Dual random fragmentation and coagulation and an application to the genealogy of Yule processes

Jean Bertoin * and Christina Goldschmidt †

March 17, 2004

Abstract

The purpose of this work is to describe a duality between a fragmentation associated to certain Dirichlet distributions and a natural random coagulation. The dual fragmentation and coalescent chains arising in this setting appear in the description of the genealogy of Yule processes.

1 Introduction

At a naive level, fragmentation and coagulation are inverse phenomena, in that a simple time-reversal changes one into the other. However, stochastic models for fragmentation and coalescence usually impose strong hypotheses on the dynamics of the processes, such as the branching property for fragmentation (distinct fragments evolve independently as time passes), and these requirements do not tend to be compatible with time-reversal. Thus, in general, the time-reversal of a coalescent process is not a fragmentation process.

Nonetheless, there are a few special cases in which time-reversal does transform a coalescent process into a fragmentation process. Probably the most important example was discovered by Pitman [16]; it is related to the so-called cascades of Ruelle and the Bolthausen-Sznitman coalescent [7], and also has a natural interpretation in terms of the genealogy of a remarkable branching process considered by Neveu, see [4] and [6].

The first purpose of this note is to point out other simple instances of such duality, which rely on certain Dirichlet and Poisson-Dirichlet distributions. Then, in the second part, we shall show that these examples are related to the genealogy of Yule processes.

*Laboratoire de Probabilités et Modèles Aléatoires and Institut universitaire de France, Université Pierre et Marie Curie, 175, rue du Chevaleret, F-75013 Paris, France.

†Laboratoire de Probabilités et Modèles Aléatoires, Université Pierre et Marie Curie, 175, rue du Chevaleret, F-75013 Paris, France.

2 Dual fragmentation and coagulation

2.1 Some notation

For every integer $n \geq 1$, we consider the simplex

$$\Delta_n := \left\{ x = (x_1, \dots, x_{n+1}) : x_i \geq 0 \text{ for every } i = 1, \dots, n+1 \text{ and } \sum_{i=1}^{n+1} x_i = 1 \right\}.$$

It will also be convenient to agree that $\Delta_0 := \{1\}$. We shall often refer to the coordinates $x_1, \dots, x_{n+1}$ of points x in Δ_n as *masses*.

We recall that the n-dimensional Dirichlet distribution with parameter $(\alpha_1, \dots, \alpha_{n+1})$ is the probability measure on the simplex Δ_n with density

$$\frac{\Gamma(\alpha_1 + \cdots + \alpha_{n+1})}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{n+1})} x_1^{\alpha_1 - 1} \cdots x_{n+1}^{\alpha_{n+1} - 1}.$$

The special case when $\alpha_1 = \dots = \alpha_{n+1} := \alpha \in]0, \infty[$ will have an important role in this work; it will be convenient to write $\mathrm{Dir}_n(\alpha)$ for this distribution. We recall the following well-known construction: let $\gamma_1, \dots, \gamma_{n+1}$ be i.i.d. gamma variables with parameters (α, c). Set $\bar{\gamma} = \gamma_1 + \cdots + \gamma_n$, so that $\bar{\gamma}$ has a gamma distribution with parameters $(\alpha(n+1), c)$. Then the $(n+1)$-tuple

$$(\gamma_1/\bar{\gamma}, \dots, \gamma_{n+1}/\bar{\gamma})$$

has the distribution $\mathrm{Dir}_n(\alpha)$ and is independent of $\bar{\gamma}$.

We also define the (ranked) infinite simplex

$$\Delta_\infty := \left\{ x = (x_1, \dots) : x_1 \geq x_2 \geq \dots \geq 0 \text{ and } \sum_{i=1}^{\infty} x_i = 1 \right\}$$

and recall that the Poisson-Dirichlet distribution with parameter $\theta > 0$, which will be denoted by $\mathrm{PD}(\theta)$ in the sequel, is the law of the random sequence

$$\xi := \left(\frac{a_1}{\sum_{i=1}^\infty a_i}, \frac{a_2}{\sum_{i=1}^\infty a_i}, \dots \right),$$

where $a_1 \geq a_2 \geq \dots > 0$ are the atoms of a Poisson random measure on $]0, \infty[$ with intensity $\theta y^{-1} \mathrm{e}^{-y} dy$. We also recall that ξ is independent of $\sum_{i=1}^\infty a_i$, and that the latter has the gamma distribution with parameters $(\theta, 1)$. By the celebrated Lévy-Itô decomposition of subordinators, we may also rephrase this construction as follows: if $\gamma = (\gamma(t), t \geq 0)$ is a standard gamma process and, for each fixed $\theta > 0$, $\delta_1 \geq \delta_2 \geq \dots$ denotes the sequence of sizes of the jumps of γ on the time interval $[0, \theta]$, then

$$\left(\frac{\delta_1}{\gamma(\theta)}, \frac{\delta_2}{\gamma(\theta)}, \dots \right)$$

has the $\mathrm{PD}(\theta)$ distribution and is independent of $\gamma(\theta)$.

2.2 Two dual random transformations

We now define two random transformations:

$$\mathrm{Frag}_k : \Delta_n \to \Delta_{n+k} \quad \text{and} \quad \mathrm{Coag}_k : \Delta_{n+k} \to \Delta_n \,,$$

where k, n are integers.

