MRCA and bottleneck in an elementary size-varying population model

JEAN-FRANÇOIS DELMAS

http://cermics.enpc.fr/~delmas

CERMICS, Univ. Paris-Est

Polytechnique, January 2010

Outline

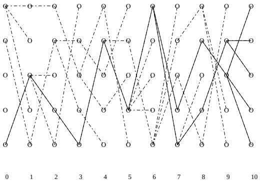
Work in progress in collaboration with Y.-T. Chen.

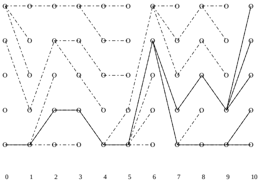
Model for constant size population

- Finite population: Moran process (1958) or Wright-Fisher's model (1930-1931).
- Infinite population: Fleming-Viot (1979) process.
- Coalescent (genealogical) tree (Moran process and its infinite population limit): Kingman (1982).
- Coalescent tree with multiple collisions: Pitman (1999) and Sagitov (1999) (finite population: Cannings' model (1974); infinite population: Fleming-Viot historical process from Dawson-Perkins (1991), Bertoin-Le Gall (2003) stochastic flows).
- Representation for the genealogy of the Fleming-Viot process using the look-down process from Donnelly-Kurtz (1999).

Wright-Fisher model

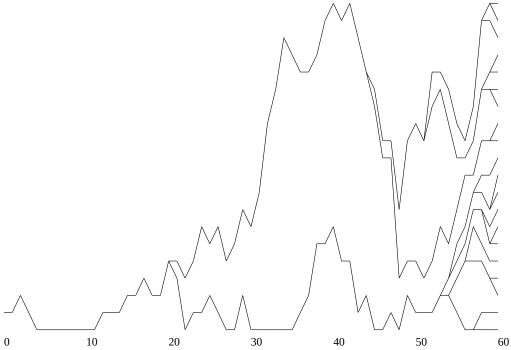
Genealogy with $N = 5$ individuals over 10 generations.





Wright-Fisher model

Genealogy with $N = 20$ individuals over 60 generations.



Model for random size population

- Finite population: Galton-Watson process (1873).
- Infinite population.
  - Population size is a Continuous State Branching process: Jirina (1958).
  - Population genealogy given by Dawson (1975)- Watanabe (1968) historical super-process.
  - Quadratic case (2 individuals merge together)
    - Conditionally on having a constant population size, the Dawson-Watanabe super-process is a Fleming-Viot process: Etheridge-March (1991).
    - Use a time change (with speed proportional to the inverse of the population size) to get Fleming-Viot process from a Dawson-Watanabe super-process: Perkins (1991).
  - In the α-stable case, see Birkner-Blath-Capaldo-Etheridge-Möhle-Schweinsberg-Wakolbinger (2005), a time change of the super-process genealogical tree gives a coalescent tree.
- See also Kaj-Krone (2003) or Jagers-Sagitov (2004) for other random size population models (finite to infinite population), where one recovers the Kingman coalescent through a random time change.

Model for stationary random size population

Let Z_t be the size of the population at time t.

- Neutral population $\Longrightarrow$ **BRANCHING PROCESS**. But:
  - Galton-Watson process or CSBP (or Dawson-Watanabe process) exhibit either **extinction** ($\lim_{t\to+\infty} Z_t = 0$) or exponential explosion ($\lim_{t\to+\infty} Z_t/\rho^t$ exists and is non trivial for some $\rho > 1$).
  - **Critical** (Z_t constant in mean) or **sub-critical** (Z_t decreasing in mean) case provides a.s. extinction. **Super-critical** (Z_t increasing in mean) case provides explosion with positive probability.
  - The starting time of the process plays an important role but its distribution for a current population is not clear.
- Instead, one can use:
  - The Yaglom (1947) distribution, that is the limit distribution of Z_t in the sub-critical or critical case conditionally on $\{Z_t > 0\}$ as $t \to +\infty$. See also the theory of **quasi-stationary distributions**.
  - The Q**-process**, that is the limit distribution of Z_t in the sub-critical or critical case conditionally on non-extinction (that is in most case conditionally on $\{Z_{t+T} > 0\}$ as $T \to +\infty$). See Roelly-Rouault (1989), see also Lambert (2007).

