# A model of branching process with immigration and non-neutral mutations

JEAN-FRANÇOIS DELMAS

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

CERMICS, Univ. Paris-Est

Stochastic Models in Ecology, Evolution and Genetics
December 2013

# Outline

1. The neutral model: stationary GW or CSBP
2. Mild bottleneck effect at the MRCA (neutral case)
3. Non neutral immigration
4. Speed of coming down from infinity and number of families

In collaboration with Y.-T. Chen (Ann. Prob. 2012) and with H. Bi (ArXiv 2013).

# Models for population (asexual and with no competition)

- Constant size population:
  - Finite population: Moran process (1958) or Wright-Fisher's model (1930-1931).
  - Infinite population: Fleming-Viot (1979) process.
  - Coalescent (genealogical) tree (for the infinite population): Kingman (1982), Pitman (1999) and Sagitov (1999).
- Random size population:
  - Finite population: Galton-Watson process (1873).
  - Infinite population.
    - Population size is a Cont. State Branching Process (CSBP): Jirina (1958); see also Dawson (1975)- Watanabe (1968) process.
    - Population genealogy given by Lévy trees: Duquesne-Le Gall (2005).
  - Links with constant size population (infinite population, stable branching).
- Our aim: study a model for stationary random size (infinite) population (continuous time) with non-neutral mutations.

# Galton-Watson (GW) process

$Y'_t$ be the size of the (finite) population at time/generation $t \in \mathbb{N}$.

- Offspring distribution = distribution of $\xi$.
- Each individual has an independent random number of children with distribution $\xi$.
- $\boxed{Y'_{t+1} = \sum_{i=1}^{Y'_t} \xi_{t+1,i}}$, with $(\xi_{i,t})$ indep. distributed as $\xi$.
- Asymptotic behavior:
  - Sub-critical: $\mathbb{E}[\xi] < 1$. Then a.s. $Y'_t = 0$ for $t$ large.
  - Critical: $\mathbb{E}[\xi] = 1$. Then (if $\mathbb{P}(\xi = 1) < 1$) a.s. $Y'_t = 0$ for $t$ large.
  - Super-critical: $\mathbb{E}[\xi] > 1$. Then a.s. $\lim_{t\to+\infty} Y'_t \in \{0, +\infty\}$.
- No stationary regime.

# Stationary random size population using GW process

Consider the sub-critical GW process conditioned on non-extinction.

- ***Q*-process**: the limit distribution of $Y' = (Y'_t, t \in \mathbb{N})$ in the sub-critical case conditionally on $\{Y'_{t+s} > 0\}$ as $s \to +\infty$.
- The $Q$-process can be extended into a stationary process $Z' = (Z'_t, t \in \mathbb{Z})$.
- Description of $Z'$:
  - $Z'$ corresponds to an **immortal individual** with size-biased offspring distribution, and other individuals have the offspring distribution given by $\xi$.
  - $Z' - 1$ corresponds to GW process $Y'$ with **immigration** given by the size-biased offspring distribution.

# Stationary random size population using GW process II

**Questions:**

- Distribution of $A'$, the TMRCA for the pop. living at time $t = 0$.
- Joint distribution of $Z'_0$ and $Z'_{-A'}$ (mild bottleneck effect).

**Generalizations:**

- Change the immigration distribution.
- Use multitype GW process to get (a finite set of) non-neutral mutations.
- **Change the offspring distribution** of each of the immigrant population to take into acount (infinitely many) **non-neutral mutations**.

# CSBP as limit of GW processes

We shall focus only on the quadratic case. Let $\theta > 0$. Assume

$$\mathrm{Var}(\xi) = 2 < +\infty.$$

We have in the sub-critical case ($\mathbb{E}[\xi] = 1 - 2\theta/n$ and $Y'_0 = [nx]$):

$$\left(\frac{1}{n} Y'_{[nt]}, t \geq 0\right) \xrightarrow[n\to\infty]{\text{(d)}} Y^\theta = (Y^\theta_t, t \geq 0).$$

- The process $Y^\theta$ is a sub-critical continuous branching process (CSBP) or a Feller diffusion:

  $$dY^\theta_t = \sqrt{2Y^\theta_t}\, dW_t - 2\theta Y^\theta_t\, dt.$$

  Branching mechanism: $\psi(\lambda) = \lambda^2 + 2\theta\lambda$.
- We have: $\mathbb{E}\left[Y^\theta_t\right] = \mathrm{e}^{-2\theta t}$.
- The lifetime: $\zeta^\theta = \inf\{t > 0; Y^\theta_t = 0\}$ is a.s. finite.

