H. STEINHAUS (Wrocław)

LICZBY PRZETASOWANE

W praktyce statystycznej nieraz posługujemy się losowaniem. Tak na przykład, gdy trzeba wnosić o populacji z próbki, musimy w jakiś sposób zdecydować, które indywidua (sztuki) mają wejść do próbki i podlegać badaniu. Niech populacja składa się z N sztuk zaopatrzonych w numery fabryczne od 1 do N (np. od 1 do 10000); chcemy pobrać z niej n sztuk (np. 25) w sposób bezstronny w celu obiektywnego oznaczenia wartości całej populacji (partii) z wyniku zbadania wartości 25-sztukowej próbki. Gdyby N było małe (np. $N=100$), można by wypisać na N jednakowych biletach liczby od 1 do N, przetasować gruntownie bilety, wyłożyć pierwszych 25 biletów i pobrać z partii sztuki zaopatrzone w numery tych biletów. Ten sposób już dla $N=1000$ staje się bardzo niewygodny. Dla uniknięcia mozołu posługują się statystycy tablicami tak zwanych *liczb przypadkowych*. Wyobraźmy sobie kostkę graniastą o 10 ścianach zaopatrzonych w napisy $0,1,2,\ldots,9$, której oś pozwala puścić ją w ruch wirowy. Za każdym upadkiem notujemy cyfrę widoczną na górnej ścianie. Każda czwórka kolejnych rzutów da nam liczbę czterocyfrową (np. 8205 lub 0734 itd.). Te liczby ustawione w wiersze i kolumny dają tablicę „liczb przypadkowych", z której możemy je odczytywać w tej kolejności, w jakiej powstały; 40000 rzutów utworzy tablicę 10000 liczb czterocyfrowych. Wprawdzie trud zbudowania takiej tablicy jest znaczny, ale jest jednorazowy: po wydrukowaniu tablicy praktyk, który chce wybrać beztendencyjnie 25 numerów czterocyfrowych, otwiera książeczkę na chybił-trafił i — zaczynając w dowolnym miejscu — odczytuje 25 kolejnych liczb.

Znanych jest kilka tablic liczb przypadkowych, np. tablica R. A. Fishera i F. Yatesa. Tablica umieszczona w Polskiej Normie PN/N-03001 jest przedrukiem części tablicy wydanej w 1946 r., której autorami są M.G. Kendall i S. B. Babington. Trudność w ułożeniu takiej tablicy polega na zapewnieniu jednakowej szansy dla każdej z cyfr 0-9 przy wyznaczaniu każdej kolejnej cyfry i na zapewnieniu niezależności kolejnych rzutów. W tym celu proponowano wyrafinowane metody, na przykład użycie liczników Geigera-Müllera: aparat składałby się z 10 jednakowych rur, po jednej dla każdej cyfry, a cząsteczka kosmiczna wpadając do rury wyzwalałaby przekaźnik i doprowadzała prąd do odpowiedniego klawisza maszyny piszącej na taśmie. Jest to zupełne uwolnienie się od interwencji człowieka.

Nie wchodząc w technikę, której użyto w celu zachowania warunków równouprawnienia cyfr i niezależności kolejnych wyborów, i przyjmując, że przy konstrukcji tablicy spełniono obydwa te postulaty z precyzją dostateczną dla praktyka, zwróćmy uwagę na założenia tkwiące w samym pomyśle takiej tablicy:

1° Zamiast odwoływać się do przypadku przy każdym pobieraniu próbki, dopuszczamy go do głosu tylko przy konstrukcji tablicy; potem będzie on już tylko powołany do skromnej roli wybierania początkowego miejsca w tablicy.

2° Tablica złożona z N liczb nie zawiera N *różnych* liczb. Tak więc np. tablica złożona z 10000 liczb czterocyfrowych nie zawiera wszystkich liczb od 0000 do 9999: rachunek prawdopodobieństwa uczy, że przy konstrukcji tablicy należy oczekiwać 6321 różnych liczb; pozostałe 3679 miejsc w tablicy wypełnią dublety, tryplety, itd. owych liczb, spośród 10000 liczb czterocyfrowych zaś 3679 liczb nie pojawi się w tablicy (nie będzie tak dosłownie: obliczenie daje tylko rezultaty oczekiwane, a nie pewne).

