Serdica

Bulgariacae mathematicae publicationes

Сердика

Българско математическо списание

AN INVARIANCE PRINCIPLE FOR REDUCED FAMILY TREES OF CRITICAL SPATIALLY HOMOGENEOUS BRANCHING PROCESSES
(with discussion)*

KLAUS FLEISCHMANN, RAINER SIEGMUND-SCHULTZE

Consider a critical spatially homogeneous branching process (Φ_n) starting at time $n=0$ with a single particle. A finer description can be given by means of family trees. Then to each particle in the n-th generation Φ_n there corresponds an ancestry line, including the positions of the ancestors. The collections of all these ancestry lines is called reduced (family) tree of the n-th generation. To each of these ancestry lines there corresponds a polygonal function on the time interval $[0, n]$ (just like a Donsker path of a random walk). Under suitable moment assumptions and contractions of space and time (similarly to those in Donsker's invariance principle) the collection Ψ_n of all those polygonal functions given that Ψ_n is non-empty, converges in distribution. Moreover, the limit has a very simple form. It corresponds to a certain binary branching Brownian motion model defined by the following properties. All particles living at the same moment $t<1$ develop independently and independent of the past. Each particle living at t moves rilatively to its position according to a standard Brownian motion up to a time point uniformly distributed in the interval $[t, 1]$. Then it splits up into exactly two particles.

That invariance principle is a refinement of the corresponding result in Fleischmann, Siegmund-Schultze (1977) for critical Galton-Watson processes. It is distantly related to a result of L. G. Gorostiza and A. R. Moncayo, where ancestry lines chosen at random of the n-th generation of certain supercritical processes are investigated. Besides, the spatially homogeneous branching model is so general that the branching random walk model in Euclidean spaces is covered.

1. Introduction. Let D be any fixed distribution of a *cluster* χ, i. e. a random finite population χ of particles with positions in any fixed d-dimensional Euclidean space R^d. The positions of particles in χ need not have any special properties as conditional independence or identical distributions for given total number χ^+ of particles in χ. Such D can be considered in the following way as characteristic for a *spatially homogeneous clustering mechanism* in discrete time. All particles develop independently and independent of the past. If at time n any particle has position x then its progeny χ_x at time $n+1$ coincides in distribution with $T_x\chi$, the population χ translated by x.

At time $n=0$ we start with a single particle lying in the origin. Then D yields according to the clustering mechanism described above a *spatially homogeneous branching process* (Φ_n), where Φ_n is the n-th generation, i. e. the population of descendents at time n of the initial particle. But for our purpose we need a finer description.

* Delivered at the "Wissenschaftliche Sitzung zur Stochastik", Jena, 17, Nov. 1977, Prof J. Mecke in the chair.

SERDICA Bulgaricae mathematicae publicationes. Vol. 4, 1978, p. 111—

Let be given the whole spatial family tree arising from the initial particle by the spatially homogeneous clustering mechanism up to a fixed time point n. Omitting in this tree all particles which do not have descendents at time n we get the so-called reduced tree of the n-th generation. This reduced tree can also be identified with the collection of all ancestry lines of particles of the n-th generation.

To each such ancestry line there corresponds a polygonal function on the interval $[0, n]$ with values in R^d. Our aim is now to investigate the asymptotic structure properties of the collection Ψ_x of all polygonal functions corresponding to the reduced tree of the n-th generation.

We assume that the Galton-Watson process $(\Phi_n{}^+)$ corresponding to (Φ_n) is critical and that its offspring distribution has a finite second moment. If ξ is the position of a particle chosen "at random" of the population χ, where χ with large total numbers χ^+ are preferred proportionally to χ^+, then we assume that the expectation $E\xi$ equals the zero vector and that the covariance matrix Cov ξ equals the unit matrix. Moreover, we assume that the expectation $E\eta$ exists where η is the position of a particle of χ chosen "at random" by preferring χ proportionally to $(\chi^+)^2$. Note that by the other assumptions $E\eta$ always exists in the case of a branching random walk model, i. e. if for given χ^+, the positions of particles in χ are independent and have a common distribution λ_D independent of χ^+.

Let $\mathcal{C}_{n^{-1}}\Psi_n$ mean that in all polygonal functions in Ψ_n the "time interval" $[0, n]$ is contracted by the factor n^{-1} and the space R^d is contracted by the factor $n^{-1/2}$. Our main result is now that $\mathcal{C}_{n^{-1}}\Psi_n$ given that Ψ_n is non-empty converges in distribution. Moreover, the limit is independent of the choice of D and coresponds to the following simple, temporarily inhomogeneous clustering model. At time $t=0$ a single initial particle starts with a standard Brownian motion. In a time point t uniformly distributed in the interval [0,1] it splits up into exactly two particles which move independently and, relatively to their positions, again according to the standard Brownian motion. But their (independent) live times are now uniformly distributed in the interval $[t, 1]$. Then again two particles arise, etc. Of course, this branching model can be obtained from the usual binary branching Brownian motion model on $[0, +\infty)$ by a suitable transformation of space and time.

By the way, that invariance principle also remains true in the case of critical Markovian spatially homogeneous branchig processes in continuous time, i. e. if the semi-group $\{D^{[n]}\}_{n=0,1\ldots}$ of cluster powers is replaced by suitably chosen spatially homogeneous semi-group $\{D^{[t]}\}_{t\geq 0}$ of critical cluster distributions.

That limit theorem is based on the following single effects.

The reduced tree of the n-th generation, given that it is non-empty, has again a branching structure (proposition 3.3). More precisely, it coincides in distribution with a (non-reduced) family tree of a certain spatially homogeneous clustering model on the "time set" $\{0, \ldots, n\}$ which is, however, inhomogeneous in time. In this model we can distinguish between clusters with at least two particles the so-called source-bunches and those with exactly one particle. Along any ancestry line the one-particle clusters between two following source bunches are called traces. Consequently, the non-empty reduced tree can be constructed by clustering alternatively sufficiently often traces and source bunches (proposition 6.4.).

In the Galton-Watson case (if the offspring distribution is critical and has a finite second moment) the source bunches have in the limit exactly two particles (cf. [3]), and in the spatial model these particles are lying at the same position by the contractions (lemma 8.2.).

Again in the Galton-Watson case, the "length" of contracted traces is distributed according to certain uniform distributions (cf. [9; 3]). In the spatial model, by the clustering mechanism, the traces form an inhomogenous random walk killed after an independent time. Hence, like in the invariance principle of Donsker and Prohorov the contracted traces are in the limit standard Brownian motions killed after a uniformly distributed time (lemma 11.2.)

Finally, by continuity properties of the clustering (proposition 2.1) the single convergence assertions can be summarized to the desired convergence theorem 7.1.

In section 2 we outline the general description of particle systems in complete separable metric spaces A by counting measures φ. In this concept clustering mechanisms are described by cluster fields. The continuity proposition 2.1, says roughly spoken that in bounded spaces any clustered distribution P_k depends continuously on both P and the cluster field k. This is a generalization of a continuity proposition in [6]. Moreover, a simple proof is given. Thus, it is of a certain independent interest although it is only a rudiment for such continuity theorems in unbounded spaces.

A cluster χ is called simple if there are not two particles in χ having the same position. If in a family tree of the n-th generation there is a non-simple cluster than in general the collection Ψ'_n of polygonal functions does not contain all information on family connections in the corresponding reduced tree, for instance the information on the last common ancestor ("source particle") of the n-th generation. This situation could be changed if we would work additionally with suitable independent continuously distributed interpolation curves instead of determined lines to connect the positions of corresponding particles. These considerations led us to the point of view to define in section 5 a cluster χ directly as a random finite counting measure on the σ-algebra $\mathfrak{U}_1$ corresponding to the set A_1 of R^d-valued continuous functions $a(t)$ on the interval [0,1] "vanishing" in $t=0$. With respect to the original problem we could now assume without loss of generality that this χ is simple almost surely. By the way there is also another argument not to work directly with lines. The random continuous functions (instead of lines) can be interpreted in many ways as a given spatial motion of particles [4].

In order tn realize this concept, in section 5 the set A of killed functions on $[0,+\infty)$ is introduced. In section 6 conditionally homogeneous cluster fields are defined which are suitable to express the temporary inhomogeneity in reduced trees. The main result is formulated in 7. The convergence considerations for the proof follow in the last sections including a proof for the "source time theorem" 11.1. simplier as earlier ones.

For notations and results in Galton-Watson theory, in the theory of weak convergence, and in point process theory we refer to Sevastianov [10], Billingsley [1], and Kerstan, Matthes, Mecke [6], respectively.

I^+ always denotes the set of all non-negative integers. For the convergence of sequences in topological spaces we write for simplicity uniformly $\to$ or more precisely $\underset{n\to\infty}{\longrightarrow}$.

2. A continuity theorem for clustering. Let A be any complete separable, metric space and $\mathfrak{A}$ the σ-algebra of all its Borel subsets.

Let M be the set of all counting measures on $\mathfrak{A}$, i. e. measures being integer-valued on bounded Borel subsets. Each φ in M has a representation as sum of Dirac measures $\varphi=\Sigma_i \delta_{a_i}$, where $(a_i)_{i \in I}$ is an at most countable family of elements in A [6, 1.1.2.]. Thus φ can be interpreted as a population of particles which are placed in the state space A at the positions a_i, $i \in I$. (Note, some of the particles may have the same positions.) We write φ^+ for the total number $\varphi(A)$ of particles in φ.

Equipped with the vague topology, M is a Polish space (cf. [6], section 1.15.). The σ-algebra of all Borel subsets of M is denoted by $\mathfrak{M}$.

Let $\mathbf{P}$ be the set of all distributions on $\mathfrak{M}$. Endowed with the weak topology, $\mathbf{P}$ is a Polish space (cf. [6], section 3.1.) Each $\mathbf{P}$ describes a random population of particles placed in A. The σ-algebra of Borel subsets of $\mathbf{P}$ is denoted by $\boldsymbol{P}$.

The mapping $[\varphi_1, \varphi_2] \curvearrowright \varphi_1+\varphi_2$ of the direct product $M \times M$ onto M is continuous and transforms each distribution $P_1 \times P_2$, where P_1, $P_2 \in \mathbf{P}$, into a distribution $P_1 * P_2$ on $\mathfrak{M}$, the convolution of P_1 and P_2. The operation $*$ in $\mathbf{P}$ is associative, commutative, continuous (cf. [6], 3.1.10) and has unit element δ_o, where o is the zero measure on $\mathfrak{A}$.