First, we fix $x = (x_1, \ldots, x_{n+1}) \in \Delta_n$ and pick an index $I \in \{1, \ldots, n+1\}$ at random according to the distribution

$$\mathbb{P}(I = i) \;=\; x_i \,, \qquad i = 1, \ldots, n+1 \,,$$

so that x_I is a size-biased pick from the sequence x. Let $\eta = (\eta_1, \ldots, \eta_{k+1})$ be a random variable with values in Δ_k which is distributed according to $\mathrm{Dir}_k(1/k)$ and independent of I. Then we split the Ith mass of x according to η and we obtain a random variable in Δ_{n+k}:

$$\mathrm{Frag}_k(x) \;:=\; (x_1, \ldots, x_{I-1}, x_I\eta_1, \ldots, x_I\eta_{k+1}, x_{I+1}, \ldots, x_{n+1}) \;.$$

Second, we fix $x = (x_1, \ldots, x_{n+k+1}) \in \Delta_{n+k}$ and pick an index $J \in \{1, \ldots, n+1\}$ uniformly at random. We merge the $k+1$ masses $x_J, x_{J+1} \ldots, x_{J+k}$ to form a single mass $\sum_{i=J}^{J+k} x_i$ and leave the other masses unchanged. We obtain a random variable in Δ_n:

$$\mathrm{Coag}_k(x) \;=\; \left(x_1, \ldots, x_{J-1}, \sum_{i=J}^{J+k} x_i, x_{J+k+1}, \ldots, x_{n+k+1} \right) \,.$$

Remark. Consider the following alternative random coagulation of $x = (x_1, \ldots, x_{n+k+1}) \in \Delta_{n+k}$. Pick $k+1$ indices $i_1, \ldots, i_{k+1}$ from $\{1, \ldots, n+k+1\}$ uniformly at random without replacement, merge the masses $x_{i_1}, \ldots, x_{i_{k+1}}$, leave the other masses unchanged and let $\widetilde{\mathrm{Coag}}_k(x)$ be the sequence obtained by ranking the resulting masses in decreasing order. Write also $\mathrm{Coag}_k^{\downarrow}(x)$ for the sequence $\mathrm{Coag}_k(x)$ re-arranged in decreasing order. Then the pairs $(\xi, \mathrm{Coag}_k^{\downarrow}(\xi))$ and $(\xi, \widetilde{\mathrm{Coag}}_k(\xi))$ have the same distribution. This remark applies in particular to the case when ξ has the law $\mathrm{Dir}_{n+k}(1/k)$, and can thus be combined with forthcoming Proposition 1.

The starting point of this work lies in the observation of a simple relation of duality which links these two random transformations via Dirichlet laws.

Proposition 1. *Let $k, n \geq 1$ be two integers, and ξ, ξ' two random variables with values in Δ_n and Δ_{n+k}, respectively. The following assertions are then equivalent:*

(i) ξ has the law $\mathrm{Dir}_n(1/k)$ and, conditionally on ξ, ξ' is distributed as $\mathrm{Frag}_k(\xi)$.

(ii) ξ' has the law $\mathrm{Dir}_{n+k}(1/k)$ and, conditionally on ξ', ξ is distributed as $\mathrm{Coag}_k(\xi')$.

It has been observed by Kingman [12] that for $k = 1$, if ξ' is uniformly distributed on the simplex Δ_{n+1} (i.e. has the law $\mathrm{Dir}_{n+1}(1)$), then $\mathrm{Coag}_1(\xi')$ is uniformly distributed on Δ_n. Clearly, this agrees with our statement.

Proof: Let $\gamma_1, \gamma_2, \ldots, \gamma_{n+1}$ be independent $\mathrm{Gamma}(1/k, 1)$ random variables and set

$$\bar{\gamma} \;=\; \sum_{i=1}^{n+1} \gamma_i \quad \text{and} \quad \xi \;=\; \left(\frac{\gamma_1}{\bar{\gamma}}, \ldots, \frac{\gamma_{n+1}}{\bar{\gamma}} \right) \,,$$

so that ξ has law $\mathrm{Dir}_n(1/k)$ and is independent of $\bar{\gamma}$. Suppose that η is a $\mathrm{Dir}_k(1/k)$ random variable which is independent of the γ_i's, and let $\Phi : \mathbb{R}^{n+k+1} \to \mathbb{R}$ be a bounded measurable function. Let I be an index picked at random from $\{1,\dots,n+1\}$ according to the conditional distribution

$$\mathbb{P}(I=i \mid \gamma_1,\dots,\gamma_{n+1}) \;=\; \gamma_i/\bar\gamma\,, \qquad i=1,\dots,n+1\,,$$

and denote by $\mathrm{Frag}_k(\xi)$ the random sequence obtained from ξ after the fragmentation of its Ith mass according to η. We have

$$\mathbb{E}\left(\Phi(\mathrm{Frag}_k(\xi)), I=i\right) = \mathbb{E}\left[\frac{\gamma_i}{\gamma}\Phi\left((\gamma_l/\bar\gamma)_{l<i}, \gamma_i\eta/\bar\gamma, (\gamma_l/\bar\gamma)_{l>i}\right)\right].$$

Now, using the independence of $\bar\gamma$ and ξ and the fact that $\bar\gamma$ has the law $\mathrm{Gamma}((n+1)/k,1)$, we see that the last expression is equal to

$$\begin{aligned}
& \frac{k}{n+1}\mathbb{E}\left[\gamma_i\Phi\left((\gamma_l/\bar\gamma)_{l<i}, \gamma_i\eta/\bar\gamma, (\gamma_l/\bar\gamma)_{l>i}\right)\right] \\
= & \frac{k}{n+1}\mathbb{E}\int_0^\infty x\Phi\left(\frac{(\gamma_l)_{l<i}}{x+\sum_{j\neq i}\gamma_j}, \frac{x\eta}{x+\sum_{j\neq i}\gamma_j}, \frac{(\gamma_l)_{l>i}}{x+\sum_{j\neq i}\gamma_j}\right)\frac{1}{\Gamma(1/k)}x^{1/k-1}\mathrm{e}^{-x}dx \\
= & \frac{1}{n+1}\mathbb{E}\left[\Phi\left(\frac{(\gamma_l)_{l<i}}{\gamma'+\sum_{j\neq i}\gamma_j}, \frac{\gamma'\eta}{\gamma'+\sum_{j\neq i}\gamma_j}, \frac{(\gamma_l)_{l>i}}{\gamma'+\sum_{j\neq i}\gamma_j}\right)\right]
\end{aligned}$$

where $\gamma' \sim \mathrm{Gamma}((k+1)/k,1)$, independently of η and $(\gamma_j)_{j\neq i}$. But then $\gamma'\eta$ is a collection of $k+1$ independent $\mathrm{Gamma}(1/k,1)$ random variables, so $\mathrm{Frag}_k(\xi)$ has the law $\mathrm{Dir}_{n+k}(1/k)$ and is independent of the random index I which is uniformly distributed on $\{1,\dots,n+1\}$. Since we can recover ξ from $\mathrm{Frag}_k(\xi)$ and I by an obvious coagulation, this completes the proof. □

