Pergamon

Mathl. Comput. Modelling Vol. 23, No. 4, pp. 93–104, 1996

0895-7177/96 $15.00 + 0.00

S0895-7177(96)00005-2

Stochastic Processes Determined by a General Success-Breeds-Success Principle

L. EGGHE
LUC, Universitaire Campus, B-3590, Diepenbeek, Belgium*
and
UIA, Informatie- en Bibliotheekwetenschap, Universiteitsplein, 1, B-2610, Wilrijk, Belgium

R. ROUSSEAU
KIHWV, Zeedijk 101, B-8400 Oostende, Belgium*
and
UIA, Informatie- en Bibliotheekwetenschap, Universiteitsplein, 1, B-2610, Wilrijk, Belgium

(Received June 1994; accepted March 1995)

Abstract—The general "success-breeds-success" (SBS) principle as introduced in a previous paper extends the classical SBS principle in that the allocation of items over sources is determined by a more general rule than in the classical case. In this article we study the time evolution of the total number of sources, the average number of items per source and the number of sources with n items at time t, in the general SBS framework. Conditional as well as absolute expectations are calculated. Moreover, we investigate if and when these processes are martingales, supermartingales or submartingales. Stability results for the stochastic processes are obtained in the sense that we are able to determine when these processes converge. The article also studies the evolution of the expected average number of items per source.

Keywords—Success-Breeds-Success, Stochastic process, Martingale, Time evolution.

1. INTRODUCTION

In [1] we introduced a general "success-breeds-success" (SBS) principle extending the classical and well-known SBS-principle as described in [2–5] and others. Such a process generates information production processes (IPPs), i.e., generalized bibliographies, of sources producing items (e.g., authors writing articles, or journals publishing papers). For more information on IPPs the reader is referred to [6–9].

This general SBS-principle is determined as follows. An IPP is regulated by a parameter $t \in \mathbb{N}_0$. At every step $(t \to t+1)$ an item enters the system. Note that the parameter t denotes time as well as the number of items in the system. The introduction of a new item at time $t+1$ leads to the following alternatives:

(i) source creation: with a probability $\alpha(t) \in]0,1[$ this item is produced by a new source, i.e., a source that was not active or did not exist up to time t;

(ii) pure SBS: if the new item is produced by an already existing source (which occurs with a probability equal to $1-\alpha(t)$), there is a chance $x(t,n)$ that this item is produced by a source that has already n items ($n \leq t$, $n \in \mathbb{N}_0$). Of course,

$$\sum_{n=1}^{t} x(t,n) = 1.$$

*Permanent address.

In [1] the general SBS-principle has been studied in terms of the average probability, denoted as $E(P(t,n))$ that at time t, a source has $n \in \mathbb{N}_0$ items. The exact study of $E(P(t,n))$ as a function of n or t is very difficult; even in the classical case only approximate results are known. Using a so-called quasi-steady-state assumption we were able to generate several well-known frequency distributions as the outcome of an SBS-scheme based on (i) and (ii) [1].

This article studies the processes $T(t)$ (the number of sources at time t), $\mu(t)$ (the average number of items per source at time t) and, for every $n \in \mathbb{N}_0$, the processes $X_t(n)$ (the number of sources with n items at time t) for the general SBS-principle. Without any approximation, only exact probabilistic arguments and formulae are presented.

In the next section the first basic probability space (Ω, F, P) is constructed, and on this probability space the processes $T(t)$ and $\mu(t)$ are defined as stochastic processes (or adapted sequences). For the terminology and notation taken from probability theory, the reader is referred to [10–12].

The third section characterizes (super), (sub)martingale properties of $T(t)$ and $\mu(t)$. Such properties provide important information on the expected increase from time t to time $t+1$, and possibly also about the limit distributions $T_\infty(t)$ and $\mu_\infty(t)$. Convergence theorems about (super), (sub)martingales play a prominent role in these derivations [11].

