

BADANIA WPLYWU WIELKOŚCI ODCHYLENIA OD PRAWA HARDY-WEINBERGA NA DOKŁADNOŚĆ ESTYMACJI CZĘSTOŚCI GENÓW W POPULACJI

STANISŁAW GNOT, ANDRZEJ MICHALSKI

Katedra Matematyki, Ośrodek Obliczeniowy Akademii Rolniczej we Wrocławiu

Praca wpłynęła 1 kwietnia 1985; w wersji ostatecznej 12 grudnia 1985

Gnot S., Michalski A., 1987. Degree of departure from the Hardy-Weinberg Law on efficiency of estimators of gene frequencies in a population. Listy Biometryczne XXIII z.1. Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza w Poznaniu (Adam Mickiewicz University Press), pp. 3-16, 6 ryc., 3 tabl. ISBN 83-232-0090-4, ISSN 0458-0036.

In the paper a geometrical and genetical interpretation of a departure from the Hardy-Weinberg Law is given. Simulation methods are used to study the effect of the departure on efficiency of estimators of gene frequencies in a population.

1. WSTĘP

Jednym z podstawowych zagadnień statystycznych w genetyce populacyjnej jest estymacja częstości genów w badanej populacji. W modelach bez genu recesywnego estymatory tych parametrów można otrzymać bezpośrednio z zaobserwowanych w próbie liczebności fenotypów. W modelach z genem recesywnym estymacja częstości genów jest możliwa jedynie przy dodatkowych założeniach. Najczęściej przyjmowanym założeniem jest losowa asocjacja gamet i w konsekwencji równowaga Hardy-Weinberga (HW). Stąd w rozważaniach praktycznych estymację częstości genów poprzedza zazwyczaj testowanie hipotezy o równowadze HW. Test dokładny dla sprawdzenia tej hipotezy w dowolnym modelu genetycznym podali Ledwina i Gnot (1980). Jeżeli liczba elementów w próbie jest dostatecznie duża, powszechnie stosowaną statystyką testową jest statystyka χ^2 . Zdarza się jednakże, że rozkład czę-

stości fenotypów w badanej populacji odbiega nieco od stanu równowagi i wówczas odrzucenie tej hipotezy zależy jedynie od wielkości próby. W przypadku odrzucenia hipotezy dalsze postępowanie związane z estymacją częstości genów nie jest określone jednoznacznie. Tymczasem należy się spodziewać, że małe odstępstwo od równowagi ma niewielki wpływ na dokładność estymatorów obliczonych przy założeniu równowagi i może być pominięte. Z praktycznego punktu widzenia wydaje się zatem usprawiedliwione testowanie hipotezy o przybliżonej zgodności z równowagą HW. Dwie techniki testowania takiej hipotezy dla małych odchyień zostały przedstawione w pracach Bednarskiego i in. (1980, 1982). Jedna z nich oparta jest na tradycyjnej statystyce χ^2 , druga polega na wyborze pewnej miary odległości od krzywej równowagi i zastosowaniu tzw. N-testu. W obu tych pracach przedstawione są wyniki teoretyczne dotyczące testowania hipotez statystycznych. W tej pracy szczególną uwagę zwracamy na interpretację genetyczną i praktyczne wykorzystanie wyników teoretycznych. W szczególności podajemy genetyczną i geometryczną interpretację odchyień od równowagi HW. Badamy również wpływ wielkości zakłócenia równowagi na dokładność estymacji częstości genów.

2. ZAKŁÓCENIA RÓWNOWAGI HW W MODELU Z DWOMA ALLELAMI

Rozważmy pojedyncze miejsce w chromosomie, w którym mogą pojawić się dwa allele A i a. Niech $p_1(AA)$, $p_2(Aa)$ oraz $p_3(aa)$ będą częstościami poszczególnych fenotypów w badanej populacji. Przy założeniu losowej asocjacji gamet częstości te dane są wzorami

$$p_1 = \theta^2, p_2 = 2\theta(1-\theta), p_3 = (1-\theta)^2, \quad (2.1)$$

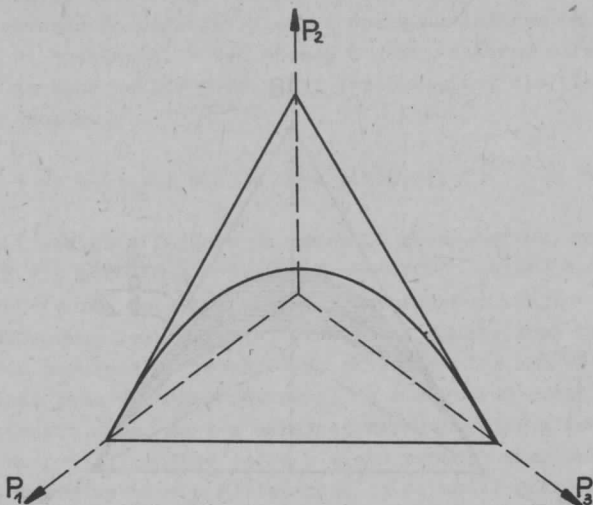
gdzie $\theta = p_1 + \frac{1}{2}p_2$ jest częstością genu A w badanej populacji. Populację z częstościami genotypów spełniającymi (2.1) nazywać będziemy populacją w stanie równowagi HW, a hipotezę

