ON COMPOUND POISSON POPULATION MODELS

Martin Möhle, University of Tübingen

(joint work with Thierry Huillet, Université de Cergy-Pontoise)

Workshop on Probability, Population Genetics and Evolution

Centre International de Rencontres Mathématiques (CIRM)

Marseille-Luminy, France

June 11, 2012

Exchangeable population models (Cannings)

Non-overlapping generations $r \in \mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$

Population size N, i.e. N individuals (genes, particles) in each generation

$\nu_i^{(r)}$:= number of offspring of individual i of generation r, $\nu_1^{(r)} + \cdots + \nu_N^{(r)} = N$

Additional assumptions on the offspring

Exchangeability: $(\nu_{\pi 1}^{(r)}, \ldots, \nu_{\pi N}^{(r)}) \overset{d}{=} (\nu_1^{(r)}, \ldots, \nu_N^{(r)}) \qquad \forall\, \pi$

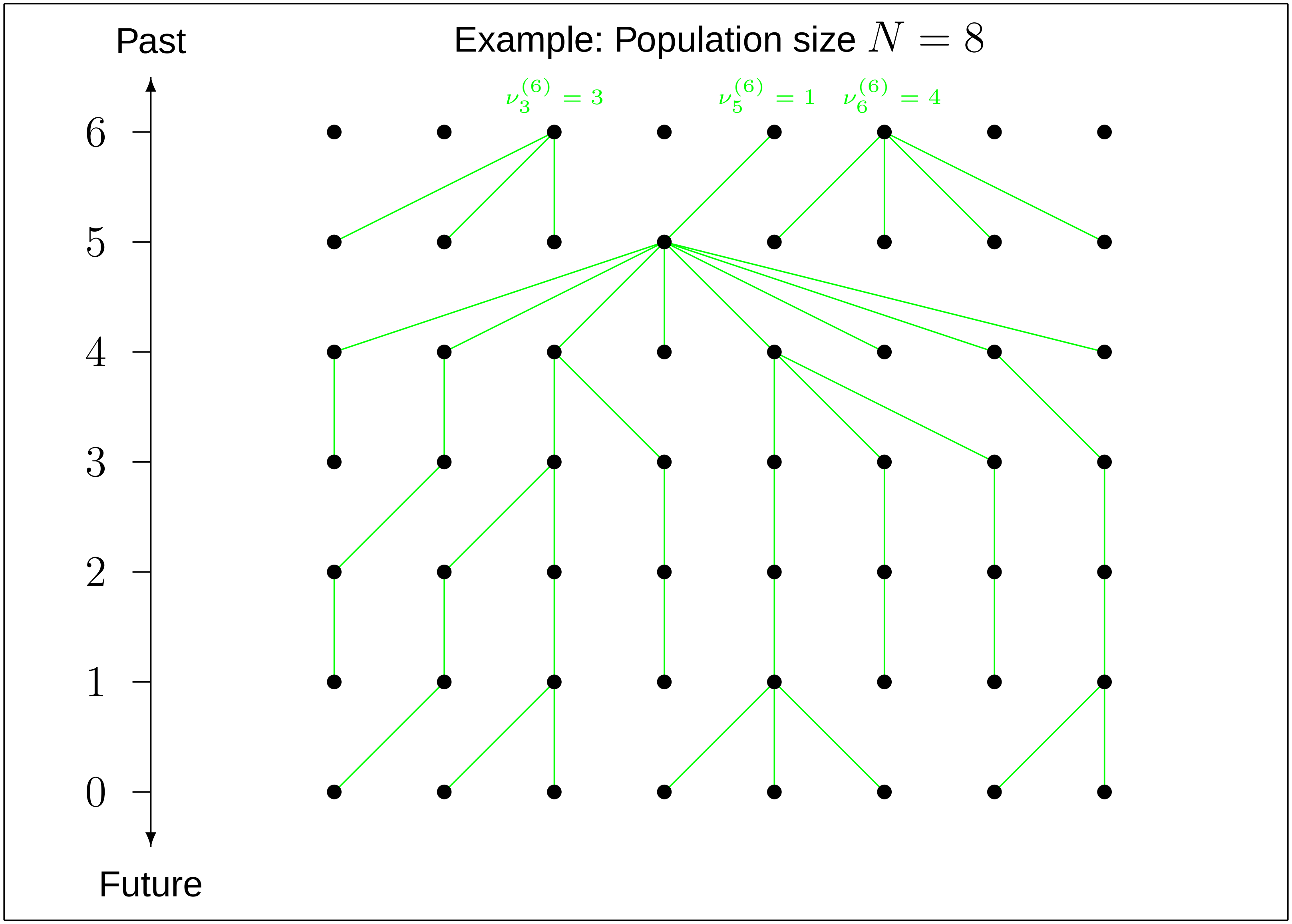
Homogeneity: $(\nu_1^{(r)}, \ldots, \nu_N^{(r)})_r$ i.i.d.