In the fourth section we study a formula describing $E(\mu(t+1))$ as a function of $E(\mu(t))$. Further, a necessary and sufficient condition is derived in order to have an increasing sequence $(E(\mu(t)))_{t\in\mathbb{N}_0}$.

The last section constructs the probability space (Ω', Σ, P') on which the processes $X_t(n)$ act (one for each $n \in \mathbb{N}_0$). This probability space is a refinement of the first one, (Ω, F, P). Then the (super), (sub)martingale properties of $(X_t(n))_{t\in\mathbb{N}_0}$ are characterized and limit theorems are obtained. These are stability properties for the behavior of every $X_t(n)$, for t large ($n \in \mathbb{N}_0$ fixed). Note that nowhere assumptions on $x(t,n)$ are made.

SBS does not only occur in IPPs (in informetrics). Also in linguistics and computer science, SBS is important. We refer to [13,14] for applications of this (in the form of Zipf's law) in the estimation of program length and in speech recognition. We refer also to [9] for an application to storage and text retrieval in a computer.

2. THE STOCHASTIC PROCESSES $T(t)$ AND $\mu(t)$

Since both $T(t)$, the number of sources at time t, and $\mu(t)$, the average number of items per source at time t, are solely determined by the allocation of items to old and new sources—and not by which (old) source is active—these processes are determined by $\alpha(t)$, and not by the $x(t,n)$ (cf. condition (ii) in the definition of the general SBS-principle).

Hence, at each step ($t \to t+1$), the situation at time t switched over to one of two possible situations at time $t+1$: either the new item is produced by a new source (which happens with a probability equal to $\alpha(t)$), or the new item is produced by an old source (which happens with a probability equal to $1-\alpha(t)$). Hence, this process can be illustrated by a dyadic tree (Figure 1). Of course, the root of this tree consists of one source producing one item at time 1. Note further that in this article we assume that items are allocated to exactly one source; the more general situation where an item can be allocated to several sources will be dealt with in a follow-up paper [15].

In Figure 1, 0 denotes a new source, and 1 denotes an old source. For $t > 1$, we will denote by (Ω_t, F_t, P_t) the elementary probability space where $\Omega_t = \{0,1\}$, $P_t(0) = \alpha(t)$, $P_t(1) = 1-\alpha(t)$ and F_t is the σ-algebra, i.e., the set of measurable sets, $\{\phi, \{0\}, \{1\}, \Omega_t\}$. For $t = 1$, $\Omega_t = \{0\}$, $P_t(0) = 1$, and $F_t = \{\phi, \{0\}\}$. Define then (Ω, F, P) as the product probability space of the spaces (Ω_t, F_t, P_t), $t \in \mathbb{N}_0$ (see [16, Chapter VII]).

$T(t)$ and $\mu(t)$ act on Ω as follows: $T(t)(\omega)$ denotes the number of sources in the IPP ω at time t; similarly, $\mu(t)(\omega)$ denotes the number of items per source in the IPP ω at time t.

