Elect. Comm. in Probab. **13** (2008), 540–547

ELECTRONIC COMMUNICATIONS in PROBABILITY

GENERAL BRANCHING PROCESSES CONDITIONED ON EXTINCTION ARE STILL BRANCHING PROCESSES

PETER JAGERS[1]
Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-412 96 Gothenburg, Sweden
email: `jagers@chalmers.se`

ANDREAS N. LAGERÅS[2]
Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden
email: `andreas@math.su.se`

Submitted April 28, 2008, *accepted in final form* October 9, 2008.

AMS 2000 Subject classification: 60J80
Keywords: Supercritical, Subcritical, Extinction, Multi-type branching process, General branching process, Crump-Mode-Jagers process.

Abstract

It is well known that a simple, supercritical Bienaymé-Galton-Watson process turns into a subcritical such process, if conditioned to die out. We prove that the corresponding holds true for general, multi-type branching, where child-bearing may occur at different ages, life span may depend upon reproduction, and the whole course of events is thus affected by conditioning upon extinction.

1 Introduction

The theory of branching processes was born out of Galton's famous family extinction problem. Later, interest turned to populations not dying out and their growth and stabilisation. In more recent years, extinction has retaken a place in the foreground, for reasons from both conservation and evolutionary biology. The time and path to extinction of subcritical general populations was studied in [4]. Here, time structure is crucial, and life spans and varying bearing ages cannot be condensed into simple, generation counting Bienaymé-Galton-Watson processes. Thus, the question arises whether (non-critical) general branching populations (also known as Crump-Mode-Jagers, or CMJ, processes) bound for extinction must behave like subcritical populations.

We answer this in the affirmative: a general, multi-type branching process conditioned to die out, remains a branching process, but one almost surely dying out. If the original process was supercritical but with a strictly positive risk of extinction, the new process is subcritical.

[1]SUPPORTED BY THE SWEDISH RESEARCH COUNCIL
[2]SUPPORTED BY THE SCIENCE FACULTY OF THE UNIVERSITY OF GOTHENBURG THROUGH THE PLATFORM FOR THEORETICAL BIOLOGY

Formulated in such a loose manner, this fact belongs to the folklore of branching, but actually it has been proved only for Bienaymé-Galton-Watson processes, [1], p. 52. A moment's afterthought tells us that it remains true for age-dependent branching processes of the Bellman-Harris type, where individuals have i.i.d. life spans, and split into i.i.d. random numbers of children, independently of life span, the time structure thus not being affected by the conditioning.

But what if the flow of time is no longer independent of reproduction? Even the simplest case, that of a splitting reproduction at death, but not independently of age at death/splitting, would seem to offer difficulties, and the same certainly applies to the more realistic general processes where reproduction occurs as a point process during life, thus mimicking the yearly births of wildlife, or the even more erratic reproduction pattern of humans.

The conceptual framework is intuitive. Starting from an Eve, individuals live and give birth independently of one another. At birth each individual inherits a *type* from her mother. The type, in its turn determines the probability measure over all possible life careers, including a life span and a marked point process which reports the successive ages at bearing, and the types of children at the various bearings. Note that multiple births are not excluded. The branching property can be summarised into the fact that given her type and birth time, the daughter process of any individual born is independent of all individuals not in her progeny (into which she herself is included).

We set out to prove that this branching property also holds for processes conditioned to die out. Initially, we shall not mention supercriticality, and only ask that the probability of extinction is non-zero for any starting type. (If that probability is one, the conditioning does not change anything!) Largely, the proof is a matter of conceptual clarity or discipline, which unfortunately forces us into the somewhat burdensome notation of probabilities on tree spaces, obscuring the essential simplicity of the matter.

The main idea behind the proof is, however, easily outlined. Indeed, consider an individual, and condition upon her progeny ultimately dying out. Her own life career is then affected precisely through her only being able to have daughters whose progeny in their turn must ultimately face extinction. In all other respects her life is independent of all others, once her type is given. This reestablishes the branching character, but with a suitably amended probability measure over her life career, which clearly is non-supercritical in the sense that the probability of ultimate extinction is one, from any starting type that can be realised.

2 Notation

Throughout this paper we use the notation of [3], which may also serve as a reference for the reader interested in further properties of general multi-type branching processes.

