

C. R A J S K I (Warszawa)

PORÓWNYWANIE POPULACYJ GENERALNYCH NA PODSTAWIE TWIERDZENIA BAYES'A

1. Sformułowanie zagadnienia. Celem niniejszej pracy jest przedstawienie nowej metody statystycznego porównywania dwóch populacji generalnych. Nazwijmy je pierwszą i drugą. Przypuśćmy, że każda z nich zawiera pewną nieznaną liczbę sztuk w jakiś sposób wyróżnionych. Oznaczmy stosunek liczby sztuk wyróżnionych w pierwszej populacji do liczebności tej populacji przez w_1 , a analogiczny stosunek w drugiej populacji — przez w_2 . Każdą z wielkości w_1 , w_2 będziemy nazywali *frakcją* sztuk wyróżnionych lub krócej *frakcją* danej populacji generalnej. Założmy następnie, że obie populacje są na tyle liczne, że nie popełnimy w dalszych rozważaniach znacznego błędu zakładając, iż w_1 i w_2 mogą przyjmować wszelkie wartości w przedziale $(0,1)$.

Przypuśćmy, że z pierwszej populacji generalnej pobraliśmy próbkę o liczebności n_1 i znaleźliśmy w niej r_1 sztuk wyróżnionych, a z drugiej pobraliśmy próbkę o liczebności n_2 i znaleźliśmy w niej r_2 sztuk wyróżnionych. Najogólniej sformułowane zagadnienie porównywania dwóch populacji generalnych polega na znajdowaniu na podstawie czwórki liczb n_1 , r_1 , n_2 , r_2 prawdopodobieństw prawdziwości różnych hipotez, o wzajemnym ustosunkowaniu się frakcji rozpatrywanych populacji. Jednym z możliwych użytków, jakie można zrobić z tych liczb, jest określanie frakcji populacji generalnych metodą największej wiarogodności. Rozwiązanie przedstawione w niniejszej pracy opiera się na twierdzeniu Bayes'a i nie korzysta z metody największej wiarogodności.

2. Ogólne rozwiązanie zagadnienia. Ograniczmy się chwilowo do pierwszej populacji generalnej. Prawdopodobieństwo prawdziwości hipotezy, że frakcja sztuk wyróżnionych w tej populacji zawiera się w granicach od x_1 do $x_1 + dx_1$, możemy określić na podstawie prawa Bayes'a. Oznaczmy prawdopodobieństwo prawdziwości

wymienionej hipotezy przez $P\{x_1 < w_1 < x_1 + dx_1\}$, prawdopodobieństwo proste, tzn. prawdopodobieństwo otrzymania r_1 sztuk wyróżnionych w próbie o licznosci n_1 , pobranej z populacji generalnej, w której frakcja sztuk wyróżnionych jest x_1 — przez $p(x_1, n_1, r_1)$ i wreszcie gęstość rozkładu *a priori* — przez $\psi_1(x_1)$. Wówczas na podstawie prawa Bayes'a mamy

$$(1) \quad P\{x_1 < w_1 < x_1 + dx_1\} = \frac{\psi_1(x_1) p(x_1, n_1, r_1) dx_1}{\int_0^1 \psi_1(y_1) p(y_1, n_1, r_1) dy_1},$$

gdzie y_1 jest zmienną pozorną.

Prawdopodobieństwo prawdziwości hipotezy, że frakcja sztuk wyróżnionych w drugiej populacji jest zawarta w granicach od x_2 do $x_2 + dx_2$ określa analogiczny wzór

$$(2) \quad P\{x_2 < w_2 < x_2 + dx_2\} = \frac{\psi_2(x_2) p(x_2, n_2, r_2) dx_2}{\int_0^1 \psi_2(y_2) p(y_2, n_2, r_2) dy_2},$$

w którym oznaczenia różnią się tylko wskaźnikiem od oznaczeń przyjętych we wzorze (1).

