ON ASYMPTOTICS OF EXCHANGEABLE COALESCENTS WITH MULTIPLE COLLISIONS

Martin Möhle

(joint work with Alex Iksanov and Alexander Gnedin)

CIRM, Marseille Luminy, May 26, 2009

Coalescent with multiple collisions (Pitman 1999, Sagitov 1999)

Fix $n \in \mathbb{N}$. Let Λ be a finite measure on $[0,1]$.

Definition. (Coalescent with multiple collisions, Λ-coalescent)

The n-coalescent with multiple collisions is a Markov process $\Pi_n = (\Pi_n(t))_{t \geq 0}$ with state space $\mathcal{P}_n$, the set of partitions of $\{1,\ldots,n\}$, and generator $Q = (q_{\pi\pi'})_{\pi,\pi' \in \mathcal{P}_n}$ with rates

$$\boxed{q_{\pi\pi'} = \int_{[0,1]} x^{k-2}(1-x)^{b-k}\,\Lambda(dx) =: \lambda_{b,k}}$$

for $\pi \prec_k \pi'$, $k \geq 2$, where

$b := |\pi| :=$ number of blocks of π

$\pi \prec_k \pi' :\Longleftrightarrow$ Exactly k blocks of π merge to form a single block of π'

Examples.

$\Lambda = \delta_0$ (Dirac measure in 0). $\Rightarrow$ Kingman coalescent (1982)

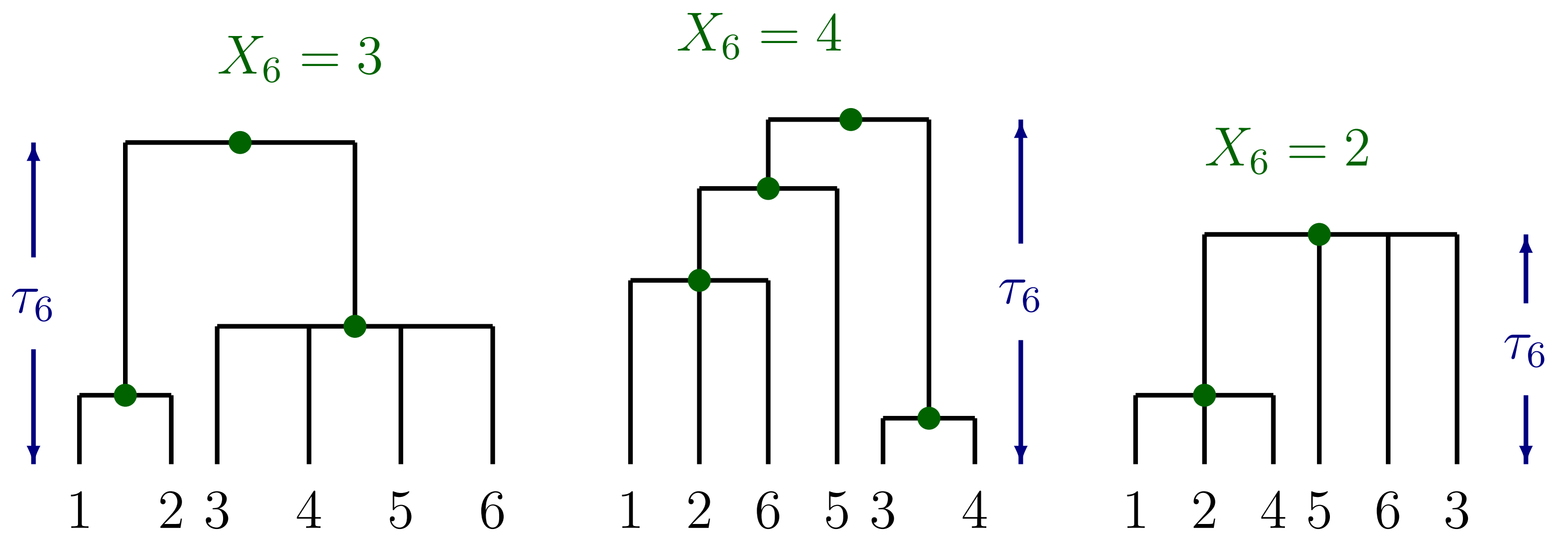
$\Lambda = U_{[0,1]}$ (uniform on $[0,1]$). $\Rightarrow$ Bolthausen-Sznitman coalescent (1998)