$$H: p_1(\theta) = \theta^2, p_2(\theta) = 2\theta(1-\theta), p_3(\theta) = (1-\theta)^2 \quad \theta \in (0,1),$$

hipotezą HW. Punkty równowagi HW w układzie trzech współrzędnych (p_1, p_2, p_3) wyznaczają parabolę równowagi przedstawioną na rycinie 1. Parabola ta dana jest równaniem równoważnym

$$p_2^2 = 4p_1p_3. \quad (2.2)$$

Podstawa sympleksu $p_2=0$ charakteryzuje populację będącą w skrajnej nierównowadze, dla której heterozygoty Aa zostały całkowicie wyeliminowane. Populację z częściowym zakłóceniem równowagi charakteryzują punkty leżące wewnątrz sympleksu, ale poza parabolą.



Ryc. 1. Parabola równowagi HW dla układu z dwoma allelami

Przykład. Selekcja naturalna jest jednym z możliwych czynników mogących spowodować zakłócenie równowagi HW. Przypuśćmy, że prawdopodobieństwami przeżycia genotypów AA, Aa oraz aa są odpowiednio w_1, w_2, w_3 . Niech θ i $1-\theta$ będą częstościami genów A i a. Przy założeniu losowej asocjacji gamet średnią częstością przeżycia jest

$$w = \theta^2 w_1 + 2\theta(1-\theta)w_2 + (1-\theta)^2 w_3,$$

a częstościami genotypów po selekcji są

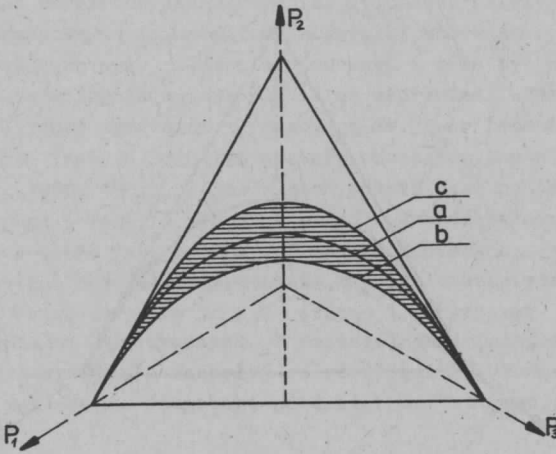
$$p_1 = \theta^2 w_1 / w, \quad p_2 = 2\theta(1-\theta)w_2 / w, \quad p_3 = (1-\theta)^2 w_3 / w.$$

Spełniają one następujące równanie

$$p_2^2 = 4kp_1p_3, \quad k = \frac{w_2^2}{w_1w_3}. \quad (2.2)$$

Równanie (2.2) jest równaniem:

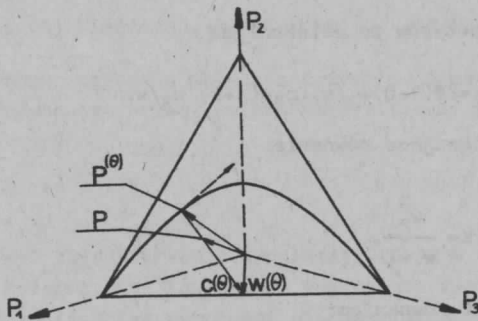
- paraboli równowagi gdy $k=1$,
- elipsy gdy $k < 1$,
- hiperboli gdy $k > 1$, (por. ryc.2).



Ryc. 2. Zakłócenia równowagi HW spowodowane selekcją z $k=2$ i $k=.5$

3. GENETYCZNA I GEOMETRYCZNA INTERPRETACJA ZAKŁÓCENIA RÓWNOWAGI HW

Przypuśćmy, że na skutek działania pewnych czynników w badanej populacji wektor $p = (p_1, p_2, p_3)$ częstości genotypów nie leży na paraboli równowagi, lecz w jej pewnym otoczeniu (ryc. 3).