Nazwijmy *hipotezę elementarną* przypuszczenie, że obie wymienione hipotezy są prawdziwe. W wielu zastosowaniach praktycznych prawdziwość jednej hipotezy jest niezależna od prawdziwości drugiej. Wówczas prawdopodobieństwo $P\{x_1 < w_1 < x_1 + dx_1, x_2 < w_2 < x_2 + dx_2\}$ prawdziwości hipotezy elementarnej jest iloczynem prawdopodobieństw określonych wzorami (1) i (2):

$$(3) \quad \begin{aligned} &P\{x_1 < w_1 < x_1 + dx_1, x_2 < w_2 < x_2 + dx_2\} = \\ &= P\{x_1 < w_1 < x_1 + dx_1\} \cdot P\{x_2 < w_2 < x_2 + dx_2\}. \end{aligned}$$

Wygodnie jest nadać rozpatrywanemu zagadnieniu interpretację geometryczną. Odłóżmy w układzie współrzędnych prostokątnych na osi odciętych frakcję pierwszej populacji generalnej, a na osi rzędnych frakcję drugiej populacji generalnej. Na wykresie takim punkt o współrzędnych w_1, w_2 reprezentuje takie dwie populacje, że pierwsza ma frakcję sztuk wyróżnionych równą w_1 , a druga — równą w_2 . Punkt taki będziemy nazywali *punktem reprezentacyjnym*.

Poprowadźmy w tym układzie cztery proste o równaniach $x_1 = \text{const}$, $x_1 + dx_1 = \text{const}$, $x_2 = \text{const}$ i $x_2 + dx_2 = \text{const}$ (rysunek 1).

Ograniczają one pewne prostokątne pole dS . Hipoteza elementarna odpowiada przypuszczeniu, że punkt reprezentacyjny leży wewnątrz pola dS . Każdej hipotezie H dotyczącej frakcyj obu populacji odpowiada pewien zbiór możliwych położenia punktu reprezentacyjnego. W pewnych rodzajach hipotez zbiór ten wypełnia jakieś pole S , które będziemy nazywali polem hipotezy H . Pole S możemy podzielić zawsze na rozłączne pola dS odpowiednio dobranych hipotez elementarnych; zatem każdą hipotezę H możemy uważać za sumę logiczną wzajemnie wykluczających się hipotez elementarnych. Wobec tego prawdopodobieństwo $P\{H\}$ prawdziwości hipotezy H jest sumą prawdopodobieństw prawdziwości hipotez elementarnych rozciągniętą na pole S :

$$(4) \quad P\{H\} = \iint_S P\{x_1 < w_1 < x_1 + dx_1, x_2 < w_2 < x_2 + dx_2\}.$$

Wzór ten stanowi ogólne rozwiązanie zagadnienia postawionego na wstępie.

3. Uproszczenie rozwiązania ogólnego. Aby zastosować wzory poprzedniego rozdziału, trzeba wiedzieć coś bliższego o funkcjach figurujących po prawych stronach (1) i (2). Jeśli chodzi o prawdopodobieństwa proste, to z założenia, że obie populacje są bardzo liczne, wynika, że obie te funkcje mało różnią się od dwumianowych. Jeśli pobieranie sztuk do próbek odbywa się ze zwracaniem, mamy dokładnie

$$(5) \quad p(x_1, n_1, r_1) = \binom{n_1}{r_1} x_1^{r_1} (1-x_1)^{n_1-r_1},$$

$$(6) \quad p(x_2, n_2, r_2) = \binom{n_2}{r_2} x_2^{r_2} (1-x_2)^{n_2-r_2}.$$

Gęstości rozkładów *a priori* — jeśli są znane — są na ogół w każdym przypadku inne, co uniemożliwia wyciągnięcie ogólnych wniosków z przytoczonych dotychczas wzorów. Przeważnie gęstości nie są znane, z czym sobie dajemy radę w ten sposób, że zakładamy rozkłady równomierne. Wówczas funkcje $\psi_1(x_1)$ i $\psi_2(x_2)$, występujące po prawych stronach (1) i (2), są stałe i ulegają skróceniu.