Mathematical model for **branching** process

- Branching mechanism:

$$\psi(\lambda) = \alpha\lambda + \beta\lambda^2 + \int_{(0,+\infty)} \pi(d\ell)\ (\mathrm{e}^{-\lambda\ell} - 1 + \lambda\ell),$$

  $\alpha \in \mathbb{R}$, $\beta \geq 0$ and $\int_{(0,+\infty)} \pi(d\ell)\ \ell \wedge \ell^2 < +\infty$.
- We assume $\beta > 0$ or $\pi \neq 0$.
- ψ is convex, C^∞ on $(0, +\infty)$, $\psi'(0) \in [-\infty, +\infty)$, $\lim_{\lambda\to+\infty} \psi(\lambda)/\lambda = +\infty$.
- ψ-CSBP $(Z_t, t \geq 0)$ is a Markov process s.t. $Z_0 = x$ under $\mathbb{P}_x$ and

$$\mathbb{E}_x[\mathrm{e}^{-\lambda Z_t}] = \mathrm{e}^{-xu(\lambda,t)} \quad \text{and} \quad \int_{u(\lambda,t)}^{\lambda} \frac{dv}{\psi(v)} = t.$$

- Excursion measure: $\mathbb{N}[1 - \mathrm{e}^{-\lambda Z_t}] = \lim_{x\to 0} \frac{1}{x}\, \mathbb{E}_x[1 - \mathrm{e}^{-\lambda Z_t}] = u(\lambda, t)$.
- $\mathbb{E}_x[Z_t] = x\,\mathrm{e}^{-\psi'(0)t}$. So the CSBP is sub-critical if $\psi'(0) > 0$, critical if $\psi'(0) = 0$ or super-critical if $\psi'(0) < 0$.

Williams' decomposition and Q-process

We consider the sub-critical case with strong extinction: $\tau = \inf\{t; Z_t = 0\}$ finite a.s. or equivalently $\int^{+\infty} \frac{dv}{\psi(v)} < +\infty$. We set

$$c(t) = \mathbb{N}[\tau > t] \quad \text{that is} \quad \mathbb{P}_x(\tau \leq t) = \mathrm{e}^{-xc(t)} .$$

- Williams' decomposition (Abraham-D (2009)): cond. on $\tau = h$,
  - one individual is alive up to time h,
  - at rates $2\beta dt$ it gives birth to a population starting with an infinitesimal mass and distributed as $\mathbb{N}[dZ; \tau \leq h - t]$,
  - at rates dt it gives birth to a population distributed as $\mathbb{E}_\ell[dZ; \tau \leq h - t]$, where the initial mass ℓ is distributed as $\ell\, \mathrm{e}^{-\ell c(h-t)} \;\pi(d\ell)$.
- Q-process: fix t and let $h \to +\infty$ to get an immortal individual which
  - at rates $2\beta dt$ gives birth to a population starting with an infinitesimal mass and distributed as $\mathbb{N}[dZ]$,
  - at rates dt gives birth to a population distributed as $\mathbb{E}_\ell[dZ]$, where the initial mass ℓ is distributed as $\ell\, \pi(d\ell)$.

Q-process and stationary distribution

- The Q process has a stationary distribution. We consider the Q-process $(\bar{Z}_t, t \in \mathbb{R})$ under its stationary measure.
- Stationary distribution $\Longrightarrow$ remove t in all formula: $\bar{Z}$ for $\bar{Z}_t$.

$$\mathbb{E}\left[\mathrm{e}^{-\lambda \bar{Z}}\right] = \mathrm{e}^{-\int_0^{+\infty} \tilde{\psi}'(u(\lambda,r))\, dr},$$

  with $\tilde{\psi}(\lambda) = \psi(\lambda) - \alpha\lambda$.
- Interpretation: an immortal individual gives birth to sub-critical population (can also be seen as an immigration).
- Notice: genealogy of Fleming-Viot process given by look-down process has also an immortal individual which gives birth to sub-population with extinction.

Quadratic case (2 individuals merge together) I

Let $(E_k, k \in \mathbb{N}^*)$ be i.i.d. exponential r.v. with mean 1.
We have

- $\psi(\lambda) = \beta\lambda^2 + 2\beta\theta\lambda, \quad \theta > 0.$
- Extinction probability for the sub-critical CSBP:

$$c(t) = \mathbb{N}[\tau > t] = \frac{2\theta}{e^{2\theta\beta t} - 1}.$$

- Feller diffusion:

$$d\bar{Z}_t = \sqrt{2\beta\bar{Z}_t}\, dW_t + 2\beta(1 - \theta\bar{Z}_t)\, dt.$$

- Stationary distribution: $\bar{Z} \overset{(d)}{=} \frac{1}{2\theta}(E_1 + E_2)$
- $\mathbb{E}[\bar{Z}] = 1/\theta.$

Time to the MRCA, population size at the MRCA

- A=time to the MRCA of the population (at fixed time t).
- $\bar{Z}^A$ size of the population at the MRCA time (that is $\bar{Z}^A = \bar{Z}_{t-A}$).
- Explicit formula for the distribution of $(\bar{Z}, A, \bar{Z}^A)$.
- Conditionally on A, $\bar{Z}$ and $\bar{Z}^A$ are independent.
- $\mathbb{P}(A \in [0,t]) = \mathbb{E}[\mathrm{e}^{-c(t)\bar{Z}}]$ and density $f_A(t) = \tilde{\psi}'(c(t))\mathbb{E}[\mathrm{e}^{-c(t)\bar{Z}}]$.
- Distribution of $\bar{Z}^A$.

