Processus à valeurs mesures et évolution démo-génétique d'une population diploïde

Camille Coron

LMO, Université Paris-Sud

Journées MAS
Toulouse
28 Août 2014

Motivation

- Modèles de génétique de population: taille constante
- Modèles de démographie: un seul type.

Comprendre les interactions entre

- la taille d'une population
- sa composition génétique.

Modèle

- Individus diploïdes
- 1 gène, 2 allèles, A et a: génotypes AA, Aa et aa.
- Processus de naissance et mort à 3 types, compétition et reproduction mendélienne:

$$(Z_t, t \geq 0) = ((N_t^{AA}, N_t^{Aa}, N_t^{aa}), t \geq 0).$$

- Taille de population au temps t: $N_t = N_t^{AA} + N_t^{Aa} + N_t^{aa}$.

Taux de naissance et mort

Si $Z_t = z = (n^{AA}, n^{Aa}, n^{aa}) \in (\mathbb{Z}_+)^3$ avec $n = n^{AA} + n^{Aa} + n^{aa} > 0$,

- Reproduction mendélienne:

$$\lambda_1(z) = \frac{b_1}{n}\left(n^{AA} + \frac{n^{Aa}}{2}\right)^2 = b_1 n (p^A)^2$$

- Mort naturelle et par compétition:

$$\mu_1(z) = n^{AA}(d_1 + c_{11}n^{AA} + c_{12}n^{Aa} + c_{13}n^{aa})$$

Échelle considérée

- Grande taille de population.
- Population représentée par le processus de sauts purs: $(Z_t^K)_{t\geq 0} = (Z_t/K)_{t\geq 0} \in (\mathbb{Z}_+/K)^3$, $K \longrightarrow +\infty$.
- Ordre de grandeur des paramètres démographiques:

$$\begin{aligned} b_i^K &= \gamma K + \beta_i \\ d_i^K &= \gamma K + \delta_i \\ c_{ij}^K &= \frac{\alpha_{ij}}{K} \\ Z_0^K &\underset{K\to\infty}{\longrightarrow} Z_0 \quad \text{en loi,} \end{aligned}$$

$$\text{il existe } C \geq 0 \text{ tel que pour tout } K \in \mathbb{N}^*, \mathbb{E}\left((N_0^K)^3\right) \leq C,$$

où $\gamma > 0$ et Z_0 est une variable aléatoire à valeur dans $(\mathbb{R}_+)^3$.

Structure de Hardy-Weinberg

- Déviation depuis la structure de Hardy-Weinberg:

$$\begin{aligned} Y_t^K = \frac{4N_t^{AA,K}N_t^{aa,K} - (N_t^{Aa,K})^2}{4N_t^K} &= N_t^K(p_t^{AA,K} - (p_t^{A,K})^2) \\ &= N_t^K(2p_t^{A,K}p_t^{a,K} - p_t^{Aa,K}) \\ &= N_t^K(p_t^{aa,K} - (p_t^{a,K})^2) \end{aligned}$$

- $2N_t^{AA,K} + N_t^{Aa,K} = N_t^{A,K} =$ nombre d'allèles A divisé par K,
- $2N_t^{3,K} + N_t^{2,K} = N_t^{a,K} =$ nombre d'allèles a divisé par K.

$$(N_t^{AA,K}, N_t^{Aa,K}, N_t^{aa,K}) \longleftrightarrow (N_t^{A,K}, N_t^{a,K}, Y_t^K)$$

Dynamique lente-rapide

Proposition

Pour tous temps $s, t > 0$, $\sup_{t \leq u \leq t+s} \mathbb{E}((Y_u^K)^2) \longrightarrow 0$ *quand* $K \to \infty$.

Théorème

La suite de processus stochastiques $\{((N_t^{A,K}, N_t^{a,K}), t \in [0,T])\}_{K \geq 0}$ *converge en loi dans* $\mathbb{D}([0,T], (\mathbb{R}_+)^2)$ *vers un processus de diffusion* $(N_t^A, N_t^a)_{0 \leq t \leq T}$.

