ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria III: MATEMATYKA STOSOWANA XXI (1982)

RYSZARD ZIELIŃSKI (Warszawa)

Statystyczna analiza jakości produkcji morfologia krwi, optymalizacja wieloekstremalna czyli o estymacji rozkładu wielomianowego*

(*Praca wpłynęła do Redakcji 14.11.1980*)

1. Przykłady wprowadzające. *Statystyczna analiza jakości produkcji.* Analiza jakości procesu produkcyjnego, odgrywająca szczególnie dużą rolę w przypadku nowo wprowadzanych technologii, polega często na sporządzaniu listy wad pojawiających się w tym procesie i ocenie częstości występowania poszczególnych typów tych wad. Kolejne jednostki pojawiające się na wyjściu procesu produkcyjnego (np. kolejne śruby z automatu tokarskiego, kolejne odcinki tkaniny z warsztatu tkackiego, kolejne odbiorniki telewizyjne opuszczające taśmę montażową) poddaje się szczegółowym badaniom i rejestruje się rodzaje wad oraz częstość ich występowania. Typowy protokół z takiego postępowania może mieć postać następującej tabelki:

Rodzaj wady	Ilość wad danego rodzaju wśród	
	pierwszych 25 wadliwych sztuk	pierwszych 50 wadliwych sztuk
1	5 (20%)	11 (22%)
2	8 (32%)	15 (30%)
3	10 (40%)	15 (30%)
4	2 (8%)	6 (12%)
5		2 (4%)
6		1 (2%)

Naturalne pytania, na które należy tu odpowiedzieć, brzmią: 1) ile różnych typów wad może pojawiać się w danym procesie produkcyjnym oraz 2) jaka jest „struktura procentowa" tych wad. Dalsze pytania mogą mieć postać: 1) czy wykryliśmy wszystkie możliwe typy wad? 2) jakie jest prawdopodobieństwo tego, że badając 200

* Wykład plenarny na IX Ogólnopolskiej Kursokonferencji Zastosowań Matematyki, Burzenin, 13–22 października 1980.

kolejnych sztuk wadliwych wykryjemy wszystkie możliwe typy wad? 3) ile sztuk należy zbadać aby prawdopodobieństwo wykrycia wszystkich typów wad było dostatecznie duże, np. równe co najmniej 99%?

Morfologia krwi. Z pewnością każdy z nas przynajmniej raz w życiu miał przeprowadzone tzw. badanie morfologii krwi. Oto przykładowy „wzór krwinek białych" otrzymany w wyniku takiego badania

Granulocyty podzielone	46
Granulocyty kwasochłonne	3
Limfocyty	49
Monocyty	2

Tu znowu powstaje kilka pytań, na przykład 1) jak bardzo ten obraz zmieniłby się, gdyby przejrzano 200 lub 500 zamiast 100 białych krwinek? 2) czy w badanej krwi nie ma w ogóle granulocytów zasadochłonnych czy tylko „z prawdopodobieństwem p frakcja takich granulocytów nie przekracza f" i ewentualnie ile wynosi to p oraz f?

Optymalizacja wieloekstremalna. Na danym zbiorze $\mathcal{X}$ określona jest funkcja f i należy znaleźć punkt x^*, taki że $f(x^*) \geqslant f(x)$ dla każdego $x \in \mathcal{X}$. Jeżeli funkcja f ma dokładnie jedno maksimum lokalne, zadanie nie przedstawia większych trudności merytorycznych (chociaż — zależnie od postaci funkcji f i sposobu opisu zbioru $\mathcal{X}$ — może być bardzo skomplikowane z technicznego punktu widzenia). Jeżeli funkcja f ma $k > 1$ maksimów lokalnych i jeżeli w dodatku liczba k nie jest znana, zadanie staje się znacznie trudniejsze. Co więcej, z punktu widzenia zastosowań, podanie wtedy jednego tylko z punktów x^* o interesującej nas własności, może być nie wystarczające i prawdopodobnie chcielibyśmy uzyskać bardziej kompletne rozwiązanie zadania wieloekstremalnego. Oto jeden z możliwych sposobów postępowania w opisanej sytuacji (por. [8]). Oznaczmy przez $x_1^*, \ldots, x_k^*$ punkty (tu i wszędzie dalej mamy oczywiście na myśli punkty z danego zbioru $\mathcal{X}$), w których funkcja f osiąga swoje maksima lokalne. Dla uproszczenia założymy, że wszystkie te maksima są izolowane, tj. że dla każdego punktu x_j^* istnieje otoczenie U_j takie, że jeżeli $x \in U_j$, to $f(x_j^*) > f(x)$. Niech A będzie ustalonym algorytmem iteracyjnym poszukiwania maksimum lokalnego (np. algorytmem gradientowym) i niech zapis $A(x) = x_j^*$ oznacza, że startując z punktu x algorytm A doprowadza nas do maksimum lokalnego x_j^*. Dokonajmy podziału zbioru $\mathcal{X}$ na podzbiory $\mathcal{X}_j^* = \{x \in \mathcal{X}: A(x) = x_j^*\}$ — są to tzw. *zbiory przyciągania* poszczególnych maksimów lokalnych; zakładamy, że są to zbiory rozłączne.