Next we turn our attention to the infinite ranked simplex and define two random transformations, $\mathrm{Frag}_\infty : \Delta_\infty \to \Delta_\infty$ and $\mathrm{Coag}_a : \Delta_\infty \to \Delta_\infty$, where $a \in [0,1]$ is some parameter. The fragmentation transformation on the infinite simplex simply mimics that on the finite simplex; in this direction, recall that the Poisson-Dirichlet $\mathrm{PD}(1)$ arises as the weak limit as $k\to\infty$ of sequence of $\mathrm{Dir}_k(1/k)$ variables after obvious reordering. More precisely, given $x=(x_1,\dots)\in\Delta_\infty$, we pick a mass x_I at random by size-biased sampling and split x_I using an independent variable $\eta=(\eta_1,\dots)$ with law $\mathrm{PD}(1)$. In other words, $\mathrm{Frag}_\infty(x)$ is the ranked sequence with unordered terms $x_1,\dots,x_{I-1},x_I\eta_1,x_I\eta_2,\dots,x_{I+1},\dots$.

Next, consider a sequence $U_1,U_2,\dots$ of i.i.d. uniform random variables and $a\in[0,1]$. Starting again from some fixed $x\in\Delta_\infty$, we merge the masses x_i for which $U_i\le a$ into a single mass and leave the others unchanged. We denote by $\mathrm{Coag}_a(x)$ the random sequence obtained by putting the resulting masses in decreasing order. We then have the following analogue of Proposition 1, which is reminiscent of Corollary 13 of Pitman [16].

Proposition 2. *Let ξ,ξ' be two random variables with values in Δ_∞. For every $\theta>0$, the following assertions are equivalent:*

(i) ξ has the law $\mathrm{PD}(\theta)$ and, conditionally on ξ, ξ' is distributed as $\mathrm{Frag}_\infty(\xi)$.

(ii) ξ' has the law $\mathrm{PD}(\theta+1)$ and, conditionally on ξ', ξ is distributed as $\mathrm{Coag}_{1/(\theta+1)}(\xi')$.

Proof: Let $\gamma=(\gamma(t),t\ge0)$ be a standard gamma process and set

$$D_t = \gamma((\theta+1)t)/\gamma(\theta+1),$$

for $0 \le t \le 1$, so that $(D_t, 0 \le t \le 1)$ is a Dirichlet process of parameter $\theta + 1$. (The vector of ordered jumps of this Dirichlet process has the PD$(\theta+1)$ distribution.) Consider the following alternative way of thinking of the random coagulation operator $\mathrm{Coag}_{1/(\theta+1)}$: pick a point V uniformly in $[0,1]$ and define a new process $(D'_t, 0 \le t \le 1)$ by

$$D'_t = \begin{cases} D_{\theta t/(\theta+1)} & \text{if } t < V \\ D_{(1+\theta t)/(\theta+1)} & \text{if } t \ge V. \end{cases}$$

As the times of the jumps of D are uniformly distributed on $[0,1]$, this picks a proportion $1/(\theta+1)$ of them and coalesces them into a single jump (say $\beta^* = D_{(1+\theta V)/(\theta+1)} - D_{\theta V/(\theta+1)}$) at V. Let $\beta_1 \ge \beta_2 \ge \ldots > 0$ be the sequence of other jumps of D' and $U_1, U_2, \ldots$ the corresponding jump times. Let $\beta'_1 \ge \beta'_2 \ge \ldots > 0$ be the sequence of jumps of D in the interval $[\theta V/(\theta+1), (1+\theta V)/(\theta+1)]$, so that $\beta^* = \sum_{i=1}^{\infty} \beta'_i$. We wish to show that D' is a Dirichlet process with parameter θ, so that the vector $(\beta^*, \beta_1, \beta_2, \ldots)$ of its jumps (re-arranged in the decreasing order) has the PD(θ) distribution. We will also show that the mass β^* resulting from the coalescence constitutes a size-biased pick from this vector.

Let

$$\gamma^1(t) = \begin{cases} \gamma(t) & \text{if } t < V\theta \\ \gamma(t+1) - (\gamma(V\theta+1) - \gamma(V\theta)) & \text{if } V\theta \le t \le \theta \end{cases}$$
$$\gamma^2(t) = \gamma(V\theta + t) - \gamma(V\theta) \qquad \text{for } 0 \le t \le 1.$$

Then γ^1 and γ^2 are independent processes with $\gamma^1 \overset{d}{=} (\gamma(t), 0 \le t \le \theta)$ and $\gamma^2 \overset{d}{=} (\gamma(t), 0 \le t \le 1)$, independently of V. Write $\delta_1 \ge \delta_2 \ge \ldots$ for the ordered sequence of jumps of γ^1 and $T_1, T_2, \ldots$ for the corresponding times of these jumps. Write $\delta'_1 \ge \delta'_2 \ge \ldots$ for the ordered sequence of jumps of γ^2. Then

(i) $U_1 = T_1/\theta, U_2 = T_2/\theta, \ldots$ are i.i.d. U$[0,1]$,

(ii) $\beta^* = \gamma^2(1)/\gamma(1+\theta)$ and so has a Beta$(\theta, 1)$ distribution,

(iii) $\frac{1}{\beta^*}(\beta'_1, \beta'_2, \ldots) = \frac{1}{\gamma^2(1)}(\delta'_1, \delta'_2, \ldots)$ and so has the PD(1) distribution,

(iv) $\frac{1}{1-\beta^*}(\beta_1, \beta_2, \ldots) = \frac{1}{\gamma^1(\theta)}(\delta_1, \delta_2, \ldots)$ and so has the PD(θ) distribution.

Furthermore, the random variables in (i) to (iv) above are independent. The fact that β^* is a size-biased pick from $(\beta^*, \beta_1, \beta_2, \ldots)$ and the PD(θ) distribution of the latter follow from (i) and (iii) and the stick-breaking scheme (see, for instance, Definition 1 in Pitman and Yor [18]). That D' is a Dirichlet process of parameter θ then follows from (iv) and the independence.