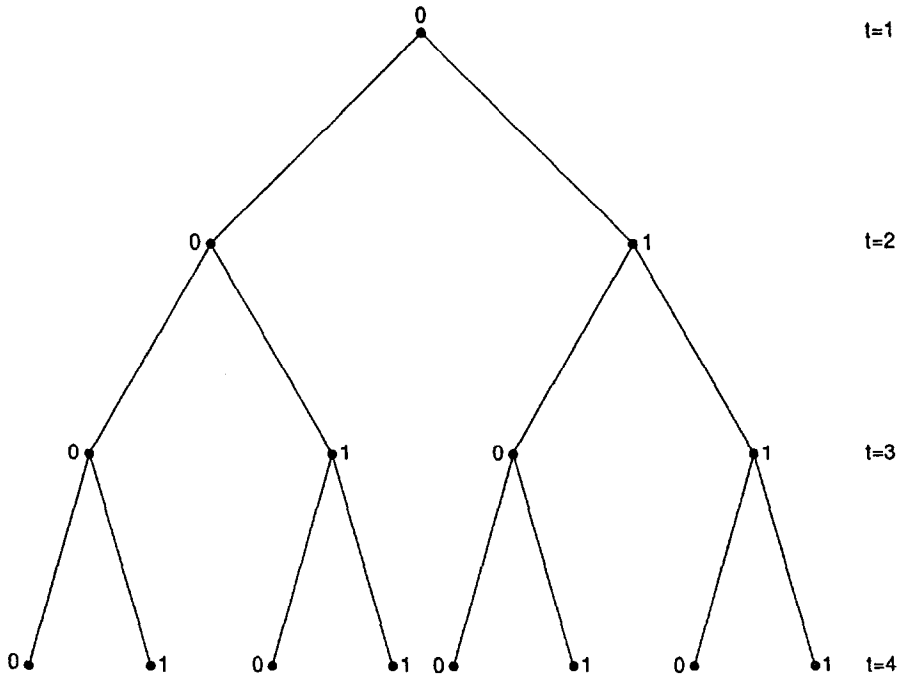

Figure 1. The basic SBS-principle as a dyadic tree.

Next, we define a new σ-algebra at the level t, denoted as G_t. This new σ-algebra is generated by the sets $\mathrm{Proj}_t^{-1}(x)$, where x is any element of $\{0,1\}^t$ and where

$$\mathrm{Proj}_t : \Omega \to \{0,1\}^t : (x_i)_{i\in\mathbb{N}_0} \to x = (x_1, \ldots, x_t)$$

denotes the projection on the first t coordinates. Since both $T(t)$ and $\mu(t)$ are clearly determined by the level t in the above dyadic tree (e.g., the sequence $\omega = 0,0,1,0$ followed by any sequence of zeros and ones, determines $T(4)(\omega) = 3$, and $\mu(4)(\omega) = 4/3$, where we have used the fact that $\mu(t) = t/T(t)$), they are both G_t-measurable.

The stochastic processes we will study first are $(T(t), G_t)_{t\in\mathbb{N}_0}$ and $(\mu(t), G_t)_{t\in\mathbb{N}_0}$. Since $T(t)$ and $\mu(t)$ are G_t-measurable (in fact, $T(t)$ and $\mu(t)$ are measurable with respect to (w.r.t.) $\{T(t) = i\}$, $i = 1, \ldots, t$), these processes are adapted sequences in the sense of [10]. This is the exact framework in which the number of sources and the average number of items per source can be studied as a function of time, using the general SBS-principle. We will denote by χ_A the indicator function of $A \subset \Omega$:

$$\chi_A = \begin{cases} 0, & \text{if } \omega \text{ does not belong to } A, \\ 1, & \text{if } \omega \text{ belongs to } A. \end{cases}$$

Then obviously, for every $t \in \mathbb{N}_0$ and every $\omega \in \Omega$,

$$T(t)(\omega) = \sum_{i=1}^{t} i\chi_{\{T(t)=i\}}(\omega), \quad \text{and} \tag{1}$$

$$\mu(t)(\omega) = \sum_{i=1}^{t} \frac{t}{i}\chi_{\{T(t)=i\}}(\omega) = \frac{t}{T(t)(\omega)}. \tag{2}$$

The stochastic processes $(T(t), G_t)$ and $(\mu(t), G_t)$ denote the totality of possible evolutions in time that can take place for $T(t)$ and $\mu(t)$.

So far we have treated $\alpha(t)$ as a deterministic probability, depending on t. Now we will introduce a further generalization: $\alpha(t)$ also will be considered as a G_t-adapted stochastic process $(\alpha(t), G_t)$. Considering $\alpha(t)$ as a stochastic process means that this sequence depends on time, but also on

the actual situation $\omega \in \Omega$. This is logical from a practical point of view. Symbolically,

$$\alpha(t)(\omega) = \sum_{i=1}^{t} \alpha(t,i)\chi_{\{T(t)=i\}}(\omega) = \alpha(t, T(t)). \tag{3}$$

NOTE. The above processes, described by Figure 1, are analogous, but not the same, to the time evolution of a gambler's fortune, symbolized by tossing a coin and betting on heads or tails.

