UDC 519.24

D. KOROLIOUK, V.S. KOROLIUK

EQUILIBRIUM IN WRIGHT–FISHER MODELS OF POPULATION GENETICS

Abstract. For multivariant Wright–Fisher models in population genetics, we introduce equilibrium states, expressed by fluctuations of probability ratio, in distinction of the traditionally used fluctuations, expressed by the difference between the current value of the random process and its equilibrium value. Then the drift component of the gene frequencies dynamic process, primarily expressed as a ratio of two quadratic forms, is transformed into a cubic parabola with a certain normalization factor.

Keywords: Wright–Fisher model, population genetics, evolutionary process, equilibrium state, fluctuations of probability ratio.

INTRODUCTION

The population genetics models by Wright–Fisher are defined by regression functions which are determined by a ratio of two quadratic forms. In the present work, the models of population genetics of genotypes interaction are determined by difference evolution equations with regression functions of increments for the frequency probabilities of genotypes. In this case, the equilibrium state of the probabilities frequency is given by the equilibrium of the regression function of increments, which is postulated by the form of such a function.

1. REGRESSION FUNCTION OF INCREMENTS

The probabilities of genotype frequencies at each stage $k \ge 0$ are determined by the evolutionary process $P(k) = (P_m(k), 0 \le m \le M)$ with $M+1$ $(M \ge 1)$ finite number of the state set $E = e_0, e_1, \dots e_M$.

The dynamics of the frequency probabilities at the next $(k+1)$-th stage $(k \ge 0)$ is given by the regression function [1, Ch.10]:

$$P_m(k+1) = W_m(p)/W(p), \ 0 \le m \le M, \ k \ge 0, \tag{1}$$

$$W_m(p) = p_m \sum_{n=0}^{M} W_{mn} p_n, \ 0 \le m \le M,$$

$$W(p) = \sum_{n=0}^{M} W_n(p). \tag{2}$$

The probabilities of frequencies obey the usual restrictions $0 \le p_m \le 1$, $\sum_{n=0}^{M} p_n = 1$. The corresponding restrictions for the survival parameters are $0 \le W_{mn} \le 1$, $0 \le m, n \le M$.

The increment of probability at each stage

$$\Delta P_m(k+1) := P_m(k+1) - P_m(k), \ 0 \le m \le M, \ k \ge 0,$$

is given by the incremental regression function [2, 3]:

$$\Delta P_m(k+1) = W_0^{(m)}(p), \ 0 \le m \le M, \tag{3}$$

$$W_0^{(m)}(p) = V_0^{(m)}(p)/W(p), \ V_0^{(m)}(p) := W_m(p) - p_m W(p), \ 0 \le m \le M. \tag{4}$$

© D. Koroliouk, V.S. Koroliuk, 2019

Let us introduce new parameters of survival:

$$V_{mn} := 1 - W_{mn}\ ,\ 0 \le m, n \le M.$$

Then the numerator of incremental regression function (4) is transformed to the form:

$$V_0^{(m)}(p) = p_m \left[\sum_{n=0}^{M} p_n (V_n,\, p) - (V_m,\, p) \right],\ 0 \le m \le M, \tag{5}$$

and the normalizing denominator (2) has the form

$$W(p) = 1 - \sum_{n=0}^{M} p_n (V_n,\, p)\ . \tag{6}$$

where the scalar product

$$(V_m,\, p) := \sum_{n=0}^{M} V_{mn} p_n\ ,\ 0 \le m \le M. \tag{7}$$

Let us introduce the equilibriums of incremental regression functions (5) by the relations

$$(V_m, \rho) = \pi\ ,\ 0 \le m \le M\ ,\ \pi := \prod_{n=0}^{M} \rho_n. \tag{8}$$

The normalized constant π is also generated by equilibriums $\rho = (\rho_m,\, 0 \le m \le M)$.

