkala*, Zast. Matem. 5 (1960), str. 289-297.

[61] B. Kopociński i L. Zubrzycka, *Uwaga o podziale zespołu cech na podzespoły zgodne*, Zast. Matem. 7 (1964), str. 317-321.

Ad § 7

[62] K. Florek, J. Łukaszewicz, J. Perkal, H. Steinhaus et S. Zubrzycki, *Sur la liaison et la division des points d'un ensemble fini*, Coll. Math. 2 (1951), str. 282-285.

[63] K. Florek, J. Łukaszewicz, J. Perkal, H. Steinhaus i S. Zubrzycki, *Taksonomia Wrocławska*, Przegląd Antropologiczny 17 (1951), str. 193-211.

[64] O. Borůvka, *O jistém problému minimálním*, Práce Moravské Přírodovědecké Společnosti 3 (1926), str. 37-58.

[65] J. B. Kruskal, Jr., *On the shortest spanning subtree of a graph and the travelling salesman problem*, Proc. Amer. Math. Soc. 7 (1956), str. 48-50.

[66] P. C. Prim, *Shortest connection networks and some generalizations*, Bell System Techn. J. 36 (1957), str. 1389-1401.

[67] J. Perkal, *Taksonomia Wrocławska*, Przegląd Antropologiczny 19 (1953), str. 82-96.

[68] S. Zubrzycki, *O łańcuszkach gwiezdnych*, Zast. Matem. 1 (1954), str. 197-205.

[69] A. Kelus i J. Łukaszewicz, *Porządkowanie populacji ludzkich według częstości grup krwi*, Przegląd Antropologiczny 20 (1954), str. 24-63.

[70] F. Szczotka, *Porządkowanie i klasyfikacja odmian pszenicy na podstawie ich farynogramów*, Zast. Matem. 2 (1955), str. 123-132.

[71] B. Nowicki, T. Olbrycht i L. Zubrzycka, *Badania nad związkiem cech pokrojowych z użytkowymi metodą taksonomii wrocławskiej*, Zast. Matem. 5 (1961), str. 333-340.

[72] E. Marczewski and H. Steinhaus, *On a certain distance of sets and the corresponding distance of functions*, Coll. Math. 6 (1958), str. 319-327.

[73] E. Marczewski i H. Steinhaus, *O odległości systematycznej biotopów*, Zast. Matem. 4 (1959), str. 195-203.

[74] J. Mikiewicz, *O poziomach ufności w taksonomii wrocławskiej*, Zast. Matem. 7 (1963), str. 1-40.

[75] H. Steinhaus, *Sur la division des corps materiels en parties*, Bull. Ac. Polon. Sc., Cl. III, 4 (1956), str. 801-804.

[76] B. Kopociński, *O podziale terytorium Polski na części*, Zast. Matem. 5 (1960), str. 173-177.

Praca wpłynęła 29. 1. 1965

С. ЗУБЖИЦКИ (Вроцлав)

О НЕКОТОРЫХ РЕЗУЛЬТАТАХ СЕМИНАРА ПО ПРИЛОЖЕНИЯМ МАТЕМАТИКИ

РЕЗЮМЕ

Автор даёт осмотр некоторых достижений семинара по приложениям математики, введенного с 1948 года во Вроцлавском Отделении Математического Института Польской Академии Наук. Настоящая работа является расширением доклада читанного автором 11 февраля 1964 года на 500-ом заседании семинара.

S. ZUBRZYCKI (Wrocław)

ON SOME RESEARCH WORK OF THE SEMINAR FOR APPLICATIONS OF MATHEMATICS

SUMMARY

The author gives a review of some results obtained and discussed on the Seminar for Applications of Mathematics held since 1948 at the Mathematical Institute of the Polish Academy of Sciences in Wrocław. The paper is an extended version of the authors address presented to the 500-th meeting of the seminar on February 11, 1964.