Nie uzasadniamy bliżej takiego postępowania, bo naświecił je niedawno H. Steinhaus w artykule *Prawdopodobieństwo, wiarogodność, możliwość*¹⁾. Zauważmy tylko, że wzory wyprowadzone w założeniu

¹⁾ Zastosowania Matematyki I (1953-54), str. 149 - 172.

równomiernego rozkładu *a priori* stosują się także do pewnej klasy doświadczalnych rozkładów *a priori*. Stwierdzono mianowicie²⁾, że istnieją takie rozkłady, dla których najlepszą aproksymacją są funkcje dwumianowe. Wówczas w liczniku i w mianowniku prawych stron każdego z równań (1) i (2) występują iloczyny funkcji dwumianowych. Otóż iloczyn dwóch funkcji dwumianowych można zawsze przedstawić w postaci iloczynu trzeciej funkcji dwumianowej przez pewien wyraz niezależny od frakcji sztuk wyróżnionych w populacji generalnej. Ponieważ wyraz tego rodzaju występuje zarówno w liczniku, jak i w mianowniku prawych stron każdego z równań (1) i (2), można przezeń skrócić, po czym otrzymuje się wzory równokształtne z tymi, które powstają przy założeniu równomiernych rozkładów *a priori*. Z tego wynika, że wzory są ważne dla wymienionej klasy empirycznych rozkładów *a priori* pod warunkiem, że symbole n_1 , r_1 , n_2 , r_2 nie będą już oznaczały liczności próbek i liczby znalezionych w nich sztuk wyróżnionych, ale pewne parametry uwzględniające zarówno wyniki zbadania próbek, jak i doświadczalnie stwierdzone rozkłady *a priori*.

Przy założeniu równomiernych rozkładów *a priori* można scałkować mianowniki prawych stron (1) i (2) i przedstawić ogólne rozwiązanie postawionego na wstępie zagadnienia w postaci

$$(7) \quad P\{H\} = \iint_S f(x_1, x_2) dx_1 dx_2,$$

gdzie $f(x_1, x_2)$ oznacza gęstość rozkładu prawdopodobieństw prawdziwości hipotez elementarnych

$$(8) \quad f(x_1, x_2) = (n_1 + 1)(n_2 + 1) \binom{n_1}{r_1} x_1^{r_1} (1 - x_1)^{n_1 - r_1} \binom{n_2}{r_2} x_2^{r_2} (1 - x_2)^{n_2 - r_2}.$$

4. Hipoteza różnicy ujemnej. Jako najprostsze zastosowanie wyprowadzonych wzorów rozpatrzmy rozwiązanie następującego zagadnienia: Jakie jest prawdopodobieństwo $P\{w_1 < w_2\}$ prawdziwości hipotezy, że frakcja pierwszej populacji generalnej jest mniejsza od frakcji drugiej populacji generalnej (rysunek 2)? Odpowiedź na to pytanie daje wzór

$$(9) \quad P\{w_1 < w_2\} = \int_0^1 dx_1 \int_{x_1}^1 f(x_1, x_2) dx_2.$$

²⁾ J. Oderfeld, *Statystyczny odbiór towarów klasyfikowanych według alternatywy*, GUS, Warszawa 1950.