$\Lambda = \delta_1$ (Dirac measure in 1). $\Rightarrow$ star-shaped coalescent

Functionals of coalescent processes

$X_n :=$ number of collisions

$\tau_n := \inf\{t \geq 0 \,:\, |\Pi_n(t)| = 1\} \;=\;$ time back to MRCA $\;=\;$ absorption time



Graphical representation of three realisations of a 6-coalescent with multiple collisions

Example: Bolthausen-Sznitman coalescent

Proposition 1. (Asymptotics of X_n, Iksanov and M. 2007, Drmota, Iksanov, Rösler and M. 2007, 2009)

$$\boxed{Y_n := \frac{(\log n)^2}{n} X_n - \log n - \log\log n \;\xrightarrow{d}\; Y}$$

where Y is 1-stable with characteristic function $\boxed{\mathrm{E}(e^{i\lambda Y}) = e^{i\lambda \log|\lambda| - \frac{\pi}{2}|\lambda|}}$, $\lambda \in \mathbb{R}$.

Proposition 2. (Asymptotics of τ_n, Goldschmidt and Martin 2005, Freund and M. 2007)

$$\boxed{\tau_n - \log\log n \;\xrightarrow{d}\; \tau}$$

where τ is Gumbel distributed.

Remark. Analytic proofs are based on singularity analysis. Probabilistic proofs use relations to random recursive trees and/or to random walks with a barrier.

Summary: Number of collisions of beta(a,1)-coalescent (2007/2008)

coalescent	parameter a	number $\mathbf{X_n}$ of collisions
Kingman	$a \to 0$	$X_n = n-1$
	$0 < a < 1$	$\frac{X_n - n(\alpha-1)}{(\alpha-1)n^{1/\alpha}} \xrightarrow{d} Y_\alpha$ (α-stable, $\alpha := 2-a$)
Bolthausen-Sznitman	$a = 1$	$\frac{(\log n)^2}{n} X_n - \log(n \log n) \xrightarrow{d} Y$ (1-stable)
	$1 < a < 2$	$\frac{X_n}{\Gamma(a) n^{2-a}} \xrightarrow{d} \int_0^\infty e^{-U_t} dt$ ($U =$ subord.)
	$a = 2$	$\frac{X_n - (2m_1)^{-1} \log^2 n}{(\frac{m_2}{3m_1^2} \log^3 n)^{1/2}} \xrightarrow{d} N$ (standard normal)
	$2 < a < \infty$	$\frac{X_n - \mu^{-1} \log n}{(\sigma^2 \mu^{-3} \log n)^{1/2}} \xrightarrow{d} N$ (standard normal)
star-shaped	$a \to \infty$	$X_n = 1$

Intuition: High mass of Λ near 0 $\widehat{=}$ Π_n has many small jumps $\widehat{=}$ increase of X_n

Assumptions

Assume that $\nu(dx) := x^{-2}\Lambda(dx)$ is a probability measure on $(0,1)$ such that

1. the support of ν is not contained in $\{1 - \delta\gamma^n : n = 0, 1, \ldots\}$ for some $\delta > 0$ and $\gamma \in (0,1)$,
2. $\int_{(0,1)} |\log x|\, \nu(dx) < \infty$.

The more general case of a finite measure ν can be reduced to the case of a probability measure by a linear time change of the coalescent.

The annihilator

There is a simpler process $A_n := (A_n(t))_{t\geq 0}$ with state space $\{0,\dots,n\}$ moving from m to $m-k$ with rate $\boxed{\binom{m}{k}\lambda_{m,k}}$, $1 \leq k \leq m \leq n$.

Remark. Total rates are $\displaystyle\sum_{k=1}^{m}\binom{m}{k}\lambda_{m,k} = \int_{[0,1]}(1-(1-x)^m)\,\nu(dx)$, $m \in \{0,\dots,n\}$.

Interpretation.

When there are m particles, any k-tuple of them collides and annihilates at rate $\lambda_{m,k}$.

Pitman's Poisson process construction.

At the generic time s_j of a unit Poisson process, a random variable $1-\eta_j$ is sampled from ν, and each of the remaining particles is marked 'head' with probability $1-\eta_j$ or 'tail' with probability η_j. The particles marked 'head' are removed.

A coupling

Define the coalescent Π_n and the annihilator A_n using the same unit Poisson process and the same sample $1-\eta_1, 1-\eta_2, \ldots$ from ν.

Call the initial n particles primary and their followers resulting from mergers secondary.