Avoid the trivial model ($\nu_i^{(r)} \equiv 1$). Define $(\nu_1, \ldots, \nu_N) := (\nu_1^{(0)}, \ldots, \nu_N^{(0)})$.

Examples. **Wright-Fisher:** $(\nu_1, \ldots, \nu_N) \overset{d}{=} \mathrm{Multinomial}(N, \frac{1}{N}, \ldots, \frac{1}{N})$

Moran: $(\nu_1, \ldots, \nu_N) =$ random permutation of $(2, \underbrace{1, \ldots, 1}_{N-2}, 0)$

Exchangeable population models (graphical representation)



Conditional branching population models

Let $\xi_1, \xi_2, \ldots$ be independent random variables taking values in $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$. Assume that $\mathbb{P}(\xi_1 + \cdots + \xi_N = N) > 0$ for all $N \in \mathbb{N} := \{1, 2, \ldots\}$.

Perform the following two steps:

1. Conditioning: Let $\mu_1, \ldots, \mu_N$ be random variables such that the distribution of them coincides with that of $\xi_1, \ldots, \xi_N$ conditioned on the event that $\xi_1 + \cdots + \xi_N = N$.

2. Shuffling: Let $\nu_1, \ldots, \nu_N$ be the $\mu_1, \ldots, \mu_N$ randomly permutated.

Conditional branching process models are particular Cannings models with offspring variables $\nu_1, \ldots, \nu_N$ constructed as above (Karlin and McGregor, 1964).

Compound Poisson population models

Compound Poisson population models are particular conditional branching process models for which ξ_n has p.g.f.

$$\mathbb{E}(x^{\xi_n}) \;=\; \exp\big(-\theta_n(\phi(z)-\phi(zx))\big), \qquad |x| \le 1, n \in \mathbb{N},$$

with parameters $0 < \theta_n < \infty$ and with a power series ϕ of the form $\phi(z) = \sum_{m=1}^{\infty} \phi_m \frac{z^m}{m!}$, $|z| < r$, with positive radius r of convergence and $\phi_m \ge 0$, $\phi_1 > 0$.

Distribution and factorial moments of μ

Notation: Taylor expansion $\exp(\theta\phi(z)) = \sum_{k=0}^{\infty} \frac{\sigma_k(\theta)}{k!} z^k$, $|z| < r$.

(The coefficients $\sigma_k(\theta)$ can be computed recursively.)

Distribution of μ: $\boxed{\mathbb{P}(\mu_1 = j_1, \ldots, \mu_N = j_N) = \frac{N!}{j_1! \cdots j_N!} \frac{\sigma_{j_1}(\theta_1) \cdots \sigma_{j_N}(\theta_N)}{\sigma_N(\sum_{n=1}^{N} \theta_n)}}$

($j_1, \ldots, j_N \in \mathbb{N}_0$ with $j_1 + \cdots + j_N = N$)

Factorial moments of μ:

$$\boxed{\mathbb{E}((\mu_1)_{k_1} \cdots (\mu_N)_{k_N}) = \frac{N!}{\sigma_N(\sum_{n=1}^{N} \theta_n)} \sum_{\substack{j_1 \geq k_1, \ldots, j_N \geq k_N \\ j_1 + \cdots + j_N = N}} \frac{\sigma_{j_1}(\theta_1) \cdots \sigma_{j_N}(\theta_N)}{(j_1 - k_1)! \cdots (j_N - k_N)!}}$$

($k_1, \ldots, k_N \in \mathbb{N}_0$)

A subclass of compound Poisson models

We focus on the subclass of compound Poisson models satisfying

$$\frac{\sigma_{k+1}(\theta)}{\sigma_k(\theta)} + \frac{\sigma_{k'+1}(\theta')}{\sigma_{k'}(\theta')} = \frac{\sigma_{k+k'+1}(\theta+\theta')}{\sigma_{k+k'}(\theta+\theta')} \qquad k, k' \in \mathbb{N}_0, \theta, \theta' \in (0,\infty). \qquad (*)$$