Po prostych przekształceniach otrzymujemy

$$(10) \quad P\{w_1 < w_2\} = \frac{n_1 + 1}{n_1 + n_2 + 2} \sum_{j=0}^{r_2} \frac{\binom{n_1}{r_1} \binom{n_2 + 1}{j}}{\binom{n_1 + n_2 + 1}{r_1 + j}}.$$

Tablica 1 zawiera prawdopodobieństwa $P\{w_1 < w_2\}$ obliczone dla $n_1 = n_2 = 40$. W zasadzie należałoby ułożyć podobne tablice również dla innych licznosci próbek, lepiej jest jednakże posłużyć się wzorem asymptotycznym powstającym przy założeniu, że licznosci próbek, pozostając jednakowe, dążą do nieskończoności. Kładąc we wzorze (10) $n_1 = n_2 = n$ oraz $n \rightarrow \infty$ otrzymujemy

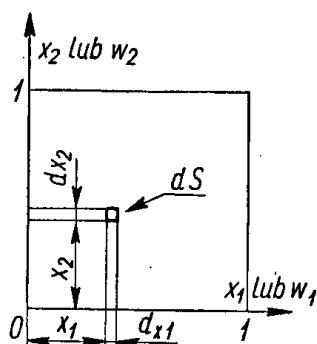
$$(11) \quad P\{w_1 < w_2\} = \sum_{j=0}^{r_2} \frac{1}{2^{r_1 + j + 1}} \binom{r_1 + j}{r_1}.$$

Na zasadzie tego wzoru obliczono dane tablicy 2. Wzór (11) ułatwia zarówno obliczanie jak i korzystanie z wyników weryfikacji omawianej hipotezy, ponieważ użytkujący ma do czynienia tylko z jedną tablicą, a nie z wieloma.

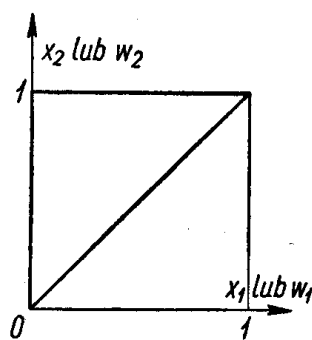
Tablica 1

Prawdopodobieństwa $P\{w_1 < w_2\}$ w promillach, obliczone z wzoru (10) dla $n_1 = n_2 = 40$

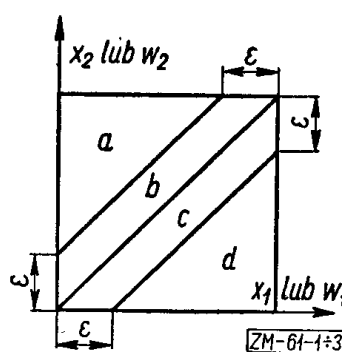
$r_2 \downarrow$	$P\{w_1 < w_2\}$			
9	1000	996	987	980
8	999	993	976	956
7	997	985	956	917
6	994	971	922	844
5	987	946	868	759
4	973	899	784	644
3	942	820	662	500
2	880	692	500	338
1	753	500	308	180
0	500	247	120	058
$r_1 \rightarrow$	0	1	2	3



Rys. 1



Rys. 2



Rys. 3

Dokładność wzoru (11) jest zupełnie dobra. Można się o tym przekonać np. przez porównanie tablicy 2 z tablicą 1. W inny sposób dokładność wzoru (11) ilustruje tablica 3. Zawiera ona prawdopodobieństwa $P\{w_1 < w_2\}$ obliczone z wzoru (10) dla $r_1=0$ i $r_2=2$ i dla różnych licznosci próbek.

Tablica 2

Prawdopodobieństwa $P\{w_1 < w_2\}$ w promillach obliczone z wzoru (11)