Finally, we recall the definitions of a martingale, a submartingale and a supermartingale (see, e.g., [10,11]).

DEFINITIONS. *Let $(X_t, G_t)_{t\in\mathbb{N}_0}$ be a stochastic process. For every $t \in \mathbb{N}_0$, $E^{G_t}X_{t+1}$ denotes the conditional expectation of X_{t+1} w.r.t. G_t, i.e., the unique function such that*

$$\int_A E^{G_t} X_{t+1}\, dP = \int_A X_t\, dP,$$

for every $A \in G_t$.

We say that $(X_t, G_t)_{t\in\mathbb{N}_0}$ is a martingale if

$$E^{G_t}X_{t+1} = X_t, \qquad P\text{-a.e.},$$

for every $t \in \mathbb{N}_0$. "P-a.e." means "P-almost everywhere", i.e., the above equality is true except on a measurable set A for which $P(A) = 0$. When the equality sign is replaced by $\geq$ we have the definition of a submartingale. When this sign is replaced by $\leq$ we have a supermartingale. The classical interpretation of these definitions in gambling theory is of the evolution of the gambler's fortune X_t over time t. In case of a martingale, this game is fair in the sense that the gambler can expect to keep his capital after gambling. In case of a submartingale, the gambler is expected to win; casinos will not allow this: at least to cover their expenses, the games are usually supermartingales in which case the gambler is expected to lose.

3. PROPERTIES OF THE ADAPTED SEQUENCES $(T(t), G_t)_{t\in\mathbb{N}_0}$ AND $(\mu(t), G_t)_{t\in\mathbb{N}_0}$

3.1. Properties of $(T(t), G_t)_{t\in\mathbb{N}_0}$

LEMMA 1. *$T(1) = 1$ and for every $t = 2, 3, \ldots$*

$$E(T(t)) = E(T(t-1)) + E(\alpha(t-1)), \tag{4}$$

$$E(T(t)) = 1 + \sum_{i=1}^{t-1} E(\alpha(i)). \tag{5}$$

PROOF. $T(1) = 1$ is clear. Let then $t \geq 2$. We have that

$$E(T(t)) = \sum_{i=1}^{t} iP(T(t) = i) = \sum_{i=1}^{t}\sum_{j=1}^{t} iP(T(t) = i \mid T(t-1) = j)P(T(t-1) = j),$$

by the principle of total chance. Due to the item-source allocation in this case we have

$$\begin{aligned} E(T(t)) &= \sum_{i=1}^{t} i[P(T(t)) = i \mid T(t-1) = i)P(T(t-1) = i) \\ &\quad + P(T(t) = i \mid T(t-1) = i-1)P(T(t-1) = i-1)] \\ &= \sum_{i=2}^{t-1} i[(1 - \alpha(t-1,i))P(T(t-1) = i) + \alpha(t-1, i-1)P(T(t-1) = i-1)] \\ &\quad + (1 - \alpha(t-1,1))P(T(t-1) = 1) + t\alpha(t-1, t-1)P(T(t-1) = t-1) \end{aligned}$$

(by definition (3))

$$= \sum_{i=1}^{t-1} i[(1-\alpha(t-1,i))P(T(t-1)=i)] + \sum_{i=2}^{t} i\alpha(t-1,i-1)P(T(t-1)=i-1)$$
$$= \sum_{i=1}^{t-1} i[(1-\alpha(t-1,i))P(T(t-1)=i)] + \sum_{i=1}^{t-1}(i+1)\alpha(t-1,i)P(T(t-1)=i)$$
$$= \sum_{i=1}^{t-1} iP(T(t-1)=i) + \sum_{i=1}^{t-1}\alpha(t-1,i)P(T(t-1)=i).$$