Niech λ będzie miarą w zbiorze $\mathcal{X}$ (np. miarą Lebesgue'a, gdy $\mathcal{X}$ jest obszarem w skończenie-wymiarowej przestrzeni liczb rzeczywistych lub miarą liczącą, gdy $\mathcal{X}$ jest zbiorem skończonym). Załóżmy dalej, że $\sum_{j=1}^{k} \lambda(\mathcal{X}_j^*) = \lambda(\mathcal{X})$ i niech $\theta_j = \lambda(\mathcal{X}_j^*)/\lambda(\mathcal{X})$. Przez *pełne rozwiązanie* wieloekstremalnego zadania optymalizacji będziemy rozumieli zbiór par $\{(x_j^*, \theta_j): j = 1, \ldots, k\}$.

Teraz już z łatwością spostrzegamy, że znowu możemy sprowadzić zadanie do zadania takiego samego rodzaju, jak w omawianych poprzednio przypadkach statystycznej analizy jakości produkcji lub morfologii krwi. Mianowicie: wybieramy „na chybił-trafił" n punktów $x_1, \ldots, x_n$ w zbiorze $\mathscr{X}$ i dla każdego z nich wyznaczamy maksimum lokalne $A(x_i)$, $i = 1, \ldots, n$. W wyniku tego otrzymamy listę maksimów lokalnych oraz częstości, z jakimi te maksima pojawiły się w naszych obliczeniach. Pytania, zredagowane przy okazji poprzednich przykładów, przyjmują teraz postać np. 1) czy wykryliśmy wszystkie maksima lokalne? 2) jak duże powinno być n, aby wszystkie maksima lokalne zostały wykryte z dostatecznie dużym prawdopodobieństwem?, itp.

2. Model statystyczny. Każdy z opisanych wyżej problemów możemy formalnie opisać w następujący sposób. Obserwujemy ciąg $\xi_1, \xi_2, \ldots, \xi_n$ dodatnich, całkowitoliczbowych, niezależnych zmiennych losowych o jednakowym rozkładzie. Niech η_j oznacza liczbę tych wyrazów obserwowanego ciągu, które przyjęły wartość $j (j = 1, 2, \ldots)$. Prawdopodobieństwo wyniku $\{\eta_1 = n_1, \ldots, \eta_k = n_k\}$ opisujemy za pomocą wzoru

$$(1) \qquad P\{\eta_1 = n_1, \ldots, \eta_k = n_k\} = \binom{n}{n_1, \ldots, n_k} \theta_1^{n_1} \ldots \theta_k^{n_k}$$

$$(0 \leqslant n_j \leqslant n,\ j = 1, \ldots, k;\ \sum_{j=1}^{k} n_j = n),$$

gdzie k oraz $\theta = (\theta_1, \ldots, \theta_k)$ są nie znanymi parametrami, natomiast $\binom{n}{n_1, \ldots, n_k}$ jest skróconym zapisem wyrażenia $n!/n_1! \ldots n_k!$. Prawdopodobieństwo (1) będziemy oznaczali krótko przez $P_{(\theta_1, \ldots, \theta_k)}(n_1, \ldots, n_k)$. Mówiąc najogólniej, zadanie polega na tym, żeby na podstawie obserwacji $(n_1, \ldots, n_k)$ oszacować k oraz $\theta = (\theta_1, \ldots, \theta_k)$.

