The general coalescent with asynchronous mergers of ancestral lines *

Serik Sagitov

Department of Mathematics
Chalmers University of Technology and Göteborg University
S-412 96 Göteborg, Sweden
serik@math.chalmers.se

August 31, 1998

Abstract

Take a sample of individuals in the fixed size population model with exchangeable family sizes. Follow the ancestral lines for the sampled individuals backward in time to observe the ancestral process. We describe a class of asymptotic structures for the ancestral process via a convergence criterium. One of the basic conditions of the criterium prevents simultaneous mergers of ancestral lines. Another key condition implies that the marginal distribution of the family size is attracted by an infinitely d ivisible distribution. If the latter is normal the coalescent allows only for the pairwise mergers (Kingman's coalescent). Otherwise multiple mergers happen with positive probability.

AMS 1991 subject classification: 92D25

Keywords: coalescent, infinitely divisible distribution, de-Finetti's theorem, reduced branching process

*This work is part of the Bank of Sweden Tercentenary Foundation project "Dependence and Interaction in Stochastic Population Dynamics"

1 Introduction

Consider a one-parent population model with non-overlapping generations. Let each generation be of size N and allow no further dependence between different generations. Assume that the family sizes $\nu_1, \ldots, \nu_N$ are exchangeable random values wit h the fixed sum $\nu_1 + \ldots + \nu_N = N$.

Sample n individuals from the current generation and trace their ancestral lines back in time. The resulting ancestral tree is described by the Markov chain $\{\mathcal{R}_r, r \geq 1\}$. The chain state $\mathcal{R}_r$ is the following equivalence relati on for the sampled individuals: $(i, j) \in \mathcal{R}_r$ iff the i-th and the j-th individuals have a common ancestor in the r-th generation backwards in time. Denote by

$$p_{\xi,\eta}(N) = P[\mathcal{R}_{r+1} = \eta | \mathcal{R}_r = \xi]$$

the transition probabilities of this chain and put

$$\sigma^2(N) = E(\nu_1 - 1)^2; \ T_N = N\sigma^{-2}(N).$$

According to Kingman (1982b) the moment condition

$$sup_N E[\nu_1^k] < \infty, \ \ k \geq 1 \tag{1}$$

implies that the asymptotic relation (all asymptotics throughout are meant to hold as $N \to \infty$):

$$p_{\xi,\eta}(N) = \delta_{\xi,\eta} + T_N^{-1} q_{\xi,\eta} + o(T_N^{-1}) \tag{2}$$

holds with

$$q_{\xi,\eta} = \begin{cases} -\begin{pmatrix} |\xi| \\ 2 \end{pmatrix} & \text{if } |\xi| = |\eta| \\ 1 & \text{if } |\xi| = |\eta| + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The symbols $|\xi|$ and $|\eta|$ stand for the number the equivalence classes produced by the equivalence relations ξ and η. The equality $|\xi| = |\eta|$ means that $\xi = \eta$, while the equality $|\xi| = |\eta| + 1$ implies that the state η is t he result of a 2-merger happened at the state ξ.

Definition. The pair (ξ, η) is called a *k-merger* if all but one equivalence classes of η are inherited from ξ without a change. The exceptional class of η is the union of k classes of ξ, so that $|\xi| = |\eta| + k - 1$.

The formula (2) was used by Kingman (1982a) to establish an asymptotic structure describing the ancestral lines merging pairwise when followed backwards in time. Recently Möhle (1998) extended this convergence result to a wider class of models .

We take up the issue and prove a convergence criterium involving two major conditions on the joint distribution of the family sizes (cf. Sections 2 and 3). This criterium leads to a wider class of the parameters $q_{\xi,\eta}$. It is char acterised by an arbitrary probability measure $F(dx)$ on the interval [0,1]:

$$q_{\xi,\eta} = \begin{cases} -\int_0^1 [1 - (1-x)^{|\xi|} - |\xi| x (1-x)^{|\xi|-1}] x^{-2} F(dx), & \text{if } \xi = \eta; \\ \int_0^1 x^{k-2} (1-x)^{|\xi|-k} F(dx), & \text{if } (\xi, \eta) \text{ is a } k\text{-merger}; \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Notice that the equality (3) is the particular case of (4) with $F(dx) = \delta_0(dx)$.