Let A' be another state space, i.e. any complete separable matric space (not necessarily defferent from A), and let $\mathfrak{A}'$, M', etc. be the expressions corresponding to this state space.

A measurable mapping $a \curvearrowright k_{(a)}$ from A into $\mathbf{P}'$ (measurable with respect to $\mathfrak{A}$ and $\boldsymbol{P}'$) is called a cluster field from A into A' (cf. [6], section 4.1.). Let A_k be the (measurable) set of all a in A where k is defined. If χ_a is distributed by $k_{(a)}$ then χ_a is interpreted as a cluster of particles placed in A' and generated by a particle placed in a. Thus k describes a very general branching mechanism.

Now we assume in addition that A is bounded. Then all populations φ in M are finite and the vague topology in M coincides with the weak one.

Let M_k be the set of all φ in M satisfying $\varphi(A \setminus A_k)=0$. If any φ in M_k has the form $\varphi=\Sigma_i \delta_{a_i}$ then the finite convolution $\ast_i k_{(a_i)}$ is denoted by $k_{(\varphi)}$. The mapping $\varphi \curvearrowright k_{(\varphi)}$ from M into $\mathbf{P}'$ is measurable (cf. [6], 4.1.3.) $k_{(\varphi)}$ describes the immediate progeny of the population φ if each particle in φ is clustered independently of the others according to k.

For P in $\mathbf{P}$ with $P(M_k)=1$ we write

$$P_k = \int P(d_\varphi) k_{(\varphi)}(\,.\,).$$

The clustered distribution P_k describes the immediate progeny of a random population φ distributed by P and clustered according to k.

Let k_0, $k_1, \ldots$ be cluster fields from A into A'. We write $k_n \to k_0$ if

$$(k_n)_{(a_n)} \xrightarrow[n \to \infty]{} (k_0)_{(a_0)}$$

whenever $a_n \in A_{k_n}$, $n \in I'^+$, with $a_n \to a_0$. Now, we have the following generalization 4.7.1. in [6].

2.1. Proposition. *Let k_0, $k_1, \ldots$ be cluster fields from A into A' satisfying $k_n \to k_0$. Further suppose that P_0, $P_1, \ldots$ are distributions on $\mathfrak{M}$ such that*

$$P_n \to P_0 \textit{ and } P_n(M_{k_n}) = 1, \; n \in \Gamma^+ .$$

Then

$$(P_n)_{k_n} \xrightarrow[n\to\infty]{} (P_0)_{k_0} .$$

Remarks. a) The assumption "A bouded" cannot be dropped, however it can be weakened, cf. [6], section 4.7., where the case $k_n \equiv k$ is treated.

b) Instead of clausters being counting measures we could work without difficulties with clusters being Radon measures on $\mathfrak{U}'$.

Proof. Let $\varphi_n \in M_{k_n}$, $n \in \Gamma^+$, with $\varphi_n \longrightarrow \varphi_0$. Then we have for some k in Γ^+ and almost all n the relation $\varphi_n^+ = \varphi_0^+ = k$. Moreover, for these n, there exist representations $\varphi_n = \Sigma_{1 \leq i \leq k} \delta_{a_{ni}}$ with $a_{ni} \longrightarrow a_{0i}$, $n \to \infty$, for each i. Thus, $k_n \longrightarrow k_0$ and the continuity of the convolution together yield

$$\mathop{*}_{1 \leq i \leq k} (k_n)_{(a_{ni})} \xrightarrow[n\to\infty]{} \mathop{*}_{1 \leq i \leq k} (k_0)_{(a_{0i})} ,$$

.i e. $(k_n)_{(\varphi_n)} \longrightarrow (k_{0(\varphi_0)}$. Hence (cf. [5], A 4.2. and [1], theorem 5.5.) from $P_n \longrightarrow P_0$ it follows $P_n((k_n)_{(\varphi)} \in (\,.\,)) \longrightarrow P_0((k_0)_{(\varphi)} \in (\,.\,))$ in $\mathbf{P}'$, i. e. we have $\int P_n(d_\varphi) f((k_n)_{(\varphi)}) \longrightarrow \int P_0(d_\varphi) f((k_0)_{(\varphi)})$ for each bounded continuous real function f on $\mathbf{P}'$. Setting $f(P') = \int P'(d\chi) h(\chi)$, $P' \in \mathbf{P}'$, where h is any bounded continuous real function on M', we obtain $\int P_n(d\varphi) \int (k_n)_{(\varphi)} (d\chi) h(\chi) \longrightarrow \int P_0(d\varphi) \times \int (k_0)_{(\varphi)} (d\chi) h(\chi)$, i. e. $\int (P_n)_{k_n}(d\chi) h(\chi) \longrightarrow \int (P_0)_{k_0}(d\chi) h(\chi)$.

3. A structure theorem for powers of homogeneous cluster fields. Let k and ω be cluster fields in A (that means cluster fields from A into A) such that $k_{(a)}(M_\omega) = 1$, $a \in A_k$. Then by $a \frown (\omega \circ k)_{(a)} = (k_{(a)})_\omega$, $a \in A_k$, we get a new cluster field $\omega \circ k$. The operation $\circ$ is associative (cf. [6], 4.3.5.) such that if $k \circ k$ makes sense, we can define the powers k^n of a cluster field k by $k^n = k \circ k^{n-1}$, $n = 1, 2, \ldots$, where k^0 is the unit element $(k^0)_{(a)} = \delta_{\delta_a}$, $a \in A_k$.

Now we assume in addition that in the bounded complete separable metric space A an associative operation $\oplus$ with zero element $\mathbf{o}$ is given. Moreover, the maping $[a_1, a_2] \frown a_1 \oplus a_2$ of the direct product $A \times A$ onto A is assumed to be continuous. It transforms the product measure $\nu_1 \times \nu_2$ of any measures ν_1, ν_2 on $\mathfrak{U}$ into the convolution $\nu_1 * \nu_2$. The convolution is associative and has unit element δ_0. Moreover, the convolution and addition of measures are distributive.

For each a in A we define a translation operator T_a in M by $T_a\varphi = \int \varphi(d_x) \delta_{a \oplus x}(\,.\,)$, $\varphi \in M$, and get the translation semi-group $(T_a)_{a \in A}$ with unit element T_0: $T_a T_b = T_{a \oplus b}$, $a, b \in A$. Now, the continuity of the addition in A and theorem 5.5. in [1] yield (cf. [6], 6.1.1.).

3.1. The mapping $[a, \varphi] \frown T_a \varphi$ of $A \times M$ onto M is continuous.

A cluster field k on A (i. e. a measurable mapping k of A into $\mathbf{P}$) is called homogeneous, if there is a distribution D on $\mathfrak{M}$ such that $k_{(a)} = D(T_a \chi \in (\,.\,))$, $a \in A$ holds. In this case we write $k = [D]$.

The mapping $D \frown [D]$ of $\mathbf{P}$ onto the set of all homogeneous cluster fields (on A) is one-to-one. Moreover, by 3.1. and theorem 5.5 in [1], we have $D_n \to D$ if and only if $[D_n] \to [D]$.

If $[D]$ and $[G]$ are homogeneous cluster fields then $[G] \circ [D]$ is again homogeneous and we have $[G] \circ [D] = [D_{[G]}]$.

If ν is any distribution on $\mathfrak{A}$ then we write Q_ν for the distribution $\int \nu(da)\delta_{\delta_a}(\,.\,)$ of the population δ_a consisting only of the particle with position a is distributed according to ν.

If $(p_i)_{i\in I}$ is a family of non-negative numbers such that $\Sigma_i p_i = 1$ and ν_i, for each i in I, a distribution on $\mathfrak{A}$ then we have the simple identity

3.2. $$(Q_\nu)_{[\sum_i p_i Q_{\nu_i}]} = \sum_i p_i Q_{\nu * \nu_i}.$$

Proof. Obviously,

$$(Q_\nu)_{[\sum_i p_i Q_{\nu_i}]} = \int Q_\nu(d\varphi)[\sum_i p_i Q_{\nu_i}]_{(\varphi)}(\,\cdot\,)$$

$$= \int \nu(da)[\sum_i p_i Q_{\nu_i}]_{(a)} = \int \nu(da)\sum_i p_i Q_{\nu_i}(T_a\varphi\in(\,.\,))$$

$$= \sum_i p_i \int \nu(da) \int \nu_i(da_i)\delta_{T_a\delta_{a_i}}(\,\cdot\,) = \sum_i p_i \int \nu(da) \int \nu_i(da_i)\delta_{\delta_{a\oplus a_i}}(\,\cdot\,)$$

$$= \sum_i p_i \int (\nu * \nu_i)(da)\delta_{\delta_a}(\,\cdot\,) = \sum_i p_i Q_{\nu * \nu_i}.$$

If c belongs to the interval $[0, 1]$ then by $a \curvearrowright (1-c)\delta_0 + c\delta_{\delta_a}$, $a \in A$, we get a certain homogeneous cluster field k_c. Instead of P_{k_c}, where P belongs to $\mathbf{P}$, we also write $\mathcal{D}_c P$ and call $\mathcal{D}_c$ the thinning operator with survival probability c (cf. [6], section 1.13.).

Let D be any distribution on $\mathfrak{M}$ different from δ_0. For each n in Γ^+ we define $D^{[n]} = ([D]^n)_{(\mathbf{o})}$, $z_n = D^{[n]}, \chi \neq 0$, and ${}_{n+1}D = (\mathcal{D}_{z_n} D)((\,\cdot\,) \mid \chi \neq 0)$. $D^{[n]}$ can be nterpreted as distribution of the n-th generation of descendants of a parent particle which was placed in $\mathbf{o}$ and was n times clustered homogeneously according to the cluster distribution D. The following proposition says that those distributions $D^{[n]}$ can also be obtained with the help of clustering without extinction.