The coagulation operator used here can be re-phrased as follows: starting with $x \in \Delta_\infty$, take a sequence $V, V_1, V_2, \ldots$ of i.i.d. U$[0,1]$ random variables, merge the masses x_i for which $V_i \in [\theta V/(\theta+1), (1+\theta V)/(\theta+1)]$, leave the other masses unchanged and, finally, rank the resulting sequence in decreasing order. Call this operator $\widetilde{\mathrm{Coag}}_{1/(\theta+1)}$. Then it is clear that whenever ξ' is a random exchangeable sequence in Δ_∞, $(\xi', \mathrm{Coag}_{1/(\theta+1)}(\xi'))$ and $(\xi', \widetilde{\mathrm{Coag}}_{1/(\theta+1)}(\xi'))$ have the same distribution. Our claim follows now readily from these results. $\square$

Remark. It may be interesting to check Proposition 2 as follows. Consider Poisson random measure M on $(0, \infty)$ with intensity $\theta x^{-1} \mathrm{e}^{-x} dx$. Let $a_1, a_2, \ldots$ be the atoms of

M in decreasing order, so that

$$\left(\frac{a_1}{\sum_{j=1}^{\infty} a_j}, \frac{a_2}{\sum_{j=1}^{\infty} a_j}, \dots\right)$$

has distribution $\mathrm{PD}(\theta)$, independently of $\sum_{j=1}^{\infty} a_j$. Let $\eta \sim \mathrm{PD}(1)$, independently of M and suppose that $\Phi : \Delta_\infty \to \mathbb{R}$ is any symmetric bounded measurable function. Then if $\xi \sim \mathrm{PD}(\theta)$, using independence we have that

$$\mathbb{E}\left[\Phi(\mathrm{Frag}_\infty(\xi))\right] = \frac{1}{\mathbb{E}\left[\sum_{j=1}^{\infty} a_j\right]} \mathbb{E}\left[\sum_{i=1}^{\infty} a_i \Phi\left(\frac{a_i \eta}{\sum_{j=1}^{\infty} a_j}, \frac{(a_l)_{l \neq i}}{\sum_{j=1}^{\infty} a_j}\right)\right].$$

By the Palm formula, this is equal to

$$\begin{aligned} & \frac{1}{\theta}\mathbb{E}\int_0^\infty x\Phi\left(\frac{x\eta}{x + \sum_{j=1}^{\infty} a_j}, \frac{(a_l)_{l=1}^{\infty}}{x + \sum_{j=1}^{\infty} a_j}\right)\theta x^{-1}\mathrm{e}^{-x}dx \\ = \; & \mathbb{E}\left[\Phi\left(\frac{a'\eta}{a' + \sum_{j=1}^{\infty} a_j}, \frac{(a_l)_{l=1}^{\infty}}{a' + \sum_{j=1}^{\infty} a_j}\right)\right] \end{aligned}$$

where $a' \sim \mathrm{Exp}(1)$, independently of M and η. But then $a'\eta$ has the distribution of the atoms of a Poisson random measure with intensity $x^{-1}\mathrm{e}^{-x}dx$ arranged in decreasing order and so we see that taking these atoms together with those of M, we obtain a Poisson random measure of intensity $(\theta+1)x^{-1}\mathrm{e}^{-1}dx$. Hence, $\mathrm{Frag}_\infty(\xi)$ has the law $\mathrm{PD}(\theta+1)$.

2.3 Dual fragmentation and coagulation chains

The dual fragmentation and coagulation operators that were defined in the preceding section incite us to introduce Markov fragmentation and coagulation chains in duality by time-reversal. Specifically, we consider for each integer $k \geq 1$ a chain

$$X^{(k)}(0), X^{(k)}(1), X^{(k)}(2), \dots,$$

where $X^{(k)}(n)$ is a random variable with values in Δ_{nk} (in particular $X^{(k)}(0) = 1$), and the conditional distribution of $X^{(k)}(n+1)$ given $X^{(k)}(n) = x$ is the law of $\mathrm{Frag}_k(x)$. We deduce from Proposition 1 by induction that for each n, $X^{(k)}(n)$ has the distribution $\mathrm{Dir}_{nk}(1/k)$. The time-reversed coagulation chain

$$\dots, X^{(k)}(n+1), X^{(k)}(n), \dots, X^{(k)}(1), X^{(k)}(0)$$

is also Markov; more precisely, the conditional distribution of $X^{(k)}(n)$ given $X^{(k)}(n+1) = x$ is the law of $\mathrm{Coag}_k(x)$. Note that for $k = 1$, this has the distribution of the jump chain of Kingman's coalescent [12].

Analogously, for $k = \infty$, we can define a Markov fragmentation chain on Δ_∞,

$$X^{(\infty)}(0), X^{(\infty)}(1), X^{(\infty)}(2), \dots,$$

such that the conditional distribution of $X^{(\infty)}(n+1)$ given $X^{(\infty)}(n) = x$ is the law of $\mathrm{Frag}_\infty(x)$. We deduce by induction from Proposition 2 that for every $\theta > 0$, if the distribution of the initial state $X^{(\infty)}(0)$ is $\mathrm{PD}(\theta)$ then, for every integer n, $X^{(\infty)}(n)$ has the distribution $\mathrm{PD}(\theta+n)$. Moreover, in this case, the time-reversed coagulation chain

$$\ldots, X^{(\infty)}(n+1), X^{(\infty)}(n), \ldots, X^{(\infty)}(1), X^{(\infty)}(0)$$

is also Markov; more precisely, the conditional distribution of $X^{(\infty)}(n)$ given $X^{(\infty)}(n+1) = x$ is the law of $\mathrm{Coag}_{1/(n+1+\theta)}(x)$.

Remarks. (a) Recall that the parameter θ can be recovered from a sample ξ of a $\mathrm{PD}(\theta)$ random variable as follows:

$$\theta = \lim_{\varepsilon\to 0+} \frac{1}{\log 1/\varepsilon} \max\{n : \xi_n > \varepsilon\} .$$

This shows that the above description for the reversed coagulation chain is indeed Markovian.