Dynamique lente

Proposition

Dans le cas neutre où $\beta_i = \beta$, $\delta_i = \delta$, $\alpha_{ij} = \alpha$ *pour tous* $i, j \in \{1, 2, 3\}$*:*

$$dN_t^A = \sqrt{\frac{4\gamma}{N_t^A + N_t^a}} N_t^A dB_t^1 + \sqrt{2\gamma \frac{N_t^A N_t^a}{N_t^A + N_t^a}} dB_t^2 + \left(\beta - \delta - \alpha \frac{N_t^A + N_t^a}{2}\right) N_t^A dt$$

$$dN_t^a = \sqrt{\frac{4\gamma}{N_t^A + N_t^a}} N_t^a dB_t^1 - \sqrt{2\gamma \frac{N_t^A N_t^a}{N_t^A + N_t^a}} dB_t^2 + \left(\beta - \delta - \alpha \frac{N_t^A + N_t^a}{2}\right) N_t^a dt$$

Comportement démo-génétique

$$dN_t = (\beta - \delta - \alpha N_t)N_t dt + \sqrt{2\gamma N_t} dB_t^1$$
$$dX_t = \sqrt{\frac{\gamma X_t(1 - X_t)}{N_t}} dB_t^2.$$

Proposition (Changement de temps)

Soit $(\tau_t, t \geq 0)$ *tel que* $t = \int_0^{\tau_t} \frac{1}{N_s} ds$ *et* $\tilde{X}_t = X_{\tau_t}$ *pour tout* $t \geq 0$. *Alors*

$$d\tilde{X}_t = \gamma \tilde{X}_t(1 - \tilde{X}_t) dB_t$$

Comparaison avec le cas haploïde (1)

Population diploïde:

$$dN_t = (\beta - \delta - \alpha N_t)N_t dt + \sqrt{2\gamma N_t} dB_t^1$$
$$dX_t = \sqrt{\frac{\gamma X_t(1 - X_t)}{N_t}} dB_t^2.$$

Diffusion de Lotka-Volterra haploïde (P. Cattiaux & S. Méléard (2010)):

$$dN_t^h = (\beta - \delta - \alpha N_t^h)N_t^h dt + \sqrt{2\gamma N_t^h} dW_t^1$$
$$dX_t^h = \sqrt{\frac{2\gamma X_t^h(1 - X_t^h)}{N_t^h}} dW_t^2.$$

Comparaison avec le cas haploïde (2)

Population diploïde:

$$dN_t^A = \sqrt{\frac{4\gamma}{N_t^A + N_t^a}} N_t^A dB_t^1 + \sqrt{2\gamma \frac{N_t^A N_t^a}{N_t^A + N_t^a}} dB_t^2 + ...$$

$$dN_t^a = \sqrt{\frac{4\gamma}{N_t^A + N_t^a}} N_t^a dB_t^1 - \sqrt{2\gamma \frac{N_t^A N_t^a}{N_t^A + N_t^a}} dB_t^2 + ...$$

Diffusion de Lotka-Volterra haploïde (P. Cattiaux & S. Méléard (2010)):

$$dN_t^{A,h} = \sqrt{2\gamma N_t^{A,h}} dB_t^1 + (\beta - \delta - \alpha(N_t^{A,h} + N_t^{a,h}))N_t^{A,h} dt$$

$$dN_t^{a,h} = \sqrt{2\gamma N_t^{a,h}} dB_t^2 + (\beta - \delta - \alpha(N_t^{A,h} + N_t^{a,h}))N_t^{a,h} dt$$

Comportement en temps long de (N, X)

- $T_0^N < \infty$ p.s. (Cattiaux et al. 2009)