Lemma 1. The equilibria of incremental regression functions are given by the relation

$$\rho_m = \pi \overline{V}_m,\ 0 \le m \le M,\ \pi := \prod_{n=0}^{M} \rho_n, \tag{9}$$

where $\overline{V}_m := \sum_{n=0}^{M} \overline{V}_{mn}$, $0 \le m \le M$, with the summands which are the elements of inverse matrix, $\mathbb{V}^{-1} := [\overline{V}_{mn}, 0 \le m, n \le M]$ with respect to the directing parameters matrix; $\mathbb{V} := [\overline{V}_{mn}, 0 \le m, n \le M]$, under the additional normalization condition $\sum_{m=0}^{M} \overline{V}_m = \sum_{m,n=0}^{M} \overline{V}_{mn} = \pi^{-1}$.

Proof. Relation (8) means that

$$\mathbb{V}\rho = \pi 1,\ \pi 1 := (\pi, 0 \le n \le M)\ .$$

Hence the vector of equilibriums has the following representation:

$$\rho = \pi \mathbb{V}^{-1} 1,\ \rho_n = \overline{V}_n \pi\ ,\ 0 \le n \le M\ ,$$

that is, the assertion of Lemma (9).

Corollary 1. Equilibria (9) provide the equilibrium state of the probability frequency (4)

$$V_0^{(m)}(\rho) \equiv 0,\ 0 \le m \le M. \tag{10}$$

Corollary 2. Equilibria (9) generate a representation of the scalar products (7) by fluctuations of the probability ratio

$$(V_m,\, p) = \pi p_m / \rho_m,\ 0 \le m \le M. \tag{11}$$

The normalizing constant π is defined in (8).

First of all, note that relation (9) coincides with the definition of the equilibrium (8), under additional assumption that the directing parameters matrix $\mathbb{V} = V_m \delta_m$; $0 \le m, n \le M$, is diagonal. Hence, we have the following lemma.

Lemma 2. There takes place the following relation:

$$\pi\rho^{-1} = \mathbb{V}1, \tag{12}$$

which coincides with formula (11).

Now the incremental regression functions (4)–(6) with the relations (10) generate the following proposition.

Proposition 1. The incremental regression functions with Wright–Fisher normalization are given by the relations

$$W_0^{(m)}(p) = V_0^{(m)}(p) / W(p),$$

$$V_0^{(m)}(p) = \pi\rho_m \left[\sum_{n=0}^{M} p_n^2 / \rho_n - p_m / \rho_m \right], \tag{13}$$

$$W(p) = 1 - \pi \sum_{n=0}^{M} p_n^2 / \rho_n .$$

The global balance condition $$\sum_{m=0}^{M} V_0^{(m)}(p) = 0$$

is obvious and has the following scalar form:

$$\sum_{m=0}^{M} p_m \sum_{n=0}^{M} p_n^2 / \rho_n - \sum_{m=0}^{M} p_m^2 / \rho_m \equiv 0 .$$

2. EQUILIBRIUM STATE

The presence of equilibrium state is provided by the local balance condition

$$V_0^{(m)}(\rho) = 0, \; 0 \le m \le M, \; \rho = (\rho_m, 0 \le m \le M) .$$

The normalizing Wright–Fisher factor has the form

$$W(\rho) = 1 - \pi, \; \pi = \prod_{m=0}^{M} \rho_m . \tag{14}$$

The equilibrium generated by the state $\rho = (\rho_m, \; 0 \le m \le M)$, is interpreted by the convergence of evolutionary processes [4, 5].

Theorem 1. For any initial data $0 < P_m(0) < 1$, $0 \le m \le M$, the evolutionary processes $P_m(k)$, $0 \le m \le M$, $k \ge 0$, which are determined by solutions of difference evolutionary equation (3) with the incremental regression function (12)–(13), converge to equilibrium as $k \to \infty$:

$$\lim_{k \to \infty} P_m(k) = \rho_m , \; 0 \le m \le M.$$

Proof. The property of the main components is used, which is specified by the sum

$$\sum_{n=0}^{M} p_n^2 / \rho_n = \sum_{n=0}^{M} p_n (p_n / \rho_n) .$$

This means averaging the fluctuations of the probability ratio p_n / ρ_n, $0 \le n \le M$, on the distribution of frequencies at the current stage. In this case, the fluctuations of the ratios are equal to one for $p_n = \rho_n$, $0 \le n \le M$, and at the same time, the main component of the incremental regression function is also equal to one.

