Total internal and external lengths of the Bolthausen-Sznitman coalescent

Juan Carlos Pardo

CIMAT, Mexico

joint work with A. Siri-Jégousse and G. Kersting
Bolthausen-Sznitman coalescent

The Bolthausen-Sznitman (BS) coalescent ($\Pi_t, t \geq 0$), is a continuous time Markov chain with values in the set of partitions of $\mathbb{N}$, starting with an infinite number of blocks/individuals.
Bolthausen-Sznitman coalescent

The Bolthausen-Sznitman (BS) coalescent ($\Pi_t, t \geq 0$), is a continuous time Markov chain with values in the set of partitions of $\mathbb{N}$, starting with an infinite number of blocks/individuals.

It is an example of an exchangeable coalescent with multiple collisions.
Total internal and external lengths of the Bolthausen-Sznitman coalescent

Introduction

Bolthausen-Sznitman coalescent

The Bolthausen-Sznitman (BS) coalescent \((\Pi_t, t \geq 0)\), is a continuous time Markov chain with values in the set of partitions of \(\mathbb{N}\), starting with an infinite number of blocks/individuals.

It is an example of an exchangeable coalescent with multiple collisions.

It was first introduced in physics, in order to study spin glasses but it has also been thought as a limiting genealogical model for evolving populations with selective killing at each generation.
Formal description
Introduction

Formal description

Let \( n \in \mathbb{N} \), then the restriction \((\Pi_t^{(n)}, t \geq 0)\) of \((\Pi_t, t \geq 0)\) to \([n] = \{1, \ldots, n\}\) is a Markov chain with values in \(\mathcal{P}_n\), the set of partitions of \([n]\), with the following dynamics:
Formal description

Let $n \in \mathbb{N}$, then the restriction $(\Pi_t^{(n)}, t \geq 0)$ of $(\Pi_t, t \geq 0)$ to $[n] = \{1, \ldots, n\}$ is a Markov chain with values in $\mathcal{P}_n$, the set of partitions of $[n]$, with the following dynamics:

Whenever $\Pi_t^{(n)}$ is a partition consisting of $b$ blocks, any particular $k$ of them merge into one block at rate

$$\lambda_{b,k} = \frac{(k-2)!(b-k)!}{(b-1)!},$$

so the next coalescence event occurs at total rate

$$\lambda_b = \sum_{k=2}^{b} \binom{b}{k} \lambda_{b,k} = b - 1.$$
Goal: determine the asymptotic behaviour of the total external length $E(n)$ of the BS coalescent restricted to $\mathcal{P}_n$, when $n \to \infty$, and relate it to its total length $L(n)$ (the sum of lengths of all external and internal branches).
Goal: determine the asymptotic behaviour of the total external length $E^{(n)}$ of the BS coalescent restricted to $P_n$, when $n \to \infty$, and relate it to its total length $L^{(n)}$ (the sum of lengths of all external and internal branches).

In the case of coalescents without proper frequencies, M"ohle (2010) proved that after a suitable scaling the asymptotic distributions of $E^{(n)}$ and $L^{(n)}$ are the same.
Goal: determine the asymptotic behaviour of the total external length $E^{(n)}$ of the BS coalescent restricted to $P_n$, when $n \to \infty$, and relate it to its total length $L^{(n)}$ (the sum of lengths of all external and internal branches).

In the case of coalescents without proper frequencies, Möhle (2010) proved that after a suitable scaling the asymptotic distributions of $E^{(n)}$ and $L^{(n)}$ are the same.

