Competition and condensation in some population models

Linglong Yuan

University of Liverpool

Summer School 2022: Stochastic population models

2022.08.24, AMSS

Outline

- Kingman's model
- Analogy to the Bose-Einstein condensation
- Mapping Bose-Einstein condensation with preferential attachment model with fitness
- A unifying approach: branching process with reinforcement
- Random permutation model

- Kingman's model

Population characteristics

Consider a population that has

- infinite size
- discrete generations
- haploidy (one gender)
- selection and mutation

What would a suitable population model look like?

Main ideas

Fitness and fitness distribution

- an individual is represented by its fitness value[1] $x \in [0, 1]$
- the population at the n-th generation is represented by the fitness distribution P_n on $[0, 1]$

[1] can be considered as the reproduction ability

Main ideas

Fitness and fitness distribution

- an individual is represented by its fitness value[1] $x \in [0,1]$
- the population at the n-th generation is represented by the fitness distribution P_n on $[0,1]$

Mutation

- an individual is born as mutant with probability $b \in (0,1)$
- the fitness value of a mutant is drawn independently from the same mutant distribution Q on $[0,1]$

[1] can be considered as the reproduction ability

Main ideas

Fitness and fitness distribution

- an individual is represented by its fitness value[1] $x \in [0,1]$
- the population at the n-th generation is represented by the fitness distribution P_n on $[0,1]$

Mutation

- an individual is born as mutant with probability $b \in (0,1)$
- the fitness value of a mutant is drawn independently from the same mutant distribution Q on $[0,1]$

Selection

- individuals with larger fitness values will produce more offspring in the next generation

[1] can be considered as the reproduction ability

Kingman's model (1978)

The model has three parameters (P_0, Q, b) and is defined as:

$$P_{n+1}(dx) = (1-b)\underbrace{\frac{xP_n(dx)}{\int_0^1 yP_n(dy)}}_{\text{selection}} + b\underbrace{Q(dx)}_{\text{mutation}}, \quad n \geq 0,$$

where

- Q is the mutant distribution
- P_n is the fitness distribution at the n-th generation for $n \geq 0$
- $b \in (0,1)$ is the deterministic mutation probability

Kingman's model (1978)

The model has three parameters (P_0, Q, b) and is defined as:

$$P_{n+1}(dx) = (1-b)\underbrace{\frac{xP_n(dx)}{\int_0^1 yP_n(dy)}}_{\text{selection}} + b\underbrace{Q(dx)}_{\text{mutation}}, \quad n \geq 0,$$

where

- Q is the mutant distribution
- P_n is the fitness distribution at the n-th generation for $n \geq 0$
- $b \in (0,1)$ is the deterministic mutation probability

Question

- Will P_n converge? What does the limit look like?

Kingman's result

Let $h := \sup\left\{x : Q([x,1]) + P_0([x,1]) > 0\right\}$. So h is interpreted as "the largest fitness value of the population."

Define $\zeta(b) := 1 - b\int \frac{Q(dy)}{1-y/h}$.

Kingman's result

Let $h := \sup\left\{x : Q([x,1]) + P_0([x,1]) > 0\right\}$. So h is interpreted as "the largest fitness value of the population."

Define $\zeta(b) := 1 - b\int \frac{Q(dy)}{1-y/h}$.

Lemma

$\zeta(b) \leq 0$ if and only if there exists a unique solution θ of the equation

$$\int \frac{b\theta Q(dx)}{\theta - (1-b)x} = 1, \quad \theta \geq (1-b)h$$

Kingman's result

Theorem (Kingman, 1978)

Democracy regime.

If $\zeta(b) \leq 0$, then $(P_n)_{n\geq 0}$ converges strongly to

$$\frac{b\theta Q(dx)}{\theta - (1-b)x},$$

with θ being the unique solution of $\int \frac{b\theta Q(dx)}{\theta-(1-b)x} = 1$.

Kingman's result

Theorem (Kingman, 1978)

Democracy regime.
If $\zeta(b) \leq 0$*, then* $(P_n)_{n\geq 0}$ *converges strongly to*

$$\frac{b\theta Q(dx)}{\theta - (1-b)x},$$

with θ *being the unique solution of* $\int \frac{b\theta Q(dx)}{\theta-(1-b)x} = 1$*.*

Meritocracy/ Aristocracy regime.
If $\zeta(b) > 0$*, then* $(P_n)_{n\geq 0}$ *converges weakly to*

$$\frac{bQ(dx)}{1 - x/h} + \zeta(b)\delta_h(dx),$$

here $\delta_h(dx)$ *is the Dirac measure at* h*. Condensation occurs.*

Interplay of selection and mutation

Democracy regime (no condensation): $b\int \frac{Q(dy)}{1-y/h} \geq 1$

- high mutation probability
- fit mutation distribution

That is, mutation dominates selection.