In generalization of [3], proposition 1.1., we have

3.3. Proposition. *For each n in Γ^+,*

$$D^{[n]} = (\mathcal{D}_{z_n}\delta_{\delta\mathbf{o}})_{[{}_1D]\,\mathrm{o}\,\ldots\,\mathrm{o}\,[{}_nD]}.$$

In particular, the conditional distributions $D^{(n)} = D^{[n]}((\,\cdot\,) \mid \chi \neq 0)$ have again a branching structure:

$$D^{(n)} = ([{}_1D]\,\mathrm{o}\,\ldots\,\mathrm{o}[{}_nD])_{(\mathbf{o})}.$$

Proof. Obviously, the assertion is valid for $n = 0$. If it is true for n, then the identity

$$[D^{[n]}] = [(1-z_n)\delta_0 + z_n D^{(n)}] = [(\mathcal{D}_{z_n}\delta_{\delta\mathbf{o}})_{[D^{(n)}]}] = [D^{(n)}]\mathrm{o}[\mathcal{D}_{z_n}\delta_{\delta_\mathbf{o}}]$$

yields

$$D^{[n+1]} = D_{[D^{[n]}]} = D_{[D^{(n)}\mathrm{o}[\mathcal{D}_{z_n}\delta_{\delta_\mathbf{o}}]} = (\mathcal{D}_{z_n} D)_{[D^{(n)}]} = (\mathcal{D}_{z_n} D)_{[{}_1D]\,\mathrm{o}\,\ldots\,\mathrm{o}\,[{}_nD]}.$$

We still have to show that $\mathcal{D}_{z_n} D = (\mathcal{D}_{z_{n+1}}\delta_{\delta_\mathbf{o}})_{[{}_{n+1}D]}$. But $(\mathcal{D}_{z_n} D)(\chi \neq \mathrm{o}) = z_{n+1}$ and $(\mathcal{D}_{z_n} D)((\,\cdot\,) \mid \chi \neq \mathrm{o}) = {}_{n+1}D$, such that

$$\mathcal{Q}_{z_n} D = (1-z_{n+1})\delta_0 + z_{n+1}\ {}_{n+1}D = ((1-z_{n+1})\delta_0 + z_{n+1}\delta_{\delta_0})_{[_{n+1}D]'}$$

and 3.3 holds by induction.

4. The space of killed functions. Let R^d, $d\geq 1$, be any fixed d-dimensional Euclidean space. For each non-negative number t, let A_t be the set of all continuous mappings a of the closed interval $[0, t]$ into R^d satisfying $a(0)=[0, \dots, 0]$, where $[0, \dots, 0]$ denotes the origin in R^d. Now we will introduce a metric ϱ in the set $A=\bigcup_{t\geq 0} A_t$ of killed functions. We have the idea that $a\in A_s$ and $b\in A_t$ are near one another if, on the one hand, s and t are such and, on the other hand, $a(x)$ and $b(y)$ do not differ essentially if x and y are neighbouring.

For a in A_s and b in A_t with $s\leq t$, put $\varrho(b, a)=\varrho(a, b)=\max\{\sup_{x\in[0,s]} \|a(x)-b(x)\|, \sup_{x\in[s,t]} \|[s, a(s)]-[x, b(x)]\|\}$, where $\|\cdot\|$ denotes the d- respectively $(d+1)$-dimensional Euclidean norm. Now we will show that $\varrho(a_2, a_2)\leqq\varrho(a_1, a_3)+\varrho(a_3, a_2)$ for any a_i in A_{t_i}, $i=1, 2, 3$. Without loss of generality we may assume that $t_1\leq t_2$. Let $\varrho(a_1, a_2)=\|a_1(x)-a_2(x)\|$ for some $x\leq t_1$. If $x\leqq t_3$ then

$$\varrho(a_1, a_2)\leq\|a_1(x)-a_3(x)\|+\|a_3(x)-a_2(x)\|\leq\varrho(a_1, a_3)+\varrho(a_3, a_2),$$

whereas in the opposite case

$$\begin{aligned}\varrho(a_1, a_2)=\|[x, a_1(x)]-[x, a_2(x)]\|&\leq\|[x, a_1(x)]-[t_3, a_3(t_3)]\|\\ &+\|[t_3, a_3(t_3)]-[x, a_2(x)]\|\leq\varrho(a_1, a_3)+\varrho(a_3, a_2).\end{aligned}$$

Now let $\varrho(a_1, a_2)=\|[t_1, a_1(t_1)]-[x, a_2(x)]\|$ for some x with $t_1\leq x\leqq t_2$. Then distinguish again between $x\leqq t_3$ and $x>t_3$ and estimate in a similar manner to get that ϱ is a metric.

Let t be any fixed non-negative number. To each a in A_s with $0\leq s\leq t$ there corresponds an element a^t in A_t defined by $a^t(x)=a(x)$ for $0\leq x\leq s$ and $a^t(x)=a(s)$ else. Thus, we have a mapping $a\curvearrowright a^t$ of the union $A_{[0,t]}$ of all sets A_s, $0\leq s\leq t$, onto A_t.

4.1. *For any $a_i\in A_{s_i}\subseteq A_{[0,t]}$, $i=1, 2$, we have*

$$\varrho(a_1, a_2)\leq|s_1-s_2|+\varrho(a_1^t, a_2^t),$$

$$\varrho(a_1, a_2)\geqq\max\{|s_1-s_2|, \varrho(a_1^t, a_2^t)\}.$$

P r o o f. These are immediate consequences of the inequalities

$$\|[s_1, a_1(s_1)]-[x, a_2(x)]\|\leq|s_1-x|+\|a_1(s_1)-a_2(x)\|,$$

$$\|[s_1, a_1(s_1)]-[x, a_2(x)]\|\geqq\max\{|s_1-x|, \|a_1(s_1)-a_2(x)\|\},$$

where $s_1\leqq x\leq s_2$.

To each s in $[0, t]$ there corresponds a killing operator K_s on A_t defined by

$$(K_s a)(x)=a(x) \qquad (a\in A_t, \quad 0\leq x\leq s).$$

4.2. *The mapping $[s, a]\curvearrowright K_s a$ of the direct product $[0, t]\times A_t$ onto $A_{[0,t]}$ is continuous.*

(Of course, in $[0, t]$ we use the Euclidean metric and in A_t and $A_{[0,t]}$ the metric ϱ.)

Proof. Let $s_n \to s$ and $a_n \to a$. Then by 4.1.

$$\varrho(K_{s_n}a_n, K_s a) \leq |s_n - s| + \varrho((K_{s_n}a_n)^t, (K_s a)^t).$$

For each v in the interval $I_n = [\min\{s_n, s\}, \max\{s_n, s\}]$ we have

$$\|(K_{s_n}a_n)^t(v) - (K_s a)^t(v)\| \leq \sup_{x \in I_n} \|a_n(x) - a(x)\| + \sup_{x,y \in I_n} \|a(x) - a(y)\|.$$

Hence

$$\varrho((K_{s_n}a_n)^t, (K_s a)^t) \leq \varrho(a_n, a) + \sup_{x,y \in I_n} \|a(x) - a(y)\|.$$

But the last term converges to zero as n tends to infinity because a is uniformly continuous.

4.3. *The metric space $[A, \varrho]$ is separable and complete.*

Proof. For the separability it is enough to show that all $A_{[0,t]}$ are separable. But this follows from 4.2. because $[0, t] \times A_t$ is separable.

Let (a_n) be any fundamental sequence in A. Then (a_n) is such in $A_{[0,t]}$ for some t. Now 4.1. yelds that (a_n^t) is also a fundamental sequence in A_t and has therefore a limit b in A_t.

Each a_n belongs to some A_{s_n} and by 4.1. (s_n) has a limit s in $[0, t]$. Applying 4.2. we obtain finally

$$a_n = K_{s_n}a_n^t \xrightarrow[n\to\infty]{} K_s b.$$

Let $[a, b] \in A_s \times A_t$ for any $s, t \geq 0$. Setting $(a \oplus b)(x) = a(x)$ for $0 \leq x \leq s$ and $(a \oplus b)(x) = a(s) + b(x - s)$ for $s \leq x \leq s + t$, we get an element $a \oplus b$ in A_{s+t}. Thus, we have an associative operation $\oplus$ in A with zero element $\mathbf{o}$, where $\mathbf{o}$ is the single element of A_0. Moreover,

4.4. *The mapping $[a, b] \curvearrowright a \oplus b$ of the direct product $A \times A$ onto A is continuous.*

Proof. Let $a_n \to a \in A_s$ and $b_n \to b \in A_t$. Then we have $a_n \in A_{s_n}$ for some sequence (s_n) converging to s and

$$\varrho(a_n \oplus b_n, a \oplus b) \leq \varrho(a_n \oplus b_n, a_n \oplus b) + \varrho(a_n \oplus b, a \oplus b) \leq \varrho(b_n, b) + \varrho(a_n, a) + |s_n - s|$$
$$+ \sup\{\|b(x) - b(y)\| : x, y \in [0, t];\ |x - y| \leq |s_n - s|\}.$$

However, the last term tends to zero because b is uniformly continuous.

Let c be in $[0, 1]$ and a be in A_t for some $t \geq 0$. Putting $(C_c a)(s) = \sqrt{c}a(c^{-1}s)$ for $c > 0$ and $0 \leq s \leq ct$ and $C_c a = \mathbf{o}$ for $c = 0$ we get an element $C_c a$ in A_{ct}. Thus, we have contraction operators C_c in A satisfying $C_c(a \oplus b) = C_c a \oplus C_c b$. Moreover

4.5. *The mapping $[c, a] \curvearrowright C_c a$ of the direct product $[0, 1] \times A$ onto A is continuous.*

Proof. Let $a_n \to a \in A_t$ and $c_n \to c$. If both c and all c_n are positive then $\varrho(C_{c_n}a_n, C_{c_n}a) \leq \varrho(a_n, a) \to 0$. On the other side

$$\varrho(C_{c_n}a, C_c a) = \varrho(\sqrt{c_n}a(c_n^{-1}(\cdot)), \sqrt{c}a(c^{-1}(\cdot)))$$

$$\leq \varrho(a(c_n^{-1}(\cdot)),\ a(c^{-1}(\cdot))) + \varrho(\sqrt{c_n}a(c^{-1}(\cdot)),\ \sqrt{c}a(c^{-1}(\cdot))).$$

The last term equals $|\sqrt{c_n}-\sqrt{c}|\ \sup\{\|a(x)\| : x\in[0, t]\}$ and vanishes if n tends to infinity. The other one can be estimated above by

$$\leq \sup\{\|a(x)-a(y)\| : 0\leq x, y\leq t;\ |x-y|\leq t\,|1-c^{-1}c_n|\},$$

which also tends to zero as $n\to\infty$.

In the case $c=0$ we get

$$\varrho(C_{c_n}a_n, \mathbf{o})\leq\sqrt{c_n}\varrho(a_n, \mathbf{o}) \to 0.$$

Setting $\varrho_A(a, b)=(1+\varrho(a, b))^{-1}\varrho(a, b)$, $a, b\in A$, we get a bounded complete metric ϱ_A in A which defines the same topology as ϱ. For the sake of abbreviation from now on the letter A always denotes this bounded complete separable metric space $[A, \varrho_A]$ in which we have defined a continuous and associative operation $\oplus$. Thus, the notations and results of the preceding sections can be applied to this A.