2.1 The Ulam-Harris family space

We choose to work within the classical Ulam-Harris framework identifying individuals with sequences of natural numbers so that $x = (x_1, x_2, \dots, x_n)$ denotes the x_nth child of the ... of the x_2th child of the x_1th child of the ancestor. The ancestor is denoted by an "empty" sequence e (mnemonic for "empty" or "Eve"), and the set of all possible individuals is

$$\mathbb{T} = e \cup \bigcup_{n \in \mathbb{N}} \mathbb{N}^n.$$

The concatenation of $x, y \in \mathbb{T}$ is xy, and thus $ex = xe = x$ for all $x \in \mathbb{T}$.

For any $e \neq x = (x_1, x_2, \dots, x_n)$ x's *mother* is $mx = (x_1, \dots, x_{n-1})$, her *rank* in the sibship is $rx = x_n$, and x's *generation* $g(x) = n$. We agree that $me = re = e$ and $g(e) = 0$. Hence, $mxrx = x$ for $x \in \mathbb{T}$, and m can be iterated so that $m^n x$ is x's nth grandmother, provided $g(x) > n$.
Clearly x *stems from* y, usually written $x \succeq y$, if $m^n x = y$ for some $n \in \mathbb{N} \cup \{0\}$, or equivalently if there exists a $z \in \mathbb{T} : x = yz$. In this terminology, x stems from herself, $x \preceq x$. In other words, $(\mathbb{T}, \preceq)$ is a partially ordered set (a semilattice). We define

$$x \sim y \text{ if } x \succeq y \text{ or } x \preceq y,$$

i.e. x and y are in direct line of descent. ($\sim$ is not an equivalence relation.)
For $A, B \subseteq \mathbb{T}, x \in \mathbb{T}$, we write $x \succeq A$ if there exists a $y \in A$ such that $x \succeq y$, and $A \preceq B$ if $x \succeq A$ for all $x \in B$. The *progeny* of $A \subseteq \mathbb{T}$ is defined as $\mathrm{Pr} A = \{x \in \mathbb{T} : x \succeq A\}$.
We call a set $L \subset \mathbb{T}$ a stopping line, or *line* for short, if no two members of L are in direct line of descent:

$$x, y \in L, x \neq y \Rightarrow x \not\sim y.$$

We say that a line L is a *covering line* if for all $x \in \mathbb{T}$ there exists a $y \in L$ such that $x \sim y$.

2.2 Life space and population space

Let $(\Omega_\ell, \mathscr{A}_\ell)$ be a *life space* so that $\omega \in \Omega_\ell$ is a possible life career of individuals. Any individual property, such as mass at a certain age or life span, is viewed as a measurable function (with respect to the σ-algebra $\mathscr{A}_\ell$) on the life space. This should be rich enough to support, at least, the functions $\tau(k), \sigma(k)$ for $k \in \mathbb{N}$. Here $\tau(k) : \Omega_\ell \to \mathbb{R}_+ \cup \{\infty\}$ is the mother's age at the kth child's birth, $0 \leq \tau(1) \leq \tau(2) \leq \cdots \leq \infty$. If $\tau(k) = \infty$, then the kth child is never born. $\sigma(k) : \Omega_\ell \to \mathscr{S}$ is the child's type, obtained at birth. The type space $\mathscr{S}$ has a (countably generated) σ-algebra $\mathscr{S}$. The whole reproduction process is then the marked point process ξ with

$$\xi(A \times B) = \#\{k : \sigma(k) \in A, \tau(k) \in B\}$$

for $A \in \mathscr{S}, B \in \mathscr{B}$, the Borel algebra on $\mathbb{R}_+$.
The population space is defined as $(\Omega, \mathscr{A}) = (\mathscr{S} \times \Omega_\ell^{\mathbb{T}}, \mathscr{S} \otimes \mathscr{A}_\ell^{\mathbb{T}})$. U_M is the projection $\Omega \to \Omega_\ell^M$, for $M \subseteq \mathbb{T}$. For simplicity $U_x = U_{\{x\}}$ and similarly $\mathrm{Pr}\, x = \mathrm{Pr}\{x\}$. The following σ-algebras are important:

$$\mathscr{F}_L = \mathscr{S} \otimes \sigma(U_x : x \not\succeq L) = \mathscr{S} \otimes \sigma(U_x : x \notin \mathrm{Pr}\, L),$$