Własność 1° ma znaczenie teoretyczne — nie widać bezpośrednio jej praktycznych konsekwencji. Własność 2° uniemożliwia bezpośrednie korzystanie z tablicy, co łatwo zobaczyć na przykładzie: gdy z partii obejmującej 10000 sztuk chcemy pobrać 25 sztuk próbnych, musimy się liczyć z ewentualnością, że dostawca mający drugi egzemplarz drukowanej tablicy złożonej z 10000 liczb czterocyfrowych umieści wszystkie złe sztuki pod tymi numerami, których w tablicy nie ma (może to zrobić nawet wtedy, gdy takich sztuk jest np. 30%). Dlatego przepisy dołączone do tablic każą brać 25 liczb kolejnych poczynając od dowolnie obranego miejsca i drugich 25 kolejnych liczb od innego miejsca, wypisać je w dwóch kolumnach obok siebie, przekreślić w pierwszej kolumnie trzecie i czwarte cyfry, w drugiej pierwsze i drugie i złożyć z pozostałych cyfr 25 nowych liczb czterocyfrowych. Utrudnia to użycie zwykłych tablic, np. tablicy PKN. Można zapytać, jaki jest cel budowania tablic o własności 2°. Wzięły się one z takich planów badania, które obliczono przy założeniu, że badany kolektyw nie ulega zmianie wskutek badania. To założenie wymaga zwracania sztuk zbadanych do partii, tak by zachowały szansę pojawienia się ponownego. W przykładzie 25 biletów należy zatem każdy wyciągnięty bilet wkładać z powrotem do talii — gdy pojawi się po raz wtóry, to notuje się po raz wtóry obok numeru wynik starego badania, tak jak gdyby to była nowa sztuka. Ten sposób, z pewnością dziwny dla praktyka, przyjął się dlatego, że matematyczna teoria niezmiennego kolektywu jest prostsza od teorii kolektywu ulegającego zmianie podczas i wskutek badania; znalazł on wyraz w tablicach „liczb przypadkowych”, które będziemy odtąd nazywać *tablicami liczb bezładnych* (TLB). Praktyk ma rację, gdy uważa taki sposób za wadliwy. Rzeczywiście, do niepewności wynikającej z tego, że od przypadku zależy, jakie sztuki wejdą do próbki, dołącza się tu niepewność, ile sztuk do niej wejdzie — może ich wejść 25, ale i mniej.

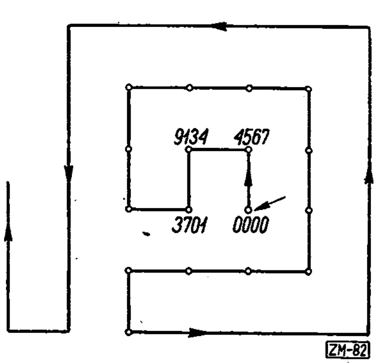
Pokażemy — bez wchodzenia w szczegóły — na przykładzie wyceny, jak można zastąpić plan badania o wyżej wytkniętej wadzie przez plan od niej wolny. Niech umowa orzeka, że o wartości partii złożonej z 10 000 sztuk decyduje wartość próbki ze 100 sztuk, pobranej według tablicy liczb bezładnych. Nietrudny rachunek dowodzi, że wśród 100 liczb tablicy liczb bezładnych złożonej z 10 000 liczb czterocyfrowych należy oczekiwać 99 różnych. Wobec tego zmieńmy umowę: zamiast 100 numerów z tablicy liczb bezładnych będziemy brali 99 różnych numerów i będziemy wyceniali partię według wartości próbki odpowiadającej tym 99 numerom. Wyobraźmy sobie, że przez wiele lat systematycznie postępujemy tak z partiami, które nadchodzą np. co tydzień. Jaka jest różnica praktyczna między starą a nową umową? Koszt badania jest ten sam, bo teraz badamy *zawsze* po 99 sztuk, a dawniej badaliśmy *średnio* 99 sztuk. Prawdziwa informacja, którą kupujemy za pieniądze wydane na badanie, jest też średnio ta sama, bo średnia ilość sztuk zbadanych nie zmieniła się. Ale przy dawnym sposobie pewne partie badaliśmy dokładniej, inne mniej dokładnie, bo liczba sztuk naprawdę zbadanych była co tydzień inna, a teraz jest stała. Wobec tego obecnie wszystkie partie wyceniamy z jednakową precyzją, a dawniej wycenialiśmy niektóre partie dokładniej, inne mniej dokładnie lub tak samo dokładnie jak obecnie; co gorsza, jedynie kaprys losu wpływał na nierówność traktowania. Jest jasne, że zmiana umowy ulepszyła ją pod każdym względem.

Nasuwa się prosty pomysł. Czy nie można by wydrukować tablicy, która by zawierała wszystkie liczby od 0000 do 9999, każdą raz i tylko raz? Taka tablica służyłaby przy planach ulepszonych według zasady przed chwilą wyjaśnionej: moglibyśmy z niej brać 99 kolejnych liczb (zaczynając w dowolnym miejscu) i mielibyśmy same różne; oczywiście składanie liczb z dwóch kolumn potrzebne przy tablicach liczb bezładnych będzie teraz zbędne: dostawca nie może teraz schować złych sztuk pod numerami, których brak w tablicy, gdyż (przy partiach po 10 000 sztuk) tablica ma wszystkie liczby czterocyfrowe, a więc numery wszystkich sztuk. Taką ulepszoną tablicę będziemy nazywali *tablicą liczb przetasowanych* (TLP).