Meritocracy/Aristocracy regime (condensation):$b\int \frac{Q(dy)}{1-y/h} < 1$

- low mutation probability
- less fit mutation distribution

That is, selection dominates mutation.

A main gradient in the proof

Let $w_n = \int xP_n(dx), \mu_n = \int x^n Q(dx), m_n = \int x^n P_0(dx)$. Let

$$W_n = w_0 w_1 \cdots w_{n-1}$$

A main gradient in the proof

Let $w_n = \int xP_n(dx), \mu_n = \int x^n Q(dx), m_n = \int x^n P_0(dx)$. Let

$$W_n = w_0 w_1 \cdots w_{n-1}$$

Then (W_n) satisfies

$$W_n = \sum_{i=1}^{n-1} W_{n-i} \times (1-b)^i b\mu_i + (1-b)^n m_n$$

Condensation wave

Theorem (Dereich and Mörters 2013)

Assume $m_n/\mu_n \to 0$ *and there exists* $\alpha > 1$ *such that*

$$Q(1-t,1) \sim t^{\alpha}, \quad t \to 0.$$

Condensation wave

Theorem (Dereich and Mörters 2013)

Assume $m_n/\mu_n \to 0$ *and there exists* $\alpha > 1$ *such that*

$$Q(1-t,1) \sim t^{\alpha}, \quad t \to 0.$$

If $\zeta(b) > 0$ *(condensation), then*

$$\lim_{n\uparrow\infty} P_n(1-x/n,1) = \frac{\zeta(b)}{\Gamma(\alpha)} \int_0^x y^{\alpha-1} e^{-y}\, dy, \qquad \text{for any } x > 0.$$

Condensation wave

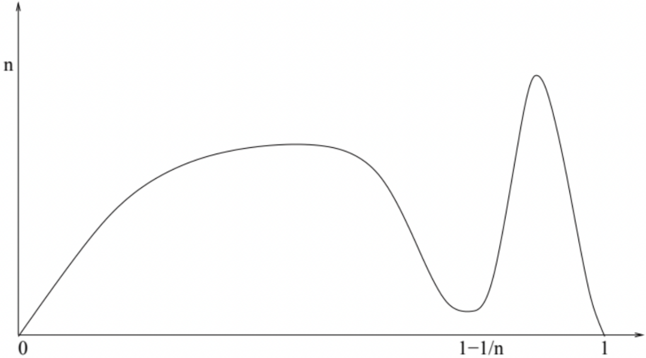

Figure: Dereich and Mörters 2013

Conjectures

1. Dereich and Mörters proposed a conjecture that the Gamma-shape condensation wave is universal in Kingman-like models.

Conjectures

1. Dereich and Mörters proposed a conjecture that the Gamma-shape condensation wave is universal in Kingman-like models.

2. In Kingman's model, if we replace b by a sequence of i.i.d. mutation probabilities (β_n) for all generations with $\mathbb{E}[\beta_n] = b$,

- how will that affect the condensation compared to the original model?
- will the same effects apply to Kingman-like models?

Similar models

- Lenski experiment (Y, 2017). Fix $\lambda > 0$. Consider a population model as follows

$$P_{n+1}(dx) = (1-b)\frac{e^{t_n x} P_n(dx)}{\lambda} + bQ(dx), \quad n \geq 0,$$

where t_n is a number such that

$$\int e^{t_n x} P_n(dx) = \lambda.$$

Similar models

- Continuous-time model (Betz, Dereich and Mörters, 2017).

  Let M be a certain set of nonnegative finite measures on $\mathbb{R}_+$.
  Let $B : M \mapsto C(\mathbb{R}_+), C : M \mapsto C(\mathbb{R}_+)$. Fix $\alpha > 0$ and define

$$\partial_t P_t(dx) = B[P_t]P_t(dx) + x^\alpha C[P_t]dx.$$

Similar models

- Continuous-time model (Betz, Dereich and Mörters, 2017).