$r_2 \downarrow$	$P\{w_1 < w_2\}$															
15	1000	1000	999	998	994	987	974	953	924	885	837	779	715	646	573	500
14	1000	1000	999	996	990	979	961	933	895	846	788	721	650	575	500	427
13	1000	1000	998	994	985	968	942	905	857	797	729	655	578	500	425	354
12	1000	999	996	989	976	952	916	868	808	738	661	581	500	422	350	285
11	1000	998	994	982	962	928	881	820	748	668	584	500	419	345	279	221
10	1000	997	989	971	941	895	834	760	676	588	500	416	339	271	212	163
9	999	994	981	954	910	849	773	685	593	500	412	332	262	203	154	115
8	998	989	967	927	867	788	696	598	500	407	324	252	192	143	105	076
7	996	980	945	887	806	709	605	500	402	315	240	180	132	095	067	047
6	992	965	910	828	726	613	500	395	304	227	166	119	084	048	039	026
5	984	938	855	746	623	500	387	291	212	151	105	072	048	032	021	013
4	969	891	773	637	500	377	274	194	133	090	059	038	024	015	010	006
3	938	812	656	500	363	254	172	113	073	046	029	018	011	006	004	002
2	875	688	500	344	227	145	090	055	033	019	011	006	004	002	001	001
1	750	500	312	188	109	062	035	020	011	006	003	002	001	000	000	000
0	500	250	125	062	031	016	008	004	002	001	000	000	000	000	000	000
$r_1 \rightarrow$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Tablica 3

Prawdopodobieństwa $P\{w_1 < w_2\}$ obliczone z wzoru (10)

$r_1=0, r_2=0, n_1=n_2=n.$

n	2	10	40	100	500	∞
$P\{w_1 < w_2\}$	0,9500	0,8900	0,8795	0,8768	0,8753	0,8750

Jak widzimy, przy zmianie liczby sztuk w próbce od 10 do nieskończoności, prawdopodobieństwo $P\{w_1 < w_2\}$ zmienia się zaledwie o 1,5‰.

5. Inne hipotezy. Pewnym uogólnieniem hipotezy rozpatrzonej w poprzednim rozdziale są hipotezy, w których zakładamy nie tylko istnienie różnicy między frakcjami sztuk wyróżnionych dwóch populacji generalnych, ale również narzucamy tej różnicy pewne ograniczenia. Możemy na przykład postawić następujące pytania (rysunek 3):

a. Jakie jest prawdopodobieństwo $P\{w_1 + \varepsilon < w_2\}$ prawdziwości hipotezy, że frakcja drugiej populacji generalnej jest większa od frakcji pierwszej populacji generalnej o więcej niż o ε , gdzie ε jest liczbą dodatnią?

b. Jakie jest prawdopodobieństwo $P\{w_1 < w_2 < w_1 + \varepsilon\}$ prawdziwości hipotezy, że frakcja drugiej populacji generalnej jest większa od frakcji pierwszej populacji generalnej o mniej niż o ε ?

c. Jakie jest prawdopodobieństwo $P\{w_1 - \varepsilon < w_2 < w_1\}$ prawdziwości hipotezy, że frakcja drugiej populacji jest mniejsza od frakcji pierwszej populacji generalnej o mniej niż o ε ?

d. Jakie jest prawdopodobieństwo $P\{w_2 < w_1 - \varepsilon\}$ prawdziwości hipotezy, że frakcja drugiej populacji generalnej jest mniejsza od frakcji pierwszej populacji generalnej o więcej niż o ε ?

Wymienione prawdopodobieństwa są określone następującymi wzorami:

$$(12) \quad P\{w_1 + \varepsilon < w_2\} = \int_0^{1-\varepsilon} dx_1 \int_{x_1+\varepsilon}^1 f(x_1, x_2) dx_2,$$

$$(13) \quad P\{w_1 < w_2 < w_1 + \varepsilon\} = \int_0^{1-\varepsilon} dx_1 \int_{x_1}^{x_1+\varepsilon} f(x_1, x_2) dx_2 + \\ + \int_{1-\varepsilon}^1 dx_1 \int_{x_1}^1 f(x_1, x_2) dx_2,$$

$$(14) \quad P\{w_1 - \varepsilon < w_2 < w_1\} = \int_{\varepsilon}^1 dx_1 \int_{x_1-\varepsilon}^{x_1} f(x_1, x_2) dx_2 + \\ + \int_0^{\varepsilon} dx_1 \int_0^{x_1} f(x_1, x_2) dx_2,$$

$$(15) \quad P\{w_2 < w_1 - \varepsilon\} = \int_{\varepsilon}^1 dx_1 \int_0^{x_1-\varepsilon} f(x_1, x_2) dx_2.$$