Consequently, the possible values of the frequency probabilities can be split into three zones:

$$(+)\ \ p_n < \rho_n; \quad (-)\ \ p_n > \rho_n; \quad (0)\ \ p_n = \rho_n .$$

The signs of the incremental regression functions (12)–(13) in such zones are the same: in zone (+) the probabilities increase, in zone (–) they decrease.

Therefore, there exists a limit (14) whose value is ensured by the necessary condition for the existence of a limit:

$$\lim_{k\to\infty} \Delta P_m(k+1) = 0 .$$

3. BINARY EVOLUTIONARY PROCESS

The binary evolutionary process $P_\pm(k)$, $k \ge 0$, is determined by the following regression functions [1]:

$$P_\pm(k+1) = W_\pm(p)/W(p),\ \ k \ge 0, \tag{15}$$

$$W_\pm(p) = P_\pm(W_\pm p_\pm + p_\mp),$$

$$W(0) = W_+(p_+) + W_-(p_-) = W_+ p_+^2 + 2p_+ p_- + W_- p_-^2 .$$

The frequency probabilities at k-th stage satisfy the usual conditions $0 \le p_\pm \le 1$, $p_+ + p_- = 1$. The survival parameters are also limited by the relation $0 < W_\pm < 1$.

For probability increments

$$\Delta P_\pm(k+1) := P_\pm(k+1) - P_\pm(k),\ \ k \ge 0,$$

the corresponding regression functions of increments can be represented as follows:

$$W_0^\pm(p_\pm) = W_\pm(p_\pm)/W(p) - p_\pm ,$$

or, equivalently,

$$W_0^\pm(p_\pm) = V_0^\pm(p_\pm)/W(p), \tag{16}$$

$$V_0^\pm(p_\pm) = W_\pm(p_\pm) - p_\pm W(p) .$$

Now, let us introduce the direction parameters based on the survival ones: $V_\pm := 1 - W_\pm$. The relative equilibriums are $\rho_\pm = V_\pm^{-1}$ with the normalization condition $V_+ + V_- = 1$. Then the numerators (15) of the regression function have the following form:

$$W_\pm(p_\pm) = p_\pm(1 - \pi p_\mp / \rho_\pm),\ \ \pi := \rho_+ \rho_- .$$

Therefore, the numerator (16) transforms into the following:

$$V_0^\pm(p_\pm) = p_\mp W_\pm(p_\pm) - p_\pm W_\mp(p_\mp) = p_+ p_-(\rho_\pm p_\mp - \rho_\mp p_\pm) .$$

The linear component has the following representation:

$$\rho_+ p_- - \rho_- p_+ = -(p_+ - \rho_+) = p_- - \rho_- .$$

So the regression functions of the increments of binary evolutionary processes are represented by the probability of fluctuations

$$V_0^\pm(p_\pm) = -p_+ p_-(p_\pm - \rho_\pm) .$$

The Wright–Fisher normalizing factor has the following representation:

$$W(p) = 1 - \pi[p_+^2 / \rho_+ + p_-^2 / \rho_-] = 1 - [\rho_- p_+^2 + \rho_+ p_-^2].$$

The global balance condition is also evident:

$$V_0^+(p_+) + V_0^-(p_-) \equiv 0.$$

CONCLUSIONS

The considered in the present work evolution processes serve as predictable components of stochastic models in population genetics and are represented by conditional mathematical expectations:

$$P_m(k+1) := E[S_N^{(m)}(k+1) | S_N(k) = P(k)], \ 0 \le m \le M, \ k \ge 0.$$

The stochastic models in population genetics are determined by averaged sums

$$S_N(k) := \frac{1}{N} \sum_{n=1}^{N} \delta_n(k), \ k \ge 0. \tag{17}$$

of random sample variables $\delta_n(k)$, $1 \le n \le N$, which take values in a finite set with $M+1$ $(M \ge 1)$ states $E = \{e_0, e_1, \ldots, e_M\}$ (see [6]). So the stochastic models (17) are defined by the sum of two components:

$$S_N(k+1) = V(S_N(k)) + \Delta\mu_N(k+1), \ k \ge 0. \tag{18}$$

The first, predictable component is generated by conditional mathematical expectations:

$$V_m(P_m(k)) = P_m(k) + V_0^{(m)}(P_m(k)) / W(P_m(k)), \ 0 \le m \le M, \ k \ge 0. \tag{19}$$