Section 4 presents a transparent illustration of the multiple coalescent structure corresponding to (4). It is shown how de-Finetti's theorem could be used to explain the role of the probability measure $F(dx)$. A special case with

$$F([0, x]) = x^{1-\alpha}, \ \alpha \in (0, 1)$$

is discussed in Section 5. A close link with the reduced branching process with infinite variance is established.

2 Three necessary conditions for the convergence to the coalescent

We start this section by recalling the Kingman formula for the transition probability $p_{\xi,\eta}(N)$. The computation of the transition probabilities for the ancestral process is done in two steps. Let the equivalence relation η split the set of s ampled individuals in a classes and suppose that the j-th class of η covers b_j equivalence classes produced by ξ. Denote by $b = b_1 + \ldots + b_a$ the total number of equivalence classes for the relation ξ. Conditional on the family sizes $(\nu_1, \ldots, \nu_N)$ forming the generation r this transition probability is

$$\sum (\nu_{j_1})_{b_1} \ldots (\nu_{j_a})_{b_a} / (N)_b,$$

where the summation extends over all distinct vectors $(j_1, \ldots, j_a) \in [1, N]^a$ and $(N)_b$ stands for the product $N(N-1)\ldots(N-b+1)$. Taking the expectation and using the exchangeability property we obtain the Kingman formula

$$p_{\xi,\eta}(N) = \frac{(N)_a}{(N)_b} E(\nu_1)_{b_1} \ldots (\nu_a)_{b_a}. \tag{5}$$

This formula is the starting point of our analysis of the asymptotic relation

$$p_{\xi,\eta}(N) = \delta_{\xi,\eta} + V_N q_{\xi,\eta} + o(V_N), \tag{6}$$

with arbitrary V_N and $q_{\xi,\eta}$ satisfying $V_N \to 0$ and $\sum_\eta q_{\xi,\eta} = 0$. Due to (5) the relation (6) with $\xi \neq \eta$ says that

$$N^{a-b} E(\nu_1)_{b_1} \ldots (\nu_a)_{b_a} \sim V_N q_{\xi,\eta}. \tag{7}$$

For $a = 1$ rewrite (7) as

$$N^{1-k} E(\nu_1)_k \sim V_N \phi_k, \;\; k \geq 2. \tag{8}$$

Put $k = 2$ in (8) to see that $N^{-1}\sigma^2(N) \sim V_N \phi_2$. Without loss of generality assume that $\phi_2 = 1$ and get $V_N \sim T_N^{-1}$. This says that T_N is the proper time scale factor for the asymptotic analysis of the ancestral proce ss. Our *first necessary condition*:

$$\sigma^2(N) = o(N),\ V_N \sim T_N^{-1} \tag{9}$$

ensures that with this choice the condition $V_N \to 0$ is satisfied.

Now we can further transform (8) into

$$N^{1-k}T_N E(\nu_1)_k \to \phi_k,\ k \geq 2.$$

Together with the decomposition

$$(\nu_1)_k = (\nu_1 - 1)^k + c_1(\nu_1 - 1)^{k-1} + \ldots + c_{k-1}(\nu_1 - 1)$$

this gives

$$N^{1-k}T_N E(\nu_1 - 1)^k \to \phi_k,\ k \geq 2$$

or in terms of $W_N = (\nu_1 - 1)/N$:

$$NT_N EW_N^k \to \phi_k,\ k \geq 2. \tag{10}$$

We claim that (10) is equivalent to the weak convergence for the sum of i.i.d. copies of W_N:

$$\frac{1}{N}\sum_{i=1}^{NT_N} W_{Ni} \to W. \tag{11}$$

Indeed, due to the theory of infinitely divisible distributions (cf. Feller(1966), chapter 17, section 8, Theorem 2) the last convergence is equivalent to the weak convergence of measures on [0,1]:

$$NT_N x^2 P^{W_N}(dx) \to F(dx),$$

where the limit measure is necessarily a probability measure on [0,1] due the choice of the summands number NT_N. It is known also (cf. Feller(1966), chapter 8, section 1) that for such limit measure the weak convergence is equivalent to the convergenc e of moments (10). Thus we conclude that

$$N^{1-k} T_N E(\nu_1 - 1)^k \to \int_0^1 x^{k-2} F(dx), \;\; k \geq 2. \tag{12}$$

Our *second necessary condition* is equivalent to (11) and reads: there is such a probability measure $F(dx)$ that the convergence

$$NT_N P[\nu_1 > Nx] \to \int_x^1 y^{-2} F(dy), \tag{13}$$

holds at all points $x \in (0,1)$ where the limit function is continuous. Notice that the limit W has the standard normal distribution if $F(dx) = \delta_0(dx)$. Otherwise the limit distribution is infinitely divisible with a non-trivial Poisson component.