Między wymienionymi czterema prawdopodobieństwami oraz prawdopodobieństwami rozważanymi w poprzednim rozdziale zachodzą następujące zależności:

$$(16) \quad P\{w_1 + \varepsilon < w_2\} + P\{w_1 < w_2 < w_1 + \varepsilon\} = P\{w_1 < w_2\},$$

$$(17) \quad P\{w_1 - \varepsilon < w_2 < w_1\} + P\{w_2 < w_1 - \varepsilon\} = P\{w_2 < w_1\},$$

$$(18) \quad P\{w_1 < w_2\} + P\{w_2 < w_1\} = 1.$$

Ponieważ $P\{w_1 < w_2\}$ znaleźliśmy poprzednio, zatem dla określenia prawdopodobieństw prawdziwości wymienionych czterech hipotez wystarczy znaleźć po jednym wyrazie lewych stron (16) i (17). Przy użyciu rozwinięcia na szereg Taylora możemy otrzymać z (13)

$$(19) \quad P\{w_1 < w_2 < w_1 + \varepsilon\} = (n_1 + 1)(n_2 + 1) \sum_{q=0}^{n_2} \frac{\varepsilon^{q+1}}{(q+1)!} \cdot A_q,$$

gdzie

$$A_q = \frac{n_2!}{(n_2 - q)!(n_1 + n_2 + 1 - q)!} \sum_i (-1)^{q-i} \frac{\binom{q}{i} \binom{n_1}{r_1} \binom{n_2 - q}{r_2 - i}}{\binom{n_1 + n_2 - q}{r_1 + r_2 - i}} -$$

$$- \sum_l (-1)^{n_1 + n_2 - r_1 - r_2} \frac{n_1! n_2!}{(n_1 - l)!(n_2 + l + 1 - q)!} \binom{l}{l + r_1 - n_1} \times$$

$$\times \binom{q - l - 1}{q - l - 1 + r_2 - n_2},$$

przy czym

$$\sup(0, q + r_2 - n_2) \leq i \leq \inf(q, r_2), \quad n_1 - r_1 \leq l \leq \inf(n_1, q + r_2 - n_2 - 1).$$

W podobny sposób można otrzymać z (14)

$$(20) \quad P\{w_1 - \varepsilon < w_2 < w_1\} = (n_1 + 1)(n_2 + 1) \sum_{q=0}^{n_1} \frac{\varepsilon^{q+1}}{(q+1)!} B_q,$$

gdzie

$$B_q = \frac{n_1!}{(n_1 - q)!(n_1 + n_2 + 1 - q)!} \sum_i (-1)^{q-1} \frac{\binom{q}{i} \binom{n_2}{r_2} \binom{n_1 - q}{r_1 - i}}{\binom{n_1 + n_2 - q}{r_1 + r_2 - i}} -$$

$$- \sum_l (-1)^{n_1 + n_2 - r_1 - r_2} \frac{n_1! n_2!}{(n_2 - l)!(n_1 + l + 1 - q)!} \binom{l}{l + r_2 - n_2} \times$$

$$\times \binom{q - l - 1}{q - l - 1 + r_1 - n_1},$$

przy czym

$$\sup(0, q + r_1 - n_1) \leq i \leq \inf(q, r_1), \quad n_2 - r_2 \leq l \leq \inf(n_2, q + r_1 - n_1 - 1).$$