According to Drmota et al. (2007) the asymptotic behaviour of the total length of the BS coalescent is given as follows

$$\frac{(\log n)^2}{n} L^{(n)} - \log n - \log \log n \xrightarrow[n \to \infty]{} Z,$$  \hspace{1cm} (1)$$

where $Z$ is a strictly stable r.v. with index 1, i.e. its characteristic exponent satisfies

$$\Psi(\theta) = -\log E\left[e^{i\theta Z}\right] = \frac{\pi}{2} |\theta| - i\theta \log |\theta|, \quad \theta \in \mathbb{R}.$$
Recently, Dhersin and Möhle (2013) showed

$$\frac{E^{(n)}}{L^{(n)}} \xrightarrow{\mathbb{P}} 1.$$
Recently, Dhersin and Möhle (2013) showed

\[ \frac{E^{(n)}}{L^{(n)}} \xrightarrow{\mathbb{P}} 1. \]

Thus one might guess that \( E^{(n)} \) satisfies the same asymptotic relation with the same scaling.
Internal and external lengths.

Let us consider \((\Pi_t^{(n)}, t \geq 0)\). We denote by
Internal and external lengths.

Let us consider \((\Pi_t^{(n)}, t \geq 0)\). We denote by

- \(U_k^{(n)} = \) the size of the \(k\)-th jump,
Internal and external lengths.

Let us consider \((\Pi_{t}^{(n)}, t \geq 0)\). We denote by

- \(U_{k}^{(n)}\) = the size of the \(k\)-th jump,
- \(X_{k}^{(n)}\) = the number of blocks after \(k\) coalescence events.
Internal and external lengths.

Let us consider \((\Pi_t^{(n)}, t \geq 0)\). We denote by

- \(U_k^{(n)}\) = the size of the \(k\)-th jump,
- \(X_k^{(n)}\) = the number of blocks after \(k\) coalescence events.

Observe that \(X_0^{(n)} = n\) and \(X_k^{(n)} = X_{k-1}^{(n)} - U_k^{(n)} = n - \sum_{i=1}^{k} U_i^{(n)}\).
Internal and external lengths.

Let us consider \( \Pi_t^{(n)}, t \geq 0 \). We denote by

- \( U_k^{(n)} \) = the size of the \( k \)-th jump,
- \( X_k^{(n)} \) = the number of blocks after \( k \) coalescence events.

Observe that \( X_0^{(n)} = n \) and \( X_k^{(n)} = X_{k-1}^{(n)} - U_k^{(n)} = n - \sum_{i=1}^{k} U_i^{(n)} \).

Let \( \tau^{(n)} \) be the number of coalescence events. More precisely

\[
\tau^{(n)} = \inf \left\{ k, X_k^{(n)} = 1 \right\}.
\]
**Internal and external lengths.**

Let us consider $(\Pi_t^{(n)}, t \geq 0)$. We denote by

- $U_k^{(n)}$ the size of the $k$-th jump,
- $X_k^{(n)}$ the number of blocks after $k$ coalescence events.

Observe that $X_0^{(n)} = n$ and $X_k^{(n)} = X_{k-1}^{(n)} - U_k^{(n)} = n - \sum_{i=1}^{k} U_i^{(n)}$.

Let $\tau^{(n)}$ be the number of coalescence events. More precisely

$$\tau^{(n)} = \inf \left\{ k, X_k^{(n)} = 1 \right\}.$$

According to Iksanov and Möhle (2007), $\tau^{(n)}$ satisfies the following asymptotic behaviour

$$\frac{(\log n)^2}{n} \tau^{(n)} - \log n - \log \log n \xrightarrow{n \to \infty} Z.$$  \hspace{1cm} (2)
Let $Y_k^{(n)}$ be the number of internal branches after $k$ coalescence events. Note that $Y_0^{(n)} = 0$. 
Let $Y^{(n)}_k$ be the number of internal branches after $k$ coalescence events. Note that $Y^{(n)}_0 = 0$.

We denote by $I^{(n)}$ for the total intern length.
Let $Y_k^{(n)}$ be the number of internal branches after $k$ coalescence events. Note that $Y_0^{(n)} = 0$.

We denote by $I^{(n)}$ for the total intern length.

Let $(e_k, k \geq 1)$ be a sequence of i.i.d. standard exponential r.v. which are independent of $X^{(n)}$ and $Y^{(n)}$, thus

$$I^{(n)} \overset{d}{=} \frac{\tau^{(n)} - 1}{\sum_{k=1}^{\tau^{(n)} - 1} \frac{Y_k^{(n)} e_k}{X_k^{(n)} - 1}}.$$
Let $Y_k^{(n)}$ be the number of internal branches after $k$ coalescence events. Note that $Y_0^{(n)} = 0$.