Théorème

Il existe une distribution ν *sur* $\mathbb{R}_+^* \times [0,1]$ *telle que pour tout* $(n,x) \in \mathbb{R}_+^* \times]0,1[$,

$$\lim_{t\to\infty} \mathbb{P}_{n,x}((N_t, X_t) \in E | N_t > 0) = \nu(E).$$

Surdominance

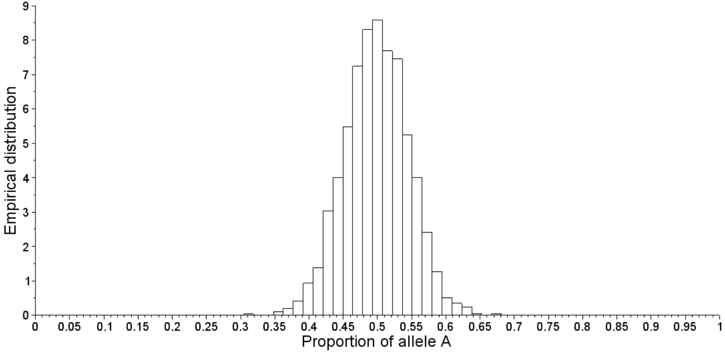

Figure: Distribution de la proportion X_t d'allèles A dans un cas de surdominance, sachant que $N_t \neq 0$. Ici $\beta_i = 1$ pour tout $i \neq 2$, $\beta_2 = 5$, $\delta_i = 0$ pour tout i, $\alpha_{ij} = 0.1$ pour tout (i, j), et $t = 100$.

Continuum d'allèles

- L alleles, $i \in [\![1, L]\!]$, $L(L+1)/2$ génotypes différents.
- Dynamique lente-rapide: diffusion $(N_t, X_t^{1,L}, X_t^{2,L}, ...X_t^{L-1,L})_{t\geq 0}$.
- Représentation par le processus

$$\zeta_t^L = \left(N_t^L, \sum_{i=1}^{L} X^{i,L}(t)\delta_{\frac{i}{L}}\right).$$

- Convergence lorsque $L \to \infty$:

$$(\zeta_t^L, t \in [0,T]) \underset{L\to\infty}{\longrightarrow} (N_t, \mathbf{p}_t)_{t\in[0,T]}$$

Problème de martingale

$$H(n,\mathbf{p}) = f(n,\langle \mathbf{p},G\rangle), \qquad \langle \mathbf{p},G\rangle = \int_0^1 G(x)d\mathbf{p}(x).$$

$$m_t^H = H(N_t,\mathbf{p}_t) - H(N_0,\mathbf{p}_0) - \int_0^{t\wedge T_\epsilon} \mathcal{L}H(N_s,\mathbf{p}_s)ds$$

est une martingale continue, avec dans le cas neutre

$$\mathcal{L}H(n,\mathbf{p}) = n(\beta-\delta-\alpha n)\,\partial_1 f(n,\langle \mathbf{p},G\rangle) + \gamma n\,\partial_{11}^2 f(n,\langle \mathbf{p},G\rangle) \\ + \frac{\gamma}{n}\left(\langle \mathbf{p},G^2\rangle - \langle \mathbf{p},G\rangle^2\right)\,\partial_{22}^2 f(n,\langle \mathbf{p},G\rangle).$$

Références

- Cattiaux, P., Collet, P., Lambert, A., Martinez, S., Méléard, S,. et San Martín, J. (2009): Quasi-stationarity distributions and diffusion models in population dynamics. Ann. Probab. 37(5):1926–1969.
- Cattiaux, P., et Méléard, S. (2009): Competitive or weak competitive stochastic Lotka-Volterra systems conditioned on non-extinction. JMB 60:797–829.
- C.C. (2013): Slow-fast stochastic diffusion dynamics and quasi-stationary distributions for diploid populations. Arxiv: 1309.3405.
- C.C. et S. Méléard (2014): Infinitely-many allele model of diploid populations with stochastically varying size. Pré-pré-publication.
- Ethier, S. N. et Nagylaki, T. (1988): Diffusion Approximations of Markov Chains with Two Time Scales and Applications to Population Genetics, II. Adv. Appl. Probab. 20(3):525–545.