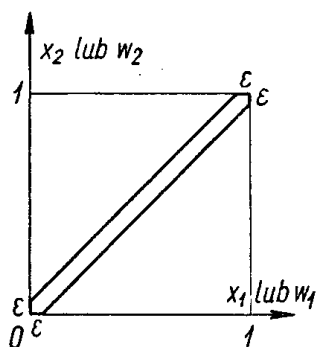
Hipoteza o ważnym znaczeniu praktycznym jest zawarta w następującym pytaniu (rysunek 4):

e. Jakie jest prawdopodobieństwo $P\{w_1 - \varepsilon < w_2 < w_1 + \varepsilon\}$ prawdziwości hipotezy, że frakcje obu populacyj różnią się od siebie mniej niż o ε ? Odpowiedź na to pytanie jest określona wzorem

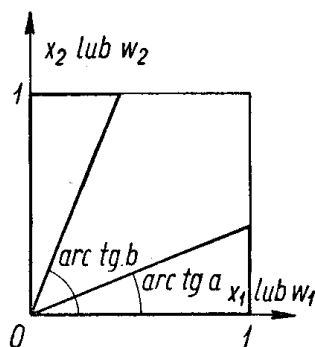
$$(21) \quad P\{w_1 - \varepsilon < w_2 < w_1 + \varepsilon\} = \int_0^\varepsilon dx_1 \int_0^{x_1 + \varepsilon} f(x_1, x_2) dx_2 +$$

$$+ \int_\varepsilon^{1-\varepsilon} dx_1 \int_{x_1 - \varepsilon}^{x_1 + \varepsilon} f(x_1, x_2) dx_2 + \int_{1-\varepsilon}^1 dx_1 \int_{x_1 - \varepsilon}^1 f(x_1, x_2) dx_2.$$

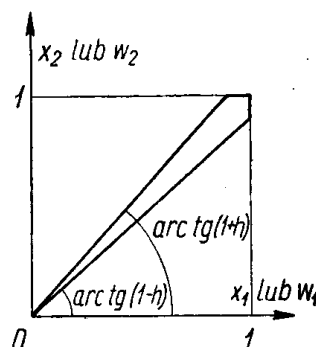
Prawdopodobieństwo to jest równe sumie prawdopodobieństw prawdziwości hipotez zawartych w pytaniach b i c.



Rys. 4



Rys. 5



Rys. 6

ZM-62-4+6

Hipotezy od a do e dotyczą różnicy frakcyj, jednakże w wielu przypadkach może nas interesować stosunek frakcyj. Wówczas możemy stawiać następujące trzy pytania:

f. Jakie jest prawdopodobieństwo $P\{w_2/w_1 < a\}$, że stosunek frakcji drugiej populacji generalnej do frakcji pierwszej populacji generalnej jest mniejszy od pewnego a , przy czym $a < 1$ (rysunek 5)?

g. Jakie jest prawdopodobieństwo $P\{w_2/w_1 > b\}$, że stosunek frakcji drugiej populacji generalnej do frakcji pierwszej populacji generalnej jest większy od pewnego b , przy czym $b > 1$ (rysunek 5)?

h. Jakie jest prawdopodobieństwo $P\{1-h < w_2/w_1 < 1+h\}$, że stosunek frakcji drugiej populacji generalnej do frakcji pierwszej populacji generalnej różni się od jedności mniej niż o h (rysunek 6)?

Odpowiedzi na te pytania dają następujące wzory:

$$(22) \quad P\{w_2/w_1 < a\} = \int_0^1 dx_1 \int_0^{ax_1} f(x_1, x_2) dx_2,$$

$$(23) \quad P\{w_2/w_1 > b\} = \int_0^{1/b} dx_1 \int_{bx_1}^1 f(x_1, x_2) dx_2,$$

$$(24) \quad P\{1-h < w_2/w_1 < 1+h\} = \int_0^{1/(h+1)} dx_1 \int_{(1-h)x_1}^{(1+h)x_1} f(x_1, x_2) dx_2 + \\ + \int_{1/(1+h)}^1 dx_1 \int_{(1-h)x_1}^1 f(x_1, x_2) dx_2.$$