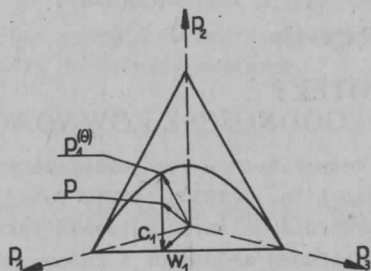


Ryc. 3. Geometryczna interpretacja zakłócenia

Niech c będzie wektorem prostopadłym do paraboli równowagi, przechodzącym przez koniec wektora p . Dla ustalenia uwagi przyjmujemy, że c jest wektorem skierowanym do podstawy $p_2=0$, o początku leżącym na paraboli i końcu leżącym na podstawie. W ten sposób c jest wektorem określonym jednoznacznie, a dowolny wektor p może być przedstawiony dla pewnego $\theta \in (0,1)$ w następujący sposób

$$p = p(\theta) + \alpha c = (1-\alpha) p(\theta) + \alpha w, \quad \alpha \in (0,1), \quad (3.1)$$

gdzie $p(\theta)$ jest wektorem leżącym na paraboli równowagi HW, tzn. spełniającym warunek (2.1), natomiast $w = w(\theta) = p(\theta) + c(\theta)$. Wzór (3.1) umożliwia następującą genetyczną interpretację zakłócenia. Populacja z wektorem częstości genotypów p jest w swojej strukturze genetycznej taka sama jak mieszanka dwóch populacji o procentowym składzie $(1-\alpha) \times 100\%$ i $\alpha \times 100\%$, z których pierwsza jest w stanie równowagi HW z wektorem częstości genotypów $p(\theta)$, natomiast druga jest w skrajnej nierównowadze z wektorem częstości $w(\theta)$. Wzór (3.1) podaje jedno z wielu przedstawień wektora p jako kombinacji liniowej wektora $p(\theta)$ leżącego na paraboli oraz wektora $w(\theta)$ o wierzchołku leżącym na podstawie sympleksu. Inne przedstawienie, dla którego można znaleźć również pewną interpretację genetyczną, jest wyznaczone przez wektor c_1 przechodzący przez koniec wektora p i prostopadły do podstawy sympleksu (ryc. 4).



Ryc. 4. Inne interpretacje zakłócenia

Wówczas $p = (1-\alpha_1)p_1(\theta) + \alpha_1 w_1$, gdzie $p_1(\theta)$ jest wektorem leżącym na paraboli równowagi oraz $w_1 = p_1(\theta) + c_1$. Warto zauważyć, że spośród wszystkich możliwych przedstawień minimalną wartość α określa przedstawienie (3.1).

Ten wybór przedstawienia wektora p jest ponadto ściśle związany z definicją otoczeń paraboli równowagi HW, przedstawioną w punkcie 4 niniejszego artykułu. Zależny od θ wektor c jest prostopadły do wektora pochodnych

$$b_0(\theta) = \left[\frac{dp_1(\theta)}{d\theta}, \frac{dp_2(\theta)}{d\theta}, \frac{dp_3(\theta)}{d\theta} \right] = 2 \left[\theta, 1-2\theta, -(1-\theta) \right].$$

Ponadto suma jego składowych jest równa 0, ponieważ suma składowych wektora $w = p(\theta) + c$ oraz $p(\theta)$ jest równa 1. Z powyższych rozważań oraz z warunku $p_2(\theta) + c_2 = 0$ otrzymujemy następujący układ równań:

$$\begin{aligned} c_1 + c_2 + c_3 &= 0, \\ c_1\theta + c_2(1-2\theta) - c_3(1-\theta) &= 0, \\ c_2 &= -2\theta(1-\theta). \end{aligned}$$

Jednoznacznym rozwiązaniem tego układu jest

$$\begin{aligned} c_1 &= 2\theta(1-\theta)(2-3\theta), \\ c_2 &= -2\theta(1-\theta), \\ c_3 &= 2\theta(1-\theta)(3\theta-1). \end{aligned} \tag{3.2}$$

Ponieważ $w(\theta) = p(\theta) + c(\theta)$, więc składowymi wektora $w(\theta)$ są

$$\begin{aligned} w_1 &= \theta^2 + 2\theta(1-\theta)(2-3\theta), \\ w_2 &= 0, \\ w_3 &= (1-\theta)^2 + 2\theta(1-\theta)(3\theta-1). \end{aligned} \tag{3.3}$$

Stąd $p_1 = \theta^2 + 2\alpha\theta(1-\theta)(2-3\theta)$, $p_2 = 2(1-\alpha)\theta(1-\theta)$,
 $p_3 = (1-\theta)^2 + 2\alpha\theta(1-\theta)(3\theta-1)$.