Oto dokładniejsza prezentacja tego modelu. Liczba k jest (nie znaną) liczbą możliwych rodzajów wad w rozważanym na wstępie zagadnieniu statystycznej analizy jakości produkcji, liczbą wszelkich możliwych typów białych krwinek w drugim przykładzie lub liczbą maksimów lokalnych w zadaniu wieloekstremalnym. Przypuśćmy, że te typy wad (rodzaje białych krwinek, maksima lokalne) zostały w pewien ustalony sposób ponumerowane; wtedy zdarzenie $\{\xi_j = i\}$ oznacza, że w wyniku j-tej obserwacji otrzymaliśmy „wadę i-tego typu". Liczby $\theta_1, \ldots, \theta_k$ interpretujemy jako „teoretyczne częstości" (prawdopodobieństwa występowania) wad poszczególnych typów; spełniają one oczywisty warunek $0 \leqslant \theta_j \leqslant 1, j = 1, \ldots, k$, oraz $\sum_{j=1}^{k} \theta_j = 1$.

W dalszym ciągu będziemy rozróżniali dwa przypadki. W pierwszym z nich założymy, że znane jest górne ograniczenie dla liczby klas k; oznaczymy je przez K. W tej sytuacji, przy ustalonym n, mamy do czynienia z rodziną rozkładów wielomianowych (1) o przetrzeni parametrów

$$(2) \qquad \left\{\theta = (\theta_1, \ldots, \theta_K):\ 0 \leqslant \theta_j \leqslant 1, j = 1, \ldots, K;\ \sum_{j=1}^{K} \theta_j = 1\right\}.$$

Tutaj estymacja parametru $\theta = (\theta_1, \ldots, \theta_K)$ nie nastręcza żadnych trudności; nie-

obciążonym estymatorem o minimalnej wariancji (osiągającej dolne ograniczenie Craméra–Rao, a więc estymatorem efektywnym) jest

$$\hat{\theta} = (n_1/n, \ldots, n_K/n) \tag{3}$$

(por. np. [6]). Przypomnijmy jednak, że K jest tylko górnym oszacowaniem liczby k klas, i że ta ostatnia liczba nie jest znana. W tej sytuacji możemy 1) spróbować dobrać na tyle duże n, żeby „prawie na pewno" (dokładniej: z dostatecznie dużym prawdopodobieństwem) wszystkie klasy zostały wykryte i wtedy posłużyć się estymatorem (3) lub 2) szacować k. Wrócimy za chwilę do tych spraw.

Drugi przypadek ma miejsce wtedy, gdy nie znamy liczby klas k i nie potrafimy wskazać górnego ograniczenia K dla tej liczby. Tutaj mamy do czynienia znowu z rodziną rozkładów (1), ale przestrzeń parametrów ma teraz postać

$$\left\{\theta = (\theta_1, \ldots, \theta_k): 0 < \theta_j \leqslant 1, j = 1, \ldots, k; \sum_{j=1}^{k} \theta_j = 1; k = 1, 2, \ldots\right\}. \tag{4}$$

Zagadnienie estymacji k w takiej sytuacji, w pewnym bardzo specjalnym przypadku, zostało rozwiązane przez Lewontina i Prouta [4]; rozwiązanie to można również znaleźć w znanym podręczniku Kendalla i Stuarta [3], rozdz. 18. Mianowicie, jeżeli $\theta_j = 1/k$, $j = 1, \ldots, k$ (tzn. jeżeli wszystkie klasy są jednakowo prawdopodobne) i jeżeli w n doświadczeniach zaobserwujemy w różnych klas, to estymator największej wiarogodności liczby klas k jest rozwiązaniem (względem k) równania

$$\frac{n}{k} = \sum_{j=k-w+1}^{k} \frac{1}{j}$$

(dla ułatwienia podajemy, że prawa strona tego równania jest w przybliżeniu równa $\log k - \log(k-w+1)$).

W ogólnym (lub choćby tylko ogólniejszym) przypadku odpowiednie estymatory nie są mi znane. Pewne podejście do zagadnienia estymacji od strony bayesowskich reguł decyzyjnych przedstawiamy nieco dalej.