Związek pomiędzy ostatnimi prawdopodobieństwami można otrzymać przyjmując

$$(25) \quad a = 1 - h,$$

$$(26) \quad b = 1 + h$$

i korzystając z tożsamości

$$(27) \quad P\{w_2/w_1 < 1-h\} + P\{1-h < w_2/w_1 < 1+h\} + P\{1+h < w_2/w_1\} = 1.$$

Rozwijając prawe strony (23) i (24) na szeregi Taylora otrzymujemy

$$(28) \quad P\{w_2/w_1 < a\} =$$

$$= 1 - \frac{n_1+1}{n_1+n_2+2} \binom{n_1}{r_1} \sum_{j=0}^{r_2} \binom{n_2+1}{j} a^j \sum_{k=0}^{n_2+1-j} (1-a)^k \frac{\binom{n_2+1-j}{k}}{\binom{n_1+n_2+1}{r_1+j+k}}.$$

$$(29) \quad P\{w_2/w_1 > b\} =$$

$$= 1 - \frac{n_2+1}{n_1+n_2+2} \binom{n_2}{r_2} \sum_{j=0}^{r_1} \binom{n_1+1}{j} \left(\frac{1}{b}\right)^j \sum_{k=0}^{n_1+1-j} \left(1 - \frac{1}{b}\right)^k \frac{\binom{n_2+1-j}{k}}{\binom{n_1+n_2+1}{r_2+j+k}}.$$

Rozwinięcie prawej strony (24) można łatwo otrzymać z (27), (28) i (29). Dowody wzorów (19), (20), (28) i (29), jak również wzorów (10) i (11) z poprzedniego rozdziału, pomijamy, ponieważ są one zbyt długie.

6. Możliwości uogólnień. Przedstawioną metodę można uogólnić w kilku kierunkach, z których dwa wydają się godne zaznaczenia. Przede wszystkim można byłoby weryfikować hipotezy statystyczne dotyczące liczby populacji generalnych większej od dwóch. Innym uogólnieniem byłaby weryfikacja hipotez dotyczących dwóch takich populacji generalnych, w których poszczególne sztuki różniłyby się od siebie nie brakiem lub posiadaniem jakiejś właściwości, lecz wartością liczbową tej właściwości.

Instytut Matematyczny Polskiej Akademii Nauk

(Praca wpłynęła dnia 13. 12. 1952 r.)

Ч. РАЙСКИЙ (Варшава)

СРАВНИВАНИЕ ГЕНЕРАЛЬНЫХ СОВОКУПНОСТЕЙ НА ОСНОВЕ ТЕОРЕМЫ БЕЙЕСА

РЕЗЮМЕ

Предлагаемая работа содержит решение следующей задачи: Даны две генеральные совокупности, каждая из коих содержит неизвестный процент выделяющихся некоторым образом штук. Из каждой совокуп-

ности вынимаем наудачу по одной выборке и отсчитываем находящиеся в них выделяющиеся штуки. Задача заключается в том, чтобы на основе результатов исследования выборок проверить любую гипотезу о взаимоотношении процентов штук выделяющихся в обеих совокупностях. Работа содержит общее решение задачи и разложение в ряд Тэйлора формул определяющих вероятность гипотез относительно разницы и относительно соотношения процентов выделяющихся штук. Частные формулы основаны на предположении равномерного распределения априорных вероятностей.

C. RAJSKI (Warszawa)

*COMPARING GENERAL POPULATIONS ON THE BASIS
OF BAYES' RULE*

SUMMARY

This paper contains a solution of the following problem: Two general populations are given each containing an unknown fraction of items marked in a certain way. Two samples are drawn at random, one from each population, and the numbers of marked items in each sample are checked. Now the problem is how to verify on the basis of the samples any hypothesis which can be made about the relation between fractions of marked items in both populations. The paper contains the general solution of this problem and Taylor's expansions of the probabilities of the hypothesis relating either to the difference or to the quotient of the fractions of the marked items. The particular formulae are based on the assumption of a uniform distribution of the prior probabilities.