The *third necessary condition* controls the correlation between family sizes within a generation:

$$N^{-a} \sigma^{-2}(N) E(\nu_1 - 1)^2 \ldots (\nu_a - 1)^2 \to 0, \;\; a \geq 2. \tag{14}$$

It comes from the requirement excluding simultaneous mergers of ancestral lines. In terms of the parameters $q_{\xi,\eta}$ this requirement means that $q_{\xi,\eta} = 0$ for all pairs (ξ, η) with $a \geq 2$, $b_1 \geq 2$, $b_2 \geq 2$, $b_3 \geq 1, \ldots$, $b_a \geq 1$. With this agreement it follows from (7) that

$$N^{a-b} T_N E(\nu_1)_{b_1} \ldots (\nu_a)_{b_a} \to 0, \;\; a \geq 2, \;\; b_1, b_2 \geq 2, \;\; b_3, \ldots, b_a \geq 1. \tag{15}$$

To deduce (14) we use the representation

$$E(\nu_1 - 1)^2 \ldots (\nu_a - 1)^2 = E(\nu_1)_2 \ldots (\nu_a)_2$$

$$+ \sum_{m=0}^{a} (-1)^{a-m} \binom{a}{m} E(\nu_1)_2 \ldots (\nu_m)_2 (\nu_{m+1} - 1) \ldots (\nu_a - 1)$$

resulting from the substitution $(\nu_j - 1)^2 = (\nu_j)_2 - \nu_j + 1$. This representation in view of (15) reduces the problem down to

$$N^{-a} T_N E(\nu_1)_2 (\nu_2 - 1) \dots (\nu_a - 1) \to 0, \ a \geq 2; \tag{16}$$

plus

$$N^{-a} T_N E(\nu_1 - 1) \dots (\nu_a - 1) \to 0, \ a \geq 2. \tag{17}$$

It is easy to see that (16) follows from (7) with $b_1 = 2$, $b_2 = \dots = b_a = 1$, $a \geq 1$.

We prove (17) by induction over a using the equality

$$(\nu_1 - 1) + \dots + (\nu_N - 1) = 0 \tag{18}$$

and the exchangeability of $\nu_1, \dots, \nu_N$. Notice that the exchangeability and (18) give

$$(N-1) E(\nu_1 - 1)(\nu_2 - 1)$$

$$= E(\nu_1 - 1)[(\nu_2 - 1) + \dots + (\nu_N - 1)] = -E(\nu_1 - 1)^2.$$

Thus

$$T_N E(\nu_1 - 1)(\nu_2 - 1) \to -1 \tag{19}$$

which proves (17) for $a = 2$. Now suppose (17) is proven for $2 \leq a \leq l$. For $a = l + 1$ we have

$$(N - l) E(\nu_1 - 1) \dots (\nu_l - 1)(\nu_{l+1} - 1)$$

$$= E(\nu_1 - 1) \dots (\nu_l - 1)[(\nu_{l+1} - 1) + \dots + (\nu_N - 1)]$$

$$= -E(\nu_1 - 1) \dots (\nu_l - 1)[(\nu_1 - 1) + \dots + (\nu_l - 1)]$$

$$= -l E(\nu_1 - 1)^2 (\nu_2 - 1) \dots (\nu_l - 1).$$

Apply (16) to finish the prove of (17) and the deduction of (14).

3 A convergence-to-coalescent criterium

In the previous section we proved the "only if" part of the following convergence-to-coalescent criterium.

Theorem 3.1 *The asymptotic relation (6) holds with* $V_N \to 0$ *and* $\sum_\eta q_{\xi,\eta} = 0$ *only if the conditions (9), (13), (14) are valid.*

On the other hand, the conditions (13) and (14) imply (2) with (4). If, furthermore, $\sigma^2(N) = o(N)$*, then* $T_N \to \infty$ *and (6) holds.*

Remark 1. Notice that under the Kingman condition (1) all three conditions (9), (13) and (14) hold. The condition (13) is due to the central limit theorem and the condition (14) follows from Hölder's ine quality.

Remark 2. If $b = a = 2$, then $q_{\xi,\eta} = 1$ for any probability measure $F(dx)$.