$K_n :=$ number of transitions of A_n as the process proceeds from n to 0

$K_{n,1} :=$ number of jumps of A_n of size 1

$\sigma_n := \inf\{t \geq 0 : A_n(t) = 0\}$ (absorption time, first time when there are no primary particles anymore)

$K_{n,0} :=$ number of epochs $s_j < \sigma_n$ when none of the primary particles are marked 'head'

$U_n := |\Pi_n(\sigma_n)|$ (remaining secondary particles)

Lemma 1. The following stochastic order relations hold:

$$\boxed{K_n - K_{n,1} \ \le_d \ X_n \ \le_d \ K_n + K_{n,0} + X_{U_n}}$$

where (on the r.h.s.) $(K_n, K_{n,0}, U_n)$ is independent of $(X_1, X_2, \ldots)$.

Proof. Coupling. If ≥ 2 primary particles collide, then both processes A_n and Π_n jump. The number of jumps of Π_n up to time σ_n does not exceed the number $K_n + K_{n,0}$, and after time σ_n the coalescent evolves with U_n particles. $\square$

Lemma 2. The sequence of distributions of X_{U_n}, $n \in \mathbb{N}$, is tight.

Sketch of proof. Use $X_{U_n} < U_n$ and study the number Q_j of secondary particles at time s_j. Given all the η_k's, $(Q_j)_j$ is a Markov chain. Introduce $N_n := \inf\{k \geq 1 : \eta_1 \cdots \eta_k \leq 1/n\}$ and use renewal theory to show that $(Q_{N_n})_n$ is tight. Now replace the fixed drop level $1/n$ by the appropriate random drop level associated with the last primary particle disappearing at time σ_n in order to show that $|Q_{N_n} - U_n|$ is stochastically bounded (technical ...) $\square$

Number of collisions $\mathbf{X_n}$: Asymptotics

Theorem 1. The following two assertions are equivalent.

(i) There exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that $(X_n - a_n)/b_n$ weakly converges to some non-degenerate and proper distribution.

(ii) The distribution of $(-\log \eta)$ either belongs to the domain of attraction of a stable law or the function $x \mapsto P(-\log \eta > x)$ slowly varies at ∞.

There are five different regimes (A) - (E) of convergence.

Possible limiting laws are normal, α-stable with $\alpha \in [1, 2)$ and (scaled) Mittag-Leffler.

Convergence Regimes

Logarithmic moments: $\mu := \mathrm{E}(-\log\eta)$ and $\sigma^2 := \mathrm{Var}(\log\eta)$

(A) If $\sigma^2 < \infty$ then, with $b_n := \mu^{-1}$ and $a_n := (\mu^{-3}\sigma^2 \log n)^{1/2}$, the limiting law of $(X_n - a_n)/b_n$ is standard normal.

(B) If $\sigma^2 = \infty$ and

$$\int_{(x,1)} (\log y)^2 \nu(dy) \sim L(-\log x) \quad \text{as } x \to 0$$

for some L slowly varying at ∞, then, with $b_n := \mu^{-1}\log n$ and $a_n := \mu^{-3/2} c_{[\log n]}$, where c_n is any sequence satisfying $\lim_{n\to\infty} nL(c_n)/c_n^2 = 1$, the limiting law of $(X_n - b_n)/a_n$ is standard normal.

(C) Assume that, for some function L slowly varying at ∞, the relation

$$P(\eta \leq x) \;\sim\; (-\log x)^{-\alpha} L(-\log x) \quad \text{as } x \to 0 \qquad (*)$$

holds with $\alpha \in [1,2)$ and that $\mu < \infty$ if $\alpha = 1$. Then, with $b_n := \mu^{-1} \log n$ and $a_n := \mu^{-(\alpha+1)/\alpha} c_{[\log n]}$, where c_n is any sequence satisfying $\lim_{n\to\infty} nL(c_n)/c_n^{\alpha} = 1$, the limiting law of $(X_n - a_n)/b_n$ is α-stable with characteristic function

$$t \;\mapsto\; \exp\Big\{ -|t|^{\alpha}\Gamma(1-\alpha)\Big(\cos\Big(\frac{\pi\alpha}{2}\Big) + i\sin\Big(\frac{\pi\alpha}{2}\Big)\mathrm{sgn}(t)\Big)\Big\}, \qquad t \in \mathbb{R}.$$

The remaining two regimes (D) and (E) cover the case when $\mu = \infty$.