3. Wybór liczby obserwacji. Niech, jak poprzednio, W będzie liczbą zaobserwowanych klas. Elementarne rozważania kombinatoryczne prowadzą do wzoru

$$P_{(\theta_1, \ldots, \theta_k)}\{W = w\} = \sum_{[n_w]} \sum_{J_{w,k}} \binom{n}{n_1, \ldots, n_w} \theta_{i_1}^{n_1} \ldots \theta_{i_w}^{n_w},$$

gdzie $\sum_{[n_w]}$ oznacza sumowanie rozciągnięte na wszystkie ciągi $(n_1, \ldots, n_w)$ dodatnich liczb całkowitych spełniających warunek $\sum_{j=1}^{w} n_j = n$ oraz $\sum_{J_{w,k}}$ oznacza sumowanie rozciągnięte na wszystkie podzbiory $\{i_1, \ldots, i_w\}$ zbioru $\{1, \ldots, k\}$. W szczególności, prawdopodobieństwo wykrycia wszystkich k klas wyraża się wzorem

$$P_{(\theta_1, \ldots, \theta_k)}\{W = k\} = \sum_{\substack{n_1=1 \\ n_1+\ldots+n_k=n}}^{n} \ldots \sum_{n_k=1}^{n} \binom{n}{n_1, \ldots, n_k} \theta_1^{n_1} \ldots \theta_k^{n_k}. \tag{5}$$

Dla każdej ustalonej wartości parametru θ można oczywiście znaleźć takie n_0, żeby prawdopodobieństwo (5) było większe od danej z góry liczby P_0, jeżeli tylko $n \geqslant n_0$.

Z drugiej jednak strony, nawet dla bardzo dużych wartości n, to prawdopodobieństwo będzie bardzo małe jeżeli choćby jedna z liczb θ_j będzie odpowiednio mała. Ponieważ $\theta = (\theta_1, \ldots, \theta_k)$, podobnie jak samo k, również nie jest znane, wydaje się, że znaleźliśmy się w sytuacji bez wyjścia. Okazuje się jednak, że można tu znaleźć pewne rozwiązanie.

Zajmijmy się najpierw przypadkiem, gdy $\theta_1 = \ldots = \theta_k = 1/k$. Wzór (5) przyjmuje teraz prostą postać

$$(6) \qquad P_{(\theta_1, \ldots, \theta_k)}\{W = k\} = k^{-1} \sum_{[\bar{n}_k]} \binom{n}{n_1, \ldots, n_k} = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \left(1 - \frac{j}{k}\right)^n.$$

Wzór ten, wraz z następującym wzorem przybliżonym (dla ustalonego k i dużych n)

$$(7) \qquad P_{(\theta_1, \ldots, \theta_k)}\{W = k\} \approx \exp\{-k \cdot \exp(-n/k)\}$$

można znaleźć w podręczniku Fellera [1], § IV. 2. Wartości n_0 (dla kilku wybranych wartości k) takie, że jeżeli $n \geqslant n_0$, to prawdopodobieństwo (7) jest równe co najmniej P_0, podaje następująca tabelka:

k	$P_0 = 0.90$	$P_0 = 0.95$	$P_0 = 0.99$
2	6	8	11
3	11	13	18
4	15	18	24
5	19	23	32
10	46	53	70
20	105	120	152
50	309	345	426
100	686	758	921

Przypuśćmy teraz, że $\min_{1 \leqslant j \leqslant k} \theta_j = \varepsilon$ (oczywiście $\varepsilon \leqslant 1/k$). Łatwo można sprawdzić, że prawdopodobieństwo wykrycia wszystkich klas może być teraz oszacowane z dołu przez liczbę

$$P_{(1/\varkappa, \ldots, 1/\varkappa)}\{W = \varkappa\},$$

gdzie $\varkappa = \varkappa(\varepsilon)$ jest najmniejszą liczbą całkowitą większą lub równą $1/\varepsilon$. To oszacowanie umożliwia nam posługiwanie się wzorem (6) lub (7) (lub przedstawioną wyżej tabelką) również w rozważanym teraz przypadku ogólniejszym. Praktyczna interpretacja jest następująca: jeżeli $n \geqslant n_0$, to z prawdopodobieństwem równym co najmniej P_0 wykryjemy wszystkie klasy takie, dla których $\theta_j \geqslant \varepsilon$.

4. Bayesowska estymacja liczby klas. W bieżącym paragrafie ograniczymy się do estymacji tylko liczby k; zagadnienie łącznej estymacji k i θ rozpatrzymy w następnym paragrafie 5.