Lemma 3.1 *If (14) holds, then*

$$N^{a-b} T_N E(\nu_1 - 1)^{b_1} \dots (\nu_a - 1)^{b_a} \to 0$$

for all $a \geq 2,\ b_1 \geq 2,\ b_2 \geq 2,\ b_3 \geq 1, \dots,\ b_a \geq 1.$

Proof. For $a \geq 2,\ b_1 \geq 2, \dots,\ b_a \geq 2$ the assertion follows from the inequality

$$\begin{aligned}
&|E(\nu_1 - 1)^{b_1} \dots (\nu_a - 1)^{b_a}| \\
&\leq E(\nu_1 - 1)^2 \dots (\nu_a - 1)^2 |\nu_1 - 1|^{b_1 - 2} \dots |\nu_a - 1|^{b_a - 2} \\
&\leq N^{b-2a} E(\nu_1 - 1)^2 \dots (\nu_a - 1)^2
\end{aligned}$$

and the condition (14).

When some b_j are equal to 1 we use the equality (cf. (18))

$$(N - k) E(\nu_1 - 1)^{b_1} \dots (\nu_k - 1)^{b_k} (\nu_{k+1} - 1) \tag{20}$$

$$= E(\nu_1-1)^{b_1+1}\dots(\nu_k-1)^{b_k}+\dots+E(\nu_1-1)^{b_1}\dots(\nu_k-1)^{b_k+1}.$$

After applying (20), possibly several times, we end up with terms that can be handled as previously.

Lemma 3.2 *If (13) and (14) hold, then*

$$N^{1-k}T_N E(\nu_1-1)^k(\nu_2-1)\dots(\nu_a-1)\to\int_0^1 x^{k+a-1}F(dx)$$

for $k\geq 1$ and $a\geq 4-k$.

Proof. Transform the last expectation iterating (20) until no linear components (ν_i-1) are left. Due to Lemma 3.1 we obtain that

$$N^{1-k}T_N E(\nu_1-1)^k(\nu_2-1)\dots(\nu_a-1)$$

$$= N^{2-k-a}T_N E(\nu_1-1)^{k+a-1}+o(1).$$

This and (12) prove the assertion.

Proof of the "if" part. It is enough to verify that (13) and (14) imply (15) and

$$N^{1-k}T_N E(\nu_1)_k\nu_2\dots\nu_a\to\int_0^1 x^{k-2}(1-x)^{a-1}F(dx),\ \ k\geq 2,\ \ a\geq 1; \quad (21)$$

Turn to the representation

$$(\nu_1)_{b_1}\dots(\nu_a)_{b_a}=\sum(\nu_1-1)^{i_1}\dots(\nu_a-1)^{i_a}$$

involving terms with $1\leq i_j\leq b_j,\ j=1,\dots,a$. Apply Lemma 3.1 and (19) to prove (15).

The convergence (21) is proved by induction over a. Since (21) for $a=1$ coincides with (12) we can proceed by assuming that (21) is true for all $1\leq a\leq l$, $k\geq 2$. Take $a=l+1$ and turn to the equation

$$E(\nu_1)_k(\nu_2-1)\dots(\nu_{l+1}-1)$$

$$= E(\nu_1)_k\nu_2\dots\nu_{l+1} + \sum_{j=1}^{l}(-1)^{l+1-j}\left(\begin{array}{c} l \\ j-1 \end{array}\right) E(\nu_1)_k\nu_2\dots\nu_j.$$

After applying Lemma 3.2 and the induction assumption we arrive at

$$N^{1-k}T_N E(\nu_1)_k\nu_2\dots\nu_{l+1}$$
$$\to \int_0^1 x^{k+l}F(dx) - \sum_{j=1}^{l}(-1)^{l+1-j}\left(\begin{array}{c} l \\ j-1 \end{array}\right)\int_0^1 x^{k+j-1}F(dx).$$

This proves (21) in view of the equality

$$x^{k+l} - \sum_{j=1}^{l}(-1)^{l+1-j}\left(\begin{array}{c} l \\ j-1 \end{array}\right) x^{k+j-1} = x^{k-2}(1-x)^l.$$

4 A simple model illustrating the multiple coalescent

Let $I^\infty = \{I_m, m = 1, 2, \dots\}$ be a collection of exchangeable random variables taking value 0 and 1. Imagine an infinite set of clusters merging together in the following discrete-time setting. Each time we draw an independent copy of the collection I^∞ and treat the indicators as the current cluster labels. All clusters labelled "1" by the next time form a single cluster.