for $L \subseteq \mathbb{T}$. If $L \preceq M$ then $\mathrm{Pr}\, L \supseteq \mathrm{Pr}\, M$ and thus $\mathscr{F}_L \subseteq \mathscr{F}_M$. Therefore $(\mathscr{F}_L : L \subseteq \mathbb{T})$ is a filtration under $\preceq$. In the usual manner, the definition of the σ-algebras $\mathscr{F}_L$ can be extended to σ-algebras of events preceding random lines $\mathscr{L}$ which are optional in the sense that events $\{\mathscr{L} \preceq L\} \in \mathscr{F}_L$ [3].
Functions $\xi, \tau(k)$ and $\sigma(k)$ were defined on the life space but we want to be able to speak about these quantities pertaining to a given $x \in \mathbb{T}$. We write

$$\begin{aligned}
\xi_x &= \xi \circ U_x && (x\text{'s reproduction process}),\\
\tau_x &= \tau(rx) \circ U_{mx} && (x\text{'s mother's age at } x\text{'s birth}),\\
\sigma_x &= \sigma(rx) \circ U_{mx} && (x\text{'s type}).
\end{aligned}$$

Note the difference between $\tau(k)$ and τ_k, $\sigma(k)$ and σ_k, for $k \in \mathbb{N} \subset \mathbb{T}$.

Finally, the process is anchored in real time by taking Eve to be born at time 0, and letting later birth times $t_x, x \in \mathbb{T}$ be recursively determined by $t_e = 0$ and $t_x = t_{mx} + \tau_x$ for $e \neq x \in \mathbb{T}$. The meaning of $t_x = \infty$ is that x is never born, so that

$$\mathscr{R} = \{x \in \mathbb{T} : t_x < \infty\}$$

is the set of *realised* individuals. This set is optional, $\mathscr{F}_{L\cap\mathscr{R}}$ is well defined [3], and so is the σ-algebra $\mathscr{F}_{\mathscr{R}}$ of events pertaining only to realised individuals. The probability space restricted to such events is that where a branching processes really lives, *cf.* [5], [2].

2.3 The probability measure and branching property

The setup is that for each $s \in \mathscr{S}$ there is a probability measure $P(s,\cdot)$ on the life space $(\Omega_\ell, \mathscr{A}_\ell)$, such that the function $s \to P(s,A)$ is measurable with $A \in \mathscr{A}_\ell$. For any $s \in \mathscr{S}$ this kernel (the *life kernel*) defines a *population probability* measure $\mathbb{P}_s$ on $(\Omega, \mathscr{A})$ with an ancestor of type $\sigma_e = s$ and such that given σ_x, x's life will follow the law $P(\sigma_x, \cdot)$ independently of the rest of the process, see [3].

Indeed, the basic branching property of the whole process can be characterised by a generalisation of this in terms of the mappings

$$S_x = (\sigma_x, U_{\mathrm{Pr}\,x}) : \mathscr{S} \times \mathbb{T} \to \mathscr{S} \times \mathbb{T},$$

which renders x the new Eve. Let $T_x = S_x^{-1}$ and $\{A_x : x \in L\} \subseteq \mathscr{A}$. Then,

$$\mathbb{P}_s\left(\bigcap_{x\in L} T_x A_x \middle| \mathscr{F}_L\right) = \prod_{x\in L} \mathbb{P}_{\sigma_x}(A_x) \qquad (\mathbb{P}_s\text{-a.s.}),$$

for lines L. This remains true for optional lines and in particular

$$\mathbb{P}_s\left(\bigcap_{x\in L\cap\mathscr{R}} T_x A_x \middle| \mathscr{F}_{L\cap\mathscr{R}}\right) = \prod_{x\in L\cap\mathscr{R}} \mathbb{P}_{\sigma_x}(A_x) \qquad (\mathbb{P}_s\text{-a.s.}), \tag{1}$$

where the intersection over the empty set is taken to be Ω and the empty product is ascribed the value one. The interpretation is that the daughter processes of all realised individuals x in a line are independent given the prehistory of the line with the population probability measure $\mathbb{P}_{\sigma_x}$, the only dependence upon the past thus being channelled through the type σ_x and the placing in time t_x. This is the branching property. We shall see that it remains true for processes which are bound to die out.

3 Conditioning on extinction

Denote by E the event that the branching process starting from Eve dies out, i.e. that $\mathscr{R}$ has only a finite number of elements. Let $q_s = \mathbb{P}_s(E)$ and E_x be the event that the branching process starting from x dies out, $E_x = T_x E$. Write $\tilde{\mathbb{P}}_s(\cdot) = \mathbb{P}_s(\cdot|E)$, which clearly only makes sense for $s \in \mathscr{S}$ such that $q_s > 0$, and let $\tilde{\mathbb{E}}_s$ denote expectation with respect to $\tilde{\mathbb{P}}_s$.