  Let M be a certain set of nonnegative finite measures on $\mathbb{R}_+$. Let $B : M \mapsto C(\mathbb{R}_+), C : M \mapsto C(\mathbb{R}_+)$. Fix $\alpha > 0$ and define

  $$\partial_t P_t(dx) = B[P_t]P_t(dx) + x^\alpha C[P_t]dx.$$

  It is a generalisation of continuous-time Kingman's model, note that (6) is equivalent to

  $$P_{n+1}(dx) - P_n(dx) = \underbrace{(1-b)\left(\frac{x}{\int yP_n(dy)} - 1\right)}_{B[P_n]} P_n(dx) + \underbrace{bQ(dx)}_{x^\alpha C[P_n]dx}$$

Similar models

- Unbounded fitness (Park and Krug, 2008)

  Consider the model

$$f_{n+1}(x) = (1-b)\frac{xf_n(x)}{\int yf_n(y)dy} + bg(x)$$

  where

  - $g(x) = e^{-x}1_{x\geq 0}$ is the density of Q
  - $f_n(x)$ is the density of P_n

Similar models

- Unbounded fitness (Park and Krug, 2008)

  Consider the model

$$f_{n+1}(x) = (1-b)\frac{xf_n(x)}{\int yf_n(y)dy} + bg(x)$$

  where

  - $g(x) = e^{-x}1_{x\geq 0}$ is the density of Q
  - $f_n(x)$ is the density of P_n

  Roughly, it holds that

$$f_n(x) \approx be^{-x} + (1-b)\phi_{n,n}(x)$$

  where $\phi_{n,n}$ is the density of $N(n,n)$.

Similar models

- Unbounded fitness (Park and Krug, 2008)

  Consider the model

$$f_{n+1}(x) = (1-b)\frac{xf_n(x)}{\int yf_n(y)dy} + bg(x)$$

  where

  - $g(x) = e^{-x}1_{x\geq 0}$ is the density of Q
  - $f_n(x)$ is the density of P_n

  Roughly, it holds that

$$f_n(x) \approx be^{-x} + (1-b)\phi_{n,n}(x)$$

  where $\phi_{n,n}$ is the density of $N(n,n)$.

Conjecture: is the Gaussian travelling wave universal?

- Kingman's model
- Analogy to the Bose-Einstein condensation

Boson gas[4]

- consider indistinguishable particles of an ideal[2] Boson gas in a closed box with rigid walls and fixed volume V
- at the energy level ε_i, there are $g(\varepsilon_i)$ distinguishable states corresponding to ε_i

[2] meaning no particle interaction

[3] we refer to Janson 2012 for a survey on balls-in-boxes model, simply generated trees and related condensation phenomenon

[4] thanks to my physicist friend Dr. Lingxuan Shao (SPEIT) for discussions

Boson gas[4]

- consider indistinguishable particles of an ideal[2] Boson gas in a closed box with rigid walls and fixed volume V
- at the energy level ε_i, there are $g(\varepsilon_i)$ distinguishable states corresponding to ε_i
- assume there are $n(\varepsilon_i)$ particles at the energy level ε_i, the number of configurations[3] is

$$\binom{n(\varepsilon_i) + g(\varepsilon_i) - 1}{n(\varepsilon_i)}$$

[2] meaning no particle interaction

[3] we refer to Janson 2012 for a survey on balls-in-boxes model, simply generated trees and related condensation phenomenon

[4] thanks to my physicist friend Dr. Lingxuan Shao (SPEIT) for discussions

Boson gas

- to achieve maximum entropy, we maximise

$$\prod_i \binom{n_i + g(\varepsilon_i) - 1}{n_i}$$

subject to

$$\sum_i n_i = N, \quad \sum_i \varepsilon_i n_i = U$$

with N the total number of particles and U the total energy

Boson gas

- to achieve maximum entropy, we maximise

$$\prod_i \binom{n_i + g(\varepsilon_i) - 1}{n_i}$$

subject to

$$\sum_i n_i = N, \quad \sum_i \varepsilon_i n_i = U$$

with N the total number of particles and U the total energy

- we obtain

$$n(\varepsilon) = \frac{g(\varepsilon)}{e^{(\varepsilon-\mu)/kT} - 1}$$

with T the temperature, $\mu \leq 0$ the chemical potential and k the Boltzmann constant

Approximation by the continuum setting

For energy levels within $(\varepsilon, \varepsilon + d\varepsilon)$, there are $g(\varepsilon)d\varepsilon$ states, where:

$$g(\varepsilon) = \frac{V}{4\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\sqrt{\varepsilon}$$

Approximation by the continuum setting

For energy levels within $(\varepsilon, \varepsilon + d\varepsilon)$, there are $g(\varepsilon)d\varepsilon$ states, where:

$$g(\varepsilon) = \frac{V}{4\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\sqrt{\varepsilon}$$

Then

$$\sum_i n(\varepsilon_i) = \int \frac{g(\varepsilon)}{e^{(\varepsilon-\mu)/kT} - 1} d\varepsilon = N$$

Bose-Einstein condensation

Let $\hat{\zeta} = 1 - \int \frac{g(\varepsilon)}{e^{\varepsilon/kT}-1} d\varepsilon$.