The set $\mathbf{V}$ of all distributions ν on $\mathfrak{A}$ equipped with the weak topology is a Polish space. In this space by $\mathcal{C}_c\nu=\nu(C_c a\in(\cdot))$, $c\in[0, 1]$, $\nu\in\mathbf{V}$, c o n t r a c t i o n o p e r a t o r s $\mathcal{C}_c$ are defined.

In analogy to the definitions above we define for each c in $[0, 1]$

$$C_c\varphi=\int\varphi(da)\delta_{C_c a}(\cdot),\quad \varphi\in M,$$

$$\mathcal{C}_c P=P(C_c\varphi\in(\cdot)),\quad P\in\mathbf{P}.$$

All those c o n t r a c t i o n o p e r a t o r s have the following continuity properties which follow immediately from 4.5 and [1], theorem 5.5.

4.6. *The mappings* $[c, \nu]\frown\mathcal{C}_c\nu$ *of* $[0, 1]\times\mathbf{V}$ *onto* $\mathbf{V}$, $[c, \varphi]\frown C_c\varphi$ *of* $[0, 1]\times M$ *onto* M, *and* $[c, P]\frown\mathcal{C}_c P$ *of* $[0, 1]\times\mathbf{P}$ *onto* $\mathbf{P}$ *are continuous.*

5. Trees. In the sections 2 and 3 we gave for elements in A, M, $\mathbf{P}$, etc. general interpretations. But now we use also more special interpretations as follows.

Each a in A is considered as an a n c e s t r y l i n e, as a b r a n c h, or as a t r a c e. According to the given definition, $a\oplus b$ can be interpreted as a continuation of the ancestry line a with the help of the branch b.

Each $\varphi=\Sigma_i\delta_{a_i}$ in M can be considered as a (family) t r e e with its ancestry lines a_i or as b u n c h with the branches a_i. The tree $T_a\varphi$ appears if the ancestry line a is continued with the help of the bunch φ. Note that $T_a\varphi$ only consists of the original "ancestry line" δ_a if $\varphi=\delta_{\mathbf{o}}$, whereas $T_a\varphi$ is the empty tree o if the bunch φ is such, i. e. $\varphi=0$.

For each $t\geq 0$ the subset A_t of A is closed and yields therefore again a bounded complete separable metric space. Thus, we can form the corresponding mathematical expressions M_t, $\mathbf{P}_t$, etc., indicated by t, which can be considered as subsets of the non-indicated ones.

From now on let D be any fixed element in $\mathbf{P}_1$ satisfying $D(\chi^+=0)>0$ and $D(\chi^+>1)>0$. We interprete D as distribution of a bunch φ corresponding to the first generation Φ_1 of the spatially homogeneous branching process (Φ_n) mentioned in the introduction (where the branches a in φ must not be lines). Moreover, for any n in Γ^+ we interprete $D^{[n]}$ as distribution of the random reduced tree corresponding to the n-th generation Φ_n. The term "reduced" refers to the fact that during the successive clustering an empty bunch causes that the corresponding preceding ancestry line vanishes. Thus, all ancestry lines in a random reduced tree φ_n, distributed according to $D^{(n)}$, have the "length" n.

The structure theorem 3.3. says that the distribution of the reduced tree φ_n can be obtained also by a certain clustering mechanism without empty bunches.

6. The source structure of reduced trees. Let G be a cluster field from $[0, +\infty]$ into A. Setting

$$|G|_{(a)} = G_{(t)}(T_a\varphi\in(\cdot)), \quad a\in A_t, \quad t\in[0, +\infty)_G,$$

we get a cluster field $|G|$ in A. Cluster fields in A of this form are called conditionally homogeneous. (These cluster fields are in general different from homogeneous cluster fields in marked spaces used for instance in [7] because the "mark" t may be changed by translations.) If even $G_{(t)} \equiv G$ for some G in **P** and all non-negative t then $|G|$ coincides with the homogeneous cluster field $[G]$ on A.

6.1. *Let $|G|$ and $|H|$ be conditionally homogeneous cluster fields* in A *such that $|H| \circ |G|$ exists. Then $|H| \circ |G|$ is again a conditionally homogeneous cluster field $|F|$ in* A, *and we have*

$$F_{(t)} = (G_{(t)})_{|H_{(t+(\cdot))}|} \quad t\in[0, +\infty)_G.$$

Proof. Let t be in $[0, +\infty)_G$ and a be in A_t. Then

$$(|H| \circ |G|)_{(a)} = (G_{(t)}(T_a\varphi\in(\cdot)))_{|H|} = \int G_{(t)}(d\varphi) |H|_{(T_a\varphi)}(\cdot).$$

Now

$$|H|_{(T_a\varphi)} = \mathop{*}_{\varphi(\{x\})>0} (|H|_{(a+x)})^{\varphi(\{x\})}$$

and, for $x\in A_s$, $t+s\in[0, +\infty)_H$,

$$|H|_{(a+x)} = H_{(t+s)}(T_{a+x}\Psi\in(\cdot)) = |H_{(t+(\cdot))}|_{(x)}(T_a\chi\in(\cdot)).$$

Hence

$$|H|_{(T_a\varphi)} = \mathop{*}_{\varphi(\{x\})>0} (|H_{(t+(\cdot))}|_{(x)})^{\varphi(\{x\})} (T_a\chi\in(\cdot)) = |H_{(t+(\cdot))}|_{(\varphi)}(T_a\chi\in(\cdot)),$$

so that

$$(|H| \circ |G|)_{(a)} = (G_{(t)})_{|H_{(t+(\cdot))}|}(T_a\chi\in(\cdot)).$$

Let D be as given in 5. and then ${}_nD$, $n=1, 2, \ldots$, be the distributions introduced in 3.

Let l and k be in Γ^+ with $l \leq k$. We define

$$r_l^k = \Big(\prod_{k\geq i>k-l} {}_iD(\chi^+=1)\Big)_{k-l}D(\chi^+>1),$$

where the non-defined expression ${}_0D(\chi^+>1)$ is understood as one. r_l^k can be interpreted as probability that the trunk of the reduced tree φ_k distributed according to $D^{(k)}$ has "length" l. Immediately from the definition it follows

6.2. *We have* $r_{l+1}^{k+1} = {}_{k+1}D(\chi^+ = 1) r_l^k$.

Set

$$ {}^kW_{(l)} = \sum_{0 \le l \le k-l} r_i^{k-l} Q_{\underset{k-l \ge j \ge k-l-(i-1)}{*} \nu_j}, $$

where the distributions ν_n are defined by $Q_{\nu_n} = {}_nD((\cdot) \mid \chi^+ = 1)$, $n = 1, 2, \ldots$, respectively $\nu_1 = \delta_0$ if ${}_1D(\chi^+ = 1) = 0$, i. e. $D(\chi^+ = 1) = 0$, where $\mathbf{0}$ is that function, in A_1 which is identically to zero. ${}^kW_{(l)}$ can be interpreted as distribution of a trunk of (the random) "length" i of the subtree of the reduced tree φ_k, distributed according to $D^{(k)}$, where the subtree is started at time l and is considered relatively to its "start point".

6.3. *For all natural numbers* n,

$$ {}^nW_{(0)} = {}_nD(\chi^+ > 1)\delta_{\delta_0} + {}_nD(\chi^+ = 1)(Q_{\nu_n})_{|{}^nW|}. $$

Proof. By 3.2. and 6.2.,

$$ {}_nD(\chi^+ = 1)(Q_{\nu_n})_{|{}^nW|} = {}_nD(\chi^+ = 1)(Q_{\nu_n})_{[{}^{n-1}W_{(0)}]} $$

$$ = {}_nD(\chi^+ = 1) \sum_{0 \le i \le n-1} r_i^{n-1} Q_{\nu_n * (\nu_{n-1} * \cdots * \nu_{n-1-(i-1)})} $$

$$ = \sum_{0 \le i \le n-1} r_{i+1}^n Q_{\nu_n * \cdots * \nu_{n-1-(i-1)}} = \sum_{1 \le i \le n} r_i^n Q_{\nu_n * \cdots * \nu_{n-(i-1)}}, $$

from which the lemma follows immediately.

Setting ${}^lS_{(l)} = \delta_{\delta_0}$ and ${}^kS_{(l)} = {}_{k-l}D((\cdot) \mid \chi^+ > 1)$, $0 \le l < k$, we interprete ${}^kS_{(l)}$ as distribution of a source bunch in the reduced tree φ_k distributed by $D^{(k)}$ where the source bunch appears at time l and is considered relatively to its "start point".

${}^kW$ and ${}^hS$ are cluster fields from $[0, +\infty)$ into A. Thus, the conditionally homogeneous cluster fields ${}^kW|$ and $|{}^kS|$ are well-defined. Now we have in generalization of [2, 2.1], the following structure theorem.

6.4. Proposition. *For all* n *in* Γ^+

$$ D^{(n)} = ({}^nW_{(0)})_{(|{}^nW| \circ |{}^nS|)^n}. $$

This proposition says that the non-empty reduced tree φ_n can be constructed by piecing together traces and source bunches alternatively at most n times, where the spatially homogeneity is expressed by the conditionally homogeneous cluster fields $|{}^nW|$ and $|{}^nS|$.