Lemma. A compound Poisson model satisfies $(*)$ if and only if $\phi_m = (m-1)!\phi_1(\phi_2/\phi_1)^{m-1}$ for all $m \in \mathbb{N}$. (These are essentially Wright–Fisher models and Dirichlet models.) If $(*)$ holds then μ has factorial moments

$$\boxed{\mathbb{E}((\mu_1)_{k_1} \cdots (\mu_N)_{k_N}) = (N)_{k_1+\cdots+k_N} \frac{\sigma_{k_1}(\theta_1) \cdots \sigma_{k_N}(\theta_N)}{\sigma_k(\theta_1 + \cdots + \theta_N)}}$$

$(k_1, \ldots, k_N \in \mathbb{N}_0)$

Assumption. In the following it is always assumed that $(*)$ holds.

Ancestral process

Take a sample of n ($\le N$) individuals from some generation and consider their ancestors.

$(i,j) \in \mathcal{R}_t = \mathcal{R}_t^{(n)} :\Longleftrightarrow$ individuals i and j have a common parent t generations backwards in time.

The ancestral process $\mathcal{R} := (\mathcal{R}_t)_{t=0,1,\ldots}$ is Markovian with state space $\mathcal{E}_n$ (set of equivalence relations on $\{1,\ldots,n\}$).

Transition probabilities: $\mathbb{P}(\mathcal{R}_{t+1} = \eta \,|\, \mathcal{R}_t = \xi) = \Phi_j^{(N)}(k_1,\ldots,k_j)$, $\xi,\eta \in \mathcal{E}_n$ with $\xi \subseteq \eta$ with

$$\boxed{\Phi_j^{(N)}(k_1,\ldots,k_j) \;:=\; \frac{1}{\sigma_{k_1+\cdots+k_j}(\sum_{n=1}^N \theta_n)} \sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N} \sigma_{k_1}(\theta_{n_1})\cdots\sigma_{k_j}(\theta_{n_j})}$$

where $j := |\eta| =$ number of blocks of η,

$k_1,\ldots,k_j :=$ group sizes of merging classes of ξ. ($\Rightarrow k_1+\cdots+k_j = |\xi|$)

Two basic transition probabilities

Notation: For $N, k \in \mathbb{N}$ define $\Theta_k(N) := \sum_{n=1}^{N} \theta_n^k$.

$$c_N \;:=\; \text{‘coalescence probability’} \;:=\; \mathbb{P}(2 \text{ individuals have same parent})$$

$$= \;\Phi_1^{(N)}(2) \;=\; \frac{1}{\sigma_2(\Theta_1(N))} \sum_{n=1}^{N} \sigma_2(\theta_n) \;=\; \frac{\phi_2\Theta_1(N) + \phi_1^2\Theta_2(N)}{\phi_2\Theta_1(N) + \phi_1^2(\Theta_1(N))^2}.$$

We also need

$$d_N \;:=\; \mathbb{P}(3 \text{ individuals have same parent})$$

$$= \;\Phi_1^{(N)}(3) \;=\; \frac{\phi_3\Theta_1(N) + 3\phi_1\phi_2\Theta_2(N) + \phi_1^3\Theta_3(N)}{\phi_3\Theta_1(N) + 3\phi_1\phi_2(\Theta_1(N))^2 + \phi_1^3(\Theta_1(N))^3}.$$

Exchangeable coalescent processes

- Exchangeable coalescents are discrete-time or continuous-time Markov processes $\Pi = (\Pi_t)_t$ with state space $\mathcal{E}$, the set of equivalence relations (partitions) on $\mathbb{N} := \{1, 2, \ldots\}$.
- During each transition, equivalence classes (blocks) merge together. Simultaneous multiple collisions of blocks are allowed.
- Schweinsberg (2000) characterizes these processes via a finite measure Ξ on the infinite simplex

$$\Delta := \{x = (x_1, x_2, \ldots) : x_1 \geq x_2 \geq \cdots \geq 0, |x| := \sum_{i=1}^{\infty} x_i \leq 1\}.$$

- These processes are therefore also called Ξ-coalescents.