The second component forms a martingale differences

$$\Delta\mu_N(k+1) = S_N(k+1) - V(S_N(k)), \ k \ge 0, \tag{20}$$

characterized by the first two moments:

$$E\Delta\mu_N^{(m)}(k+1) = 0, \ 0 \le m \le M, \tag{21}$$

$$E[(\Delta\mu_N^{(m)}(k+1))^2 | S_N(k)] = \sigma_m^2(S_N(k)), \ 0 \le m \le M, \ k \ge 0.$$

The conditional dispersion is determined by regression functions:

$$\sigma_m^2(p) = V_m(p)[1 - V_m(p)], \ 0 \le m \le M.$$

The asymptotical properties of stochastic models (18)–(21) as $N \to \infty$ and as $k \to \infty$ will be investigated in the authors' next paper.

REFERENCES

1. Ethier S.N., Kurtz T.G. Markov processes: Characterization and convergence. New York: Willey, 1986. 534 p.
2. Koroliouk D., Koroliuk V.S., and Rosato N. Equilibrium process in biomedical data analysis: The Wright–Fisher model. *Cybernetics and Systems Analysis.* 2014. Vol. 50, N 6. P. 890–897.
3. Koroliouk D. Two component binary statistical experiments with persistent linear regression. *Theor. Probability and Math. Statist.* 2014. Vol. 90. P. 103–114.

4. Korolyuk V.S., Koroliouk D. Diffusion approximation of stochastic Markov models with persistent regression. *Ukrainian Mathematical Journal.* 1995. Vol. 47, N 7. P. 1065–1073.

5. Skorokhod A.V., Hoppensteadt F.C., Salehi H. Random perturbation methods with applications in science and engineering. New York: Springer-Verlag, 2002. 488 p.

6. Koroliouk D. Multivariant statistical experiments with persistent linear regression and equilibrium. *Theor. Probability and Math. Statist.* 2015. Vol. 92. P. 71–78.

Надійшла до редакції 06.11.2017

Д.В. Королюк, В.С. Королюк

РІВНОВАГА В МОДЕЛЯХ ПОПУЛЯЦІЙНОЇ ГЕНЕТИКИ РАЙТА–ФІШЕРА

Анотація. Для мультиваріантних моделей Райта–Фішера в популяційній генетиці введено рівноважні стани, виражені флуктуаціями імовірнісних відношень, що відрізняються від традиційно використовуваних флуктуацій, виражених різницею між поточним значенням випадкового процесу та його рівноважним значенням. Показано, що тоді дрейфова компонента динамічного процесу генетичних частот, спочатку введена як відношення двох квадратичних форм, трансформується в кубічну параболу з певним коефіцієнтом нормалізації.

Ключові слова: модель Райта–Фішера, популяційна генетика, еволюційний процес, рівноважний стан, флуктуації ймовірнісних відношень.

Д.В. Королюк, В.С. Королюк

РАВНОВЕСИЕ В МОДЕЛЯХ ПОПУЛЯЦИОННОЙ ГЕНЕТИКИ РАЙТА–ФИШЕРА

Аннотация. Для мультивариантных моделей Райта–Фишера в популяционной генетике введены равновесные состояния, выраженные флуктуациями вероятностных отношений, в отличие от традиционно используемых флуктуаций, выражаемых разностью между текущей величиной случайного процесса и его равновесным значением. Показано, что тогда дрейфовая составляющая динамического процесса частот генов, первоначально введенная как отношение двух квадратичных форм, преобразуется в кубическую параболу с некоторым коэффициентом нормировки.

Ключевые слова: модель Райта–Фишера, популяционная генетика, эволюционный процесс, равновесное состояние, флуктуации вероятностных отношений.

Королюк Дмитро,
доктор фіз.-мат. наук, старший науковий співробітник Інституту телекомунікацій та глобального інформаційного простору НАН України, Київ, e-mail: dimitri.koroliouk@ukr.net.

Королюк Володимир Семенович,
доктор фіз.-мат. наук, академік НАН України, радник дирекції Інституту математики НАН України, Київ, e-mail: vskorol@yahoo.com.