$$\mathbb{E}\left[\mathrm{e}^{-\mu\bar{Z}^A}|A=t\right] = \frac{\mathbb{E}\left[\mathrm{e}^{-(\mu+c(t))\bar{Z}}\right]}{\mathbb{E}\left[\mathrm{e}^{-c(t)\bar{Z}}\right]}.$$

- Bottleneck effect: $\mathbb{P}(\bar{Z}^A \leq z|A=t) \geq \mathbb{P}(Z \leq z)$.

$$\boxed{\bar{Z}^A \text{ is stoch. less than } \bar{Z}.}$$

Quadratic case II: Bottleneck effect

Let $(E_k, k \in \mathbb{N}^*)$ be i.i.d. exponential r.v. with mean 1. We have:

- Cond. on $\{A = t\}$,

$$\bar{Z} \overset{(d)}{=} \frac{1}{2\theta + c(t)}(E_1 + E_2 + E_3) \quad \text{and} \quad \bar{Z}^A \overset{(d)}{=} \frac{1}{2\theta + c(t)}(E_4 + E_5).$$

- $\mathbb{E}[\bar{Z}^A | A = t] = \frac{2}{3}\mathbb{E}[\bar{Z} | A = t]$.
- $\mathbb{P}(\bar{Z}^A < \bar{Z} | A = t) = \mathbb{P}(E_4 + E_5 < E_1 + E_2 + E_3) = \frac{11}{16}$.

$$\boxed{\mathbb{E}\left[\bar{Z}^A\right] = \frac{2}{3}\mathbb{E}\left[\bar{Z}\right]} \quad \text{and} \quad \boxed{\mathbb{P}(\bar{Z}^A < \bar{Z}) = \frac{11}{16}.}$$

Quadratic case III: Time to the MRCA

See also Lambert (2003) under the quasi-stationary distribution. See Evans-Ralph (2010) for the MRCA process.
We have:

- $\mathbb{P}(A \in [0,t] | \bar{Z}) = \mathrm{e}^{-c(t)\bar{Z}}$. Cond. on $\bar{Z}$, $A \overset{(d)}{=} \dfrac{1}{2\beta\theta} \log\left(1 + \dfrac{2\theta\bar{Z}}{E_1}\right)$, with E_1 indep. of $\bar{Z}$.
- $A \overset{(d)}{=} \dfrac{1}{2\beta\theta} \log\left(1 + \dfrac{E_2 + E_3}{E_1}\right)$.
- Let A_n the time to the MRCA of n individuals taken at random and the immortal individual:

$$\mathbb{E}\left[\bar{Z}^n \, \mathrm{e}^{-\lambda\bar{Z}} \, \mathbf{1}_{\{A^n \in [0,t]\}}\right] = \frac{(n+1)! s^n}{(2\theta + \lambda s)^n} \left(\frac{2\theta}{2\theta + \lambda}\right),$$

with $s = 1 - \mathrm{e}^{-2\beta\theta t}$ and

$$\mathbb{E}\left[\frac{\bar{Z}^n}{\mathbb{E}\left[\bar{Z}^n\right]} \mathbf{1}_{\{A^n \in [0,t]\}}\right] = s^n = \left(1 - \mathrm{e}^{-2\beta\theta t}\right)^n.$$

Last coalescent event

Let $N^A + 1$ the number of individuals involved in the last coalescent event (that is the number of old families).

- In the quadratic case a.s. $N^A = 1$.
- In the general case explicit distribution of $(\bar{Z}, A, N^A)$.
- $\mathbb{E}\left[a^{N^A} | A = t\right] = 1 - \dfrac{\tilde{\psi}'((1-a)c(t))}{\tilde{\psi}'(c(t))}$.
- In the stable case $\tilde{\psi}(\lambda) = c\lambda^{1+\alpha_0}$, $\alpha_0 \in (0,1)$:
  $\mathbb{E}\left[a^{N^A} | A = t\right] = 1 - (1-a)^{\alpha_0}$.
- If $\psi''(0) < +\infty$, then $\mathbb{E}[N^A | A = t] = \psi''(0) \dfrac{c(t)}{\tilde{\psi}'(c(t))}$ is decreasing in t.
- Work in progress: distribution of the reduced tree. See Duquesne-Le Gall (2002) for reduced trees in CSBP genealogy.

Speed of coming down from infinity

See Berestycki-Berestycki-Limic (preprint) for coalescent process.
Let $N^s + 1$ the number of ancestors living at time s in the past from the current population (for s large enough, $N^s = 0$ and $\lim_{s\to 0} N^s = +\infty$).

- The following convergence holds in probability:

$$\lim_{s\to 0} \frac{N^s}{c(s)} = \bar{Z}.$$

- In the quadratic case, we have the following fluctuations

$$\sqrt{c(s)\mathbb{E}[\bar{Z}]}\,\left(\frac{N^s}{c(s)} - \bar{Z}\right) \xrightarrow[s\to 0]{(d)} \bar{Z} - \bar{Z}',$$

  where $\bar{Z}'$ is independent of $\bar{Z}$ and distributed as $\bar{Z}$.

- Work in progress: fluctuations for the α-stable case.