Proof. At first, ${}^nW_{(k)} = {}^{n+1}W_{(k+1)}$, ${}^nS_{(k)} = {}^{n+1}S_{(k+1)}$, $k \le n$. By 6.1, $(|{}^{n+1}W| \circ |{}^{n+1}S|)^n \circ |{}^{n+1}W|$ is again a conditionally homogeneous cluster field $|G|$ and we have

$$ G_{(1)} = ({}^{n+1}W_{(1)})_{(|{}^{n+1}S_{(1+(\cdot))}| \circ |{}^{n+1}W_{(1+(\cdot))}|)^n} = ({}^nW_{(0)})_{(|{}^nS| \circ |{}^nW|)^n}. $$

Obviously, the assertion in 6.4 is valid for $n = 0$. If it is true for n then, as just demonstrated, $D^{(n)} = G_{(1)}$, i. e.

$$ [D^{(n)}]_{(a)} = G_{(1)}(T_a\chi \in (\cdot)), \quad a \in A_1. $$

This yields togetger with 3.3.,

$$D^{(n+1)} = ({}_{n+1}D)_{[D^{(n)}]} = ({}_{n+1}D)_{|G|}.$$

Now

$${}_{n+1}D = {}_{n+1}D(\chi^+ > 1)\,(\delta_{\delta_0})_{|{}^{n+1}S|} + {}_{n+1}D(\chi^+ = 1)Q_{\nu_{n+1}}$$

so that

$$({}_{n+1}D)_{|G|} = {}_{n+1}D(\chi^+ > 1)\,(\delta_{\delta_0})_{(|{}^{n+1}W| \circ |{}^{n+1}S|)^{n+1}}$$

$$+ {}_{n+1}D(\chi^+ = 1)\,((Q_{\nu_{n+1}})_{|{}^{n+1}W|})_{(|{}^{n+1}W| \circ |{}^{n+1}S|)^{n}}.$$

In the last expression we may substitute the power n by $n+1$ because of $Q_{\nu_{n+1}}(M_1) = 1$. But 6.3. yields now

$$D^{(n+1)} = ({}^{n+1}W_{(0)})_{(|{}^{n+1}W| \circ |{}^{n+1}S|)^{n+1}},$$

and the structure theorem 6.4. holds by induction.

7. A binary branching Brownian motion. Let β be the distribution of the d-dimensional standard Brownian motion on the time interval $[0, 1]$. We consider β as distribution on $\mathfrak{A}_1$. For each t in $[0, 1)$, let β_t be the distribution

$$(1-t)^{-1} \int_0^{1-t} dx \int \beta(da)\delta_{K_x a}(\,\cdot\,)$$

on $\mathfrak{A}$ where K_x are the killing operators on A_1 introduced in 4. Consequently, the distribution β_t describes a Brownian motion killed in a uniformly distributed time point. Now $B_{(t)} = (Q_{\beta_t})^2$, $t \in [0, 1)$, (in the sense of convolution powers) yields a cluster field B from $[0, +\infty)$ into A, so that $|B|$ is a conditionally homogeneous cluster field in A. This cluster field $|B|$ describes the cluster mechanism of the *temporarily inhomogeneous binary branching Brownian motion* mentioned already in the introduction.

Setting

$$f(x) = \int D(d\chi)x^{\chi^+} \quad (x \in [0, 1]),$$

then f is called the *offspring probability generating function* corresponding to D. Let f' and f'' be the first and second derivatives of f, respectively.

For each natural number m let λ_D^m be the *moment measure of order* m of D, i. e.

$$\lambda_D^m = \int D(d\varphi)\varphi^{\times m}(\,.\,)$$

(powers in the sense of product measures). $\lambda_D =_{\mathrm{df}} \lambda_D^1$ is called the *intensity measure* of D.

To formulate our main result we colect all *assumptions*. D was assumed to be an element in $\mathbf{P}_1$. Let

(1) $$f'(1-) = 1 \text{ and } 0 < f''(1-) < +\infty,$$

i. e. the Galton-Watson process corresponding to f is critical and has a finite second moment. Furthermore, suppose that

(2) $$E_a(1) = [0, \ldots, 0], \operatorname{Cov} a(1) = \mathbf{I}, \quad E\|a\|^2 < +\infty,$$

where the expectation E and the covariance Cov are formed with respect to the distribution λ_D, where $\mathbf{I}$ is the d-dimensional unit matrix, and where

$$\| a \| =_{\mathrm{DF}} \sup_{0 \leq t \leq 1} \| a(t) \|, \quad a \in A_1.$$

Moreover, let

$$\int \lambda_D^2(A_1 \times (da)) \| a(1) \| < +\infty. \tag{3}$$

By the way, in view of all other assumptions (3) is always fulfilled if D is purely random (cf. [6, section 1.14.]), as it is in the branching random walk case.

7.1. Theorem. *For each k in Γ^+,*

$$\mathcal{C}_{n^{-1}}(({}^nW_{(0)})_{(|{}^nW| \circ |{}^nS|)^k}) \xrightarrow[n\to\infty]{} (Q_{\beta_0})_{|B|^k}.$$

Consequently, the "tree of the k-th source generation" converges in distribution to the tree of the k-th generation of that binary branching Brownian motion.

8. The convergence of source bunches. Let $\bar{D}$ be the distribution

$$(f''(1-))^{-1} \int D(d\chi) \int \chi(da) \int (\chi - \delta_a)(db) \delta_{\delta_a + \delta_b}(\cdot)$$

on $\mathfrak{M}$, which has almost surely realizations with exactly two "points" in A_1.

8.1. *The distributions $(\mathcal{O}_c D)((\cdot) \mid \chi^+ > 1)$, $0 < c < 1$, converge in variation to $\bar{D}$ as $c \to 0$.*

Proof. It suffices to demonstrate that the measures

$$H_c =_{\mathrm{df}} ((\mathcal{O}_c D)(\chi^+ > 1))^{-1} (\mathcal{O}_c D)((\cdot), \chi^+ = 2)$$

converge in variation to $\bar{D}$. Now

$$(\mathcal{O}_i D)(\chi^+ > 1) = 1 - (\mathcal{O}_c D)(\chi^+ = 0) - (\mathcal{O}_c D)(\chi^+ = 1)$$

$$= 1 - \int D(d\chi)(1-c)^{\chi^+} - \int D(d\chi) \int \chi(dx) c (1-c)^{\chi^+ - 1} = 1 - f(1-c) - c f'(1-c)$$

and

$$(\mathcal{O}_c D)((\,.\,), \chi^+ = 2) = \int D(d\chi) 2^{-1} \int \chi(da) \int (\chi - \delta_a)(db) c^2 (1-c)^{\chi^+ - 2} \delta_{\delta_a + \delta_b}(\cdot).$$

Thus we can estimate the variation distance $\| H_c - \bar{D} \|$ by

$$\leq \int D(d\chi) \int \chi(da) \int (\chi - \delta_a)(db) \left| x_c (1-c)^{\chi^+ - 2} - (f''(1-))^{-1} \right|,$$

where $x_c = 2^{-1} c^2 / [1 - f(1-c) - c f'(1-c)]$. By L'Hospital's rule,

$$\lim_{c\to 0} x_c = \lim_{c \to 0} (f''(1-c))^{-1} = (f''(1-))^{-1}.$$

Moreover,

$$\lim_{c \to 0} (1-c)^{\chi^+ - 2} = 1, \quad \chi^+ > 1.$$

Hence, the "integrand" $|\cdot|$ converges to zero. Moreover, it is bounded because of $0 \leq (1-c)^{\chi^+ - 2} \leq 1$, $\chi^+ > 1$. Finally,

$$\int D(d\chi)\int \chi(da)\int(\chi-\delta_a)(db)=\int D(d\chi)\chi^+(\chi^+-1)=f''(1-)$$

is finite, such that by the bounded convergence theorem the integral converges to zero.

Let k be any cluster field in A and c be in $(0, 1]$. Setting

$$(\mathcal{C}_c k)_{(a)}=\mathcal{C}_c(k_{(Cc^{-1}a)})\quad (a\in\{C_c b:\ b\in A_k\}),$$

we obtain a new cluster field $\mathcal{C}_c k$, i. e. contraction operators $\mathcal{C}_c$ for cluster fields.

Let τ be the following cluster field in A:

$$\tau_{(a)}=\delta_{2\delta_a},\quad a\in A_{[0,1]}=_{\mathrm{df}}\bigcup_{0\le t<1} A_t.$$

8.2. *We have* $\mathcal{C}_{n^{-1}}\,{}^n S\to\tau$.

Proof. For each natural number n, let a_n be in $A_{\mathcal{C}_{n^{-1}}|^n S|}\cap A_{[0,1)}$. Then a_n has the form

$$a_n=C_{n^{-1}}b_n,\quad b_n\in A_{k_n},\quad k_n\in\{0,\dots,n-1\}.$$

Thus

$$|^n S|_{(C_n a_n)}=|^n S|_{(b_n)}={}^n S_{(k_n)}(T_{b_n}\chi\in(\cdot))$$

$$={}_{n-k_n}D(T_{b_n}\chi\in(\cdot)\mid\chi^+>1)=(\mathcal{D}_{z_{n-k_n-1}}D)(T_{b_n}\chi\in(\cdot)\mid\chi^+>1).$$

Hence

$$(\mathcal{C}_{n^{-1}}|^n S|)_{(a_n)}=\mathcal{C}_{n^{-1}}(|^n S|_{(C_n a_n)})$$

$$=(\mathcal{D}_{z_{n-k_n-1}}D)(T_{a_n}C_{n^{-1}}\chi\in(\cdot)\mid\chi^+>1).$$

Now we assume that $a_n\to a_0\in A_{[0,1]}$. Then $n-k_n-1$ tends to infinity as $n\to\infty$ and we have $z_n\to 0$ (cf. [10, theorem 2. 1. 4.]). Thus, by 8.1.,

$$(\mathcal{D}_{z_{n-k_n-1}}D)((\cdot)\mid\chi^+>1)\to\bar{D}.$$

On the other side, by the continuity properties 3.1. and 4.6. from $\chi_n\to\chi\in M_1$ with $\chi^+=2$ it follows

$$T_{a_n}C_{n^{-1}}\chi_n\to T_{a_0}C_0\chi=2\sigma_{a_0},$$

so that by [1], theorem 5.5., the convergence relation in 8.2. holds.

9. The distributions ν_n. These distributions were introduced in 6. In the following we exclude the particular case ${}_1D(\chi^+=1)=0$, where ν_1 was defined formally by $\nu_1=\delta_0$.