3.1 Main result

Theorem 1. *Any branching process conditioned on extinction remains a branching process, but with extinction probability one. Its life kernel is* $\tilde{P}(s,A) := \tilde{\mathbb{P}}_s(\mathscr{S} \times A \times \Omega_\ell^{\mathbb{T}\setminus\{e\}})$ *for* $A \in \mathscr{A}_\ell$*. Thus, for any covering lines* L *and* $\{A_x : x \in L\} \subseteq \mathscr{A}$

$$\tilde{\mathbb{P}}_s\left(\bigcap_{x\in L\cap\mathscr{R}} T_x A_x \middle| \mathscr{F}_{L\cap\mathscr{R}}\right) = \prod_{x\in L\cap\mathscr{R}} \tilde{\mathbb{P}}_{\sigma_x}(A_x) \qquad (\mathbb{P}_s\text{-a.s.}). \tag{2}$$

Furthermore, the Radon-Nikodym derivative $d\tilde{\mathbb{P}}_s/d\mathbb{P}_s$ *with respect to the* σ*-algebra* $\mathscr{F}_{L\cap\mathscr{R}}$ *is given by*

$$\left.\frac{d\tilde{\mathbb{P}}_s}{d\mathbb{P}_s}\right|_{\mathscr{F}_{L\cap\mathscr{R}}} = \frac{1}{q_s}\prod_{x\in L\cap\mathscr{R}} q_{\sigma_x}. \tag{3}$$

Proof. First, note that

$$E = \bigcap_{x\in L\cap\mathscr{R}} T_x E, \tag{4}$$

for covering lines L. Indeed, since $\{L\cap\mathscr{R}=\emptyset\} \subseteq E$ and intersection over an empty index set yields the full space,

$$\begin{aligned}
E &= (E\cap\{L\cap\mathscr{R}=\emptyset\}) \cup (E\cap\{L\cap\mathscr{R}\neq\emptyset\}) \\
&= \left(\{L\cap\mathscr{R}=\emptyset\} \cap \bigcap_{x\in L\cap\mathscr{R}} T_x E\right) \cup \left(\{L\cap\mathscr{R}\neq\emptyset\} \cap \bigcap_{x\in L\cap\mathscr{R}} T_x E\right) \\
&= \bigcap_{x\in L\cap\mathscr{R}} T_x E.
\end{aligned}$$

The branching property (1) implies that

$$\mathbb{P}_s(E|\mathscr{F}_L) = \prod_{x\in L\cap\mathscr{R}} q_{\sigma_x} = \mathbb{P}_s(E|\mathscr{F}_{L\cap\mathscr{R}}). \tag{5}$$

Hence, for any covering line L and $A \in \mathscr{F}_L$

$$\tilde{\mathbb{P}}_s(A) = \mathbb{E}_s\left[\mathbb{P}_s(E|\mathscr{F}_L);A\right]/q_s = \mathbb{E}_s\left[\prod_{x\in L\cap\mathscr{R}} q_{\sigma_x};A\right]/q_s,$$

and thus (3) holds. Equations (4), (1) and (5) yield

$$\begin{aligned}
\mathbb{P}_s\left(E\cap\bigcap_{x\in L\cap\mathscr{R}} T_x A_x \middle| \mathscr{F}_{L\cap\mathscr{R}}\right) &= \mathbb{P}_s\left(\bigcap_{x\in L\cap\mathscr{R}} T_x(A_x\cap E) \middle| \mathscr{F}_{L\cap\mathscr{R}}\right) \\
&= \prod_{x\in L\cap\mathscr{R}} q_{\sigma_x}\tilde{\mathbb{P}}_{\sigma_x}(A_x) \\
&= \mathbb{P}_s(E|\mathscr{F}_{L\cap\mathscr{R}}) \prod_{x\in L\cap\mathscr{R}} \tilde{\mathbb{P}}_{\sigma_x}(A_x),
\end{aligned}$$

and (2) follows. ☐

Remark 1. With $X := \xi_e(\mathscr{S} \times \mathbb{R}_+)$ being Eve's total offspring and $L = \mathbb{N}$, the first generation,

$$\left.\frac{d\tilde{\mathbb{P}}_s}{d\mathbb{P}_s}\right|_{\mathscr{F}_{\mathbb{N}\cap\mathscr{R}}} = \frac{1}{q_s} \prod_{k\in\mathbb{N}:k\le X} q_{\sigma_k}, \tag{6}$$