4. TESTOWANIE HIPOTEZY O PRZYBLIŻONEJ ZGODNOŚCI Z RÓWNOWAGĄ HW

Szczegółów dotyczące rozważanych w tym rozdziale problemów przedstawione są w pracach Bednarskiego i in. (1980, 1982). Tutaj ograniczymy się jedynie do podania niezbędnych dla dalszych rozważań faktów, dotyczących testowania hipotezy o przybliżonej zgodności z równowagą HW.

Niech n_1, n_2, \dots, n_K będą liczebnościami o rozkładzie wielomianowym z wektorem prawdopodobieństw $p = (p_1, p_2, \dots, p_K)$, $\sum p_i = 1$, $\sum n_i = n$. Rozważmy hipotezę

$$H_0 : p_i = p_i(\theta), \quad \theta \in \Theta \subset \mathbb{R}^r, \quad r < K,$$

orzekającą, że p_i są pewnymi funkcjami nieznanego wektora parametrów θ , $\theta = (\theta_1, \theta_2, \dots, \theta_r)$. Niech $P_0 = \{p \in S, p = p(\theta), \theta \in \Theta\}$ będzie hipotetycznym zbiorem wektorów p . Tutaj S jest sympleksem, tzn. zbiorem wektorów p , dla których $\sum p_i = 1$. Dla ustalonego n określmy \mathcal{E} -otoczenie zbioru P_0 w następujący sposób

$$P_{\varepsilon, n} = \{ p \in S; \sum a_i^2(p, \theta) \leq \varepsilon^2/n, \theta \in \Theta \},$$

gdzie $a_i(p, \theta) = [p_i - p_i(\theta)] / \sqrt{p_i(\theta)}$, $i=1, 2, \dots, K$, natomiast wektor p jest tak dobrany, aby a był wektorem prostopadłym do wierszy macierzy pochodnych cząstkowych

$$B = \{ \sqrt{p_i(\theta)} \partial p_i(\theta) / \partial \theta_j \}.$$

Interesuje nas problem testowania hipotezy $H_{\varepsilon, n}$: $p \in P_{\varepsilon, n}$ orzekającej, że p należy do ε -otoczenia zbioru P_0 . Tak sformułowana hipoteza o ε -zgodności z modelem P_0 umożliwia znalezienie asymptotycznego rozkładu statystyki

$$\chi^2 = \sum_{i=1}^K \frac{[n_i - np_i(\hat{\theta})]^2}{np_i(\hat{\theta})}, \quad (4.1)$$

gdzie $\hat{\theta}$ jest estymatorem największej wiarygodności (NW) dla θ obliczonym przy modelu P_0 . Bednarski (1981) udowodnił, że asymptotycznym rozkładem statystyki (4.1) przy założeniu prawdziwości hipotezy $H_{\varepsilon, n}$ jest niecentralny rozkład χ^2 z $K-r-1$ stopniami swobody i parametrem niecentralności ε^2 . Dla celów praktycznych otoczenie $P_{\varepsilon, n}$ wygodnie jest przedstawić w nieco innej formie. Zauważmy, że dla $p = p(\theta) + \alpha c(\theta) = (1-\alpha)p(\theta) + \alpha w(\theta)$, gdzie $w(\theta) = p(\theta) + c(\theta)$, mamy

$$a_i = \alpha c_i(\theta) / \sqrt{p_i(\theta)},$$

a wektor $a(\theta)$ jest prostopadły do wierszy macierzy B wtedy i tylko wtedy, gdy $c(\theta)$ jest prostopadły do wierszy macierzy

$$B_0 = \left\{ \frac{\partial p_i(\theta)}{\partial \theta_j} \right\},$$

tzn. $c \in N(B_0)$, gdzie $N(B_0)$ jest jądrem macierzy B_0 . Stąd

$$P_{\varepsilon, n} = \left\{ p \in S, p = p(\theta) + \alpha c(\theta), \sum \frac{c_i^2(\theta)}{p_i(\theta)} \leq \frac{\varepsilon^2}{n\alpha^2}, c \in N(B_0) \right\}. \quad (4.2)$$

4.1. TESTOWANIE PRZYBLIŻONEJ ZGODNOŚCI Z RÓWNOWAGĄ HW W MODELU Z DWOMA ALLELAMI

W zastosowaniu powyższych rozważań do modelu z dwoma allelami, występujący w definicji otoczenia $P_{\varepsilon, n}$ wektor $c(\theta)$ oraz wektor $w(\theta) = p(\theta) + c(\theta)$ określone są wzorami (3.2) oraz (3.3). Korzystając ze wzoru (3.2) łatwo wykazać, że warunek

$$\sum \frac{c_1^2(\theta)}{p_1(\theta)} \leq \frac{\varepsilon^2}{n\alpha^2}$$

jest spełniony dla $\varepsilon \geq \varepsilon_0$, gdzie

$$\varepsilon_0^2 = n\alpha^2 [4(1-\theta)^2(2-3\theta)^2 + 2\theta(1-\theta) + 4\theta^2(3\theta-1)^2].$$