We denote by $I^{(n)}$ for the total intern length.

Let $(e_k, k \geq 1)$ be a sequence of i.i.d. standard exponential r.v. which are independent of $X^{(n)}$ and $Y^{(n)}$, thus

$$I^{(n)} \overset{d}{=} \tau^{(n)} - 1 \sum_{k=1}^{\tau^{(n)}-1} Y_k^{(n)} \frac{e_k}{X_k^{(n)} - 1}.$$ 

**Theorem**

*For the total internal length of the Bolthausen-Sznitman coalescent, we have*

$$\frac{(\log n)^2}{n} I^{(n)} \overset{\mathbb{P}}{\longrightarrow} 1.$$
Since $L^{(n)} = I^{(n)} + E^{(n)}$, we deduce the asymptotic distribution of the total external length $E^{(n)}$. 
Since $L^{(n)} = I^{(n)} + E^{(n)}$, we deduce the asymptotic distribution of the total external length $E^{(n)}$.

**Corollary**

*For the total external length of the Bolthausen-Sznitman coalescent, we have*

$$\frac{(\log n)^2}{n} E^{(n)} - \log n - \log \log n \xrightarrow{n \to \infty} Z - 1.$$
Since \( L^{(n)} = I^{(n)} + E^{(n)} \), we deduce the asymptotic distribution of the total external length \( E^{(n)} \).

**Corollary**

*For the total external length of the Bolthausen-Sznitman coalescent, we have*

\[
\frac{(\log n)^2}{n} E^{(n)} - \log n - \log \log n \xrightarrow{d} Z - 1.
\]

**Asymptotic behaviour:** in the \( Beta(2 - \alpha, \alpha) \)-coalescent with \( 0 < \alpha < 2 \).
Since $L^{(n)} = I^{(n)} + E^{(n)}$, we deduce the asymptotic distribution of the total external length $E^{(n)}$.

**Corollary**

For the total external length of the Bolthausen-Sznitman coalescent, we have

$$\frac{(\log n)^2}{n} E^{(n)} - \log n - \log \log n \xrightarrow{d} Z - 1.$$  

Asymptotic behaviour: in the $Beta(2 - \alpha, \alpha)$-coalescent with $0 < \alpha < 2$.

(0, 1) Möhle (2010) proved that $E^{(n)}/n$ converges in law to a random variable defined in terms of a driftless subordinator depending on $\alpha$. 
Since $L^{(n)} = I^{(n)} + E^{(n)}$, we deduce the asymptotic distribution of the total external length $E^{(n)}$.

**Corollary**

*For the total external length of the Bolthausen-Sznitman coalescent, we have*

$$\frac{(\log n)^2}{n} E^{(n)} - \log n - \log \log n \xrightarrow{n \to \infty} Z - 1.$$  

**Asymptotic behaviour:** in the $Beta(2 - \alpha, \alpha)$-coalescent with $0 < \alpha < 2$.

(0, 1) Möhle (2010) proved that $E^{(n)}/n$ converges in law to a random variable defined in terms of a driftless subordinator depending on $\alpha$.

(1, 2) Kersting et al. (2012) proved that $(E^{(n)} - c(\alpha)n^{2-\alpha})/n^{1/\alpha+1-\alpha}$ converges weakly to a stable r.v. of index $\alpha$. 
Since \( L^{(n)} = I^{(n)} + E^{(n)} \), we deduce the asymptotic distribution of the total external length \( E^{(n)} \).

**Corollary**

*For the total external length of the Bolthausen-Sznitman coalescent, we have*

\[
\frac{(\log n)^2}{n} E^{(n)} - \log n - \log \log n \xrightarrow{d} Z - 1.
\]

**Asymptotic behaviour:** in the \( \text{Beta}(2 - \alpha, \alpha) \)-coalescent with \( 0 < \alpha < 2 \).

\( (0, 1) \) Möhle (2010) proved that \( E^{(n)}/n \) converges in law to a random variable defined in terms of a driftless subordinator depending on \( \alpha \).