Można by taką tablicę otrzymać drogą losowania, byłoby to jednak żmudne, gdyż trzeba by przy każdej nowej liczbie czterocyfrowej badać, czy już nie było jej wcześniej. Dlatego zdecydowałem się na radykalne rozwiązanie, które polega na zupełnie określonym procesie matematycznym. Kolejność liczb przetasowanych nie weźmie się z żadnego aparatu fizycznego ani z innego źródła niematematycznego, nad którym nie panujemy i które nie pozwala określić wynik jednoznacznie. W tym miejscu krytyczny czytelnik zauważy, że sporządzenie tablicy liczb czterocyfrowych, zawierającej każdą z nich raz i tylko raz i określonej matematycznie, jest bardzo proste: wystarczy napisać liczby 0000,0001,...,9999 w naturalnej kolejności. Dlaczego musimy odrzucić ten projekt? Aby na to pytanie odpowiedzieć, trzeba uprzytomnić sobie cel, któremu służą tablice. Zwykle przy produkcji masowej automat produkcyjny numeruje sztuki towaru w kolejności ich powstawania. Wyobraźmy sobie, że automat się zepsuł; zanim to spostrzeżono, wyrzucił kilkaset sztuk wadliwych, po czym go naprawiono i znowu produkował same dobre sztuki. Gdy zastosujemy tablicę liczb kolejnych i pobierzemy według niej próbkę 25-sztukową, to będą one prawdopodobnie albo niemal wszystkie dobre, albo niemal wszystkie złe; prawdopodobieństwo, że próbka da obraz partii, czyli — jak się mówi — będzie *reprezentatywna*, będzie małe. Jest jasne, że np. przy wycenie partii według próbki nie można polegać na próbce złożonej ze sztuk kolejno numerowanych. Stąd właśnie wziął się pomysł liczb przypadkowych. Powiedziano sobie, że trzeba w tablicy wyłączyć wszelki rytm, który by mógł odpowiadać rytmowi zjawiska badanego statystycznie. Tak np., gdyby w tablicy na niektórych stronicach znalazły się same liczby parzyste a na innych same nieparzyste, można by się obawiać, że próbki nie będą reprezentatywne; łatwo sobie wyobrazić w automacie wytwórczym defekt, który psuje właśnie co drugą sztukę; to dałoby nam próbki złożone bądź z samych dobrych, bądź też z samych złych sztuk, zależnie od stronicy, z której odczytalibyśmy numery. Idąc po tej drodze doszło się do poglądu, że tablice przydatne do losowania próbek można uzyskać tylko przez wylosowanie liczb i wydrukowanie ich w tej kolejności, w jakiej los je ustawił. Chcąc zapewnić każdej sztuce jednakowe prawdopodobieństwo wejścia do próbki użyto bądź to takich aparatów do losowania, które zapewniają każdej z liczb (np. z liczb 0000-9999) równe prawdopodobieństwo znalezienia się w tablicy, bądź też wzięto liczby z innych źródeł, w których założenie *jednakowego prawdopodobieństwa* jest naturalne. Postarano się także o to, by wynik któregokolwiek losowania nie wpływał na wynik innego. Ta cecha *niezależności* jest istotna, gdy dążymy do reprezentatywności próbek.

Temu rozumowaniu można postawić poważne zarzuty. Nieporozumienie polega na złudzeniu, jakoby grę przypadku można było uchwycić na gorącym uczynku i utrwalić na papierze w postaci ciągu liczb przypadkowych. Jest to iluzja, bo z chwilą, gdy ciąg zostaje *wylosowany*, przestaje być *losowy*. Nie trzeba zbytniej finezji, żeby to udowodnić. Zanim zabrano się do budowy tablicy liczb bezładnych, każda z liczb 0000-9999 miała tę samą szansę wejścia do tablicy, co każda inna, ale gdy tablicę wydrukowano, zabrakło w niej paru tysięcy liczb a inne pojawiły się wielokrotnie; szanse przestały być równe i — jak już wspomnieliśmy wyżej — utrudniło to używanie tablic liczb bezładnych w praktyce. Ponadto niezależność, która tkwiła w aparacie (np. w kostce 10-ściennej) zginęła w tablicy: po liczbie 6345 następuje w niej 7099 i tylko ta liczba, natomiast aparat mógł po 6345 dać równie dobrze każdą liczbę czterocyfrową jak każdą inną. Toteż zdarzały się przypadki przy konstrukcji tablic, że wydawcy poprawiali je według swego gustu, uważając, że w niektórych miejscach nadmiar liczb parzystych lub nieparzystych nie zgadza się z tym, czego oni oczekiwali. Krótko mówiąc, ciąg przypadkowy był zanadto przypadkowy. Te sprzeczności nie są jednak niczym innym jak wyrazem faktu, że albo żaden, albo każdy ciąg 10000 liczb czterocyfrowych zasługuje na miano przypadkowego, gdyż dla idealnego aparatu do losowania nie ma żadnej różnicy między którymkolwiek ciągiem a innym.