Hence,

$$E(T(t)) = E(T(t-1)) + E(\alpha(t-1)).$$

Now, (5) follows, by recursion. ∎

PROPOSITION 1. *The adapted sequence $(T(t), G_t)_{t\in\mathbb{N}_0}$ is a submartingale. It is L^1-bounded if and only if*

$$\sum_{i=1}^{\infty} E(\alpha(i)) < \infty, \tag{6}$$

in which case there exists $T_\infty \in L^1(\Omega, F, P)$ such that $\lim_{t\to\infty} T(t) = T_\infty$, P-a.e. (i.e., except on a set of P-measure zero).

PROOF. By (1), for every $t \in \mathbb{N}_0$,

$$T(t) = \sum_{i=1}^{t} i\chi_{\{T(t)=i\}}.$$

By definition of the general SBS-principle,

$$E^{G_t}T(t+1) = \sum_{i=1}^{t}(\alpha(t)(i+1) + (1-\alpha(t))i)\chi_{\{T(t)=i\}} = \sum_{i=1}^{t}(\alpha(t)+i)\chi_{\{T(t)=i\}} \geq T(t).$$

Hence $(T(t), G_t)_{t\in\mathbb{N}_0}$ is a submartingale.

The L^1-boundedness requires that

$$\sup_{t\in\mathbb{N}_0} E(T(t)) = \sup_{t\in\mathbb{N}_0} \int_\Omega T(t) < \infty.$$

The previous lemma shows that this is true if and only if condition (6) holds. Now, invoke the submartingale convergence theorem of Doob (see, e.g., [11]), yielding in case of (6), an integrable function $T_\infty \in L^1(\Omega, F, P)$ such that $\lim_{t\to\infty} T(t) = T_\infty$, P-a.e. ∎

NOTE. The above proposition is very important since it gives a sufficient condition (6) to end up with a finite number of sources T_∞, when $t \to \infty$, *no matter what the outcome of our process is*. This result is much stronger than the result that $\sup_{t\in\mathbb{N}_0} E(T(t)) < \infty$; here one can have situations where a stable T_∞ does not exist. The submartingale property, however, protects us against such events.

3.2. Properties of $(\mu(t), G_t)_{t\in\mathbb{N}}$

We first investigate when $(\mu(t), G_t)_{t\in\mathbb{N}_0}$ is a (super), (sub)martingale.

PROPOSITION 2. *The adapted sequence $(\mu(t), G_t)_{t\in\mathbb{N}_0}$ is a martingale if and only if*

$$\alpha(t) = \frac{T(t)+1}{t+1}. \tag{7}$$

It is a supermartingale if and only if

$$\alpha(t) \geq \frac{T(t)+1}{t+1}. \tag{8}$$

It is a submartingale if and only if

$$\alpha(t) \leq \frac{T(t)+1}{t+1}. \tag{9}$$

This condition boils down to

$$\alpha(t) \leq \frac{2}{t+1},$$

in case $\alpha(t)$ is constant in $\omega \in \Omega$ (i.e., if $\alpha(t)$ only depends on t but $\alpha(t)$ is not a random variable).

In case of (8), there exists an integrable $\mu_\infty \in L^1(\Omega, F, P)$ such that $\lim_{t\to\infty} \mu(t) = \mu_\infty$, P-a.e., and furthermore,

$$E^{G_t}(\mu_\infty) \leq \mu(t),$$

for every $t \in \mathbb{N}_0$.