Bose-Einstein condensation

Let $\hat{\zeta} = 1 - \int \frac{g(\varepsilon)}{e^{\varepsilon/kT}-1} d\varepsilon$. Then

- if $\hat{\zeta} \le 0$ (i.e., $T > T_c$), the particle distribution is

$$n(\varepsilon)d\varepsilon = \frac{g(\varepsilon)}{e^{(\varepsilon-\mu)/kT} - 1} d\varepsilon,$$

  where μ is the unique solution of $\int \frac{g(\varepsilon)}{e^{(\varepsilon-\mu)/kT}-1} d\varepsilon = N$

- if $\hat{\zeta} > 0$ (i.e., $T < T_c$), the particle distribution is

$$\frac{g(\varepsilon)}{e^{\varepsilon/kT} - 1} d\varepsilon + \hat{\zeta}\delta_0(d\varepsilon)$$

- Kingman's model
- Analogy to the Bose-Einstein condensation
- Mapping Bose-Einstein condensation with preferential attachment model with fitness

Preferential attachment model with fitness

Bianconi and Barabási (2000) introduced the preferential attachment model with fitness

- discrete: the network grows in discrete times

Preferential attachment model with fitness

Bianconi and Barabási (2000) introduced the preferential attachment model with fitness

- discrete: the network grows in discrete times
- addition: at each time step we add a new node.
  - a fitness value η_n is assigned to the n-th node, sampled independently from a common distribution on $(0, 1)$

Preferential attachment model with fitness

Bianconi and Barabási (2000) introduced the preferential attachment model with fitness

- discrete: the network grows in discrete times
- addition: at each time step we add a new node.
  - a fitness value η_n is assigned to the n-th node, sampled independently from a common distribution on $(0, 1)$
- connection: the $n-$th node is connected to the $j-$th node with probability

$$\frac{k_j \eta_j}{\sum_{i=1}^{n-1} k_i \eta_i}$$

  where k_i is the degree (number of links) of the $i-$th node

Mapping

Define $\varepsilon_n = -T \log \eta_n$, which is mapped to an energy level in a Boson gas

- adding the n–th node into the network corresponds to
  - adding a new energy level ε_{n+1} and
  - 2 non-interacting particles to the system

Mapping

Define $\varepsilon_n = -T \log \eta_n$, which is mapped to an energy level in a Boson gas

- adding the n–th node into the network corresponds to
  - adding a new energy level ε_{n+1} and
  - 2 non-interacting particles to the system
- for the 2 particles added to the system
  - one particle sits at the level ε_n, and
  - the other one at level ε_j with probability

$$\frac{k_j \eta_j}{\sum_{i=1}^{n-1} k_i \eta_i}$$

Bose-Einstein condensation

Let $g(\varepsilon)$ be the density of the distribution from which ε_n is drawn.

Let $\bar{\zeta} = 1 - \int \frac{g(\varepsilon)}{e^{\varepsilon/T}-1} d\varepsilon$. Then in the limit $n \to \infty$

- Fit-get-rich regime.
  If $\bar{\zeta} \leq 0$ (i.e., $T > T_c$), the particle (link) distribution is

$$\frac{g(\varepsilon)}{e^{(\varepsilon-\mu^*)/T} - 1} d\varepsilon,$$

  where μ^* is the unique solution of $\int \frac{g(\varepsilon)}{e^{(\varepsilon-\mu^*)/T}-1} d\varepsilon = 1$

- Winner-takes-all regime.
  If $\bar{\zeta} > 0$ (i.e., $T < T_c$), the particle distribution is

$$\frac{g(\varepsilon)}{e^{\varepsilon/T} - 1} d\varepsilon + \bar{\zeta}\delta_0(d\varepsilon)$$

- Winner-takes-all regime.
  If $\bar{\zeta} > 0$ (i.e., $T < T_c$), the particle distribution is

$$\frac{g(\varepsilon)}{e^{\varepsilon/T} - 1} d\varepsilon + \bar{\zeta}\delta_0(d\varepsilon)$$

Bianconi and Barabási (2000):

The fittest node is not only the largest, but despite the continuous emergence of new nodes that compete for links, it always acquires a finite fraction of links.