(D) Assume that $\mu = \infty$ and that ($*$) in (C) holds with $\alpha = 1$. Let c be any positive function satisfying $\lim_{x\to\infty} \dfrac{xL(c(x))}{c(x)} = 1$, and set $\psi(x) := x \displaystyle\int_{e^{-c(x)}}^{1} \frac{P(\eta \le y)}{y}\, dy$.

Let b be any positive function satisfying $b(\psi(x)) \sim \psi(b(x)) \sim x$, $x \to \infty$.

Then, with $b_n := b(\log n)$ and $a_n := b(\log n)c(b(\log n))/\log n$, the limiting law of $(X_n - a_n)/b_n$ is 1-stable with characteristic function

$$t \;\mapsto\; \exp\Big(-|t|\Big(\frac{\pi}{2} - i\log|t|\mathrm{sgn}(t)\Big)\Big), \qquad t \in \mathbb{R}.$$

The last regime:

(E) If $(*)$ in (C) holds with $\alpha \in [0,1)$ then, with $a_n := \dfrac{\log^\alpha n}{L(\log n)}$, the limiting law of $\dfrac{X_n}{a_n}$ is scaled Mittag-Leffler θ_α (exponential if $\alpha = 0$) with moments

$$\int_{[0,\infty)} x^n \, \theta_\alpha(dx) \; = \; \frac{n!}{\Gamma^n(1-\alpha)\Gamma(1+n\alpha)}, \qquad n \in \mathbb{N}.$$

Basic ideas of the proof

- Use the coupling $K_n - K_{n,1} \le_d X_n \le_d K_n + K_{n,0} + X_{U_n}$ and the tightness of X_{U_n} to show that $(X_n - b_n)/a_n$ weakly converges if and only if $(K_n - b_n)/a_n$ weakly converges to the same distribution.

- Interpret K_n as the number of non-empty boxes in an occupancy scheme, where n balls are dropped in box 1 with probability $\xi_1 := 1 - \eta_1$, the remaining balls are dropped in box 2 with probability $\xi_2 := 1 - \eta_2$ and so on (Bernoulli sieve).

  (Gnedin, Iksanov, Negadajlov, Rösler 2008)

Example 1. Number of collisions for beta coalescents

An application of Regime (A) to beta coalescents with

$$\nu(dx) = \frac{\Lambda(dx)}{x^2} = cx^{a-3}(1-x)^{b-1}dx$$

($a > 2$ and $b, c > 0$) shows that

$$\boxed{\frac{X_n - \mu^{-1}\log n}{(\sigma^2\mu^{-3}\log n)^{\frac{1}{2}}} \xrightarrow{d} N(0,1)}$$

where

$$\begin{aligned} \mu &= \Psi(a-2+b) - \Psi(b), \\ \sigma^2 &= \Psi'(b) - \Psi'(a-2+b) \end{aligned}$$

and Ψ is the logarithmic derivative of the gamma function.

Example 2

Suppose that $P(\eta \leq x) = \dfrac{1}{(1-\log x)^{\alpha}}$, $x \in (0,1)$, for some $\alpha > 0$.

Then all five regimes occur.

Parameter α	$\mu := \mathrm{E}(-\log \eta)$	$\sigma^2 := \mathrm{Var}(\log \eta)$	**Regime**
$0 < \alpha < 1$	∞	∞	(E) Mittag-Leffler
1	∞	∞	(D) 1-stable
$1 < \alpha < 2$	$\dfrac{1}{\alpha - 1}$	∞	(C) α-stable
2	1	∞	(B) normal
$2 < \alpha < \infty$	$\dfrac{1}{\alpha - 1}$	$\dfrac{\alpha}{(\alpha-1)^2(\alpha-2)}$	(A) normal

Absorption times

Lemma. The following relation holds:

$$\boxed{\tau_n \stackrel{d}{=} \sigma_n + \tau'_{U_n}}$$

where, on the right hand side, $\tau'_j \stackrel{d}{=} \tau_j$, $j \in \mathbb{N}$, and (σ_n, U_n) and $(\tau'_j)_j$ are independent.

Proof. Coupling. When all the primary particles disappear at time σ_n, there are U_n secondary particles left.

Absorption times: Asymptotics

Theorem 2. Assume that condition (ii) of Theorem 1 holds. Then, with X_n replaced by τ_n, condition (i) of Theorem 1 holds.

There are again five different regimes of convergence.