Domain of attraction

For $n \in \mathbb{N}$ let $\varrho_n : \mathcal{E} \longrightarrow \mathcal{E}_n$ denote the restriction of $\mathcal{E}$ to $\mathcal{E}_n$, the set of equivalence relations on $\{1, \ldots, n\}$.

Definition.

- We say that the considered population model is in the domain of attraction of a continuous-time coalescent $\Pi = (\Pi_t)_{t \geq 0}$, if, for each sample size $n \in \mathbb{N}$, the time-scaled ancestral process $(\mathcal{R}^{(n)}_{[t/c_N]})_{t \geq 0}$ weakly converges to $(\varrho_n \circ \Pi_t)_{t \geq 0}$ as $N \longrightarrow \infty$.
- We say that the considered population model is in the domain of attraction of a discrete-time coalescent $\Pi = (\Pi_t)_{t=0,1,\ldots}$, if, for each sample size $n \in \mathbb{N}$, the ancestral process $(\mathcal{R}^{(n)}_t)_{t=0,1,\ldots}$ weakly converges to $(\varrho_n \circ \Pi_t)_{t=0,1,\ldots}$ as $N \longrightarrow \infty$.

Results (Regime 1)

Theorem 1. Suppose that $(*)$ holds. If $\sum_{n=1}^{\infty} \theta_n < \infty$, then the compound Poisson population model is in the domain of attraction of a discrete-time Ξ-coalescent.

Characterization of Ξ. There exists a consistent sequence $(Q_j)_{j\in\mathbb{N}}$ of probability measures Q_j on the j-simplex $\Delta_j := \{(x_1,\ldots,x_j) \in [0,1]^j : x_1+\cdots+x_j \leq 1\}$ uniquely determined via their moments

$$\boxed{\int_{\Delta_j} x_1^{k_1}\cdots x_j^{k_j}\, Q_j(dx_1,\ldots,dx_j) \;=\; \frac{\sigma_{k_1}(\theta_1)\cdots\sigma_{k_j}(\theta_j)}{\sigma_{k_1+\cdots+k_j}(\sum_{n=1}^{\infty}\theta_n)}, \qquad k_1,\ldots,k_j \in \mathbb{N}_0.}$$

Let Q denote the projective limit of $(Q_j)_{j\in\mathbb{N}}$, let $X_1, X_2, \ldots$ be random variables with joint distribution Q, and let ν be the joint distribution of the ordered variables $X_{(1)} \geq X_{(2)} \cdots$. Then, Ξ has density $x \mapsto (x,x) := \sum_{n=1}^{\infty} x_n^2$ with respect to ν. The measure Ξ is concentrated on the subset Δ^* of points $x = (x_1, x_2, \ldots) \in \Delta$ satisfying $|x| := \sum_{n=1}^{\infty} x_n = 1$.

Regime 1 (continued)

Remark. The proof of Theorem 1 is based on general convergence theorems for ancestral processes of Cannings models (M. and Sagitov 2001) and on the moment problem for the j-dimensional simplex (Gupta).

Examples. Suppose that $\theta := \sum_{n=1}^\infty \theta_n < \infty$.

Wright-Fisher models. If $\phi(z) = \phi_1 z$, then $\sigma_k(\theta) = \theta^k \phi_1^k$. In this case ν is the Dirac measure at $p = (\theta_1/\theta, \theta_2/\theta, \ldots) \in \Delta^*$. The measure Ξ assigns its total mass $\Xi(\Delta) = (p,p) = (\sum_{n=1}^\infty \theta_n^2)/\theta^2$ to the single point p.

Dirichlet models. If $\phi(x) = -\log(1-x)$, then $\phi_m = (m-1)!$, $m \in \mathbb{N}$, and $\sigma_k(\theta) = [\theta]_k := \theta(\theta+1)\cdots(\theta+k-1)$, $k \in \mathbb{N}_0$. The limiting coalescent is the discrete-time Dirichlet-Kingman coalescent with parameter $(\theta_n)_{n\in\mathbb{N}}$.

Results (Regime 2)

Theorem 2. Suppose that $(*)$ holds. If $\sum_{n=1}^{\infty} \theta_n = \infty$ and if $\sum_{n=1}^{\infty} \theta_n^2 < \infty$, then the compound Poisson population model is in the domain of attraction of the Kingman coalescent.

Remarks.