Main proof ingredient

The rigorous proof was given later by Borgs, Chayes, Daskalakis and Roch (2007).
The main idea is to consider the process as a
Generalised Pólya urn

Main proof ingredient

The rigorous proof was given later by Borgs, Chayes, Daskalakis and Roch (2007).
The main idea is to consider the process as a
Generalised Pólya urn

- there are $q < \infty$ bins and each bin i is assigned a fitness value η_i

Main proof ingredient

The rigorous proof was given later by Borgs, Chayes, Daskalakis and Roch (2007).
The main idea is to consider the process as a
Generalised Pólya urn

- there are $q < \infty$ bins and each bin i is assigned a fitness value η_i
- each bin i is associated with a random vector $\xi_i = (\xi_{i,1}, \cdots, \xi_{i,q})$

Main proof ingredient

The rigorous proof was given later by Borgs, Chayes, Daskalakis and Roch (2007).
The main idea is to consider the process as a
Generalised Pólya urn

- there are $q < \infty$ bins and each bin i is assigned a fitness value η_i
- each bin i is associated with a random vector $\xi_i = (\xi_{i,1}, \cdots, \xi_{i,q})$
- let $X_n = (X_{n,1}, X_{n,2}, \ldots, X_{n,q})$, where $X_{n,i}$ is the number of balls in bin i at time n.

Main proof ingredient

The rigorous proof was given later by Borgs, Chayes, Daskalakis and Roch (2007).
The main idea is to consider the process as a
Generalised Pólya urn

- there are $q < \infty$ bins and each bin i is assigned a fitness value η_i
- each bin i is associated with a random vector $\xi_i = (\xi_{i,1}, \cdots, \xi_{i,q})$
- let $X_n = (X_{n,1}, X_{n,2}, \ldots, X_{n,q})$, where $X_{n,i}$ is the number of balls in bin i at time n.
- at each time n, we pick bin i with probability proportional to $\eta_i X_{n-1,i}$

Main proof ingredient

The rigorous proof was given later by Borgs, Chayes, Daskalakis and Roch (2007).
The main idea is to consider the process as a
Generalised Pólya urn

- there are $q < \infty$ bins and each bin i is assigned a fitness value η_i
- each bin i is associated with a random vector $\xi_i = (\xi_{i,1}, \cdots, \xi_{i,q})$
- let $X_n = (X_{n,1}, X_{n,2}, \ldots, X_{n,q})$, where $X_{n,i}$ is the number of balls in bin i at time n.
- at each time n, we pick bin i with probability proportional to $\eta_i X_{n-1,i}$
- if bin i is selected, we draw an independent copy ξ_i^n of ξ_i and let $X_n = X_{n-1} + \xi_i^n$

For the most general Pólya urn, see Mailler and Villemonais 2020

Stochastic approximation

- Let $(X_n)_{n\geq 0}$ be a Markov chain and $(\mathscr{G}_n)_{n\geq 0}$ the filtration
- Assume

$$\mathbb{E}[X_{n+1} - X_n \,|\, \mathscr{G}_n] = f_n(X_n),$$

Stochastic approximation

- Let $(X_n)_{n\geq 0}$ be a Markov chain and $(\mathscr{G}_n)_{n\geq 0}$ the filtration
- Assume
$$\mathbb{E}[X_{n+1} - X_n \,|\, \mathscr{G}_n] = f_n(X_n),$$
then
$$X_{n+1} - X_n = f_n(X_n) + R_{n+1} - R_n$$
where $R_n = X_n - \mathbb{E}[X_n \,|\, \mathscr{G}_{n-1}]$ and (R_n) is a martingale

Stochastic approximation

- Let $(X_n)_{n\geq 0}$ be a Markov chain and $(\mathscr{G}_n)_{n\geq 0}$ the filtration
- Assume
$$\mathbb{E}[X_{n+1} - X_n \,|\, \mathscr{G}_n] = f_n(X_n),$$
then
$$X_{n+1} - X_n = f_n(X_n) + R_{n+1} - R_n$$
where $R_n = X_n - \mathbb{E}[X_n \,|\, \mathscr{G}_{n-1}]$ and (R_n) is a martingale
- In some good cases, no matter what the almost sure limit of R_n is, X_n always converges to the same constant

Stochastic approximation

- Let $(X_n)_{n\geq 0}$ be a Markov chain and $(\mathscr{G}_n)_{n\geq 0}$ the filtration
- Assume
$$\mathbb{E}[X_{n+1} - X_n \,|\, \mathscr{G}_n] = f_n(X_n),$$
then
$$X_{n+1} - X_n = f_n(X_n) + R_{n+1} - R_n$$
where $R_n = X_n - \mathbb{E}[X_n \,|\, \mathscr{G}_{n-1}]$ and (R_n) is a martingale
- In some good cases, no matter what the almost sure limit of R_n is, X_n always converges to the same constant