Wszystkie wyżej podane wątpliwości skłoniły mnie do zastąpienia tablicy liczb bezładnych przez *tablicę liczb przetasowanych*. Ograniczyłem się do liczb czterocyfrowych. Wybrałem liczbę pierwszą niezbyt odległą od 5000 i łatwą do zapamiętania: 4567. Jej wielokrotności, to znaczy liczby $4567n$, otrzymane, gdy n przebiega liczby $0,1,2,3,\ldots,9999$, są po kolei $0000, 4567, 9134, 3701, \ldots, 5433$ — pisząc je ograniczamy się do ostatnich 4 cyfr (np. $4567\times 3=13\,701$; piszemy tylko 3701). Łatwo wykazać, że otrzymano tym sposobem wszystkie liczby od 0000 do 9999, każdą tylko raz. Tę pierwszą permutację wykonano za pomocą maszyny do rachowania: wystarczyło kolejno dodawać 4567. Liczby $4567n$ ustawiono na arkuszu podzielonym na 100×100 okienek, zaczynając od środka arkusza i idąc „drogą cygańską" (rys.1) ku brzegom. Na lewym górnym rogu arkusza znalazła się liczba 9034. Od niej rozpoczęto zbierać liczby z arkusza i przepisywać na karty, których przygotowano 25:



Rys. 1

po jednej karcie dla 400 liczb czterocyfrowych. Każda karta miała 40 wierszy, w każdym wierszu było miejsce na 10 liczb. Zbieranie z arkusza odbyło się według pewnej zasady iteracyjnej, która działa tasująco, jak procesy ergodyczne. Wpisano mianowicie liczbę 9034 w miejsce początkowe pierwszego wiersza karty; ponieważ 9034 ma środkowe cyfry 03, więc na arkuszu poszukano okienka o 10 kolumn dalej na prawo i o 3 wiersze niżej. Znaleziono w nim liczbę 2745 — wpisano ją na karcie w pierwszym wierszu obok 9034 (tj. w drugiej kolumnie karty); trzecią z kolei liczbę znaleziono na arkuszu o 7 kolumn na prawo i 4 wiersze poniżej liczby 2745 — wpisano ją obok 2745, to jest w trzeciej kolumnie pierwszego wiersza karty. Tak postępowano dalej, przekreślając za każdym razem okienko, z którego pobrano liczbę. Przy tym przy odliczaniu kroków poziomych uważano okienko przekreślone za istniejące, przy odliczaniu kroków pionowych na arkuszu za nieistniejące. Gdy przepis zmuszał do szukania liczby w kolumnie arkusza, której wszystkie pozycje już zostały wcześniej przekreślone, wracano do pierwszego wiersza i brano pierwszą z kolei nieprzekreśloną pozycję. Nie możemy tutaj wchodzić w szczegóły, chcemy jednak wytłumaczyć, dlaczego właśnie cyfry środkowe b, c liczby *abcd* wybrano do określenia liczby następującej po *abcd*. Wiadomo mianowicie, że określenie, jaka jest np. pierwsza lub ostatnia cyfra liczby 3^{100}, jest łatwe, gdyż pierwsze zadanie polega na oszacowaniu wielkości (np. logarytmowaniem), a drugie na znalezieniu okresu w ciągu reszt *modulo* 10 kolejnych potęg trójki. Trudne natomiast jest określenie cyfr bliskich środka. Dlatego wyróżniliśmy właśnie cyfry środkowe; chcieliśmy uniknąć związków między liczbą a następną, które podlegałyby widocznym regułom matematycznym.