Sample n clusters at the time $t = 0$ and denote by Z_t the number of clusters formed by the time t out of the initial n clusters. We can calculate the transition probability

$$p(i,j) = P[Z_{t+1} = j | Z_t = i]$$

of the Markov chain $\{Z_t\}$ using de-Finetti's theorem:

$$p(i, i-k+1) = \left(\begin{array}{c} i \\ k \end{array}\right)\int_0^1 x^k(1-x)^{i-k}G(dx),\ 2 \le k \le i; \tag{22}$$

$$p(i,i) = \int_0^1 [(1-x)^i + ix(1-x)^{i-1}]G(dx),\ i \ge 2. \tag{23}$$

Here $G(dx)$ is the probability measure characterising the dependence structure in the set I^∞ of exchangeable indicators.

To get a continuous-time approximation from this assume that the expectation EI_m is small, in that $G(dx) = G_N(dx)$ and $G_N(dx) \to \delta_0(dx)$. Suppose that for some infinitely increasing sequence B_N there is such a probability measure $F(dx)$ on [0,1] that the weak convergence holds

$$B_N x^2 G_N(dx) \to F(dx). \tag{24}$$

Under the condition (24) the probability (23) has the asymptotics

$$p_N(i,i) = 1 - B_N^{-1} q_i + o(B_N^{-1})$$

with

$$q_i = \int_0^1 \frac{1-(1-x)^i - ix(1-x)^{i-1}}{x^2} F(dx). \tag{25}$$

On the other hand, the transition probability (22) has the asymptotics

$$p_N(i, i-k+1) = B_N^{-1} \binom{i}{k} \int_0^1 x^{k-2}(1-x)^{i-k} F(dx) + o(B_N^{-1}).$$

We conclude that the merger size distribution (the probability of having $(i-k+1)$ clusters after a merger in a system of i clusters) is

$$\pi_k(i) = \frac{1}{q_i} \binom{i}{k} \int_0^1 x^{k-2}(1-x)^{i-k} F(dx). \tag{26}$$

This formula perfectly agrees with the limit distribution of Theorem 3.1.

Example. If I^∞ is a system of i.i.d. indicators, then $G_N(dx) = \delta_{p(N)}(dx)$ and $p(N) \to \infty$. For $B_N = 1/p(N)$ the condition (24) holds with $F(dx) = \delta_0(dx)$, $q_i = \binom{i}{2}$, $\pi_2(i) = 1$ and we arrive at Kingman's coalescent (cf. Kingman (1982c)).

5 A link to the reduced branching process

In the classical theory of the Galton-Watson processes the counterpart structure for the coalescent is called the reduced branching process (RBP). Losely speaking both the coalescent and the RBP address the same phenomenon. The crucial difference is that the branching process framework is future-oriented and the asymptotic results comes after scaling by the population age while the coalescent presumes that the past population sizes are known but the population age is unknown and the scaling factor is the current population size.

The RBP with finite variance discovered by Fleischmann and Siegmund-Shultze(1977) is a time inhomogeneous splitting process on the unit time interval [0,1). It starts with one particle which splits in two particles at a random time uniformly distributed o ver [0,1). These new particles independently of each other mimic the life career of the initial particle: split in two after a random time uniformly distributed over the interval [the birth time,1). Each new generation of particles split faster on average than the previous one and as a result by the time 1 we get the infinite number of particles. We emphasize that as with Kingman's coalescent in the RBP with finite variance the ancestral lines merge pairwise.

The corresponding RBP with infinite variance was found in Yakymiv(1980), where the critical Galton-Watson process under the Zolotarev-Slack condition was considered (cf. Zolotarev (1957) and Slack (1968)). The infinite variance case differs from the finit e variance RBP only in the offspring size of a particle: now after the uniformly distributed lifetime each RBP-particle transforms into a random number of new particles μ with the generating function given by the formula

$$E[s^{\mu}] = 1 - \frac{1+\alpha}{\alpha}(1-s) + \frac{1}{\alpha}(1-s)^{1+\alpha},\ \alpha \in (0,1). \tag{27}$$

The case $\alpha = 1$ corresponds to the finite variance case.

Remark. By taking derivatives of the generating function in (29) we calculate that

$$P[\mu = k] = \frac{(1+\alpha)(1-\alpha)\dots(k-2-\alpha)}{k!},\ k \geq 2. \tag{28}$$

Notice also that $E[\mu] = \frac{1+\alpha}{\alpha}$ and $\mathrm{Var}[\mu] = \infty$ for $0 < \alpha < 1$.