Jeżeli zatem godzimy się na odstępstwo od równowagi rzędu α , to powinniśmy testować hipotezę $H_{\varepsilon_0, n}: p \in P_{\varepsilon_0, n}$. Testowanie tej hipotezy jest jednak praktycznie niemożliwe, ponieważ ε_0 jest funkcją nieznanego parametru θ . Chcąc uwolnić się od tej zależności zauważmy, że $\min_{\theta} \varepsilon_0^2(\theta) = n\alpha^2$ i jest osiągnięte dla $\theta=1/2$. W związku z powyższym testując hipotezę

$$H_{\sqrt{n}\alpha, n}: p \in P_{\sqrt{n}\alpha, n}$$

zabezpieczamy się przed każdą z możliwych wartości parametru θ , tzn. w przypadku przyjęcia tej hipotezy nie ma podstaw, aby przypuszczać, że odstępstwo od równowagi jest większe niż to spowodowane mieszaniem populacji z wektorem $p = (1-\alpha)p(\theta) + \alpha w(\theta)$, niezależnie od wartości nieznanego parametru θ .

5. ZAKŁÓCENIE RÓWNOWAGI HW W UKŁADZIE AB-0 GRUP KRWI

Układ AB-0 grup krwi jest tu modelowym przykładem układu z trzema allelami, z których 0 jest genem recesywnym. W tym układzie przy założeniu losowej asocjacji gamet częstości p_1, p_2, p_3 i p_4 fenotypów A, B, AB i 0 spełniają warunek

$$p_1 = \theta_1^2 + 2\theta_1\theta_3, \quad p_2 = \theta_2^2 + 2\theta_2\theta_3, \quad p_3 = 2\theta_1\theta_2, \quad p_4 = \theta_3^2, \quad (5.1)$$

gdzie $\theta_1, \theta_2, \theta_3$ są częstościami genów A, B i 0, $\sum \theta_i = 1$. Można wykazać, że (5.1) jest równoważny warunkowi

$$\sqrt{p_1+p_4} + \sqrt{p_2+p_4} = 1 + \sqrt{p_4} \quad (5.2)$$

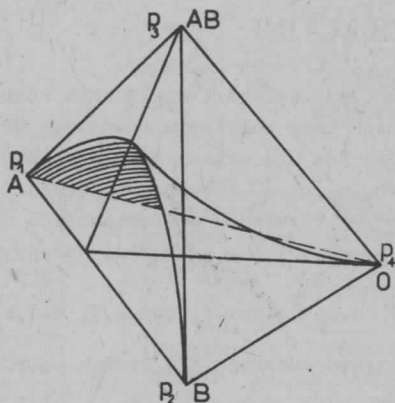
lub

$$p_3 = 2 \left[\sqrt{p_1+p_4} \sqrt{p_2+p_4} - \sqrt{p_4} \right]. \quad (5.3)$$

Powyższe równanie wraz z warunkiem $\sum p_i = 1$ opisuje powierzchnię, której wykres na sympleksie przestrzeni 4-wymiarowej przedstawia rycina 5.

Macierz B_0 pochodnych cząstkowych jest dla tego układu postaci

$$B_0(\theta) = 2 \begin{pmatrix} \theta_3 & -\theta_2 & \theta_2 & -\theta_3 \\ -\theta_1 & \theta_3 & \theta_1 & -\theta_3 \end{pmatrix}.$$



Ryc. 5. Powierzchnia równowagi HW dla układu AB-O grup krwi

Podobnie jak dla układu dwuallelowego poszukujemy wektora $c(\theta)$ prostopadłego do wierszy macierzy $B_0(\theta)$, spełniającego warunek $\sum c_i(\theta) = 0$ dla każdego θ . Łatwo wykazać, że wektor $c(\theta)$ wyznaczony jest z dokładnością do stałej przez

$$\begin{aligned}
 c_1 &= (\theta_1 - \theta_3)(\theta_2 - \theta_3) + 4\theta_2\theta_3, \\
 c_2 &= (\theta_1 - \theta_3)(\theta_2 - \theta_3) + 4\theta_1\theta_3, \\
 c_3 &= (\theta_1 - \theta_3)(\theta_2 - \theta_3) - 4\theta_3^2, \\
 c_4 &= (\theta_1 - \theta_3)(\theta_2 - \theta_3) - 4\theta_1\theta_2.
 \end{aligned}
 \tag{5.4}$$