Dalszy krok polegał na pocięciu wszystkich kart na kolumny, których powstało w ten sposób 250; kolumny pierwszej karty otrzymały numery od 1 do 10, drugiej od 11 do 20,..., ostatniej od 241 do 250. Te kolumny sklejono w nowe karty biorąc co 17-tą kolumnę. Tak powstała karta I złożona z kolumn, które przedtem miały numery 1, 18, 35,..., 137, 154, karta II z kolumn o dawnych numerach 171, 188, 205, 222, 239, 6, 23, 40, 57, 74 itd. Cały maszynopis objął 25 kart I—XXV. Skontrolowano go, czy zawiera każdą liczbę od 0000 do 9999, usunięto kilka błędów i przepisano, po czym skontrolowano ponownie. Ten drugi maszynopis nadawał się już do bezpośredniej reprodukcji typograficznej, dzięki czemu można było go wydrukować bez interwencji zecera i korektora. Dołączono do niego opis i sposób użycia i ogłoszono *Tablicę liczb przetasowanych czterocyfrowych* jako zeszyt VI Rozpraw Matematycznych. Przezwyciężenie początkowych trudności technicznych i innych zawdzięczam pomocy pp. M. Kusiakowej i L. Jakowlewy-Zubrzyckiej z Ogólnej Grupy Zastosowań I. M. PAN oraz p. J. Battka, asystenta Politechniki Wrocławskiej.

Zanim podamy sposoby posługiwania się Tablicą liczb przetasowanych — pierwszą w ogóle tablicą tego rodzaju — odpowiemy na pytania, które nam postawiono, a które zapewne nasuną się także i czytelnikowi: Jakie są kryteria, którymi posługiwaliśmy się przy wyborze

TABLICA LICZB PRZETASOWANYCH (II)

4430	2524	8665	8807	0360	4538	3552	6259	1323	2709
6556	2073	2624	6162	7308	1669	0762	0160	6854	3696
3102	3968	4613	2439	1074	4783	0089	3772	7506	3087
2955	4742	6113	4003	4111	6258	5028	5931	6231	6875
8910	1085	2486	3872	4881	7126	7227	8848	5251	8277
3857	2048	0064	9655	7068	7626	8536	8602	1714	2765
1997	8329	7327	4622	4593	5756	9451	7014	3670	7253
6137	6848	1061	9126	4415	2719	8421	3514	9416	8633
9393	0116	7309	4800	8892	9529	3057	4443	1086	6476
9982	9738	8508	3549	2629	1052	7820	0648	3708	9341
2535	2056	1986	7727	4766	5112	5944	8038	5966	6451
3115	0226	6404	6040	4371	6056	3535	0229	0959	7241
5948	5189	8265	0761	3706	8595	3606	1181	7422	6706
6235	4726	6292	4987	0552	2447	4961	2262	3673	1175
3551	4222	7909	5402	2265	3091	8071	1212	3414	7943
7638	7783	8981	8616	1781	1453	8309	9044	4877	9413
0602	5964	9928	0026	5463	6443	0032	1187	8758	2197
5879	9979	8771	0640	4553	8845	7643	6485	7843	9453
0525	6311	4235	3641	3356	9379	1592	9092	8860	9764
5135	0930	0722	2435	0739	6831	0158	8858	2492	3853
0119	7673	1989	0618	0588	1494	9737	3271	9557	4643
1605	2750	7758	4367	9317	8928	4148	0605	7239	6321
9015	9056	3395	4757	1191	1195	7825	3172	0000	0665
4232	9819	7765	5360	0318	5119	9267	6446	0899	5603
6079	1709	8004	5414	7786	4542	8385	2523	1953	6635
7109	0270	4846	7144	1979	0907	5792	1155	2842	9495
0307	4569	2171	1442	1388	4463	8098	6740	4998	7399
6655	4273	4066	6194	5507	7343	5405	2050	1139	1726
9254	3077	2584	5083	0612	9293	4230	4760	7476	1864
6755	5369	4137	2760	3255	9919	2688	1001	7615	0696
1100	6440	9847	2606	9310	3217	0336	5718	2239	5989
6951	0369	4731	8872	2248	9229	9976	3114	1616	5608
5243	2139	6906	2329	1573	3189	6043	4540	8366	5356
4050	1521	3386	2218	7355	7686	0967	1157	0431	7642
6529	6648	8479	6725	4121	2313	8111	3118	4990	4390
8308	1837	6768	9420	8374	5098	5511	3451	5326	1675
9615	4680	0215	8692	4747	8838	5754	8607	9534	2603
0069	1114	3823	7927	9904	0977	6296	1820	6419	8933
7387	6313	5049	8462	2987	3184	8964	4639	1814	3993
6335	9850	2193	0759	3206	7985	9653	0403	8935	4495