PROOF. By (2)

$$\mu(t+1) = \sum_{i=1}^{t+1} \frac{t+1}{i} \chi_{\{T(t+1)=i\}}.$$

Hence,

$$\begin{aligned} E^{G_t}\mu(t+1) &= (t+1)\sum_{i=1}^{t}\left(\frac{1-\alpha(t)}{2} + \frac{\alpha(t)}{i+1}\right)\chi_{\{T(t)=i\}} \\ &= (t+1)\left(\frac{1-\alpha(t)}{T(t)} + \frac{\alpha(t)}{T(t)+1}\right) \geq \mu(t), \end{aligned} \tag{10}$$

if and only if (by (2)),

$$\alpha(t) \leq \frac{T(t)+1}{t+1}$$

(and the same for the equality and the other inequality signs).

In case $\alpha(t)$ is constant on Ω, this condition is equivalent with

$$\alpha(t) \leq \frac{2}{t+1}.$$

In case of a supermartingale, we invoke the supermartingale theorem for positive processes (see [11]) saying that there exists a function $\mu_\infty \in L^1(\Omega, F, P)$ such that

$$\lim_{t\to\infty} \mu(t) = \mu_\infty, \qquad P\text{-a.e.},$$

and one has that $E^{G_t}(\mu_\infty) \leq \mu(t)$, for all $t \in \mathbb{N}_0$. ∎

So whenever

$$\alpha(t) \geq \frac{T(t)+1}{T+1},$$

we have a stable result for the average number of items per source in the sense that, for *any* evolution of the stochastic process, we end up with finite averages. Note that, since $0 < \alpha(t) < 1$, (10) implies

$$\frac{t+1}{T(t)+1} < E^{G_t}\mu(t+1) < \frac{t+1}{T(t)},$$

an obvious result. The number $(T(t)+1)/(t+1)$ is a turning point for the IPP. It determines whether the conditional average will increase. In the next section, we will study the same problem for the absolute expectation $E(\mu(t))$.

Note that (10) gives the following relation:

$$E^{G_t}\mu(t+1) = \frac{t+1}{t}\mu(t)\frac{1+T(t)-\alpha(t)}{T(t)+1}. \tag{11}$$

Concerning the process $\mu(t)$ itself, we have the following trivial result.

PROPOSITION 3. *For every* $t \in \mathbb{N}_0$,

$$\mu(t+1) \leq \mu(t) + \frac{1}{T(t+1)}.$$

PROOF. By definition,

$$\mu(t+1) = \frac{t+1}{T(t+1)} = \frac{t}{T(t+1)} + \frac{1}{T(t+1)} \leq \mu(t) + \frac{1}{T(t+1)},$$

since $T(t+1) \geq T(t)$, obviously. ∎

4. PROPERTIES OF $E(\mu(t))$

4.1. The Case of a Constant α

Since we can draw special conclusions in the case of a constant α (i.e., constant w.r.t. t as well as ω—hence the classical SBS case), these are presented in a special subsection.

PROPOSITION 4. *If* $\alpha \in]0,1[$ *is constant, then*

$$E(\mu(t)) = \frac{1-(1-\alpha)^t}{\alpha}, \tag{12}$$

for $t \in \mathbb{N}$.

PROOF. We will denote $1-\alpha$ by β. Formula (2) implies

$$E(\mu(t)) = \sum_{i=1}^{t} \frac{t}{i} P(T(t) = i).$$

But

$$P(T(t) = i) = \binom{t-1}{i-1} \alpha^{i-1}\beta^{t-i}, \tag{13}$$

since $\alpha \in]0,1[$ is a constant (and since $T(1) = 1$). Hence,

$$E(\mu(t)) = \sum_{i=1}^{t} \frac{t}{i} \binom{t-1}{i-1} \alpha^{i-1}\beta^{t-i}.$$