Stałą τ należy tak dobrać, aby podobnie jak w przypadku modelu z dwoma allelami wektor $w(\theta) = p(\theta) + c(\theta)$ był wektorem częstości fenotypowych w populacji będącej w skrajnej nierównowadze. Geometrycznie oznacza to, że koniec wektora $w(\theta)$ leży na jednej ze ścian sympleksu, tzn. że co najmniej jedna składowa wektora $w(\theta)$ jest zerowa, natomiast pozostałe są dodatnie i sumują się do jedności. Rozważmy dwa możliwe rozwiązania zależne od parametru θ . W pierwszym przypadku $w_3 = 0$, tzn. całkowicie wyeliminowane są fenotypy AB, a

$$\tau_1 = \frac{2\theta_1\theta_2}{2\theta_3^2 + \theta_3 - \theta_1\theta_2}.$$

W drugim przypadku $w_4 = 0$ i wyeliminowane są fenotypy O, a

$$\tau_2 = \frac{-\theta_3^2}{2\theta_3^2 - \theta_3 - 3\theta_1\theta_2}.$$

Otoczenia $P_{\xi, n}$ wyznaczają wektory $p = p(\theta) + \tau_1 c(\theta)$, gdzie wektor $c(\theta)$ dany jest wzorem (5.4).

6. BADANIE SYMULACYJNE

W celu zbadania wpływu wielkości zakłócenia równowagi HW na dokładność estymatorów obliczonych przy założeniu równowagi HW przeprowadzono następujące badania symulacyjne dla układu AB-0 grup krwi. Wygenerowano po 100 przykładów wartości n_1, n_2, n_3, n_4 , reprezentujących liczebności fenotypowe A, B, AB, 0 w próbie według rozkładów wielomianowych z całkowitą liczbą obserwacji $N = 100$ ($\sum n_i = 100$) i z wektorami prawdopodobieństw fenotypów

$$p(i, k) = p\{\theta(i)\} + \alpha_k \tau c\{\theta(i)\}, \quad i=1, 2, 3; \quad k=1, 2, \dots, 6. \quad (5.5)$$

Badania symulacyjne przeprowadzono dla trzech wektorów częstości genów

$$\theta(1) = (1/3, 1/3, 1/3),$$

$$\theta(2) = (1/4, 1/4, 1/2),$$

$$\theta(3) = (1/2, 3/10, 1/5)$$

oraz następujących wielkości zakłóceń

$$\alpha_k = k \cdot 0.05, \quad k=1, 2, \dots, 6.$$

Na bazie tak wygenerowanych liczebności obliczono metodą NW (por. Elandt-Johnson (1971)) estymatory wektora θ dla populacji o zakłóconej równowadze, o częstościach fenotypów danych wzorami (5.5). Obliczenia wykonano według zamieszczonego schematu.

W populacji będącej w równowadze HW wektor θ wyznaczony jest jednoznacznie przez wektor p częstości fenotypów w następujący sposób (por. Gnot i in. (1978)):