Rozważane teraz zagadnienie możemy rozpatrywać analogicznie do tego, jak G. Schwarz [5] rozważał zagadnienie estymacji wymiaru modelu statystycznego

w pewnej szczególnej klasie takich modeli. Będziemy rozważali rodzinę rozkładów wielomianowych (1) z przestrzenią parametrów (4). Wprowadźmy rozkład a priori na przestrzeni parametrów i niech ten rozkład ma postać $\mu = \sum_{j=1}^{\infty} \alpha_j \mu_j$, gdzie α_j jest prawdopodobieństwem a priori tego, że liczba klas jest równa j oraz μ_j jest warunkowym rozkładem a priori parametru θ, gdy $k = j$. Zauważmy, że rozkład μ_j jest skoncentrowany na pewnym $(j-1)$-wymiarowym sympleksie w j-wymiarowej przestrzeni liczb rzeczywistych.

Zdefiniujmy, jak zwykle to się robi przy budowie bayesowskich postępowań decyzyjnych, funkcję straty związaną z rozważanymi estymatorami. Przyjmiemy, że ponosimy pewną stałą stratę w wysokości $c > 0$, gdy błędnie odgadniemy liczbę k klas. Wykonując standardowe rachunki łatwo stwierdzamy, że ryzyko bayesowskie reguły decyzyjnej, która wynikowi $(n_1, \ldots, n_w)$ przyporządkowuje wartości $d = d(n_1, \ldots, d_w)$ estymatora, wynosi

$$c \cdot \sum_{w=1}^{\infty} \sum_{[\bar{n}w]} \binom{n}{n_1, \ldots, n_w} \sum_{\substack{k=w \\ k \neq d}}^{\infty} A_k(n_1, \ldots, n_w),$$

gdzie

$$A_k(n_1, \ldots, n_w) = \alpha_k \sum_{J_{w,k}} a^{(k)}_{i_1, \ldots, i_w}(n_1, \ldots, n_w)$$

oraz

$$a^{(k)}_{i_1, \ldots, i_w}(n_1, \ldots, n_w) = \int \theta_{i_1}^{n_1} \ldots \theta_{i_w}^{n_w} \mu_k(d\theta).$$

Jeżeli więc w wyniku n eksperymentów zaobserwujemy w klas z częstościami $n_1, \ldots, n_w$, to optymalną decyzją odnośnie liczby klas k jest taka decyzja d, dla której

$$A_d(n_1, \ldots, n_w) = \max_{k \geqslant w} A_k(n_1, \ldots, n_w).$$

Ten ogólny wzór pozwala nam na konstrukcję optymalnego estymatora bayesowskiego przy dowolnych rozkładach a priori (α_j) oraz $\mu_j, j = 1, 2, \ldots$ W szczególności, gdy $\alpha_j = \text{const}$ („niewłaściwy rozkład a priori" według którego każda liczba klas $k = 1, 2, \ldots$ jest a priori jednakowo prawdopodobna) oraz jeżeli μ_j są rozkładami jednostajnymi, otrzymujemy bardzo proste rozwiązanie: optymalna wartość estymatora liczby klas to taka liczba d, dla której

$$\binom{d}{w} \frac{\Gamma(d)}{\Gamma(n+d)} = \max_{k \geqslant w} \binom{k}{w} \frac{\Gamma(k)}{\Gamma(n+k)},$$

gdzie n jest liczbą wykonanych obserwacji oraz w jest liczbą wszystkich wykrytych klas. Na przykład: optymalne decyzje d dla przypadku, gdy po wykonaniu $n \geqslant 5$ eksperymentów zaobserwowano $w = 5$ różnych klas, podaje następująca tabelka:

n	d	n	d	n	d
5	∞	9	9 lub 10	15	6 lub 7
6	24 lub 25	10	8 lub 9	16–24	6
7	14 lub 15	11	8	25 lub więcej	5
8	11	12–14	7		

Spójnik „lub" w powyższej tabelce oznacza, że obie decyzje (np. 24 i 25 dla $n = 6$) dają tę samą wartość ryzyka bayesowskiego.