Reference: Robbins and Monro 1951, and Kiefer and Wolfowitz 1952, Benaïm 1999

- Kingman's model
- Analogy to the Bose-Einstein condensation
- Mapping Bose-Einstein condensation with preferential attachment model with fitness
- A unifying approach: branching process with reinforcement

Branching processes with reinforcement

Definition (Dereich, Mailler and Mörters, 2017)

- the process starts with one family of one individual whose fitness is drawn from the distribution Q
  - at time t assume there exist $M(t)$ families, and there are $Z_n(t)$ individuals of fitness F_n in the n-th family, for $1 \leq n \leq M(t)$

Branching processes with reinforcement

Definition (Dereich, Mailler and Mörters, 2017)

- the process starts with one family of one individual whose fitness is drawn from the distribution Q
  - at time t assume there exist $M(t)$ families, and there are $Z_n(t)$ individuals of fitness F_n in the n-th family, for $1 \leq n \leq M(t)$
- independently, every individual gives birth with a rate equal to its fitness,
  - or equivalently, in the n-th family birth events occur with a time-dependent rate $F_n Z_n(t)$

Branching processes with reinforcement

Definition (Dereich, Mailler and Mörters, 2017)

- the process starts with one family of one individual whose fitness is drawn from the distribution Q
  - at time t assume there exist $M(t)$ families, and there are $Z_n(t)$ individuals of fitness F_n in the n-th family, for $1 \leq n \leq M(t)$
- independently, every individual gives birth with a rate equal to its fitness,
  - or equivalently, in the n-th family birth events occur with a time-dependent rate $F_n Z_n(t)$
- when a birth even occurs in the n-th family,
  - with probability β a new family is founded, initially consisting of one individual with a fitness drawn from Q
  - with probability γ a new individual with fitness F_n is added to the n-th family

  here we require $\beta + \gamma \geq 1$

Kingman's model as a special case

- individuals give birth to new individuals with a rate equal to their fitnesses
  - with probability β the new individual is a mutant with fitness drawn from
  - with probability $1 - \beta$ the new individual is not a mutant, then it inherits the fitness of its parent

Kingman's model as a special case

- individuals give birth to new individuals with a rate equal to their fitnesses
  - with probability β the new individual is a mutant with fitness drawn from
  - with probability $1 - \beta$ the new individual is not a mutant, then it inherits the fitness of its parent

This model corresponds to $\beta + \gamma = 1$ in the general model

Preferential attachment model as a special case

Recall:

- it starts with one vertex with fitness drawn from Q
- at each time step, a new vertex is introduced, equipped with a fitness drawn from Q
- the new vertex is linked to one of the present vertices with probability proportional to *its fitness* $\times$ *its degree*

Preferential attachment model as a special case

Recall:

- it starts with one vertex with fitness drawn from Q
- at each time step, a new vertex is introduced, equipped with a fitness drawn from Q
- the new vertex is linked to one of the present vertices with probability proportional to *its fitness* $\times$ *its degree*

It corresponds to $\beta = \gamma = 1$ in the general model

- family $=$ vertex, family size $=$ its degree

Preferential attachment model as a special case

Recall:

- it starts with one vertex with fitness drawn from Q
- at each time step, a new vertex is introduced, equipped with a fitness drawn from Q
- the new vertex is linked to one of the present vertices with probability proportional to *its fitness* $\times$ *its degree*

It corresponds to $\beta = \gamma = 1$ in the general model

- family $=$ vertex, family size $=$ its degree
- at a birth event, a new family is founded
  - i.e., a new vertex is introduced

Preferential attachment model as a special case

Recall:

- it starts with one vertex with fitness drawn from Q
- at each time step, a new vertex is introduced, equipped with a fitness drawn from Q
- the new vertex is linked to one of the present vertices with probability proportional to *its fitness* $\times$ *its degree*

It corresponds to $\beta = \gamma = 1$ in the general model

- family = vertex, family size = its degree
- at a birth event, a new family is founded
  - i.e., a new vertex is introduced
- at the same time, the family that gave birth increases its size by 1
  - i.e., the degree of the selected vertex increases by 1

It is a Crump-Mode-Jagers process

The branching process with reinforcement is in fact a Crump-Mode-Jagers branching process

It is a Crump-Mode-Jagers process

The branching process with reinforcement is in fact a Crump-Mode-Jagers branching process