# Stationary CSBP

$$dY_t^\theta = \sqrt{2Y_t^\theta}\, dW_t - 2\theta Y_t^\theta\, dt.$$

Consider the sub-critical CSBP conditioned on non-extinction.

- ***Q*-process**: the limit distribution of $Y^\theta = (Y_t^\theta, t \geq 0)$ conditionally on $\{Y_{t+s}^\theta > 0\}$ as $s \to +\infty$.
- The $Q$-process can be extended into a stationary process $Z^\theta = (Z_t^\theta, t \in \mathbb{R})$:

$$dZ_t^\theta = \sqrt{2Z_t^\theta}\, dW_t + 2(1 - \theta Z_t^\theta)\, dt.$$

- Interpretation of $Z^\theta$:
  - $Z^\theta$ corresponds to an **immortal individual** with infinite birth rate.
  - $Z^\theta$ corresponds to CSBP $Y^\theta$ with (infinite rate) **immigration**.

# Excursion measure and immigration

- Excursion measure:

$$\mathbb{N}\left[Y^\theta \in \cdot\right] = \lim_{x \to 0} \frac{1}{x} \mathbb{E}\left[Y^\theta \in \cdot | Y_0^\theta = x\right].$$

- Excursion duration: $\mathbb{N}[\zeta^\theta > t] = 2\theta/(\mathrm{e}^{2\theta t} - 1)$.
- Immigration representation (with the convention: $Y_t^\theta = 0$ for $t < 0$): :

$$\boxed{Z_t = \sum_{i \in I} Y_{t-t_i}^{(i)}} \quad \text{for all } t \in \mathbb{R},$$

with $t_i$ the immigration time of $Y^{(i)}$ and $\sum_{i \in I} \delta_{Y^{(i)}, t_i}(dY, dt)$ a PPM with intensity:

$$\boxed{2\mathbb{N}\left[dY^\theta\right]\, dt}.$$

# Time to the MRCA, population size at the MRCA

Results from Chen-D. (2012).

- $A$=time to the MRCA of the population (at fixed time $t$).
- Let $Z^{(A)}$ be size of the population at the MRCA time:

$$Z^{(A)} = Z^{\theta}_{(t-A)}.$$

- Explicit formula (for general CSBP) for the distribution of

$$(Z_t^{\theta}, A, Z^{(A)}).$$

- Conditionally on $A$, $Z_t^{\theta}$ and $Z^{(A)}$ are independent.
- Mild bottleneck effect:

$$\boxed{Z^{(A)} \text{ is stoch. less than } Z_t^{\theta}.}$$

And we have:

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

# Non-neutral immigration (Bi-D., 2013)

- Coupling for $q \geq \theta$: $\mathbb{N}$-a.e. $\boxed{Y_t^q \leq Y_t^\theta}$ for all $t \geq 0$.
- The parameter $\theta$ can be seen as a fitness parameter.

**Mutation measure** $\mu(d\theta)$ on $(0, +\infty)$.

- Immigration:

$$Z_t = \sum_{i \in I} Y^{(i)}_{t-t_i} \quad \text{for all } t \in \mathbb{R},$$

  with $t_i$ the immigration time of $Y^{(i)}$ and $\sum_{i\in I} \delta_{\theta_i, Y^{(i)}, t_i}(d\theta, dY, dt)$ a PPM with intensity:

$$\boxed{2\mu(d\theta)\mathbb{N}\left[dY^\theta\right]\, dt.}$$

  Notice $\theta_i$ is the fitness parameter of $Y^{(i)}$.