(b) There is simple representation for the $k=\infty$ fragmentation chain in terms of compound bridges with exchangeable increments which is inspired by [5]. Let $U_0, U_1, \ldots$ be a sequence of independent uniform variables on $[0,1]$. For each n, we consider the elementary bridge $b_n : [0,1] \to [0,1]$ defined by

$$b_n(t) = \frac{n}{n+1}t + \frac{1}{n+1}\mathbf{1}_{\{t>U_n\}}, \qquad t \in [0,1].$$

Then is is easy to check that for every $n \in \mathbf{N}$, the sequence $b_n \circ b_{n+1} \circ \cdots \circ b_{n+i}$ converges pointwise almost surely as $i \to \infty$ to a bridge with exchangeable increments B_n which has no drift and infinitely many jumps a.s. If we write $\beta_n \in \Delta_n$ for the sequence of the sizes of the jumps of B_n ranked the decreasing order, then the chain $(\beta_n, n \in \mathbf{N})$ has the same law as $X^{(\infty)}$. We refer to [5] for the necessary technical background.

3 The genealogy of Yule processes

We shall now show that the dual fragmentation and coagulation chains which we introduced in the preceding section are naturally connected to the genealogy of Yule processes.

3.1 Discrete setting

For every integer $k \geq 1$, we write $Y^{(k)} = \left(Y_t^{(k)}, t \geq 0\right)$ for the Yule process started from $Y_0^{(k)} = 1$: $Y_t^{(k)}$ gives the number of individuals alive a time t in a branching process in which each individual lives for an exponential time of parameter 1 and gives birth at its death to $k+1$ children, which then evolve independently of one another according to the same rules as their parent. We agree to label each child of an individual by an integer in $\{1, \ldots, k+1\}$, which allows us to order individuals at any generation in a consistent way: given two distinct individuals, we may consider their most recent common ancestor. Plainly, two different children of this ancestor are ancestors of exactly one of these two individuals, and the labelling of the children of the most recent common ancestor induces the order of the individuals.

Lemma 3. *The process $\left(\exp(-kt)Y_t^{(k)}, t \geq 0\right)$ is a uniformly integrable martingale and its limit $W^{(k)}$ has the* $\mathrm{Gamma}(1/k, 1/k)$ *distribution.*

Proof: A similar limit result is stated in Athreya & Ney [1] on page 130; however, the limiting distribution given there is incorrect and so we shall provide here a detailed proof. The martingale property is classical, so we focus on the distribution of the limit $W^{(k)}$. Define

$$\Phi_t(s) := \mathbf{E}\left(s^{Y_t^{(k)}}\right).$$

The backward equation implies that

$$\frac{\partial}{\partial t}\Phi_t(s) \;=\; \Phi_t^{k+1}(s) - \Phi_t(s) \quad , \quad \Phi_0(s) = s\,.$$

This equation has solution

$$\Phi_t(s) \;=\; s\mathrm{e}^{-t}\left(1-\left(1-\mathrm{e}^{-kt}\right)s^k\right)^{-1/k}.$$

Hence, for $\theta < 0$,

$$\begin{aligned}\mathbf{E}\left(\exp\left(\theta\mathrm{e}^{-kt}Y_t^{(k)}\right)\right) &= \exp\left(\theta\mathrm{e}^{-kt}\right)\mathrm{e}^{-t}\left(1-\left(1-\mathrm{e}^{-kt}\right)\exp\left(\theta k\mathrm{e}^{-kt}\right)\right)^{-1/k}\\ &= \left[\mathrm{e}^{kt}\exp\left(-\theta k\mathrm{e}^{-kt}\right)-\mathrm{e}^{kt}+1\right]^{-1/k},\end{aligned}$$

and when $t\to\infty$, this quantity converges to

$$(1-k\theta)^{-1/k} \;=\; \left(\frac{1/k}{1/k-\theta}\right)^{1/k},$$

which is the moment generating function of a gamma random variable with parameters $(1/k, 1/k)$. □

We think of $W^{(k)}$ as the size of the terminal population. For every $t \ge 0$, by application of the branching property at time t, we may decompose the terminal population into sub-populations having the same ancestor at time t. Specifically,

$$W^{(k)} \;=\; \sum_{i=1}^{Y_t^{(k)}} W_i^{(k)}(t)\,,$$

where $W_i^{(k)}(t)$ is the size of the terminal sub-population descending from the ith individual in the population at time t. Observe that conditionally on $Y_t^{(k)}$, the variables $W_i^{(k)}(t)$ are independent and all have the same law as $\mathrm{e}^{-t}W^{(k)}$.

Finally, we define the genealogical process $G^{(k)} = \left(G^{(k)}(t), t\ge 0\right)$ associated to $Y^{(k)}$ by

$$G^{(k)}(t) \;=\; \left(W_1^{(k)}(t),\ldots,W_{Y_t^{(k)}}^{(k)}(t)\right).$$

The genealogical structure of the Yule process can be described in terms of the fragmentation chain $X^{(k)}$ of Section 2.3 as follows.

Theorem 4. *Let $N = (N_t, t\ge 0)$ be a standard Poisson process which is independent of the chain $X^{(k)}$. Then for each $w > 0$, the compound chain*

$$\left(wX^{(k)}(N_{wt}), t\ge 0\right)$$

has the same law as the time-changed process

$$\left(G^{(k)}\left(\frac{1}{k}\log(1+kt)\right), t\geq 0\right)$$

conditioned on $W^{(k)} = w$.

Remark. Theorem I of Kendall [11] states that given $W^{(1)}$, $\left(Y^{(1)}_{\log(1+t/W^{(1)})}, t\geq 0\right)$ is a Poisson process with unit parameter. This is clearly an aspect of Theorem 4. Moreover, on page 130 of Athreya & Ney [1], it is suggested that no generalization of Kendall's result to a more general continuous-time Markov branching process is known; Theorem 4 constitutes a small such generalization.