Sketch of proof. By the previous lemma and Lemma 2, σ_n and τ_n have the same asymptotics. σ_n has the same limiting law as the first passage time through level $\log n$ for a compound Poisson process with the generic jump $-\log \eta$ and intensity 1.

Example: Absorption time for beta coalescent

An application of Regime (A) to beta coalescents with

$$\nu(dx) = \frac{\Lambda(dx)}{x^2} = cx^{a-3}(1-x)^{b-1}dx$$

($a > 2$ and $b, c > 0$) shows that

$$\boxed{\frac{C\tau_n - \mu^{-1}\log n}{((\sigma^2+\mu^2)\mu^{-3}\log n)^{\frac{1}{2}}} \xrightarrow{d} N(0,1)}$$

with μ and σ^2 as before and $C := \dfrac{a-1+b}{a-1}\dfrac{a-2+b}{a-2}$.

Remarks.

○ The factor C comes from the linear time change (ν is in general not a probability measure).

○ Devroye (1999) obtains the same scaling for the depth of a random node in random trees.

Summary: Absorption time of beta(a,1)-coalescent

coalescent	parameter a	absorption time $\tau_{\mathbf{n}}$
Kingman	$a \to 0$	$\tau_n \stackrel{\text{a.s.}}{\longrightarrow} \tau$, $\mathrm{E}(\tau) = 2$
	$0 < a < 1$	$\tau_n \stackrel{\text{a.s.}}{\longrightarrow} \tau$, $\mathrm{E}(\tau) < \infty$
Bolthausen-Sznitman	$a = 1$	$\tau_n - \log\log n \stackrel{d}{\longrightarrow} \tau$ (Gumbel)
	$1 < a < 2$	Conjecture: asymptotically normal?
	$a = 2$	Conjecture: asymptotically normal?
	$2 < a < \infty$	$\frac{C\tau_n - \mu^{-1}\log n}{((\sigma^2+\mu^2)\mu^{-3}\log n)^{1/2}} \stackrel{d}{\longrightarrow} N$ (standard normal)
star-shaped	$a \to \infty$	$\tau_n \stackrel{d}{=} \mathrm{Exp}(1)$

External branch length

Z_n := length of an external branch, chosen uniformly at random from the n external branches of the coalescent Π_n

Let T_n be the time of the first jump of Π_n and let $I_n := |\Pi_n(T_n)|$ the number of blocks after that first jump.

Z_n satisfies the distributional recursion

$$Z_1 = 0 \quad \text{and} \quad Z_n \;=\; T_n + B_n Z_{I_n}, \qquad n \in \{2, 3, \ldots\},$$

where T_n is independent of $B_n Z_{I_n}$, $B_2, B_3, \ldots$ are Bernoulli random variables with $P(B_n = 1|I_n) = (I_n - 1)/n$, and, conditional on I_n, B_n and Z_{I_n} are independent.

Summary: External branch length of beta(a,1)-coalescent

coalescent	parameter a	external branch length $\mathbf{Z_n}$
Kingman	$a \to 0$	$nZ_n \xrightarrow{d} Z$ having density $x \mapsto 8/(2+x)^3$
	$0 < a < 1$	Conj.: $n^{1-a} Z_n \xrightarrow{d} Z =?$ (work in progress?)
Bolthausen-Sznitman	$a = 1$	$(\log n) Z_n \xrightarrow{d} Z \stackrel{d}{=} \mathrm{Exp}(1)$
	$1 < a < \infty$	$Z_n \xrightarrow{d} Z \stackrel{d}{=} \mathrm{Exp}(a/(a-1))$
star-shaped	$a \to \infty$	$Z_n \stackrel{d}{=} \mathrm{Exp}(1)$

Extensions to Xi-coalescents without proper frequencies

Some results can be easily extended to Ξ-coalescents (with simultaneous multiple collisions). For example, $P(Z_n > t)$ is the probability that a randomly chosen individual is still a singleton at time t. By exchangeability, we can call this individual '1', and it follows that

$$P(Z_n > 1) \;=\; P(\{1\} \text{ is a singleton block of } \Pi_n(t)).$$

If the coalescent does no have proper frequencies, i.e., if the measure Ξ on the simplex

$$\Delta := \{x = (x_1, x_2, \ldots) : x_1 \geq x_2 \geq \cdots \geq 0, |x| := \textstyle\sum_i x_i \leq 1\}$$

satisfies $\boxed{\alpha := \int_\Delta \frac{|x|}{(x,x)}\,\Xi(dx) < \infty}$, where $(x,x) := \sum_i x_i^2$, then,

$$\lim_{n\to\infty} P(Z_n > t) \;=\; P(\{1\} \text{ is a block of } \Pi_\infty(t)) \;=\; \exp(-\alpha t),$$

using an argument of Pitman (1999), Eq. (37).