\( (1, 2) \) Kersting et al. (2012) proved that \( (E^{(n)} - c(\alpha)n^{2-\alpha})/n^{1/\alpha+1-\alpha} \) converges weakly to a stable r.v. of index \( \alpha \).

\( \alpha \to 2 \) In Kingman’s case a logarithmic correction appears and the limit law is normal (Janson and Kersting, 2011).
Idea of the proof.

We first define

$$ \tilde{I}(n) = \sum_{k=1}^{\tau(n) - 1} \frac{Y_k(n)}{X_k(n)} $$

and

$$ \hat{I}(n) = \sum_{k=1}^{\tau(n) - 1} \frac{\mathbb{E}[Y_k(n) \mid X(n)]}{X_k(n)} $$.
Idea of the proof.

We first define

\[ \tilde{I}(n) = \sum_{k=1}^{\tau(n)-1} \frac{Y_k(n)}{X_k(n)} \quad \text{and} \quad \hat{I}(n) = \sum_{k=1}^{\tau(n)-1} \frac{\mathbb{E}[Y_k(n)|X(n)]}{X_k(n)}. \]

We use a similar argument used by Kersting et al. (2012) which gives a recursive formula for \( \hat{I}(n) \).
Idea of the proof.

We first define

\[
\tilde{I}(n) = \sum_{k=1}^{\tau(n)-1} \frac{Y_k(n)}{X_k(n)} \quad \text{and} \quad \hat{I}(n) = \sum_{k=1}^{\tau(n)-1} \mathbb{E}\left[\frac{Y_k(n)}{X_k(n)} \mid X(n)\right].
\]

We use a similar argument used by Kersting et al. (2012) which gives a recursive formula for \(\hat{I}(n)\).

More precisely, let
Idea of the proof.

We first define

\[ \tilde{I}^{(n)} = \sum_{k=1}^{\tau^{(n)}-1} \frac{Y_k^{(n)}}{X_k^{(n)}} \quad \text{and} \quad \hat{I}^{(n)} = \sum_{k=1}^{\tau^{(n)}-1} \frac{\mathbb{E}[Y_k^{(n)} | X^{(n)}]}{X_k^{(n)}}. \]

We use a similar argument used by Kersting et al. (2012) which gives a recursive formula for \( \hat{I}^{(n)} \).

More precisely, let

- \( Z_k^{(n)} = \) number of external branches after \( k \) jumps.
Idea of the proof.

We first define

\[ \tilde{I}^{(n)} = \sum_{k=1}^{\tau^{(n)}-1} \frac{Y_k^{(n)}}{X_k^{(n)}} \quad \text{and} \quad \hat{I}^{(n)} = \sum_{k=1}^{\tau^{(n)}-1} \frac{\mathbb{E}[Y_k^{(n)} | X^{(n)}]}{X_k^{(n)}}. \]

We use a similar argument used by Kersting et al. (2012) which gives a recursive formula for \( \hat{I}^{(n)} \).

More precisely, let

- \( Z_k^{(n)} = \) number of external branches after \( k \) jumps.
- \( Z_{k-1}^{(n)} - Z_k^{(n)} = \) number of external branches which participate to the \( k \)-th coalescence event.
Idea of the proof.

We first define

$$\tilde{I}(n) = \sum_{k=1}^{\tau(n)-1} \frac{Y_k(n)}{X_k(n)}$$ and $$\hat{I}(n) = \sum_{k=1}^{\tau(n)-1} \frac{\mathbb{E}[Y_k(n) | X(n)]}{X_k(n)}.$$

We use a similar argument used by Kersting et al. (2012) which gives a recursive formula for $\hat{I}(n)$.