A typical family

- born with one individual equipped with fitness F drawn from Q which is supported on $(0,1)$ with $Q(x,1) > 0, \forall x \in (0,1)$

It is a Crump-Mode-Jagers process

The branching process with reinforcement is in fact a Crump-Mode-Jagers branching process

A typical family

- born with one individual equipped with fitness F drawn from Q which is supported on $(0,1)$ with $Q(x,1) > 0, \forall x \in (0,1)$
- the family size process $(Y(t))_{t \geq 0}$ grows as a Yule process with rate γF

It is a Crump-Mode-Jagers process

The branching process with reinforcement is in fact a Crump-Mode-Jagers branching process

A typical family

- born with one individual equipped with fitness F drawn from Q which is supported on $(0,1)$ with $Q(x,1) > 0, \forall x \in (0,1)$
- the family size process $(Y(t))_{t \geq 0}$ grows as a Yule process with rate γF
- Given (F, Y), the birth times of mutant offspring from this family is an inhomogeneous Poisson process $(\Pi(t))_{t \geq 0}$ with intensity measure

$$\frac{\beta + \gamma - 1}{\gamma} \delta Y(t) + (1 - \gamma) F Y(t) dt$$

A tyical family is characterised by (F, Y, Π)

More notations

Let $(\phi(t))_{t\geq 0}$ be the cadlag process taking values in $\mathbb{N}_0$ that assigns a score to a family t time units after its foundation. It is a function of (F, Y, Π)

More notations

Let $(\phi(t))_{t\geq 0}$ be the cadlag process taking values in $\mathbb{N}_0$ that assigns a score to a family t time units after its foundation. It is a function of (F, Y, Π)

The n-th family is characterised by $(F_n, Y_n, \Pi_n, \phi_n)$. Let τ_n be the birth time of the n-th family

More notations

Let $(\phi(t))_{t\geq 0}$ be the cadlag process taking values in $\mathbb{N}_0$ that assigns a score to a family t time units after its foundation. It is a function of (F, Y, Π)

The n-th family is characterised by $(F_n, Y_n, \Pi_n, \phi_n)$. Let τ_n be the birth time of the n-th family

Define

$$Z^{\phi}(t) = \sum_{n:\tau_n<t} \phi_n(t-\tau_n)$$

More notations

Let $(\phi(t))_{t\geq 0}$ be the cadlag process taking values in $\mathbb{N}_0$ that assigns a score to a family t time units after its foundation. It is a function of (F, Y, Π)

The n-th family is characterised by $(F_n, Y_n, \Pi_n, \phi_n)$. Let τ_n be the birth time of the n-th family

Define

$$Z^{\phi}(t) = \sum_{n:\tau_n<t} \phi_n(t-\tau_n)$$

Reference: Nerman 1981

Lemma

The following two statements are equivalent and the resulting λ^* are the same

- *there exists an $\lambda^* \geq \gamma$, called the Malthusian exponent, such that*
$$\int_0^\infty e^{-\lambda^* s} \mathbb{E}[\Pi(ds)] = 1$$
- $\tilde{\zeta} := 1 - \frac{\beta}{\gamma} \int_0^1 \frac{x}{1-x} Q(dx) \leq 0$ *and λ^* is the unique solution of*
$$\beta \int_0^1 \frac{x}{\lambda^* - \gamma x} Q(dx) = 1$$

Convergence and condensation

Define the empirical distribution

$$\Xi_t := \frac{1}{N(t)} \sum_{n=1}^{M(t)} Z_n(t)\delta_{F_n}$$

here $N(t)$ is the total number of individuals

Convergence and condensation

Define the empirical distribution

$$\Xi_t := \frac{1}{N(t)} \sum_{n=1}^{M(t)} Z_n(t)\delta_{F_n}$$

here $N(t)$ is the total number of individuals

Theorem

Assume ϕ satisfies some conditions.
If $\tilde{\zeta} < 0$ (no condensation), then there exists a positive random variable W, not depending on ϕ such that

$$\lim_{t\to\infty} e^{-\lambda^* t} Z_t^{\phi} = W \frac{\int_0^\infty e^{-\lambda^* t}\mathbb{E}[\phi(t)dt]}{\int_0^\infty te^{-\lambda^* t}\mathbb{E}[\Pi(dt)]}$$

Thus, $\Xi_t \to \pi$ almost surely weekly with $\pi(dx) = \beta\frac{x}{\lambda^* - \gamma x}Q(dx)$
If $\tilde{\zeta} \geq 0$ (condensation), then $\Xi_t \to \pi$ almost surely weakly where

$$\pi(dx) = \frac{\beta}{\gamma}\frac{x}{1-x}Q(dx) + \tilde{\zeta}\delta_1$$

Other results

If $Q(1-h,1) = h^{\alpha}\ell(h)$ with $\alpha > 1$ and $\ell(h)$ slowly varying, then we are in the condensation regime