- Neutral case corresponds to $\mu = \delta_\theta$.

# Non-neutral immigration: existence and properties of $Z$

- The process $Z = (Z_t, t \in \mathbb{R})$ is well defined iff:

$$\int_{0+} |\log(\theta)| \; \mu(d\theta) < +\infty \quad \text{and} \quad \int^{+\infty} \frac{\mu(d\theta)}{\theta} < +\infty. \qquad (1)$$

  We assume (1) holds.
- The process $Z$ is non-Markov (to get the Markov property you need to keep track of the size of all the current families with different fitness).
- The process $Z$ is continuous.
- A.s. for all $t \in \mathbb{R}$ we have $Z_t > 0$ if $\langle \mu, 1 \rangle > 1/2$.

# Non-neutral immigration: the MRCA

- $A$=time to the oldest immigrant ($\simeq MRCA$) of the population (at fixed time $t$).
- $Z^A$ size of the population at the MRCA time:

$$Z^{(A)} = Z_{(t-A)}.$$

- Mutation type of the MRCA: $\Theta$.
- Explicit formula for the distribution of

$$(Z_t, A, \Theta, Z^{(A)}).$$

- Result: Conditionally on $A$, $(\Theta, Z_t)$ and $Z^{(A)}$ are independent.
- Result: Bottleneck effect:

$$\boxed{Z^{(A)} \text{ is stoch. less than } Z_t.}$$

# Stable mutation measure

$$\mu(d\theta) = c\theta^{\alpha-1}\mathbf{1}_{\{\theta>0\}}\, d\theta \quad \text{for some } \alpha \in (0,1).$$

- (1) holds; $\langle \mu, 1\rangle = +\infty$ and $\mathbb{E}[Z_t] = +\infty$.
- **Strong bottleneck effect**: for $\alpha \in (1/2, 1)$: $\mathbb{E}\left[Z^{(A)}\right] < +\infty$.
- Let $\Theta_*$ be the mutation type of an individual chose at random in population at time $t$. (Notice $\Theta_*$ has a size biased distribution.)
- **$\Theta$ is stoch. less than $\Theta_*$:**

$$Y^{\Theta} \text{ is stoch. larger than } Y^{\Theta_*},$$

that is the MRCA has **greater fitness** than a random individual (see also Fearnhead (JAP 2002) for similar results).

# Speed of coming down from infinity

See Berestycki-Berestycki-Limic (2010) for coalescent process.
Let $M_s$ be the number of ancestors living at time $s$ in the past from the current population ($M_s = 0$ for $s > A$, and $\lim_{s\to 0} M_s = +\infty$).

- The following convergence holds a.s.:

$$\lim_{s\to 0} sM_s = Z_0.$$

- Fluctuations under some regularity assumption on $\mu$ (for $\alpha \in (0, 1/2)$ in the stable case or for neutral immigration and general CSBP):

$$s^{-1/2} \ (sM_s - Z_{-s}) \ \xrightarrow[s\to 0]{(d)} \sqrt{Z_0}\, G,$$

with $G \sim \mathcal{N}(0,1)$ independent of $Z_0$.
- Fluctuations in the stable case with $\alpha \in (1/2, 1)$:

$$s^{-\alpha-1} \ (sM_s - Z_{-s}) \ \xrightarrow[s\to 0]{(d)} c_\alpha.$$

# Number of families

Let $N_s$ be the number of families at time $s$ in the past which have descendants in the current population:

$$N_s = \sum_{i \in I} \mathbf{1}_{\{t_i < -s,\, Y^i_{-t_i} > 0\}}.$$

- In the neutral case (stable branching mechanism):

$$\lim_{s \to 0} N_s / \log(1/s) = c > 0.$$

- In the stable (non-neutral) case:

$$\lim_{s \to 0} s^{\alpha} N_s = c > 0.$$

# Open questions

- Law and properties of the genealogical tree.
- Does this model fit (better) to some data?
- Does usual algorithms (built from the Kingman model) detect a bottleneck effect in the quadratic stationary CSBP model?