Proof: Set $\tau(t) := \frac{1}{k}\log(1+kt)$ and let T be the time of the first birth in the Yule process, which is also the time of the first dislocation of $G^{(k)}$. The $k+1$ fragments of $G^{(k)}(T)$ can be written as $\mathrm{e}^{-kT}Z_1, \ldots, \mathrm{e}^{-kT}Z_{k+1}$ where, by the branching property, $Z_1, \ldots, Z_{k+1}$ are i.i.d. Gamma$(1/k, 1/k)$ random variables, independent of T which is Exp(1). Define a change of variables by

$$\begin{aligned}
&S = \tau^{-1}(T) = (\mathrm{e}^{kT}-1)/k \\
&U_1 = \mathrm{e}^{-kT}Z_1, \quad \ldots, \quad U_k = \mathrm{e}^{-kT}Z_k, \quad W = \mathrm{e}^{-kT}(Z_1+\cdots+Z_{k+1}).
\end{aligned}$$

It is straightforward to see that the joint density of $(T, Z_1, \ldots, Z_{k+1})$ is

$$\begin{aligned}
&f(t, z_1, \ldots, z_{k+1}) \\
&= \mathrm{e}^{-t}\Gamma(1/k)^{-(k+1)}(1/k)^{(k+1)/k}(z_1z_2\ldots z_{k+1})^{-(k-1)/k}\exp(-(z_1+\cdots+z_{k+1})/k)
\end{aligned}$$

and so the joint density of $(S, U_1, \ldots, U_k, W)$ is

$$\begin{aligned}
g(s, u_1, \ldots, u_k, w) = &\, w\mathrm{e}^{-ws}\cdot(1/k)\Gamma(1/k)^{-k}w^{-1/k}(u_1u_2\ldots u_k(w-u_1-\cdots-u_k))^{1/k-1} \\
&\cdot(1/k)^{1/k}\Gamma(1/k)^{-1}w^{-(k-1)/k}\exp(-w/k).
\end{aligned}$$

Hence, $W\sim$ Gamma$(1/k, 1/k)$ (as we already knew) and, conditional on $W = w$, we have $S\sim$ Exp(w) and $(U_1, U_2, \ldots, U_k, W-U_1-\cdots-U_k)\sim w\mathrm{Dir}_k(1/k)$ independently of S. Thus, the first dislocation has the correct dynamics. But by the branching property, subsequent dislocations are independent for different sub-populations and the total rate of fragmentation is always w. Hence result. $\square$

In the terminology of [2], Theorem 4 states that the time-changed genealogical process $G^{(k)}\circ\tau$ is a self-similar fragmentation with index 1, dislocation law $\mathrm{Dir}_k(1/k)$ and erosion coefficient 0. It may be interesting to observe that in the special case $k = 1$, this result can also be derived as follows.

Consider a real Brownian motion B started from 1 and killed when it reaches 0 (at time $T_0 = \inf\{t\geq 0 : B_t = 0\}$). For every $u\in[0,1[$, let $\widetilde{Y}_u$ denote the number of excursions of B away from 1 which go below level u. Then $(Y^{(1)}_{-\log(1-u)})_{0\leq u<1}$ is a version of $(\widetilde{Y}_u)_{0\leq u<1}$. To see this, let us consider the evolution of $\widetilde{Y}$. Firstly, $\widetilde{Y}_0 = 1$, corresponding to the single excursion below 1 which reaches 0. Let $D = \sup\{t < T_0 : B_t = 1\}$, the starting time of the final excursion which hits 0, let $U = \inf_{0\leq t\leq D} B_t$ be the level reached by the deepest excursion below 1 before D and let T_U be the time at which it is reached. Then, by Williams' path decomposition theorem (Theorem VII.4.9 of Revuz and Yor [19]), U is distributed uniformly on $[0,1[$ and, conditional on U, $(B_t)_{0\le t<T_U}$ is a Brownian motion started at 1 and stopped when it first hits level U. By symmetry, $(B_{D-t})_{0\le t<D-T_U}$ is another independent Brownian motion started at 1 and stopped when it first hits level U. Thus, $\widetilde{Y}_u$ is equal to 1 on $[0,U[$, $\widetilde{Y}_U=2$ and $(\widetilde{Y}_{U+v})_{0\le v<1-U}$ evolves as the sum of two independent processes which are the same as $\widetilde{Y}$ except that the times until the first jumps are now uniform on $[0,U[$ rather than on $[0,1[$. (This is Theorem 8 of Le Gall [15], repeated here for completeness.) Time-changing $Y^{(1)}$ with $u\to-\log(1-u)$ means that its exponential inter-jump times become uniform and so we do, indeed, have $(Y^{(1)}_{-\log(1-u)})_{0\le u<1}\stackrel{d}{=}(\widetilde{Y}_u)_{0\le u<1}$.

A more elegant way of expressing the preceding argument is to say that the Brownian path encodes a continuous-state branching process with quadratic branching mechanism. The local time at level 1, $L^1_{T_0}$, satisfies

$$L^1_{T_0}=\lim_{u\to 1-}2(1-u)\widetilde{Y}_u.$$

In this context, $\frac{1}{2}L^1_{T_0}$ corresponds to the size of the population at time 1 in the continuous-state branching process generated by a single ancestor conditioned to have descendents up to time 1. The so-called reduced tree associated with the population at time 1 is described up to the deterministic time-change $u\to-\log(1-u)$ by the Yule process $Y^{(1)}$. See, for instance, Section 2.7 in Duquesne and Le Gall [8], and Fleischmann and Siegmund-Schultze [9]. Note that the well-known fact that $\frac{1}{2}L^1_{T_0}$ has an exponential distribution with mean 1 (Proposition VI.4.6 of Revuz and Yor [19]) gives another derivation of the limiting distribution in Lemma 3, since

$$W^{(1)}=\lim_{t\to\infty}e^{-t}Y^{(1)}_t=\lim_{u\to 1-}(1-u)Y^{(1)}_{-\log(1-u)}\stackrel{d}{=}\frac{1}{2}L^1_{T_0}.$$