1. Time-scaling satisfies $c_N = \Theta_2(N)/(\Theta_1(N))^2$ if $\phi_2 = 0$ and $c_N \sim \phi_2/(\phi_1^2\Theta_1(N))$ if $\phi_2 > 0$, where $\Theta_k(N) := \sum_{n=1}^{N} \theta_n^k$ for $k \in \mathbb{N}$.
2. In contrast to the situation in Theorem 1, the limiting coalescent in Theorem 2 does not depend on the function ϕ of the compound Poisson model. Theorem 2 is for example applicable if $\theta_n = n^{-\alpha}$ with $\alpha \in (\frac{1}{2}, 1]$.

Sketch of proof

The proof of Theorem 2 is based on the following technical lemma.

Lemma. If $(*)$ holds, then the following five conditions are equivalent.

(i) $\lim\limits_{N\to\infty} \dfrac{\Theta_2(N)}{(\Theta_1(N))^2} \;=\; 0.$

(ii) $\lim\limits_{N\to\infty} \dfrac{\Theta_3(N)}{\Theta_1(N)\,\Theta_2(N)} \;=\; 0.$

(iii) $\lim\limits_{N\to\infty} c_N \;=\; 0.$

(iv) $\lim\limits_{N\to\infty} \dfrac{d_N}{c_N} \;=\; 0.$

(v) The compound Poisson model is in the domain of attraction of the Kingman coalescent.

Remark. The proof that (i) - (iii) are equivalent is technical but elementary. The equivalence of (iv) and (v) and the implication ‘(iv) $\Rightarrow$ (iii)’ hold even for arbitrary Cannings models (M., 2000). The interesting point is that, for compound Poisson models, (iii) implies (iv). Note that this implication does not hold for arbitrary Cannings models.

Results (Regime 3)

Theorem 3. Suppose that ($*$) holds, that $\sum_{n=1}^{\infty}\theta_n = \infty$ and that $\sum_{n=1}^{\infty}\theta_n^2 = \infty$. Then the compound Poisson model is in the domain of attraction of the Kingman coalescent if and only if $\Theta_2(N)/(\Theta_1(N))^2 \to 0$ as $N \to \infty$. In this case the time-scaling c_N satisfies $c_N \sim \phi_2/(\phi_1^2\Theta_1(N)) + \Theta_2(N)/(\Theta_1(N))^2$.

Corollary. (unbiased case, Huillet and M., 2010) If ($*$) holds and if $\theta_n = \theta$ does not depend on n, then the compound Poisson model is in the domain of attraction of the Kingman coalescent. The time-scaling c_N satisfies $c_N \sim (1+\phi_2/(\phi_1^2\theta))/N$.

Results (Regime 3, continued)

Theorem 4. Suppose that $(*)$ holds and that all the limits

$$p_1(k) := \lim_{N\to\infty} \frac{\Theta_k(N)}{(\Theta_1(N))^k}, \qquad k \in \mathbb{N},$$

exist. Then all the limits

$$p_j(k_1,\ldots,k_j) := \lim_{N\to\infty} \frac{1}{(\Theta_1(N))^{k_1+\cdots+k_j}} \sum_{\substack{n_1,\ldots,n_j=1 \\ \text{all distinct}}}^{N} \theta_{n_1}^{k_1}\cdots\theta_{n_j}^{k_j},$$

$k_1,\ldots,k_j \in \mathbb{N}$, exist. Suppose now in addition that $\sum_{n=1}^\infty \theta_n = \infty$ and that $p_1(2) > 0$. Then, the compound Poisson model is in the domain of attraction of a discrete-time Ξ-coalescent Π. The characterizing measure $\nu(dx) := \Xi(dx)/(x,x)$ of Π is the Dirac-measure at $x = (x_1, x_2, \ldots) \in \Delta$, where $x_1 := \lim_{k\to\infty}(p_1(k))^{1/k}$ and $x_{n+1} := \lim_{k\to\infty}(p_1(k) - (x_1^k + \cdots + x_n^k))^{1/k}$, $n \in \mathbb{N}$.

Regime 3 (continued)

Remarks.

- Let Z_N be a random variable taking the value $\theta_n/\Theta_1(N)$ with probability $\theta_n/\Theta_1(N)$, $n \in \{1,\ldots,N\}$. The existence of all the limits $p_1(k)$, $k \in \mathbb{N}$, is equivalent to the convergence $Z_N \to Z$ in distribution, where Z has characteristic function $t \mapsto \sum_{k=0}^{\infty}(t^k/k!)p_1(k+1)$, $t \in \mathbb{R}$.
- In contrast to the situation in Theorem 1, the limiting discrete-time Ξ-coalescent in Theorem 4 does not depend on the function ϕ of the compound Poisson model.

