Inégalités pour la sélection génétique

Stefan Neuwirth

Modélisons la sélection génétique d'un locus donné. Considérons une population suffisamment grande pour que les fluctuations aléatoires soient négligeables. Supposons que ce locus ait k allèles $A_1, \dots, A_k$: chaque individu est porteur d'un génotype A_iA_j. Pour modéliser l'évolution de la diversité génétique de ce locus, supposons qu'un mâle rencontre une femelle au hasard ; on sait que le génotype de l'œuf contiendra au hasard un des deux allèles du gamète mâle et un des deux allèles du gamète femelle. Donc l'évolution des génotypes se réduit à l'évolution de la probabilité $p_1, \dots, p_k$ des allèles parmi la population fertile ; on a

$$p_i \geqslant 0,\ p_1 + \cdots + p_k = 1.$$

L'accouplement de deux individus aléatoires donne le génotype A_iA_j avec probabilité p_ip_j ; notons w_{ij} la probabilité de survie jusqu'à maturité d'un individu avec ce génotype. La matrice de survie w est symétrique et à coefficients positifs :

$$w_{ij} = w_{ji},\ w_{ij} \geqslant 0.$$

Donc le nombre d'individus fertiles de génotype A_iA_j sera proportionnel à $w_{ij}p_ip_j$, de sorte que la probabilité de l'allèle A_i à la génération suivante est

$$p_i' = \frac{\sum_{j=1}^k w_{ij}p_ip_j}{W},$$

où

$$W = \sum_{i=1}^k \sum_{j=1}^k w_{ij}p_ip_j$$

est la survie moyenne. La survie moyenne de la génération suivante est

$$\begin{aligned} W' &= \sum_{i=1}^k \sum_{j=1}^k w_{ij}p_i'p_j' \\ &= W^{-2} \sum_{i=1}^k \sum_{j=1}^k w_{ij} \sum_{i'=1}^k w_{i'j}p_{i'}p_j \sum_{j'=1}^k w_{ij'}p_ip_{j'} \end{aligned}$$

et Mulholland et Smith [6], ainsi que Atkinson, Watterson et Moran [1] ont montré l'inégalité $W' \geqslant W$ d'abord conjecturée par Mandel et Hughes [5]. D'autres preuves ainsi que des généralisations se trouvent dans [3, 4].

Cette inégalité se déduit aussi par approximation du cas où

$$p_1 = \cdots = p_k = 1/k,$$

pour lequel elle se réduit à

$$\mathrm{somme}(ww^{\mathrm{t}}w) \geqslant \mathrm{somme}(w)^3/k^2$$

où somme(w) est la somme des coefficients de la matrice w.

Dans ce modèle, la survie moyenne augmente donc au fil des générations et on peut montrer que les probabilités $p_1, \dots p_k$ des allèles tendent vers un maximum local de W. La diversité génétique s'explique ici par l'existence d'un maximum pour lequel toutes les probabilités p_i sont strictement inférieures à 1.

Notre contribution consiste à mieux comprendre cette inégalité et à la généraliser. Voici trois développements.

1. Dans une note inédite, nous montrons que

$$\sum_{i=1}^k \sum_{j=1}^k w_{ij} p'_i q'_j \sum_{i'=1}^k w_{i'j} p''_{i'} \sum_{j'=1}^k w_{ij'} q''_{j'} \geqslant \frac{\left(\sum_{i=1}^k \sum_{j=1}^k w_{ij} (p_i p'_i p''_i)^{1/3} (q_j q'_j q''_j)^{1/3} \right)^3}{\sum_{i=1}^k p_i \sum_{j=1}^k q_j}$$

   pour tous coefficients $p_i, p'_i, p''_i, q_j, q'_j, q''_j$ positifs.

2. Dans [2], nous montrons que si w est une matrice $k \times l$ à coefficients positifs, alors

$$\text{somme}\Big(\overbrace{w w^{\mathrm{t}} w \dots}^{n \text{ termes}}\Big) \geqslant \text{somme}(w)^n / (k^{\lfloor (n-1)/2 \rfloor} l^{\lfloor n/2 \rfloor}).$$

   Notons que w n'est plus supposée symétrique.

3. Dans [7], nous montrons que si de plus ρ et γ sont deux réels positifs tels que

$$w_{i\star} = \sum_{j=1}^l w_{ij} \geqslant 2\rho \text{ et } w_{\star j} = \sum_{i=1}^k w_{ij} \geqslant 2\gamma,$$

   alors

$$\sum_{i=1}^k \sum_{j=1}^l w_{ij} (w_{i\star} - \rho)(w_{\star j} - \gamma) \geqslant w_{\star\star} (w_{\star\star}/k - \rho)(w_{\star\star}/l - \gamma),$$

   où $w_{\star\star} = \text{somme}(w)$.

Notre intérêt pour ces inégalités vient aussi de leur application à la théorie des graphes.

Références

[1] F. V. Atkinson, G. A. Watterson, and P. A. P. Moran. A matrix inequality. *Quart. J. Math. Oxford Ser. (2)*, 11 :137–140, 1960.

[2] William Banks, Asma Harcharras, Stefan Neuwirth, and Éric Ricard. Matrix inequalities with applications to the theory of iterated kernels. *Linear Algebra Appl.*, 362 :275–286, 2003.

[3] J. F. C. Kingman. On an inequality in partial averages. *Quart. J. Math. Oxford Ser. (2)*, 12 :78–80, 1961.

[4] J. F. C. Kingman. An inequality involving Radon-Nikodym derivatives. *Proc. Cambridge Philos. Soc.*, 63 :195–198, 1967.

[5] S. P. H. Mandel and I. M. Hughes. Change in mean viability at a multi-allelic locus in a population under random mating. *Nature*, 182 :63–64, 1958.

[6] H. P. Mulholland and C. A. B. Smith. An inequality arising in genetical theory. *Amer. Math. Monthly*, 66 :673–683, 1959.

[7] Stefan Neuwirth. The size of bipartite graphs with girth eight. `http://arxiv.org/math/0102210`, 2001.