It is known from excursion theory that in the scale of the local time at level 1, the rate of excursions of B away from 1 which reach level $u\in\,]0,1[$ but do not exceed $u-du$ is $(1-u)^{-2}du$. Note that the map $s\to 1-\frac{1}{1+s}$ from $\mathbf{R}_+$ to $[0,1[$ has inverse $u\to\frac{1}{1-u}-1$ and, thus, transforms Lebesgue measure on $\mathbf{R}_+$ into the measure $(1-u)^{-2}du$ on $[0,1[$. Suppose that we split the local time at level 1 according to the occurrence of excursions exceeding level u, so that we obtain the sequence

$$\widetilde{W}(u)\;=\;\left(\widetilde{W}_1(u),\ldots,\widetilde{W}_{\widetilde{Y}_u}(u)\right),$$

where $\widetilde{W}(u)$ is the sequence of the increments of the local time at level 1 on the maximal time intervals such that at the beginning and end of each interval B is at 1 and during the interval remains above level u. Then it follows easily that the time-changed process $\left(\widetilde{W}\left(1-\frac{1}{1+s}\right),s\ge 0\right)$ is a fragmentation in which each mass, say x, splits at rate x into xU and $x(1-U)$ where U is uniform. In other words, conditionally on $\frac{1}{2}L^1_{T_0}=w$, the process $\left(\widetilde{W}\left(1-\frac{1}{1+s}\right),s\ge 0\right)$ is distributed as the compound fragmentation chain $\left(wX^{(1)}(N_{ws}),s\ge 0\right)$, where N is an independent standard Poisson process.

Finally, the composition of the two time-changes which appear in this analysis yields

$$s\to-\log\left(1-\left(1-\frac{1}{1+s}\right)\right)=\log(1+s)\,,\qquad s\in\mathbf{R}_+\,,$$

and so we recover Theorem 4 in the special case $k = 1$. Unfortunately, it does not seem that there are similar interpretations for $k \geq 2$.

Corollary 5. *We have that*

$$\left(\frac{1}{W^{(k)}} G^{(k)} \left(\frac{1}{k} \log\left(1 + ke^{-t}/W^{(k)}\right)\right), t \in \mathbf{R}\right)$$

is a time-homogeneous Markov coalescent process which is independent of $W^{(k)}$. For any $n \geq 1$, given that it is in state $x \in \Delta_{nk}$, it waits an exponential time of parameter n and then jumps to a variable distributed as $\mathrm{Coag}_k(x)$, independently of the exponential time.

Note that the case $k = 1$ of this result gives a variation of Kingman's coalescent. The jump-chains are identical, as we have already noted, but here the rate of coalescence of two blocks depends on the total number of blocks present, whereas in Kingman's coalescent it does not.

Proof: Firstly, we note that by Theorem 4,

$$\left(\frac{1}{W^{(k)}} G^{(k)} \left(\frac{1}{k} \log(1 + ke^{-t}/W^{(k)})\right), t \in \mathbf{R}\right)$$

has the same law as

$$\left(X^{(k)}(N_{e^{-t}}), t \in \mathbf{R}\right)$$

and so we will work with the latter process instead. The $k = 1$ case is essentially treated in [3] and the proof proceeds in the same way here. The jump chain clearly behaves in the correct manner and so it remains to check that the inter-jump times are as claimed. Let $0 \leq T_1 \leq T_2 \leq \ldots$ be the jump times of $(N_t)_{t\geq 0}$. Then the first instant that $X^{(k)}(N_{e^{-t}})$ has exactly $nk + 1$ terms is

$$\inf\{t \in \mathbf{R} : N_{e^{-t}} = n\} = -\log T_{n+1}.$$

The sequence of inter-jump times is

$$\ldots, \log T_{n+1} - \log T_n, \log T_n - \log T_{n-1}, \ldots, \log T_2 - \log T_1$$

and it is easily shown that this is a sequence of independent exponential random variables with parameters

$$\ldots, n, n-1, \ldots, 1$$

respectively. □

3.2 Continuous setting

Continuous-state branching processes (or CSBP's) were introduced by Lamperti [13, 14] as limits of rescaled branching processes. Typically, a CSBP is a time-homogeneous Markov process with values in $\mathbf{R}_+$,

$$Z = (Z(t, a),\ t \geq 0 \text{ and } a \geq 0),$$

(where the parameter t refers to time and the parameter a to the starting point i.e. $Z(0, a) = a$ a.s.) which fulfils the branching property: the path-valued process $(Z(\cdot, x), x \geq$

0) has independent and stationary increments. In particular, if $\widetilde{Z}(\cdot, y)$ is an independent copy of $Z(\cdot, y)$, then $Z(\cdot, x) + \widetilde{Z}(\cdot, y)$ has the law of $Z(\cdot, x+y)$. There is a simple relation connecting CSBP's and Bochner's subordination for subordinators which enables us to define their genealogy; we refer the interested reader to [4] for heuristics, detailed arguments etc.

We call a *continuous state Yule process* a CSBP

$$Y = (Y(t,a),\, t \geq 0 \text{ and } a \geq 0)\,,$$

which evolves as follows: for each $a > 0$, the process $Y(\cdot, a)$ waits an exponential time with parameter a and then jumps to $a+1$. It then evolves independently as if it had been started in state $a+1$. In terms of the genealogy, the sub-population of size 1 which is born at a jump time has a parent which is chosen uniformly at random from the population present before the jump. Note that this genealogy is easy to describe in a consistent manner for different values a of the starting population.

It is immediate that for an integer starting point $a \in \mathbf{N}$, the process $(Y(t,a), t \geq 0)$ is a Yule process $Y^{(1)}$ with 2 offspring, as considered in the preceding section. However, we stress that its genealogy is not the same as that of $Y^{(1)}$, as we are dealing with a continuous population in the first case and a discrete population in the second.