We thus obtain the conditioned life kernel

$$\tilde{P}(s,A) = \frac{1}{q_s} \int_A \prod_{k\in\mathbb{N}:k\le X(\omega)} q_{\sigma_k(\omega)} P(s,d\omega),$$

showing both that the number of children under the conditioning tends to be lower, and that the children tend to have types with higher extinction probabilities, which is quite reasonable.
For single-type processes with extinction probability q, equation (6) simplifies to

$$\left.\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\right|_{\mathscr{F}_{\mathbb{N}\cap\mathscr{R}}} = q^{X-1}, \tag{7}$$

and in the Bienaymé-Galton-Watson case the conditioned offspring distribution is thus

$$\tilde{\mathbb{P}}(X = k) = \mathbb{E}[q^{X-1}; X = k] = \mathbb{P}(X = k)q^{k-1},$$

in perfect agreement with [1, Theorem I.12.3].

Remark 2. As was pointed out to us by a referee, general multi-type branching processes are introduced in [3] as Markov fields on the set $\mathbb{T}$, and in this context $\tilde{\mathbb{P}}_s$ may be seen as an h-transform of $\mathbb{P}_s$, since the product in the right hand side of (3) is a harmonic function defined on lines $L \subseteq \mathbb{T}$.

Example 1. *(Sevastyanov processes)* Consider single-type Sevastyanov splitting processes, where individuals have a life span distribution G and at death split into k particles with probability $p_k(u)$, if the life span $L = u$. By (7) we conclude that

$$\tilde{\mathbb{P}}(X = k, L \in du) = \mathbb{E}[q^{X-1}; X = k, L \in du] = p_k(u)q^{k-1}G(du).$$

Hence

$$\tilde{G}(du) = \tilde{\mathbb{P}}(L \in du) = \sum_{k=0}^{\infty} p_k(u)q^{k-1}G(du)$$

and $\tilde{\mathbb{P}}(X = k, L \in du) = \tilde{p}_k(u)\tilde{G}(du)$, with

$$\tilde{p}_k(u) = \frac{p_k(u)q^k}{\sum_{i=0}^{\infty} p_i(u)q^i},$$

and the conditioned Sevastyanov process remains a Sevastyanov process with life span distribution $\tilde{G}$ and splitting probabilities $\tilde{p}_k(u)$.

Example 2. *(A two-type general branching process)* Consider a two-type process where life spans are exponentially distributed with expected value 1. Individuals of type 1 bear children of type 1 according to a Poisson process with intensity 3. Individuals of type 2 bear children according to a Poisson process with intensity 5 and each child is of type 1 or type 2 with equal probability independently of each other.

This process can also be reformulated due to the memorylessness of the exponential distribution. We interpret the birth of a child as the splitting of the mother into two new individuals (or particles), one being the child and one being the mother. Particles of type 1 have exponential life spans with expected value $\frac{1}{3}$ and at death either split into two new particles of type 1 with probability $p_{11} = \frac{2}{3}$ or leave no offspring with probability $p_{10} = \frac{1}{3}$. Particles of type 2 have exponential life spans with expected value $\frac{1}{5}$ and split into two particles of type 2 with probability $p_{22} = \frac{2}{5}$, two particles — one of either type — with probability $p_{21} = \frac{2}{5}$, or leave no offspring with probability $p_{20} = \frac{1}{5}$.
The extinction probabilities solve (taking expected values in (5) with $L = \mathbb{N}$)

$$\begin{aligned} q_1 &= p_{11}q_1^2 + p_{10} \\ q_2 &= p_{22}q_2^2 + p_{21}q_2q_1 + p_{20}, \end{aligned}$$

so

$$\begin{aligned} q_1 &= \frac{1}{2} \\ q_2 &= \frac{2-\sqrt{2}}{2} \doteq 0.293 \end{aligned}$$

and by (6) the conditional probabilities are

$$\begin{aligned} \tilde{p}_{11} &= \frac{1}{q_1} p_{11} q_1^2 = \frac{1}{3} \\ \tilde{p}_{10} &= \frac{1}{q_1} p_{10} = \frac{2}{3} \\ \tilde{p}_{22} &= \frac{1}{q_2} p_{22} q_2^2 = \frac{2-\sqrt{2}}{5} \\ \tilde{p}_{21} &= \frac{1}{q_2} p_{21} q_2 q_1 = \frac{1}{5} \\ \tilde{p}_{20} &= \frac{1}{q_2} p_{20} = \frac{2}{2-\sqrt{2}} \cdot \frac{1}{5} = \frac{2+\sqrt{2}}{5}. \end{aligned}$$