Using that

$$\frac{1}{i}\binom{t-1}{i-1} = \frac{1}{t}\binom{t}{i}$$

yields

$$E(\mu(t)) = \sum_{i=1}^{t} \binom{t}{i} \alpha^{i-1}\beta^{t-i} = \frac{1}{\alpha}\sum_{i=1}^{t} \binom{t}{i} \alpha^{i}\beta^{t-i} = \frac{1}{\alpha}\left((\alpha+\beta)^t - \beta^t\right),$$

$$E(\mu(t)) = \frac{1-\beta^t}{\alpha}.$$ ∎

COROLLARY 5. *If $\alpha \in [0,1[$ is a constant, then $E(\mu(t))$ increases strictly.*

PROOF. This follows readily from (12) for $\alpha \in]0,1[$ and is trivial for $\alpha = 0$. ∎

COROLLARY 6. $\liminf_{t\in\mathbb{N}_0} \mu(t) \in L^1$.

PROOF. Since (12) implies that

$$\lim_{t\to\infty} E(\mu(t)) = \frac{1}{\alpha},$$

we have

$$\sup_{t\in\mathbb{N}_0} E(\mu(t)) < \infty.$$

Invoke Fatou's lemma now (see, e.g., [10,16]), to yield that

$$\liminf_{t\in\mathbb{N}_0} \mu(t) \in L^1.$$ ∎

This result is a (partial) stability result for the process $\mu(t)$ for large t (partial because of the occurrence of lim inf; it is not clear when this lim inf is actually a limit).

We now come to the calculation of $E(\mu(t))$ for general α as in (3).

4.2. General Case

PROPOSITION 7. *For every $t \in \mathbb{N}_0$, $E(\mu(1)) = 1$ and for $t \geq 2$,*

$$E(\mu(t)) = E(\mu(t-1)) + E\left(\frac{1}{T(t-1)}\right) - tE\left(\frac{\alpha(t-1)}{T(t-1)(T(t-1)+1)}\right), \tag{14}$$

$$E(\mu(t)) = 1 + \sum_{j=1}^{t-1} E\left(\frac{1}{T(j)}\right) - \sum_{j=1}^{t-1}(j+1)E\left(\frac{\alpha(j)}{T(j)(T(j)+1)}\right). \tag{15}$$

PROOF. $E(\mu(1)) = 1$ is clear. Let then $t \geq 2$.

$$E(\mu(t)) = \sum_{i=1}^{t} \frac{t}{i} P(T(t) = i) \left(= \sum_{i=1}^{t}\sum_{j=1}^{t} \frac{t}{i} P(T(t) = i \mid T(t-1) = j) P(T(t-1) = j) \right),$$

by the principle of total chance. Due to the way items are assigned to sources in the general SBS principle, we have

$$\begin{aligned}
E(\mu(t)) &= \sum_{i=1}^{t} \frac{t}{i} [P(T(t) = i \mid T(t-1) = i) P(T(t-1) = i) \\
&\quad + P(T(t) = i \mid T(t-1) = i-1) P(T(t-1) = i-1)] \\
&= \sum_{i=2}^{t-1} \frac{t}{i} [(1 - \alpha(t-1,i)) P(T(t-1) = i) + \alpha(t-1, i-1) P(T(t-1) = i-1)] \\
&\quad + t(1 - \alpha(t-1,1)) P(T(t-1) = 1) + \alpha(t-1, t-1) P(T(t-1) = t-1) \\
&= \sum_{i=1}^{t-1} \frac{t}{i} P(T(t-1) = i) - \sum_{i=1}^{t-1} \frac{t}{i} \alpha(t-1,i) P(T(t-1) = i) \\
&\quad + \sum_{i=2}^{t} \frac{t}{i} \alpha(t-1, i-1) P(T(t-1) = i-1) \\
&= \frac{t}{t-1} E(\mu(t-1)) + t \left[- \sum_{i=1}^{t-1} \frac{1}{i} \alpha(t-1,i) P(T(t-1) = i) \right. \\
&\quad \left. + \sum_{i=1}^{t-1} \frac{1}{i+1} \alpha(t-1,i) P(T(t-1) = i) \right]
\end{aligned}$$

$$= \frac{t}{t-1}E(\mu(t-1)) - t\left[\sum_{i=1}^{t-1}\frac{1}{i(i+1)}\alpha(t-1,i)P(T(t-1)=i)\right]$$
$$= E(\mu(t-1)) + E\left(\frac{1}{T(t-1)}\right) - tE\left(\frac{\alpha(t-1)}{T(t-1)(T(t-1)+1)}\right),$$

which is (14). Applying (14) recursively yields (15) (using that $E(\mu(1)) = E(1) = 1$). ∎