We have the following analogue of Lemma 3:

Lemma 6. *For every $a \geq 0$, the process $\left(\mathrm{e}^{-t} Y(t,a), t \geq 0\right)$ is a uniformly integrable martingale. Its limit, say $\gamma(a)$, viewed as a process in the variable a, has the same finite dimensional laws as a standard gamma process.*

Proof: For $a = 1$, we see from Lemma 3 and the identity in distribution $Y(\cdot, 1) \stackrel{\mathcal{L}}{=} Y^{(1)}(\cdot)$ that $\left(\mathrm{e}^{-t} Y(t,1), t \geq 0\right)$ is a uniformly integrable martingale and that its limit has the standard exponential distribution. The proof is easily completed by an appeal to the branching property. □

Just as in the preceding section, we think of $\gamma(a)$ as the size of the terminal population when the initial population has size a. We can express $\gamma(a)$ as

$$\gamma(a) = \sum_{b \leq a} \delta_b\,,$$

where $\delta := (\delta_b, b \geq 0)$ is the jump process of γ, which corresponds to decomposing the terminal population into sub-populations having the same ancestor at the initial time. We write $G(0,a)$ for the sequence of the jumps of γ on $[0,a]$, ranked in decreasing order, and we deduce from Lemma 6 that conditionally on $\gamma(a) = g$, $G(0,a)/g$ has distribution $\mathrm{PD}(a)$.

More generally, by the branching property, we can decompose the terminal population into sub-populations having the same ancestor at any given time t. This gives

$$\gamma(a) = \sum_{b \leq Y(t,a)} \mathrm{e}^{-t} \delta_b^{(t)}\,,$$

where $\delta^{(t)} := (\delta_b^{(t)}, b \geq 0)$ is the jump process of a standard gamma process $\gamma^{(t)}$ which is independent of the Yule process up to time t, $(Y(s,c),\, s \in [0,t] \text{ and } c \geq 0)$. This enables us to define for each $a > 0$ the genealogical process associated to a Yule process $Y(\cdot, a)$,

$$G(\cdot, a) = (G(t,a), t \geq 0)\;,$$

where $e^t G(t,a)$ is the ranked sequence of the sizes of the jumps of the subordinator $\gamma^{(t)}$ on the interval $[0, Y(t,a)]$.

An easy variation of the arguments for the proof of Theorem 4 shows that the genealogical structure of the Yule process can be described in terms of the fragmentation chain $X^{(\infty)}$ of Section 2.3 as follows.

Theorem 7. *Fix $a, g > 0$ and let the chain $X^{(\infty)}$ have initial distribution* PD(a). *Introduce a standard Poisson process, $N = (N_t, t \geq 0)$, which is independent of the chain $X^{(\infty)}$. Then the compound chain*

$$\left(gX^{(\infty)}(N_{gt}), t \geq 0\right)$$

has the same law as the time-changed process

$$\left(G\left(\log(1+t), a\right), t \geq 0\right)$$

conditioned on $\gamma(a) = g$.

Likewise, the analogue of Corollary 5 is as follows.

Corollary 8. *Fix $a > 0$. Then*

$$\left(\frac{1}{\gamma(a)} G\left(\log(1 + e^{-t}/\gamma(a)), a\right), t \in \mathbf{R}\right)$$

is a time-homogeneous Markov coalescent process which is independent of $\gamma(a)$. Suppose that it is in state $x \in \Delta_\infty$ and recall Remark (a) of Section 2.3. Then if

$$\lim_{\epsilon \to 0+} \frac{1}{\log 1/\epsilon} \max\{i : x_i > \epsilon\} = n + a,$$

the process waits an exponential time of parameter n and then jumps to a variable distributed as $\mathrm{Coag}_{1/(n+a)}(x)$, *independently of the exponential time.*

References

[1] Athreya, K.B. and Ney, P.E. (1972). *Branching processes.* Springer-Verlag, Berlin-Heidelberg-New York

[2] Bertoin, J. (2002). Self-similar fragmentations. *Ann. Inst. Henri Poincaré* **38**, 319-340

[3] Bertoin, J. (2003). Random covering of an interval and a variation of Kingman's coalescent. To appear in *Random Structures Algorithms.* Also available as Preprint PMA-794 at `http://www.proba.jussieu.fr/mathdoc/preprints/index.html`

[4] Bertoin, J. and Le Gall, J.-F. (2000). The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching processes. *Probab. Theory Relat. Fields* **117**, 249-266

[5] Bertoin, J. and Le Gall, J.-F. (2003). Stochastic flows associated to coalescent processes. *Probab. Theory Relat. Fields* **126**, 261-288

[6] Bertoin, J. and Pitman, J. (2000). Two coalescents derived from the ranges of stable subordinators. *Elect. J. Probab.* **5**, 1-17. Available via `http://www.math.u-psud.fr/~ejpecp/ejp5contents.html`

[7] Bolthausen, E. and Sznitman, A.S. (1998). On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Physics* **197**, 247-276

[8] Duquesne, T. and Le Gall, J.-F. (2002). *Random trees, Lévy processes and spatial branching processes.* Astérisque **281**

[9] Fleischmann, K. and Siegmund-Schultze, R. (1977) The structure of reduced critical Galton-Watson processes. *Math. Nachr.* **79**, 233-241

[10] Kallenberg, O. (1973). Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrsch. verw. Gebiete* **27**, 23-36

[11] Kendall, D. G. (1966). Branching processes since 1873. *J. London Math. Soc.* **41**, 385-406

[12] Kingman, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13**, 235-248

[13] Lamperti, J. (1967). The limit of a sequence of branching processes. *Z. Wahrsch. verw. Gebiete* **7**, 271-288

[14] Lamperti, J. (1967). Continuous-state branching processes. *Bull. Amer. Math. Soc.* **73**, 382-386

[15] Le Gall, J.-F. (1989). Marches aléatoires, mouvement brownien et processus de branchement. *Séminaire de Probabilités XXIII*, Lecture Notes in Math., 1372, Springer, Berlin, 258-274

[16] Pitman, J. (1999). Coalescents with multiple collisions. *Ann. Probab.* **27**, 1870-1902

[17] Pitman, J. (2002). *Combinatorial Stochastic Processes.* Lecture notes for the St Flour summer school. To appear. Available at `http://stat-www.berkeley.edu/users/pitman/621.ps.Z`

[18] Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855-900

[19] Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*, Third edition. Springer-Verlag, Berlin