For each natural number n,

$${}_nD((\cdot),\chi^+=1)=\int D(d\chi)\int\chi(da)z_{n-1}(1-z_{n-1})^{\chi^+-1}\delta_{\delta_a}(\cdot),$$

in particular

$$_nD(\chi^+=1)=z_{n-1}\int D(d\chi)\chi^+(f_{n-1}(0))^{\chi^+-1}=z_{n-1}f'(f_{n-1}(0)),$$

where $f_0(x)=x$, $f_n(x)=f(f_{n-1}(x))$, $x\in[0,1]$. Hence

$$\nu_n=(f'(f_{n-1}(0)))^{-1}\int D(d\chi)(f_{n-1}(0))^{\chi^+-1}\chi(\cdot).$$

Now $(f_{n-1}(0))^{\chi^+-1}\leqq 1$, $\chi^+>0$, and (cf. [10, theorem 2.1.4.]) $f_n(0)=1-z_n\to 1$, so that $f'(f_n(0))\to 1$. Consequently,

9.1. *There is a sequence* (c_n) *of numbers converging towards* 1 *such that*

$$\nu_n=c_n\int D(d\chi)(f_{n-1}(0))^{\chi^+-1}\chi(\cdot)\leqq c_n\lambda_D\leqq c\lambda_D$$

for some constant c *(except the case* $D(\chi^+=1)=0$, *where formally* $\nu_1=\delta_0$)

These estimations yield

9.2. *For all real functions* h *on* A_1 *integrable with respect to* λ_D

$$\int\nu_n(da)h(a)\to\int\lambda_D(da)h(a).$$

Proof. We have

$$\left|\int\lambda_D(da)h(a)-\int\nu_n(da)h(a)\right|\leqq\left|(1-c_n)\int\lambda_D(da)h(a)\right|+\int(c_n\lambda_D-\nu_n)(da)\,|h(a)|.$$

Obviously, the first expression on the right-hand side of this inequality tends to zero. The second one can be estimated by

$$\leqq c_n\int_{|h(a)|>m}\lambda_D(da)\,|h(a)|+m(c_n-1)\underset{n\to\infty}{\longrightarrow}\int_{|h(a)|>m}\lambda_D(da)\,|h(a)|,$$

where m is any natural number.

10. A version of the invariance principle of Donsker and Prohorov. The aim of this section is to prove the following convergence to the Brownian motion.

10. 1. Proposition. *Let* $\nu_1,\nu_2,\ldots$ *be the distributions introduced in* 6. *Then*

$$\mathcal{C}_{n^{-1}}(\nu_n*\cdots*\nu_1)\to\beta.$$

To demonstrate this we need some notations and lemmas.

The usual addition and subtraction of functions in A_1 are continuous operations which we denote with the help of the symbols $+$ and $-$.

Puting $a^l(t)=ta(1)$, $a\in A_1$, $t\in[0,1]$, we obtain a continuous mapping $a\curvearrowright a^l$ of A_1 into itself, a certain linearization of the functions in A_1.

Writing $\nu_n^s=\nu_n(a^l-E_na^l\in(.))$, $n=1,2,\ldots$, where the expectation E_n is formed with respect to ν_n, we start with the following version of a Lindeberg codition.

10.2. *For all* $\varepsilon>0$,

$$L_n(\varepsilon)=_{\mathrm{df}}\int_{\|a(1)\|>\varepsilon\sqrt{n}}\nu_n^s(da)\,\|a(1)\|^2\underset{n\to\infty}{\longrightarrow}0.$$

Proof. By definition,

$$L_n(\varepsilon)=\int_{\|a(1)-E_na(1)\|>\varepsilon\sqrt{n}}\nu_n(da)\,\|a(1)-E_na(1)\|^2.$$

In virtue of 9.2. and (2)

$$\|E_na(1)\|=\left\|\int\nu_n(da)a(1)\right\|\longrightarrow\left\|\int\lambda_D(da)a(1)\right\|=0.$$

Consequently we have for almost all n

$$\|a(1)-E_n a(1)\| \leq \|a(1)\| + \|E_n a(1)\| \leq \|a(1)\| + 1$$

and hence in connection with 9.1. for those n

$$L_n(\varepsilon) \leq \int_{\|a(1)\|+1>\varepsilon\sqrt{n}} \nu_n(da)(\|a(1)\|+1)^2 \leq c \int_{\|a(1)\|+1>\varepsilon\sqrt{n}} \lambda_D(da)(\|a(1)\|+1)^2.$$

But in view of (2) this expression tends to zero as $n \to \infty$.

10.3. *We have* $\mathcal{C}_{n-1}(\nu_n^s * \cdots * \nu_1^s) \longrightarrow \beta$.

Proof. a) To each a in A_1 there corresponds a "reflected" a^r in A_1 defined by $a^r(t)=a(1)-a(1-t)$, $t \in [0,1]$. This mapping $a \curvearrowright a^r$ is continuous and we have $(a^r)^r=a$. If a is linear, i. e. $a^l=a$, then even $a^r=a$ holds. Therefore, the distributions ν_n^s are invariant with respect to that "reflection". On the other hand also β has this invariance property.

Consequently, 10.3. is equivalent to

$$\alpha_n =_{\mathrm{df}} \mathcal{C}_{n-1}(\nu_1^s * \cdots * \nu_n^s) \longrightarrow \beta,$$

which we shall prove.

b) Let n be any natural number. We write ${}_nt=n^{-1}[nt]$, $t \in [0,1]$, where $[x]$ denotes the integer part of the real number x. Eor each t in $[0,1]$ and $\varepsilon>0$ we have for almost all n

$$\alpha_n(\|a(t)-a({}_nt)\|>\varepsilon) \leq {}^s\nu_{[nt]+1}(\|a\|>\varepsilon\sqrt{n})$$
$$\leq \nu_{[nt]+1}(\|a(1)\|+1>\varepsilon\sqrt{n}) \leq c\lambda_D(\|a(1)\|+1>\varepsilon\sqrt{n}).$$

Now the last term tends to zero as $n \to \infty$. Thus, for all finite sequences $0=t_0<\cdots<t_m \leq 1$, we have for all $\varepsilon>0$

$$\alpha_n(\|[a(t_1),\ldots,a(t_m)]-[a({}_nt_1),\ldots,a({}_nt_m)]\|>\varepsilon) \xrightarrow[n\to\infty]{} 0,$$

(where $\|.\|$ denotes here the md-dimensional Euclidean norm).

c) By the continuity of linear transformations, the desired convergence assertion

$$\alpha_n([a({}_nt_1),\ldots,a({}_nt_m)] \in (.)) \xrightarrow[n\to\infty]{} \beta([a(t_1),\ldots,a(t_m)] \in (.))$$

is equivalent to

$$\alpha_n([a(nt_1)-a({}_nt_0),\ldots,a({}_nt_m)-a({}_nt_{m-1})] \in (.))$$
$$\xrightarrow[n\to\infty]{} \beta([a(nt_1)-a(t_0),\ldots,a(t_m)-a(t_{m-1})] \in (.)).$$

But these vectors have independent components, so that this convergence relation is equivalent to

$$\alpha_n(a({}_nt_2)-a({}_nt_1) \in (.)) \longrightarrow \beta(a(t_2)-a(t_1) \in (.)), \quad 0 \leq t_1 < t_2 \leq 1.$$

Now

$$\alpha_n(a({}_nt_2)-a({}_nt_1) \in (.))$$
$$=\nu^s_{[nt_1]+1} \times \cdots \times \nu^s_{[nt_2]}(n^{-1/2}(a_{[nt_1]+1}(1)+\cdots+a_{[nt_2]}(1)) \in (.))$$
$$\longrightarrow \beta(a(t_2-t_1) \in (.)) = \beta(a(t_2)-a(t_1) \in (.)).$$

In fact, Lindenberg's central limit theorem can be applied because of

$$Ea_i(1)=0, \quad \mathrm{Cov}\ a_i(1) \xrightarrow[i\to\infty]{} \mathbf{I}, \text{ and } 10.2.,$$

where E and Cov are formed with respect to ν_i^s and the covariances converge because of 9.2.

Summarizing, we get together with b) the convergence of all finite-dimensional distributions.

d) It remains to show that the sequence (α_n) is relatively compact. For this end it suffices to prove that for each $j=1,\ldots,d$ the sequence $(\alpha_n^j) = (\alpha_n(a^j\in(.)))$ has this property where a^j denotes the j-th component of a. In view of Prohorov's theorem we are going to demonstrate the tightness of (α_n^j).

In order to get a contradiction we assume now that there is an $\varepsilon>0$ such that for all $\eta>1$ thare is a sequence (k_n) of non-negative integers such that

$$\limsup_{n\to\infty}\ e_{k_n,n}(\eta)\geqq \eta^{-2}\varepsilon,$$

where

$$e_{k,n}(\eta)=\nu_{k+1}^s\times\cdots\times\nu_{k+1}^s(\max_{1\leqq i\leqq n}|a_{k+1}^j(1)+\cdots+a_{k+1}^j(1)|\geqq\eta\sqrt{n})\ (k\in\Gamma+).$$

Applying [1], formula 10.7., we get for all k and n

$$e_{k,n}(\eta)\leqq 2\nu_{k+1}^s\times\cdots\times\nu_{k+n}^s(|a_{k+1}^j(1)+\cdots+a_{k+n}^j(1)|\geqq\eta\sqrt{n}-\sqrt{2}\,s_{k,n}),$$

where

$$s_{k,n}=(\sum_{1\leqq i\leqq n}\mathrm{Var}\ a_{k+i}^j(1))^{1/2}.$$

By 9.2. and (2),

$$\mathrm{Var}\ a_m^j(1)\xrightarrow[m\to\infty]{}\int\lambda_D(da)(a^j(1))^2=1,$$

so that $s_{k,n}\leqq r\sqrt{n}$ for some constant r, which does not depend on k and n. Hence

$$e_{k_n,n}(\eta)\leqq 2\nu_{k_n+1}^s\times\cdots\times\nu_{k_n+n}^s(n^{-1/2}|a_{k_n+1}^j(1)+\cdots+a_{k_n+n}^j(1)|\geqq\eta-\sqrt{2}\,r).$$

Applying again Lindeberg's theorem we obtain for all sufficiently large η

$$\limsup_{n\to\infty}\ e_{k_n,n}(\eta)\leqq 2\beta(|a^j(1)|\geqq\eta-\sqrt{2}r)$$

$$\leqq 2(\eta-\sqrt{2}\,r)^{-3}\int\beta(da)\,|a^j(1)|^3<\eta^{-2}\varepsilon,$$

which contradicts the assumption above. Consequently, for all $\varepsilon>0$ there is an $\eta>1$ and an $n_0\in\Gamma+$ such that for all k in $\Gamma+$ and all $n\geqq n_0$

$$e_{k_n,n}(\eta)<\eta^{-2}\varepsilon.$$

But this yelds (cf. [1], theorem 8.4.) the desired tightness.

If ν is any distribution on $\mathfrak{U}_1$ then we write $\nu^l=\nu(a^l\in(.))$.