takiego właśnie systemu tasowania, jaki przyjęliśmy za podstawę tablicy? Na czym opieramy mniemanie, że w tablicy nie znajdą się prawidłowości matematyczne, które nie tylko pozwolą odgadnąć jej sztuczny charakter, ale doprowadzą do błędów lub nadużyć przy jej stosowaniu? Na żadne z tych pytań nie odpowiemy w sposób zadowalający matematyka, gdyż żadne z nich nie jest i nie może być sformułowane matematycznie. Wybierając system tasowania postanowiliśmy skomplikować go do tego stopnia, żeby uniknąć wszelkich ujemnych skutków niedostatecznego przetasowania. Można było komplikację posunąć jeszcze dalej, ale kryterium praktyczne orzeka, że zbędne jest dalsze tasowanie, jeżeli nikt po obejrzeniu tablicy nie zdoła wskazać prawidłowości, które mogłyby mieć ujemne skutki w praktyce. Wykonano następującą próbę: z jednej karty TLP utworzono tablicę liczb bezładnych obracając kartę o 90°; utworzono w podobny sposób tablicę liczb bezładnych z jednej ze *znanych* tablic i zwrócono się do statystyków z żądaniem, żeby odróżnili tablicę „sztuczną" od „przypadkowej": nie otrzymano odpowiedzi.

TLP składa się z 25 *kart*. Na każdej karcie jest 400 *liczb* czterocyfrowych, tj. 1600 cyfr wypełniających 40 *wierszy* i 40 *rzędów*. Każde 4 wiersze tworzą *strofę*, każde 4 rzędy *kolumnę*. Każde przecięcie strofy z kolumną tworzy *blok* złożony z 16 cyfr; na każdej karcie jest 100 bloków.

Użytek główny TLP polega na pobraniu z niej k różnych od siebie liczb czterocyfrowych. W tym celu należy wziąć na dowolnej karcie dowolną liczbę za początkową, a następnie czytać tablicę dalej sposobem książkowym, aż uzbieramy k liczb. Niech np. $k=60$. Weźmy kartę II, a na niej liczbę 8845 — jest to druga (od góry) liczba w piątej strofie i szóstej kolumnie. Czytamy po kolei: 8845, 7643, 6485, 7843, 9453, 0525, ..., 9764, 5135, ..., 3853, 0119, ..., 4643, 1605, ..., 6321, 9015, ..., 0665, 4232, 9819, 7765, 5360, 0318 (wypisaliśmy wyraźnie pierwszych 5 i ostatnich 5 liczb, a z pięciu wierszy między nimi tylko liczby początkowe i końcowe, zaznaczając inne kropkami). Gdyby trzeba było pobrać 600 liczb ($k=600$), musielibyśmy po wyczerpaniu karty II przejść na kartę III. Gdybyśmy byli wybrali za początkową jakąś liczbę karty XXV, to — przy $k=600$ — musielibyśmy po wyczerpaniu karty XXV przejść na kartę I, mianowicie na liczbę 9034, która stoi w pierwszej kolumnie i pierwszym wierszu, i czytać dalej wiersz za wierszem aż do zebrania 600 liczb.

Nie podajemy żadnych reguł określających liczbę początkową, jakkolwiek przepis, że znajduje się ją w dowolnym miejscu, jest niezadowalający. Można by uzależnić to miejsce od wskazań zegarka, a mianowicie numer karty od stanu sekundnika, numer wiersza od wskazań minutowych itd., ale taki przepis wymagałby zbyt wielu komentarzy.

Zauważmy, że w przykładzie wyżej podanym po liczbie 8845 następuje 7643. Przed pocięciem 25 pierwotnych kart liczba 8845 należała do szóstej, a liczba 7643 do 23 kolumny; oznacza to, że 7643 powstała z 8845 przez 797 operacyj, które prowadziły od liczby do liczby przy zdejmowaniu ich z pierwotnego arkusza. Ta uwaga pokazuje zupełną praktyczną niemożliwość wykrycia jakichś prostych związków między kolejnymi liczbami TLP. Nawet kompletna znajomość przepisu konstrukcyjnego TLP nie pozwoliłaby obliczyć w ciągu doby jaka liczba stoi obok 8845.

Zamiast czytać TLP książkowym sposobem, można ją czytać przeciwnie, to znaczy w lewo i w górę. Zamiast czytać liczby wprost można je czytać wspak, bo tym sposobem znowu otrzymamy każdą liczbę czterocyfrową i każdą tylko raz. Można także kombinować różne sposoby czytania liczb z różnymi sposobami posuwania się w tablicy. Można czytać liczby w kolejności kolumnowej: na karcie II po 8845 nastąpi 9379, 6831, 1494,..., 7985, a potem najwyższa liczba szóstej kolumny karty III itd. Oczywiście czytanie wspak i posuwanie się w górę jest dozwolone także w kolejności kolumnowej. Otrzymujemy więc kilkanaście sposobów czytania TLP, odnoszących się do głównego jej użytku.