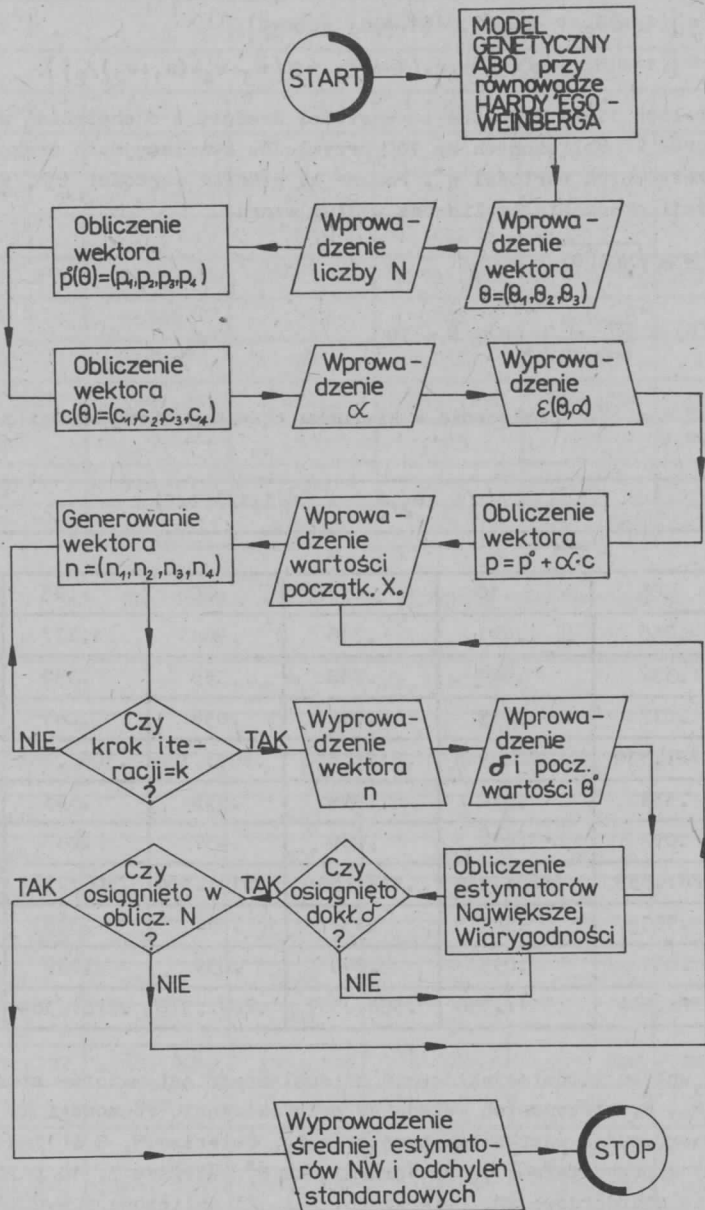
$$\begin{aligned} \theta_1 &= p_4 + p_1 - \sqrt{p_4(p_4 + p_1)} + (1/2)p_3, \\ \theta_2 &= p_4 + p_2 - \sqrt{p_4(p_4 + p_2)} + (1/2)p_3, \\ \theta_3 &= 1 - \theta_1 - \theta_2. \end{aligned} \quad (6.1)$$

Dla populacji o zakłóconej równowadze z wektorem częstości fenotypów $p = (1-\alpha)p(\theta) + \alpha w(\theta)$, częstościami genów są $\theta^* = (1-\alpha)\theta + \alpha\bar{\theta}$, gdzie θ dany jest wzorem (6.1) dla $p = p(\theta)$, natomiast $\bar{\theta}$ jest wektorem częstości genów w populacji będącej w skrajnej nierównowadze. Wektor $\bar{\theta}$ nie jest jednak określony jednoznacznie przez $w(\theta)$, a jego składowe mogą przyjmować wartości z przedziałów

$$\bar{\theta}_1 \in [w_1/2, w_1],$$

$$\bar{\theta}_2 \in [w_2/2, w_2],$$

$$\bar{\theta}_3 \in [w_3 + w_4, w_3 + w_4 + (w_1 + w_2)/2].$$



Ryc. 6. Schemat blokowy programu

Stąd teoretyczne wartości składowych wektora θ^* należą do przedziałów

$$\theta_1^* \in [(1-\alpha)\theta_1 + \alpha w_1/2, (1-\alpha)\theta_1 + \alpha w_1],$$

$$\theta_2^* \in [(1-\alpha)\theta_2 + \alpha w_2/2, (1-\alpha)\theta_2 + \alpha w_2],$$

$$\theta_3^* \in [(1-\alpha)\theta_3 + \alpha(w_3+w_4), (1-\alpha)\theta_3 + \alpha\{w_3+w_4+(w_1+w_2)/2\}].$$

W tabelach 1, 2 i 3 podane są wartości średnie i odchylenia standardowe estymatorów $\hat{\theta}$ obliczonych ze 100 przykładów symulacyjnych oraz przedziały dla teoretycznych wartości θ^* . Podane są również wartości ϵ_0 , występujące w definicji otoczenia, obliczone według wzoru

$$\epsilon_0 = \alpha \sqrt{N \cdot A(\theta)},$$

gdzie $A(\theta) = \sum_{i=1}^4 c_i^2/p_i(\theta)$, $N = 100$.

T a b l i c a 1. Zakłócenie w kierunku skrajnej nierównowagi z brakiem recesywów 0

$(\theta_1, \theta_2, \theta_3) = (1/3, 1/3, 1/3)$						
α						
	.05	.10	.15	.20	.25	.30
ϵ_0	.245	.491	.736	.981	1.227	1.472
$\hat{\theta}_1$.332	.327	.338	.335	.339	.328
SD_1	.037	.035	.034	.038	.041	.037
θ_1^*	.328:.339	.322:.344	.317:.350	.311:.356	.306:.361	.300:.367
$\hat{\theta}_2$.333	.332	.330	.335	.333	.334
SD_2	.037	.037	.038	.037	.037	.039
θ_2^*	.328:.339	.322:.344	.317:.350	.311:.356	.306:.361	.300:.367
$\hat{\theta}_3$.335	.341	.332	.330	.329	.328
SD_3	.041	.033	.040	.039	.039	.037
θ_3^*	.322:.344	.311:.356	.300:.367	.289:.378	.278:.389	.267:.400