10.4. *We have*

$$\mathcal{C}_{n-1}(\nu_n^l*\cdots*\nu_1^l)\longrightarrow\beta.$$

Proof. In view of 10.3. we have only to show that

$$\int\mathcal{C}_{n-1}(\nu_n^l*\cdots*\nu_1^l)(da)a\longrightarrow 0.$$

Because of

$$\| \int \mathcal{C}_{n-1}(\nu_n^l * \cdots * \nu_1^l)(da)a \| \leq n^{-1/2} \sum_{n \geq i \geq 1} \| \int \nu_i(da)a(1) \|$$

it suffices to prove that $\| \int \nu_n(da)a(1) \| = O(n^{-1})$, $n \to \infty$, or, for any j-th component a^j of a,

$$| \int \nu_n(da)a^j(1) | = O(n^{-1}), \quad n \to \infty.$$

In virtue of 9.1. we obtain

$$| \int \nu_{n+1}(da)a^j(1) | \leq c | \int D(d\chi)(f_n(0))^{\chi^+-1} \int \chi(da)a^j(1) | .$$

Writing formally

$$E(x) = \sum_{k>0} x^{k-1} \int_{\chi^+=k} D(d\chi) \int \chi(da)a^j(1)$$

we get a power series F satisfying in view of (2)

$$F(1) = \int \lambda_D(da)a^j(1) = 0,$$

so that F converges for $x \in [0,1]$. By assumption (3)

$$\int \lambda_D^2(A_1 \times (da))a^j(1)$$

makes sense and coincides with

$$= \int D(d\chi)\chi^+ \int \chi(da)a^j(1) = \int D(d\chi)(\chi^+ - 1) \int \chi(da)a^j(1)$$
$$= \sum_{k>0} (k-1)1^{k-2} \int_{\chi^+=k} D(d\chi) \int \chi(da)a^j(1).$$

Therefore (cf. for instance [11, 4.3.8.8°]) the left-hand derivative of F in $x=1$ exists. Hance $F(1-x) = O(x)$, $x \to 0$. Consequently

$$| \int D(d\chi)(f_n(0))^{\chi^+-1} \int \chi(da)a^j(1) |$$
$$= | \sum_{k>0} (f_n(0))^{k-1} \int_{\chi^+=k} D(d\chi) \int \chi(da)a^j(1) |$$
$$= F(1-z_n) = O(z_n) = O(n^{-1}), \quad n \to \infty.$$

In fact, (1) yields (cf. [10, theorem 1.9.1.]) $z_n = O(n^{-1})$, $n \to \infty$.

P r o o f o f p r o p o s i t i o n 10.1. If ν is any distribution on $\mathfrak{A}_1$ then $\nu^0 = \nu(a - a^l \in (.))$ is the distribution of the error of the linearization. Because of 10.4. it suffices to show that for any $\varepsilon > 0$

$$\mathcal{C}_{n-1}(\nu_n^0 * \cdots * \nu_1^0)(\| a \| > \varepsilon)$$

converges to zero as $n \to \infty$. Clearly this expression can be estimated from above by

$$\leq \sum_{n \geq i \geq 1} \mathcal{C}_{n-1}(\nu_n^0 * \cdots * \nu_1^0)(\sup_{t \in [n-i/n,\, n-i+1/n]} \| a(t) \| > \varepsilon)$$
$$= \sum_{n \geq i \geq 1} \nu_i^0(\| a \| > \varepsilon\sqrt{n}).$$

Using $\| a - a^l \| \leq \| a \| + \| a^l \| \leq 2 \| a \|$, $a \in A_1$, and 9.1. we can continue with

$$\leq \sum_{n \geq i \geq 1} \nu_i(2 \| a \| > \varepsilon\sqrt{n}) \leq cn\lambda_D(2 \| a \| > \varepsilon\sqrt{n})$$
$$= cn\,\lambda_D(4\varepsilon^{-2} \| a \|^2 > n) \leq c\,4\,\varepsilon^{-2} \int_{4\varepsilon^{-2}\| a \|^2 > n} \lambda_D(da) \| a \|^2,$$

which tends to zero by (2) as $n \to \infty$.

11. The convergence of traces. We start with the following "source time theorem" for critical Galton-Watson processes (cf. [9; 3]).

11.1. *For all ε in* [0,1)

$$\sum_{m\leq[\varepsilon n]} r_m^n \xrightarrow[n\to\infty]{} \varepsilon .$$

Consequently, the "length" of the trunk of the reduced tree φ_n distributed by $D^{(n)}$ is nearly equidistributed on $\{0,\ldots,n\}$ if n is large.

Proof. At first 6.2. yields for any natural number n

$$r_m^n = \prod_{n\geq t\geq n-[\varepsilon n]} {}_tD(\chi^+=1)\, r^{n-[\varepsilon n]-1}_{m-[\varepsilon n]-1}, \quad [\varepsilon n]<m\leq n.$$

Hence

$$q^n =_{\mathrm{df}} \sum_{[\varepsilon n]<m\leq n} r_m^n = \prod_{n\geq i\geq n-[\varepsilon n]} {}_iD(\chi^+=1).$$

Clearly,

$$\begin{aligned}{}_iD(\chi^+=1) &= (\mathcal{Q}_{z_{i-1}} D)(\chi^+=1 \mid \chi^+\neq 0)\\ &= (1-\textstyle\int D(d\chi)(1-z_{i-1})^{\chi^+})^{-1} \int D(d\chi)\int \chi(dx) z_{i-1}(1-z_{i-1})^{\chi^+-1}\\ &= z_i^{-1} z_{i-1} f'(f_{i-1}(0))\end{aligned}$$

and

$$\begin{aligned} f'(f_{i-1}(0)) &\leq (f_i(0)-f_{i-1}(0))^{-1}(f(f_i(0))-f(f_{i-1}(0)))\\ &= (z_{i-1}-z_i)^{-1}(z_i-z_{i+1}),\end{aligned}$$

which yield together

$${}_iD(\chi^+=1) \leq \frac{z_{i-1}}{z_i}\,\frac{z_i-z_{i+1}}{z_{i-1}-z_i}.$$

Thus

$$q^n \leq \frac{z_{n-[\varepsilon n]-1}}{z_n}\,\frac{z_n-z_{n+1}}{z_{n-[\varepsilon n]-1}-z_{n-[\varepsilon n]}}.$$

Using $f(1-x)=1-x+2^{-1}f''(1-)x^2+o(x^2)$, $x\to 0+$, we obtain for $n\to\infty$

$$z_n-z_{n+1} = f(1-z_n)-(1-z_n) = 2^{-1}f''(1-)z_n^2+o(z_n^2),$$

which implies now

$$q^n \leq \frac{z_{n-[\varepsilon n]-1}}{z_n}\,\frac{2^{-1}f''(1-)z_n^2+o(z_n^2)}{2^{-1}f''(1-)z^2_{n-[\varepsilon n]-1}+o(z^2{n-[\varepsilon n]-1})}.$$

But (cf. [10, theorem 2.2.4])

$$z_n(z_{n-[\varepsilon n]-1})^{-1} \xrightarrow[n\to\infty]{} 1-\varepsilon,$$

so that

$$\limsup_{n\to\infty} q^n \leq 1-\varepsilon .$$

On the other side $f'(f_i(0)) \geq (f_i(0)-f_{i-1}(0))^{-1}(f(f_i(0))-f(f_{i-1}(0)))$, which yields on a similar way

$$\liminf_{n\to\infty} q^n \geq 1-\varepsilon .$$

Let β_t, $t\in[0,1)$, be the distributions introduced in 7. Then the conditionally homogeneous cluster field $|Q_{\beta(\cdot)}|$ in A makes sense.

11.2. *We have*

$$\mathcal{C}_{n-1}|^nW| \longrightarrow |Q_{\beta(\cdot)}|.$$

Proof. For each natural number n, let a_n be in $A_{\mathcal{C}_{n-1}|^nW|}\cap A_{[0,1)}$ converging to some $a_0\in A_t\subseteq A_{[0,1)}$. Using the same notations as in the proof of 8.2., we obtain

$$|^nW|_{(C_na_n)}=|^nW|_{(b_n)}={}^nW_{(k_n)}(T_{b_n}\chi\in(\cdot))$$
$$=\sum_{0\le i\le n-k_n} r_i^{n-k_n}(\delta_{b_n}*\nu_{n-k_n}*\cdots*\nu_{n-k_n-(i-1)})(\delta_a\in(\cdot)).$$

Therefore we have

$$(\mathcal{C}_{n-1}|^nW|)_{(a_n)}=\mathcal{C}_{n-1}(|^nW|_{(C_na_n)})$$
$$=\sum_{0\le i\le n-k_n} r_i^{n-k_n}(\delta_{a_n}*\mathcal{C}_{n-1}(\nu_{n-k_n}*\cdots*\nu_1))(\delta_{K_{n^{-1}(k_n+i)}a}\in(\cdot)),$$

where $K_{n^{-1}(k_n+i)}$ are killing operators on A_1 introduced in 4.

For each n,

$$r^{n,k_n}=\sum_{0\le i\le n-k_n} r_i^{n-k_n}\delta_{n^{-1}(k_n+i)}$$

is a distribution on the interval $[n^{-1}k_n,1]$. Moreover, by 11.1., the sequence (r^{n,k_n}) converges weakly to the uniform distribution μ_t on the interval $[t,1]$.

In view of 10.1. and 4.6.

$$\gamma_n=_{\mathrm{df}}\mathcal{C}_{n-1}(\nu_{n-k_n}*\cdots*\nu_1)=\mathcal{C}_{1-n^{-1}k_n}\mathcal{C}_{(n-k_n)^{-1}}(\nu_{n-k_n}*\cdots*\nu_1)$$

converges to $\mathcal{C}_{1-t}\beta=\beta(K_{1-t}\,a\in(\cdot))$.

Applying those two convergence relations, 3.1., and 4.2. (and the separability of A) we are able to conclude that

$$(\mathcal{C}_{n-1}|^nW|)_{(a_n)}=(r^{n-k_n}\times(\delta_{a_n}*\gamma_n))(\delta_{K_xa}\in(\cdot))$$
$$\xrightarrow[n\to\infty]{}(\mu_t\times(\delta_{a_0}*\mathcal{C}_{1-t}\beta)(\delta_{K_xa}\in(\cdot))$$
$$=(\mu_t\times\beta)(\delta_{K_x(a_0\oplus K_{1-t}a)}\in(\cdot))=(\mu_t\times\beta)(T_{a_0}\delta_{K_{x-t}a}\in(\cdot))=Q_{\beta_t}(T_{a_0}\varphi\in(\cdot)).$$

12. Proof of theorem 7.1. By 11.2., the convergence relation is valid for $k=0$. We assume that is true for k, i. e.

$$P_{n,k}=_{\mathrm{df}}\mathcal{C}_{n-1}(({}^nW_{(0)})_{(|^nW|\circ|^nS|)^k})\xrightarrow[n\to\infty]{}(Q_{\beta_0})_{|B|^k}.$$

By 8.2. $\mathcal{C}_{n-1}|^nS|\longrightarrow\tau$ and by 11.2. $\mathcal{C}_{n-1}|^nW|\longrightarrow|Q_{\beta(\cdot)}|$, and we have $|Q_{\beta(\cdot)}|\circ\tau=|B|$. Hence, applying two times theorem 2.1., we obtain

$$P_{n,k+1}=((P_{n,k})_{\mathcal{C}_{n-1}|^nS|})_{\mathcal{C}_{n-1}|^nW|}$$
$$\xrightarrow[n\to\infty]{}((Q_{\beta_0})_{|B|^k})_{|B|}=(Q_{\beta_0})_{|B|^{k+1}},$$

and 7.1. holds by induction.