Praktyka wymaga nieraz zmniejszenia tablicy do liczb *trójcyfrowych*, a czasem powiększenia do *pięciocyfrowych*, a nawet sześciocyfrowych.

Pierwsze z tych zadań rozwiązujemy w ten sposób, że ograniczamy się do liczb zaczynających się od 0. Wyłowienie takich liczb z TLP jest szczególnie łatwe przy kolejności kolumnowej. Tak np. w kolumnie liczby 8845 znajdujemy 0907, 0977, a na następnej karcie 0712, 0334, 0574, 0320, 0462, 0586, 0467, 0746 — tych 10 liczb czytamy jako 907, 977, 712, ..., 746; oczywiście w razie, gdy $k>10$, użyjemy dalszych kart. Moglibyśmy także użyć do tego samego celu liczb zaczynających się od 1 lub innej cyfry, jednak ta cyfra musi być ustalona przed rozpoczęciem pobierania liczb trójcyfrowych.

Powiększenie TLP do liczb *pięciocyfrowych* wymaga wyboru k liczb czterocyfrowych (jednym ze sposobów już opisanych) i dopisaniu **na początku** każdej liczby *nowej* pierwszej cyfry; tę cyfrę bierzemy z dowolnego rzędu TLP zaczynając od dowolnej cyfry i posuwając się pionowo w dół. Ten sposób daje same różne liczby pięciocyfrowe i każda liczba pięciocyfrowa może być wybrana, jeżeli jednak weszła do próbki sztuka nr 76435, już nie wejdzie do niej sztuka nr 86435 ani nr 66435 itd. Praktycznie ta zależność jest zwykle bez znaczenia. W większości przypadków można nawet dopisywać przed pierwszą cyfrą ostatnią z bloku sąsiadującego po lewej. Tak np. na karcie II znajdziemy w miejscu 8845 liczbę 38845 i posuwając się pionowo w dół kolejno liczby 69379, 96831, 81494, ..., 78838, 40977 — zatrzymaliśmy się o dwa wiersze od dolnego brzegu karty, tzn. pobraliśmy 21 liczb. Podkreślamy jednak, że jakkolwiek ten łatwy sposób daje same różne liczby, pomija on z góry większość liczb pięciocyfrowych. I tutaj też można by osłabić ewentualny wpływ tej wady (która jest bez znaczenia poza odbiorem towarów), ale nie chcemy wchodzić w te drugorzędne szczegóły. Pozostawiamy też czytelnikowi rozszerzenie tablicy do liczb sześciocyfrowych, co wymaga sposobów analogicznych do tych, które dały liczby pięciocyfrowe.

TLP może być czytana także jak TLB, to jest jako tablica liczb bezładnych czterocyfrowych. Wystarczy w tym celu czytać bloki pionowo. Tak np. zaczynając od bloku w szóstej strofie i szóstej kolumnie karty II odczytamy kolejno liczby 1815, 4911, 9291, 4859, 9479, ..., 3153, 6706, ..., 3335; zatrzymaliśmy się na końcu karty i znaleźliśmy 180 liczb czterocyfrowych. Tym razem już nie możemy gwarantować różności tych liczb. Jeżeli potraktujemy tak otrzymaną TLB jako powstałą z losowania z urny zawierającej liczby od 0000 do 9999 ze zwracaniem numerów wylosowanych do urny (tak jak chcą autorowie tablic liczb bezładnych), to będziemy mogli obliczyć oczekiwaną liczbę L *różnych* liczb na danej karcie (np. na karcie II) jak następuje:

Prawdopodobieństwo, że pierwsza liczba karty, c_1, nie powtórzy się jest $\left(\frac{9999}{10000}\right)^{399}$ — nazwijmy je p. Ale p jest także prawdopodobieństwem, że nie powtórzy się druga liczba karty, c_2, zatem oczekiwana liczba liczb na karcie, które się na niej powtarzają, jest

$$400p = 400\left(1-\frac{1}{10000}\right)^{10000\cdot 0,0399} \approx 400\,e^{-0,04} \approx 384,3.$$