Examples. Dirac-coalescents, Poisson-Dirichlet coalescent

Discussion: A more general approach

Suppose that $\alpha < \infty$. For $t \geq 0$ let S_t denote the frequency of singletons of $\Pi_\infty(t)$ and set $X_t := -\log S_t$. Then $X = (X_t)_{t \geq 0}$ is a subordinator with Laplace exponent

$$\Phi(a) \;=\; \int_\Delta (1-(1-|x|)^a)\frac{\Xi(dx)}{(x,x)}, \qquad a \geq 0.$$

We can still define the annihilator using Schweinberg's Poisson construction or Kingman's paint-box construction (drop all particles marked head). The annihilator has total rates $\Phi(m)$, $m \in \{0,\ldots,n\}$, and the inequality $X_n \geq_d K_n - K_{n,1}$ holds true. The other bound in general does not make sense as a variable analogous to $K_{n,0}$ might be ∞. It seems plausible that, under additional conditions, X_n and $K_n - K_{n,1}$ still have the same limiting law.

Remark. This approach is for example also useful to analyse the asymptotics of the number of types M_n for Ξ-coalescents without proper frequencies and with mutation (rate $r > 0$):

$$\boxed{\frac{M_n}{n} \xrightarrow{d} \int_0^\infty re^{-rt}S_t\,dt}$$

(exponential integral of a subordinator with drift)

References I

CALIEBE, A., NEININGER, R., KRAWCZAK, M. AND RÖSLER, U. (2007) On the length distribution of external branches in coalescence trees: genetic diversity within species. *Theor. Popul. Biol.* **72**, 245–252.

DELMAS, J.-F., DHERSIN, J.-S. AND SIRI-JEGOUSSE, A. (2008) Asymptotic results on the length of coalescent trees. *Ann. Appl. Probab.* **18**, 997–1025.

DEVROYE, L. (1999) Universal limit laws for depths in random trees. *SIAM J. Computing* **28**, 409–432.

DRMOTA, M., IKSANOV, A., MÖHLE, M., AND RÖSLER, U. (2007) Asymptotic results concerning the total branch length of the Bolthausen-Sznitman coalescent. *Stoch. Process. Appl.* **117**, 1404–1421.

DRMOTA, M., IKSANOV, A., MÖHLE, M., AND RÖSLER, U. (2009) A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Struct. Algorithms* **34**, 319–336.

FREUND, F. AND MÖHLE, M. (2009) On the time back to the most recent common ancestor and the external branch length of the Bolthausen-Sznitman coalescent. *Markov Process. Related Fields* **15**, to appear.

References II

FREUND, F. AND MÖHLE, M. (2008) On the number of allelic types for samples taken from exchangeable coalescents with mutation. Preprint.

GNEDIN, A.V. (2004) The Bernoulli sieve. *Bernoulli* **10**, 79–96.

GNEDIN, A. AND YAKUBOVICH, Y. (2007) On the number of collisions in Λ-coalescents. *Electron. J. Probab.* **12**, 1547–1567.

GNEDIN, A., IKSANOV, A. AND MÖHLE, M. (2008) On asymptotics of exchangeable coalescents with multiple collisions. *J. Appl. Probab.* **45**, 1186–1195.

GNEDIN, A., IKSANOV, A. NEGADAJLOV, P. AND RÖSLER, U. (2009) The Bernoulli sieve revisited. *Ann. Appl. Probab.* **19**, to appear.

References III

GOLDSCHMIDT, C. AND MARTIN, J.B. (2005) Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron. J. Probab.* **10**, 718–745.

IKSANOV, A. AND MÖHLE, M. (2007) A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Comm. Probab.* **12**, 28–35.

IKSANOV, A. AND MÖHLE, M. (2008) On the number of jumps of random walks with a barrier. *Adv. Appl. Probab.* **40**, 206–228.

IKSANOV, A., MARYNYCH, A. AND MÖHLE, M. (2009) On the number of collisions in beta(2,b)-coalescents. *Bernoulli* **15**, to appear.

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

PITMAN, J. (1999) Coalescents with multiple collisions. *Ann. Probab.* **27**, 1870–1902.

SAGITOV, S. (1999) The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36**, 1116–1125.