It will be shown here that for the coalescent with

$$F(dx) = dx^{1-\alpha} \tag{29}$$

the merger size distribution is closely related to the RBP offspring size μ distribution:

$$\pi_k(i) = P[\mu = k|\mu \leq i],\ 2 \leq k \leq i. \tag{30}$$

It is also shown that given (29) the formula

$$q_i = \frac{i!}{(1+\alpha)(2-\alpha)\dots(i-1-\alpha)} - \frac{(1-\alpha)(i-\alpha)}{(1+\alpha)(i+1)},\ i \geq 3 \tag{31}$$

holds and the following series converges

$$\sum_{i=2}^{\infty} 1/q_i < \infty. \tag{32}$$

The last inequality demonstrates that the total coalescent time has finite average.

Proof of (30). Under (29) we have

$$\pi_k(i) q_i = (1-\alpha)\begin{pmatrix} i \\ k \end{pmatrix} B(k-1-\alpha, i-k+1),\ 2 \leq k \leq i.$$

Since

$$B(k-1-\alpha, i-k+1) = \frac{(i-k)!}{(k-1-\alpha)\dots(i-1-\alpha)}$$

it follows that

$$\pi_k(i) = \frac{(1-\alpha)\dots(k-2-\alpha)}{k!q_i} \frac{i!}{(2-\alpha)\dots(i-1-\alpha)}$$

and therefore

$$\pi_k(i) = \frac{i!P[\mu = k]}{q_i(1+\alpha)(2-\alpha)\dots(i-1-\alpha)}.$$

Summing the last equality over k we obtain that

$$1 = \frac{i!P[\mu \le i]}{q_i(1+\alpha)(2-\alpha)\dots(i-1-\alpha)}. \tag{33}$$

The last two formulae imply (30).

Proof of (31) and (32). Due to the equality

$$\sum_{i=0}^{\infty} s^i P[\mu \ge i] = \sum_{i=0}^{\infty} s^i \sum_{k \ge i} P[\mu = k]$$

$$= \sum_{k=0}^{\infty} P[\mu = k] \frac{1-s^k}{1-s} = \frac{1-h(s)}{1-s} = \frac{1+\alpha}{\alpha}(1-s) - \frac{1}{\alpha}(1-s)^{\alpha}$$

we can compute explicitly the tail probabilities

$$P[\mu \ge i] = \frac{(1-\alpha)\dots(i-1-\alpha)}{i!}.$$

Put this into (33) to arrive at (31).

To derive (32) we apply (33) :

$$\sum_{i=2}^{\infty} 1/q_i = \sum_{i=2}^{\infty} \frac{q_i(1+\alpha)(2-\alpha)\dots(i-1-\alpha)}{i!P[\mu \ge i]}$$

$$< c\sum_{i=2}^{\infty} P[\mu = i](i-1-\alpha) < cE\mu.$$

Acknowledgements. The author wishes to thank Marek Kimmel, Simon Tavare and Martin Möhle for stimulating discussions.

6 References

1. Feller, W. An introduction to probability theory and its applications, v. 2 (Wiley, New York).

2. Fleischmann, K. and Siegmund-Shultze, R.(1977) The structure of reduced Galton-Watson processes. Math. Nachr. 79, 233-241.

3. Kingman, J.F.C.(1982a) On the genealogy of large populations. J. Appl. Prob. 19A, 27-43.

4. Kingman, J.F.C.(1982) Exchangeability and the Evolution of large populations. in: Koch, G. and Spizzichino, F. Exchangeability in probability and statistics, North-Holland, 97-112.

5. Kingman, J.F.C.(1982c) The coalescent. Stoch. Process. Appl. 13, 235-248.

6. Möhle, M.(1998) Robustness results for the coalescent. J. Appl. Probab. (to appear)

7. Slack, R.S.(1968) A branching process with mean one and possibly infinite variance. Z. Wahrscheinlichkeitsthoerie verw. Geb. 9, 139-145

8. Yakymiv, A.L.(1980) Reduced branching processes. Theory Probab. Appl. 25, 584-588

9. Zolotarev, V.M. (1957) More exact statements of several theorems in the theory of branching processes. (Russian) Teor. Veroyatnost. i Primemen. 2, 256-266.