Wobec tego należy oczekiwać na karcie 15,7 dubletów, trypletów etc. Ale nietrudne oszacowanie pokazuje, że trypletów i wyższych repetycji należy oczekiwać razem mniej niż 1, tak że można je pominąć i oczekiwać na karcie 7 do 8 par liczb jednakowych, co daje $L = 391$ lub 392. Szukanie par na karcie nie wymaga więcej niż godzinę czasu — eksperymenty na kilku kartach dały liczby par mało odbiegające od spodziewanych. Celem takich prób jest stwierdzenie, że TLP czytana pionowo nie różni się empirycznie od tablicy liczb bezładnych — widzimy w tym argument za naszą tezą, iż liczby w TLP są dostatecznie przetasowane. Ale to odnosi się do badania poszczególnych kart, natomiast mógłby ktoś naszą całą tablicę liczb bezładnych odróżnić od zwykłej tablicy liczb bezładnych, gdyby zadał sobie trud porachowania wszystkich cyfr. Okazało by się, że jest w naszej tablicy dokładnie 4000 zer, tyleż jedynek, tyleż dwójek itd., a w zwykłej tablicy najprawdopodobniej te liczności byłyby wprawdzie bliskie 4000, ale nie wszystkie dokładnie równe 4000. Z naszego punktu widzenia ten sztuczny charakter naszej tablicy liczb bezładnych wcale nie jest wadą, nie widzimy bowiem, jaką zaletą mogłaby się okazać w praktyce przewaga frekwencji pewnych cyfr w tablicy: przeciwnie, gdy przy antagonistycznym użyciu tablicy jedna ze stron wie o przewadze np. cyfry 7, a druga nie, to tablica może stać się źródłem nadużyć, jak źle scentrowana ruleta; jeżeli żadna ze stron o tej przewadze nie wie, to może nastąpić systematyczne krzywdzenie jednej ze stron bez winy drugiej.

Nadmieńmy jeszcze, że do użytku TLP jako TLB nadaje się nie tylko czytanie bloków w dół, ale także w górę, a nawet na przemian.

Można by zapytać, dlaczego uważamy możliwość używania TLP jako TLB za zaletę tablicy, skoro na wstępie skrytykowaliśmy istniejące tablice liczb bezładnych? Po prostu dlatego, że mogą się znaleźć zwolennicy planów badania, ogłoszonych przez PKN lub inne instytucje, którzy chcą się trzymać ich ściśle i albo nie umieją, albo nie chcą ulepszyć owych planów w sposób wskazany przez nas na przykładzie wyceny. Nasza tablica dająca alternatywny sposób czytania zadowoli zarówno tych, których nasze wywody przekonały jak tych, którzy — z różnych względów — za nimi nie pójdą.

Streśćmy rezultaty niniejszego studium. Wykazaliśmy możliwość efektywnego określenia i praktycznej konstrukcji tablicy zawierającej każdą liczbę od 0000 do 9999 raz i tylko raz, bez uciekania się do sposobów pozamatematycznych. Wykazaliśmy, że taka tablica nadaje się lepiej do rozwiązywania zagadnień praktycznych niż dotychczasowe tablice wynikłe z losowania: choćby przez to, że daje numery czterocyfrowe bezpośrednio. W końcu wskazaliśmy prosty sposób interpretacji TLP jako tablicy liczb bezładnych, przy czym tak uzyskana TLB ma pewne zalety, których zwykłe TLB są pozbawione, a co więcej, jest określona — jak TLP — bez wprowadzania elementów niematematycznych.

INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK

Praca wpłynęła 26. 6. 1953 r.

Г. ШТЕЙНГАУЗ (Вроцлав)

ПЕРЕТАСОВАННЫЕ ЧИСЛА

РЕЗЮМЕ

До сих пор т. н. таблицы случайных чисел составлялись не математическим путем, например при помощи таких физикальных процессов как метание игральной кости.

Можно однако составить иные таблицы, названные здесь таблицами перетасованных чисел, без содействия чуждых математике приёмов. Предлагаемая работа является комментарием к такой таблице изданной Математическим Институтом Польской Академии Наук, а состоящей из всех четырех-цыфровых систем

с 0000 по 9999. Каждая такая система выступает один лишь раз, а перетасовка заключается в некотором итерационном процессе. Таблица отличается от таблицы случайных чисел содержащих те самые системы тем, что в таблице случайных чисел некоторые системы выступают больше раза, а некоторые не выступают вовсе. Устранение этого недостатка облегчает пользование этой таблицей в практике. Кроме того вертикальное чтение этой таблицы дает нам таблицу случайных чисел.

H. STEINHAUS (Wrocław)

SHUFFLED NUMBERS

SUMMARY

The so-called tables of random numbers have been obtained so far in an unmathematical way, *e. g.* from such physical processes as casting dice. However, it is possible to construct other tables, called here the tables of shuffled numbers without introducing factors which are alien to mathematics. The paper is a commentary to such a table, published by the Mathematical Institute of PAN (Polish Academy of Science) and containing all possible arrangements of four digits from 0000 to 9999. Each arrangement appears only once and shuffling consists in a certain iteration process. This table differs from the tables of random numbers containing the same number of four digit arrangements in the following respect: in the tables of random numbers certain arrangements appear more than once while other arrangements do not appear at all. The removal of this defect facilitates the practical application of the new table. Moreover, reading the new table vertically we obtain a table of random numbers.