Ocenę wpływu stopnia zakłócenia na dokładność estymatorów częstości genów $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ otrzymanych metodą NW przy założeniu równowagi HW można dokonać porównując wartości estymatorów $\hat{\theta}_i$ (wiersze 5, 8 i 11 w tabelach 1, 2 i 3) z przedziałami teoretycznymi dla θ_i^* (wiersze 7, 10 i 13), a ich odchylenia standardowe SD_i (wiersze 6, 9 i 12) obliczone z wyników symulacji z asymptotycznymi odchyleniami standardowymi estymatorów NW obliczonymi metodami analitycznymi (por. Elandt-Johnson (1971)). Dla $\theta = (1/3, 1/3, 1/3)$ (tabl. 1) średnie estymatorów uzyskanych w wyniku symulacji są dla wszystkich α bliskie środkom przedziałów, a odchylenia standardowe rzędu .040

T a b l i c a 2. Zakłócenie w kierunku skrajnej nierównowagi z brakiem kodominantów AB

$(\theta_1, \theta_2, \theta_3) = (1/4, 1/4, 1/2)$						
α						
	.05	.10	.15	.20	.25	.30
ϵ_0	.202	.404	.607	.809	1.011	1.210
$\hat{\theta}_1$.249	.239	.245	.244	.245	.234
SD ₁	.033	.034	.033	.032	.035	.035
θ_1^*	.247:.257	.244:.264	.242:.271	.239:.278	.236:.284	.233:.291
$\hat{\theta}_2$.244	.246	.242	.242	.238	.241
SD ₂	.032	.034	.032	.032	.032	.034
θ_2^*	.247:.257	.244:.264	.242:.271	.239:.278	.236:.284	.233:.291
$\hat{\theta}_3$.507	.515	.513	.515	.517	.525
SD ₃	.036	.041	.041	.040	.036	.039
θ_3^*	.486:.506	.473:.511	.459:.517	.444:.523	.431:.528	.418:.534

T a b l i c a 3. Zakłócenie w kierunku skrajnej nierównowagi z brakiem recesywów O

$(\theta_1, \theta_2, \theta_3) = (1/2, 3/10, 1/5)$						
α						
	.05	.10	.15	.20	.25	.30
ϵ_0	.106	.213	.320	.426	.533	.639
$\hat{\theta}_1$.485	.487	.506	.499	.491	.502
SD ₁	.039	.041	.039	.036	.041	.040
θ_1^*	.487:.498	.473:.497	.460:.495	.447:.494	.434:.492	.420:.491
$\hat{\theta}_2$.304	.305	.296	.304	.309	.302
SD ₂	.033	.032	.034	.034	.032	.036
θ_2^*	.291:.297	.282:.294	.273:.291	.264:.288	.255:.285	.246:.282
$\hat{\theta}_3$.212	.208	.198	.197	.200	.196
SD ₃	.035	.032	.039	.038	.036	.038
θ_3^*	.205:.222	.209:.245	.214:.267	.218:.289	.223:.311	.227:.334

bliskie odchyleniom standardowym dla estymatorów NW, które są równe odpowiednio .037, .037, .041 dla estymacji θ_1 , θ_2 , θ_3 . W tym przypadku potwierdza się wcześniejsze przypuszczenie autorów o niewielkim wpływie małych

odchylen na dokładność estymacji. Dla dużych odchylen ($\alpha > .20$) długości przedziałów teoretycznych są zbyt duże, przez co ocena dokładności jest utrudniona.

LITERATURA

- Bednarski, T. (1981). On optimal properties of the chi-square test of fit for testing ϵ -validity of parametric models. *Math. Operationsf. Stat., Ser. Statistics* 12, 31-41.
- Elandt-Johnson, R. (1971). *Probability Models and Statistical Methods in Genetics*. John Wiley, New York.
- Gnot, S., Matej, H., Szulga, T. (1978). Efficiency of the determinant method of estimation gene frequencies in three allelic models. *Arch. Immunol. Ther. Exp.* 26, 165-167.
- Bednarski, T., Gnot, S., Ledwina, T. (1980). Testing approximate validity of Hardy-Weinberg Law in population genetics. *Instytut Matematyki Politechniki Wrocławskiej, Raport* 32.
- Bednarski, T., Gnot, S., Ledwina, T. (1982). Testing approximate validity of Hardy-Weinberg Law. *W. Grossmann et al., Probability and Statistical Inference*, 35-46.
- Ledwina, T., Gnot, S. (1980). Testing for Hardy-Weinberg Equilibrium. *Biometrics* 36. 162-165.