5. Bayesowska estymacja parametru (k, θ). W przypadku, gdy chcemy jednocześnie szacować liczbę klas k i prawdopodobieństwa $\theta = (\theta_1, \ldots, \theta_k)$, postępowanie jest bardziej złożone. Wynika to przede wszystkim stąd, że sam problem staje się bardziej złożony, gdyż nie jest łatwo sformułować jednolite postulaty pod adresem łącznej estymacji obu, tak różnych od siebie, wielkości k oraz θ. Rozpatrzymy taką wersję postępowania, według której dokonuje się wyboru pomiędzy tylko dwiema decyzjami:

d_1 — „liczba klas jest równa w (tzn. jest równa liczbie klas wykrytych) oraz $\theta_j = n_j/n$, $j = 1, \ldots, w$", lub

d_2 — „liczba klas jest większa od w, prawdopodobieństwo θ_j dla j-tej wykrytej klasy jest równe $(1-\gamma)n_j/n$ oraz łączne prawdopodobieństwo wszystkich nie wykrytych klas jest równe γ".

Przyjmiemy takie same założenia o rozkładzie a priori jak w poprzednim paragrafie i skonstruujemy rozwiązanie w następujący sposób. Przypuśćmy, że ponosimy stratę $c > 0$ jeżeli podejmujemy decyzję d_1 w sytuacji, gdy $w < k$ (tzn. jeżeli nasze oszacowanie liczby klas jest zbyt niskie) oraz stratę $C = vc, v > 0$, jeżeli podejmujemy decyzję d_2 w sytuacji, gdy $w = k$ (tzn. jeżeli nasze oszacowanie liczby klas jest zbyt wysokie). Wykonując odpowiednie rachunki otrzymamy, że po zaobserwowaniu wyniku $(n_1, \ldots, n_w)$ należy podjąć decyzję d_1 wtedy i tylko wtedy, gdy

$$c \cdot \sum_{k=w+1}^{\infty} A_k(n_1, \ldots, n_w) \leqslant C \cdot A_w(n_1, \ldots, n_w),$$

co w przypadku $\alpha_j = \text{const}$ oraz jednostajnych rozkładów a priori μ_j redukuje się do reguły: podjąć decyzję d_1 wtedy i tylko wtedy, gdy

$$\frac{\Gamma(n+w)\Gamma(n-w-1)}{\Gamma(n)\Gamma(n-1)} \leqslant 1+v.$$

w	v		
	0.5	1	2
1	6	4	3
2	16	10	7
3	31	19	13
4	51	30	20
5	76	45	29
10	273	160	102
20	10 37	608	384
50	6 290	3 680	2 323
100	24 894	14 569	9 194

Dla ustalonej wartości w wyrażenie po lewej stronie powyższej nierówności maleje wraz z n i zbiega do jedności, skąd łatwo wynika, że jeżeli $v = 0$, to decyzji d_1 nigdy nie należy podejmować i że dla każdej liczby $v > 0$ oraz dla każdego w istnieje takie $n(w, v)$, że decyzja d_1 jest optymalna wtedy i tylko wtedy, gdy $n \geqslant n(w, v)$. Wartości $n(w, v)$ dla kilku wybranych w oraz v podaje następująca tabelka (p. str. 45).

Pozostaje wyznaczenie wartości γ — estymatora łącznego prawdopodobieństwa nie wykrytych klas. Konstrukcja takiego estymatora nie nastręcza większych trudności, gdy przyjmiemy — jak to się zwykle robi i co nie jest pozbawione sensu w naszym przypadku — kwadratową funkcję strat. Zakładając, jak poprzednio, że μ_j są rozkładami jednostajnymi, otrzymujemy optymalny estymator bayesowski w postaci

$$\gamma = \frac{\sum_{k=w}^{\infty} \frac{k-w}{n+k} \alpha_k \binom{k}{w} \frac{\Gamma(k)}{\Gamma(n+k)}}{\sum_{k=w}^{\infty} \alpha_k \binom{k}{w} \frac{\Gamma(k)}{\Gamma(n+k)}}.$$

W przypadku rozkładu a priori $\alpha_2 = 1$, $\alpha_j = 0$ dla $j \neq 2$ oraz $w = 1$ (jedna zaobserwowana klasa) otrzymujemy znane bayesowskie oszacowanie $\gamma = 1/(n+2)$ dla prawdopodobieństwa sukcesu w schemacie Bernoulli'ego, gdy n kolejnych eksperymentów zakończyło się porażkami (por. np. [6], zad. 10.6). Inny interesujący przypadek szczególny ma miejsce dla $\alpha_j = \text{const}$; wtedy, dla $n \geqslant w+2$, mamy $\gamma = w(w+1)/n(n-1)$.