Example. Fix $\lambda > 1$ and suppose that $\theta_n = \lambda^n$, $n \in \mathbb{N}$. Then, $p_1(k) = (\lambda-1)^k/(\lambda^k - 1) > 0$, $k \in \mathbb{N}$. In this case the measure Ξ of the limiting Ξ-coalescent assigns its total mass $\Xi(\Delta) = p_1(2) = (\lambda-1)/(\lambda+1)$ to the single point $x = (x_1, x_2, \ldots) \in \Delta^*$ defined via $x_n := (\lambda-1)/\lambda^n = (1-1/\lambda)(1/\lambda)^{n-1}$, $n \in \mathbb{N}$.

Generalization: Assume that ($*$) does not hold

Theorem. (Huillet, M. 2011)

Fix $\theta \in (0,\infty)$ and suppose that the equation $\theta z\phi'(z) = 1$ has a real solution $z(\theta) \in (0,r)$. Then $\mu_1 \to X$ in distribution as $N \to \infty$, where X has distribution

$$\boxed{\mathbb{P}(X=k) \;=\; \sigma_k(\theta)\frac{(z(\theta))^k}{k!}e^{-\theta\phi(z(\theta))}} \qquad k \in \{0,1,2,\ldots\}.$$

The associated symmetric compound Poisson model is in the domain of attraction of the Kingman coalescent. The effective population size $N_e := 1/c_N$ satisfies $N_e \sim \varrho N$ as $N \to \infty$, where $\varrho := 1/\mathbb{E}((X)_2) = 1/(1+\theta(z(\theta))^2\phi''(z(\theta))) \in (0,1]$.

Remark. Proof uses the saddle point method to establish the asymptotics of $\sigma_N(N\theta)$.

Open cases.

a) Symmetric models without a solution $z(\theta)$, condensation (work in progress).

b) Non symmetric models.

Conclusions

- Asymptotics of the ancestry for some compound Poisson population models analyzed
- Results essentially based on convergence theorems (M. 2000 and M. and Sagitov 2001) for ancestral processes of exchangeable Cannings population models
- Convergence to the Kingman coalescent if and only if $\Theta_2(N)/(\Theta_1(N))^2 \longrightarrow 0$
- Compound Poisson models satisfying $(*)$ are never in the domain of attraction of a continuous-time coalescent different from Kingman's coalescent; discrete-time Ξ-coalescents (with simultaneous multiple collisions) arise if the parameters θ_n are 'unbalanced'
- Three regimes depending on the behavior of the series $\sum_n \theta_n$ and $\sum_n \theta_n^2$; complete convergence results for the first two regimes; partial convergence results for the third regime (when both series diverge)
- Convergence to the Kingman coalescent holds even for more general symmetric compound Poisson models, which do not necessarily satisfy the restriction $(*)$

References

CANNINGS, C. (1974) The latent roots of certain Markov chains arising in genetics: a new approach, I. Haploid models. *Adv. Appl. Probab.* **6**, 260–290.

GUPTA, J. C. (1999) The moment problem for the standard k-dimensional simplex. *Sankhyā A* **61**, 286–291.

HUILLET, T. AND MÖHLE, M. (2011) Population genetics models with skewed fertilities: a forward and backward analysis. *Stochastic Models* **27**, 521–554.

HUILLET, T. AND MÖHLE, M. (2012) Correction on ‘Population genetics models with skewed fertilities: a forward and backward analysis’. *Stochastic Models*, to appear.

KARLIN, S. AND MCGREGOR, J. (1964) Direct product branching processes and related Markov chains. *Proc. Nat. Acad. Sci. U.S.A.* **51**, 598–602.

References (continued)

KINGMAN, J. F. C. (1982) The coalescent. *Stochastic Process. Appl.* **13**, 235–248.

MÖHLE, M. (2000) Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. Appl. Probab.* **32**, 983–993.

MÖHLE, M. (2011) Coalescent processes derived from some compound Poisson population models. *Electron. Comm. Probab.* **16**, 567–582.

MÖHLE, M. AND SAGITOV, S. (2001) A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* **29**, 1547–1562.

SCHWEINSBERG, J. (2000) Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5**, 1–50.