In terms of the original process, the new, conditioned process will have individuals of type 1 whose life spans are exponential with expectation $\frac{1}{3} \cdot \frac{1}{\tilde{p}_{10}} = q_1 = \frac{1}{2}$, and it will have individuals of type 2, with life spans also exponential, but the expected value $\frac{1}{5} \cdot \frac{1}{\tilde{p}_{20}} = \frac{1}{2+\sqrt{2}} = q_2 \doteq 0.293$. Children are still born according to Poisson processes with intensities 3 and 5, respectively. However, the proportion of types born by individuals of type 2 is changed: the probability of a child being of type 1 is now $\frac{\tilde{p}_{21}}{\tilde{p}_{21}+\tilde{p}_{22}} = \frac{1}{3-\sqrt{2}} \doteq 0.631$.

3.2 Super- and subcriticality

Finally, we address the question of whether a supercritical process conditioned on extinction is subcritical. In this section we leave the full generality of Theorem 1. This is partly because the notions of super- and subcriticality for general multi-type branching processes are quite involved: Several conditions on the life kernel are required in order to ascertain that the process is *Malthusian*, meaning roughly that the population grows or declines in an asymptotically exponential manner. The asymptotic growth rate is then given by the Malthusian parameter, and the process is called supercritical if this rate is positive, and subcritical if it is negative. We refer to [3] for details.

Let us consider a supercritical process such that also the conditioned process is Malthusian. Such processes clearly exist, e.g. single-type Bienaymé-Galton-Watson processes, see [1], but here we do not delve into what conditions are necessary for this to hold for general processes.

It is clear that the conditioned branching process has extinction probability one for any starting type but this would also be the case if the process were nontrivially critical, i.e. with zero growth rate.

In the single-type case it follows from (7) that the expected total number of children per individual in the conditioned process satisfies

$$\tilde{\mathbb{E}}[X] = \mathbb{E}[Xq^{X-1}] = f'(q),$$

in terms of the offspring generating function f of the embedded Bienaymé-Galton-Watson process. It is well known that $f'(q) < 1$ if the original process was supercritical, see [1], and we conclude that the conditioned process is subcritical.

For the multi-type case, besides requiring the unconditioned and conditioned processes to be Malthusian, we also need $q := \sup_{\mathscr{S}} q_s < 1$. As in the single-type case we will consider the embedded generation counting process. Let $X_n(A) = \operatorname{card}\{x \in \mathbb{N}^n \cap \mathscr{R} : \sigma_x \in A\}$ denote the number of individuals of type A in nth generation. Then

$$\tilde{\mathbb{E}}_s[X_n(\mathscr{S})] = \mathbb{E}_s\left[X_n(\mathscr{S})e^{\int_{\mathscr{S}} \log q_r X_n(dr)}\right]/q_s \le \mathbb{E}_s\left[X_n(\mathscr{S})q^{X_n(\mathscr{S})}\right]/q_s \to 0,$$

as $n \to \infty$ (by dominated convergence, since $X_n(\mathscr{S})$ must either tend to zero or to infinity). But the expected size of the embedded process tending to zero means exactly that the process is subcritical.

References

[1] ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes*. Springer-Verlag, New York. Die Grundlehren der mathematischen Wissenschaften, Band 196. MR0373040

[2] CHAUVIN, B. (1986) Sur la propriété de branchement *Annales de l'Institut H. Poincaré. Probabilités et Statistiques* **22**, 2, 233–236. MR0850758

[3] JAGERS, P. (1989). General branching processes as Markov fields. *Stochastic Processes and their Applications* **32**, 2, 183–212. MR1014449

[4] JAGERS, P., KLEBANER F. C., AND SAGITOV, S. (2007). On the path to extinction. *Proceedings of the National Academy of Science* **104**, 15, 6107–6111. DOI: 10.1073/pnas.0610816104

[5] NEVEU, J. (1986). Arbres et processus de Galton-Watson. *Annales de l'Institut H. Poincaré. Probabilités et Statistiques* **22**, 2, 199–207. MR0850765