6. Przykłady liczbowe. *Statystyczna analiza jakości produkcji.* Wróćmy do przykładu z paragrafu 1 i przypuśćmy, że należy oszacować (k, θ) przy założeniach z paragrafu 5 dla $v = 1/2$. To ostatnie założenie oznacza, że nasze straty związane z niedoszacowaniem liczby różnych typów braków oceniamy dwa razy wyżej niż straty które ponosimy wtedy, gdy oceniamy tę liczbę zbyt wysoko. W tablicy podanej w paragrafie 5 odczytujemy $n(4, 1/2) = 51$, a ponieważ w naszym przypadku $n = 25 < 51$ (por. pierwsza kolumna tabelki z paragrafu 1), podejmujemy decyzję d_2 („liczba typów wad jest większa od 4"). Na podstawie wzoru podanego na końcu poprzedniego paragrafu, udział wad nie wykrytych typów oceniamy na $\gamma = 3.3\%$. Wyniki przedstawione w drugiej kolumnie tabelki z pierwszego przykładu także prowadzą do decyzji d_2, ale teraz $\gamma = 1.7\%$. Gdybyśmy w wyniku dalszych badań znaleźli 55 (lub więcej) wad i nie odkryli wśród nich ani jednej wady nowego typu, powinniśmy uznać, że wszystkie typy zostały już wykryte ($n(6, 1/2) = 105$).

Morfologia krwi. Przypuśćmy, że z dużym prawdopodobieństwem P_0, powiedzmy $P_0 = 0.99$, chcielibyśmy wykryć wszystkie takie rodzaje białych krwinek, których częstość występowania w naszym organizmie jest równa co najmniej 1%. Ze wzoru (7) (lub odpowiedniej tabelki) znajdujemy, że w tym celu tak zwany „wzór białych krwinek" powinien być wyznaczony po obejrzeniu co najmniej $n = 921$ tych krwinek (zwykle przyjmuje się $n = 100$).

7. Kilka uwag końcowych. Wszystkie przedstawione wyżej wyniki uzyskaliśmy przy licznych założeniach uproszczających; umożliwiło to nam doprowadzenie niektórych rozwiązań do końcowych, niezbyt zawiłych wzorów lub nawet tablic, ale z drugiej strony ograniczyło zakres zastosowań. Interesujące i niezmiernie ważne dla praktyki byłoby uzyskanie analogicznych wyników dla bardziej realistycznych funkcji strat i dla bardziej realistycznych rozkładów a priori. Jeżeli chodzi o rozkłady μ_j, to stosunkowo łatwo można uogólnić przedstawione wyniki na przypadek, gdy są to rozkłady Dirichleta (rozkłady jednostajne są szczególnym przypadkiem takich rozkładów). Przypomnijmy, że rodziny rozkładów wielomianowych i rozkładów Dirichleta są rodzinami sprzężonymi w zagadnieniach bayesowskich. Szersze informacje na temat tych niezwykle pożytecznych rozkładów oraz obszerne tablice przydatne dla rozważanych tu problemów można znaleźć w książkach [2] i [7].

Prace cytowane

[1] W. Feller, *Wstęp do rachunku prawdopodobieństwa*, tom I, PWN, 1966.
[2] N. L. Johnson, S. Kotz, *Distributions in statistics: continuous multivariate distributions*, Wiley, 1972.
[3] M. G. Kendall, A. Stuart, *The advanced theory of statistics*, vol. II (istnieje tłumaczenie rosyjskie z 1973 roku).
[4] R. C. Lewontin, T. Prout, *Estimation of the number of different classes in a population*, Biometrics 12 (1956), str. 211–223.
[5] G. Schwarz, *Estimating the dimension of a model*, Ann. of Statist. 6 (1978), str. 461–464.
[6] S. D. Silvey, *Wnioskowanie statystyczne*, PWN, 1978.
[7] M. Sobel, V. R. R. Uppuluri, K. Frankowski, *Selected tables in mathematical statistics*, vol. IV, Amer. Math. Soc. 1977.
[8] R. Zieliński, *A statistical estimate of the structure of multi-extremal problem*, Math Programming 21 (1981), str. 348–356.