More precisely, let

- $Z_k^{(n)} = \text{number of external branches after} \ k \ \text{jumps}$.
- $Z_{k-1}^{(n)} - Z_k^{(n)} = \text{number of external branches which participate to the} \ k-\text{th coalescence event}$.

$$\mathcal{L}(Z_{k-1}^{(n)} - Z_k^{(n)} | X(n), Z_{k-1}^{(n)}) \sim \text{Hyp}(X_{k-1}^{(n)}, Z_{k-1}^{(n)}, 1 + U_k^{(n)})$$
Recall that $U_k^{(n)} = X_{k-1}^{(n)} - X_k^{(n)}$ denotes the size of the $k$-th jump of the block counting process.
Recall that $U_k(n) = X_{k-1}^{(n)} - X_k^{(n)}$ denotes the size of the $k$-th jump of the block counting process.

Then

$$ \mathbb{E} \left[ Z_k^{(n)} \bigg| X^{(n)}, Z_{k-1}^{(n)} \right] = Z_{k-1}^{(n)} \frac{X_k^{(n)} - 1}{X_{k-1}^{(n)}}, $$

and

$$ \mathbb{E} \left[ Z_k^{(n)} \bigg| X^{(n)} \right] = \mathbb{E} \left[ Z_{k-1}^{(n)} \bigg| X^{(n)} \right] \frac{X_k^{(n)} - 1}{X_{k-1}^{(n)}}.$$

Recall that $U_k^{(n)} = X_{k-1}^{(n)} - X_k^{(n)}$ denotes the size of the $k$-th jump of the block counting process.

Then

$$\mathbb{E}\left[ Z_k^{(n)} \bigg| X^{(n)}, Z_{k-1}^{(n)} \right] = Z_{k-1}^{(n)} \frac{X_k^{(n)} - 1}{X_{k-1}^{(n)}},$$

and

$$\mathbb{E}\left[ Z_k^{(n)} \bigg| X^{(n)} \right] = \mathbb{E}\left[ Z_{k-1}^{(n)} \bigg| X^{(n)} \right] \frac{X_k^{(n)} - 1}{X_{k-1}^{(n)}}.$$

Since $Y_k^{(n)} = X_k^{(n)} - Z_k^{(n)}$ it follows

$$\hat{I}^{(n)} = \sum_{k=1}^{\tau^{(n)}-1} \left( 1 - \prod_{i=1}^{k} \left( 1 - \frac{1}{X_i^{(n)}} \right) \right).$$
The identity from above allow us to get

$$\frac{(\log n)^2}{n} \hat{I}(n) \xrightarrow{\mathbb{P}} 1.$$
The identity from above allow us to get

$$\frac{(\log n)^2}{n} \hat{I}(n) \xrightarrow{P} 1.$$ \hfill n \to \infty

Finally the following two approximations give us the result

$$\frac{I(n) - \tilde{I}(n)}{\sqrt{n}}$$ is stochastically bounded.

$$\frac{\tilde{I}(n) - \hat{I}(n)}{\sqrt{n}}$$ is stochastically bounded.

All the asymptotics are based in a coupling argument introduced by Iksanov and Möhle (2007).
The identity from above allow us to get

$$\frac{\left(\log n\right)^2}{n} \hat{I}(n) \xrightarrow{\mathbb{P}} 1.$$ \hspace{1cm} n \to \infty

Finally the following two approximations give us the result

$$\frac{I(n) - \tilde{I}(n)}{\sqrt{n}} \text{ is stochastically bounded.}$$

$$\frac{\tilde{I}(n) - \hat{I}(n)}{\sqrt{n}} \text{ is stochastically bounded.}$$
The identity from above allow us to get

\[
\frac{(\log n)^2}{n} \hat{I}(n) \xrightarrow{\mathbb{P}} 1. \\
\]

Finally the following two approximations give us the result

\[
\frac{I(n) - \tilde{I}(n)}{\sqrt{n}} \text{ is stochastically bounded.}
\]

\[
\frac{\tilde{I}(n) - \hat{I}(n)}{\sqrt{n}} \text{ is stochastically bounded.}
\]

All the asymptotics are based in a coupling argument introduced by Iksanov and Möhle (2007).