Appendix. Condition (3) can be dropped. To demonstrate this we prove lemma 10.4. without using (3) as follows.

In view of 10.3. we have only to show that

$$\int \mathcal{C}_{n-1}(\nu_n^l * \cdots * \nu_1^l)(da)a \longrightarrow 0.$$

Because of

$$\left\| \int \mathcal{C}_{n-1}(\nu_n^l * \cdots * \nu_1^l)(da)a \right\| \leqq n^{-1/2} \sum_{n \geqq i \geqq 1} \left\| \int \nu_i(da)a(1) \right\|$$

it suffices to prove that for any j-th component a^j of a

$$n^{-1/2} \sum_{n \geqq i \geqq 1} \left| \int \nu_{i+1}(da)a^j(1) \right| \longrightarrow 0.$$

Setting for all natural numbers k

$$c_k = \int_{\chi^+ = k} D(d\chi) \int \chi(da)a^j(1)$$

and in a usual way $c_k = c_k^+ - c_k^-$ and writing formally

$$F(x) = \sum_{k>0} c_k x^{k-1}$$

we get a power series F satisifying by (2)

$$F(1) = \int \lambda_D(da)a^j(1) = 0,$$

so that F converges absolutely for $x \in (0, 1]$. Thus, the expressions

$$F_+(x) = \sum_{k>1} c_k^+(1 - x^{k-1}), \quad F_-(x) = \sum_{k>1} c_k^-(1 - x^{k-1}), \quad x \in (0, 1]$$

make sense and we have $F = F_- - F_+$. Then

$$|F(x)| \leqq \sum_{k>1} |c_k|(1 - x^{k-1}) =_{\mathrm{df}} G(x), \quad x \in (0, 1].$$

Now 9.1. yields for each natural number i

$$\begin{aligned} \left| \int \nu_{i+1}(da)a^j(1) \right| &\leqq c \left| \int D(d\chi)(f_i(0))^{\chi^+ - 1} \int \chi(da)a^j(1) \right| \\ &= c\,|F(f_i(0))| \leqq cG(f_i(0)) \end{aligned}$$

and it suffices to prove that

$$n^{-1/2} \sum_{n \geqq i \geqq 1} G(f_i(0)) \rightarrow 0.$$

By (1) there is a constant $m \geqq 1$ (cf. [10, theorem 2.2.4.]) such that $iz_i \leqq m$ holds for all natural numbers i. Hence $f_i(0) = 1 - z_i \geqq 1 - i^{-1}m$ and by the monotony of G

$$G(f_i(0)) \leqq G(1 - i^{-1}m), \quad i \geqq m.$$

Thus

$$\sum_{n \geqq i \geqq m} G(f_i(0)) \leqq \sum_{n \geqq i \geqq m} \sum_{k>1} |c_k|(1 - (1 - i^{-1}m)^{k-1}).$$

By $(1 - i^{-1}m)^{k-1} \geqq 1 - (k-1)i^{-1}m$ the last inequality can be continued with

$$\leqq \sum_{n \geqq i \geqq m} \Big(\sum_{1 < k \leqq i} |c_k|(k-1)i^{-1}m + \sum_{k>i} |c_k| \Big).$$

Now

$$\sum_{1<k\leq i} |c_k|(k-1)=\sum_{1<k} |c_k| \min\{k-1, i\}-i\sum_{k>i} |c_k|$$

and by $m\geq 1$ we can further continue with

$$\leq m \sum_{n\geq i\geq m} i^{-1} \sum_{1<k} |c_k| \min\{k-1, i\}=m \sum_{n\geq i\geq m} i^{-1} \sum_{1\leq l\leq i} \sum_{l<k} |c_k|.$$

It suffices to show that the last expression is an $o(n^{1/2})$ as $n\to\infty$. We have

$$(\sum_{l<k} |c_k|)^2\leq(\sum_{l<k} \int_{\chi^+=k} D(d\chi)\int\chi(da) \, |a^j(1)|)^2$$

$$\leq(\int_{\chi^+>l} D(d\chi)\int\chi(da)\,\|a(1)\|)^2\leq \int_{\chi^+>l} D(d\chi)\int\chi(da)\,\|a(1)\|^2 \int_{\chi^+>l} D(d\chi)\chi^+.$$

The last integral can be estimated from above by

$$l^{-1}\int_{\chi^+>l} D(d\chi)(\chi^+)^2,$$

which by (1) is an $o(l^{-1})$ as $l\to\infty$, so that by (2) we finally have

$$\sum_{l<k} |c_k|=o(l^{-1/2}), \quad l\to\infty.$$

If (d_l) is any sequence of non-negative numbers satisfying $d_l=o(l^{-1/2})$ as $\to\infty$ then

$$\sum_{1\leq l\leq i} d_l=o(\sum_{1\leq l\leq i} l^{-1/2}), \quad i\to\infty.$$

Since

$$\sum_{1\leq l\leq i} l^{-1/2}=O(i^{1/2}), \quad i\to\infty$$

we are able to conclude that

$$\sum_{1\leq l\leq i} d_l=o(i^{1/2}).$$

Applying this argument two times we find

$$\sum_{n\geq i\geq m} i^{-1} \sum_{1\leq l\leq i} \sum_{l<k} |c_k|=o(n^{1/2}), \quad n\to\infty.$$

Acknowledgement. We would like to thank O. Kallenberg in Göteborg who called in question condition (3).

REFERENCES

1. P. Billingsley. Convergence of probability measures. New York, 1968.
2. K. Fleischmann, U. Prehn. Subkritische räumlich homogene Verzweigungsprozesse. *Math. Nachr.*, **70**, 1975, 231—250.
3. K. Fleischmann, R. Siegmund-Schultze. The structure of reduced critical Galton-Watson processes. *Math. Nachr.*, 79, 1977, 233—241.

4. L. G. Gorostiza, A. R. Moncayo. Invariance principle for random processes on Galton-Watson trees. *J. Math. Anal. Appl.*, **60**, 1977, 461—476.
5. O. Kallenberg. Random measures. Berlin, 1975.
6. K. Matthes, J. Kerstan, J. Mecke. Infinitely divisible point processes. Chichester, 1978.
7. J. Kerstan, K. Matthes, U. Prehn. Verallgemeinerungen eines Satzes von Dobrushin II. *Math. Nachr.*, **51**, 1971, 149—188.
8. U. Prehn. Ein Grenzwertsatz für subkritische Verzweigungsprozesse mit endlich vielen Typen von Teilchen II. *Math. Nachr.* (to appear).
9. А. М. Зубков. Предельные распределения расстояния до ближайшего общего предка. *Теория вероятностей и ее применения*, **20**, 1975, 614—623.
10. Б. А. Севастьянов. Ветвящиеся процессы. Москва, 1971.
11. Г. М. Фихтенгольц. Курс дифференциального и интегрального исчисления, том II. Москва, 1972.

Akademie der Wissenschaften der DDR
Zentralinstitut für Mathematik und Mechanik
DDR-108 Berlin, Mohrenstrasse 39

Received 26. 5. 1977

Discussion.

J. F. C. Kingman (Oxford): The interesting results of K. Fleischmann and R. Siegmund-Schultze for critical branching processes lead me to wonder whether similar features are exhibited by another model of considerable biological interest. In a critical branching process the expected total number in each generation is constant, but the actual total fluctuates considerably, and eventually vanishes. But in many situations it makes more sense to think of the population size as being determined by external factors rather than by individual reproductive propensities.

As a simple illustration, consider a process in which each generation has fixed size N. The family structure is defined by supposing that the probability that the N particles in a particular generation have respectively $d_1, d_2, \dots, d_N$ daughters in the next generation is given by a multinomial formula $N!\,N^{-N}/d_1!\,d_2!\dots d_N!$, $d_1+d_2+\dots+d_N=N$. This determines a family tree resembling a critical branching process in some respects, but different in others. For example, the probability that all the particles in a given generation have a single common ancestor t generations back is near 1 when t/N is large, lying in fact between $1-3(1-N^{-N})^t$ and $1-(1-N^{-1})^t$.

The problem of spatial distribution of successive generations, when a daughter particle is randomly displaced from its parent, has been studied by P. A. P. Moran (and more recently by myself, see *Proc. Roy. Soc. A*, **351**, 1976, 19—31). It might well be possible to apply the methods of the present paper to give deep results for this model, which in some biological contexts is less unrealistic than the branching process model.

P. Mandl (Prague): The results of the paper are of interest for practical applications. Hence it would be useful to write a version using an intuitive process description instead of measure constructions. Such a treatment would also be welcome for probabilists. The limit process, the binary branching Brownian motion on [0, 1), in view of the results of K. Fleischmann and R. Siegmund-Schultze, is an important stochastic model. It seems that its investigation in detail is not yet done. For this aim perhaps, the proved invariance principle could be used to go over to the limit starting with more simple processes.

C. J. Mode (Philadelphia): A rather large number of interesting limit theorems, conditioned on nonextinction, are associated with the critical Galton-Watson process. In the simple Galton-Watson process, an exponential law appears; while in critical spatially homogeneous branching processes, diffusion processes appear under suitable contractions of space and time. The authors, using an invariance principle combined with advanced analysis, show that a certain binary branching Brownian motion model arises, a model that is of some interest in its own right. Throughout the paper the authors demonstate admirable mathematical expertise.

Potential applications of the theory become conspicuous, however, by their absence. The model is sufficiently general to include random walks as a special case. Consequently, the potential for application seems large. Yet, the paper does not contain one example of demonstrable interest in physics, biology or the social sciences. It is the opinion of this reviewer that the inclusion of well thought out applications of the theory in the paper would have greatly increased its intrinsic interest.