Other results

If $Q(1-h,1) = h^{\alpha}\ell(h)$ with $\alpha > 1$ and $\ell(h)$ slowly varying, then we are in the condensation regime

Moreover,

- $\frac{max_n Z_n(t)}{N(t)} \to 0, \quad t \to \infty$
- the largest families are born around $T(t) = \frac{\alpha}{\lambda^*} \log t$ (if there is condensation, then$\lambda^* = \gamma$; otherwise $\lambda^* > \gamma$)
- the largest families at time t have fitness $1 - c/t$ and size of order $e^{\gamma(t-T(t))}$

Condensation scenarios

Terminology from Berg, Lewis and Pulè (1986)

- Macroscopic occupation of the ground state: the proportion of individuals in the largest family is asymptotically positive
- Non-extensive condensation: no single family makes an asymptotically positive contribution. The condensation is a collective efforts of a growing number of families

Further questions

- does the condensation wave behave like Gamma function?
- what if fitness can be arbitrarily large?
- can we compute the genealogy and see if there is any connection between the genealogy and the condensation?

- Kingman's model
- Analogy to the Bose-Einstein condensation
- Mapping Bose-Einstein condensation with preferential attachment model with fitness
- A unifying approach: branching process with reinforcement
- Random permutation model

Random permutaion model

It is tightly connected to the Bose-Einstein condensation for Boson gas, see Betz and Ueltschi (2009, 2011)

Random permutaion model

It is tightly connected to the Bose-Einstein condensation for Boson gas, see Betz and Ueltschi (2009, 2011)

Definition

The probability of a permutation π of $\{1, 2, \ldots, n\}$ is defined as

$$\mathbb{P}_n(\pi) = \frac{\prod_{j \geq 1} \theta_j^{r_j(\pi)}}{h_n n!}$$

where

- $r_j(\pi)$ is the number of cycles of length j
- $\theta_j > 0$ is the weight for the cycle of length j
- h_n is a normalisation

Random permutaion model

It is tightly connected to the Bose-Einstein condensation for Boson gas, see Betz and Ueltschi (2009, 2011)

Definition

The probability of a permutation π of $\{1, 2, \ldots, n\}$ is defined as

$$\mathbb{P}_n(\pi) = \frac{\prod_{j \geq 1} \theta_j^{r_j(\pi)}}{h_n n!}$$

where

- $r_j(\pi)$ is the number of cycles of length j
- $\theta_j > 0$ is the weight for the cycle of length j
- h_n is a normalisation

Remarks

1. If $\theta_j = \theta$ for all j, this is the Ewens sampling formula
2. This is a problem of allocating distinguishable balls in indistinguishable boxes

Main results

- assume $\theta_j = j^\alpha \ell(j)$ for $\alpha > 0$ and ℓ slowly varying.
- let $\beta_n = \sum_{j=1}^n \theta_j$ and $\beta^{\leftarrow}(t) = min\{n : \beta_n \geq t\}$
- define the empirical cycle length distribution

$$\mu_n = \frac{1}{n} \sum_{i \geq 1} \lambda_i \delta_{\frac{\lambda_i}{\beta^{\leftarrow}(n)}}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots$ are ordered cycle lengths of a random permutation

Main results

- assume $\theta_j = j^{\alpha}\ell(j)$ for $\alpha > 0$ and ℓ slowly varying.
- let $\beta_n = \sum_{j=1}^{n} \theta_j$ and $\beta^{\leftarrow}(t) = min\{n : \beta_n \geq t\}$
- define the empirical cycle length distribution

$$\mu_n = \frac{1}{n} \sum_{i \geq 1} \lambda_i \delta_{\frac{\lambda_i}{\beta^{\leftarrow}(n)}}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots$ are ordered cycle lengths of a random permutation

Then

$$\lim_{n \to \infty} \mu_n[0, x] = (\gamma + 1) \int_0^x y^{\alpha} e^{-\Gamma(\alpha+2)^{\frac{1}{\alpha+1}} y} dy$$

Conclusions

- many models exhibit condensation phenomena with universal characteristics
- finer properties of condensation are still missing:
  - dominant players
  - condensation/travelling wave
  - random environment
  - genealogy vs condensation, etc
- new and more general models to explore (achieving different condensation scenarios)

Thank you for your attention!