COROLLARY 8. *For every* $t \in \mathbb{N}_0$,

$$E(\mu(t+1)) > E(\mu(t)),$$

if and only if

$$E\left(\frac{\alpha(t)}{T(t)(T(t)+1)}\right) < \frac{E\,(1/T(t))}{t+1}. \tag{16}$$

If $\alpha(t)$ *is constant in* $\omega \in \Omega$, *then this condition boils down to*

$$\alpha(t) < \frac{E\,(1/T(t))}{(t+1)E\,(1/T(t)(T(t)+1))}. \tag{17}$$

PROOF. Apply (14) for t replaced by $t+1$. ∎

NOTE. This result is in accordance with Corollary 5. Indeed, if α is constant (in t and ω), then by Proposition 4,

$$\frac{E\,(1/T(t))}{(t+1)E\,(1/T(t)(T(t)+1))} = \alpha\frac{1-\beta^t}{1-\beta^t-t\beta^t(1-\beta)} > \alpha.$$

This can be seen from (13) and the fact that

$$E\left(\frac{1}{T(t)}\right) = \sum_{i=1}^{t}\frac{1}{i}P(T(t)=i),$$
$$E\left(\frac{1}{T(t)(T(t)+1)}\right) = \sum_{i=1}^{t}\frac{1}{i(i+1)}P(T(t)=i).$$

NOTE. Corollary 8 above gives a general answer to the problem of the evolution of $E(\mu(t))$ over time t. This answers a problem raised in [17] where it was found, experimentally, some evolutions of $\mu(t)$ over time t (in the more restricted SBS as described in the introduction), and asked for an explanation.

NOTE. Equation (14) yields (divide by t)

$$E\left(\frac{1}{T(t)}\right) = E\left(\frac{1}{T(t-1)}\right) - E\left(\frac{\alpha(t-1)}{T(t-1)(T(t-1)+1)}\right),$$

and after recursion (and using $E(1/T(1)) = 1$)

$$E\left(\frac{1}{T(t)}\right) = 1 - \sum_{j=1}^{t-1}E\left(\frac{\alpha(j)}{T(j)(T(j)+1)}\right).$$

Hence, multiplying by t,

$$E(\mu(t)) = t - t\sum_{j=1}^{t-1}E\left(\frac{\alpha(j)}{T(j)(T(j)+1)}\right). \tag{18}$$

PROBLEM. Determine a condition (in terms of α) under which

$$\sup_{t\in\mathbb{N}_0} E(\mu(t)) < \infty.$$

If this condition is compatible with (9), then we have the a.e. convergence of the submartingale (as we have already proved in case $(\mu(t), G_t)_{t\in\mathbb{N}_0}$ is a supermartingale).

5. THE STOCHASTIC PROCESSES $(X_t(n))_{t\geq n}$ FOR ALL $n \in \mathbb{N}_0$

To characterize $X_t(n)$ (the number of sources with n items at time t), we need the full definition of SBS now ((i) and (ii) of Section 1—for $T(t)$ and $\mu(t)$, only (i) was needed). In the same way Ω was constructed in Section 2, we now have at any step $t \to t+1$ a division into many parts (according to what is happening to the $(t+1)^{\text{st}}$ item, it belongs to a new source or is added to a source with n items $(n = 1, \ldots, t)$). This generalization, along the lines of the construction of (Ω, F, P), leads us to the probability space (Ω', Σ, P'), a refinement of the former one. Indeed, now (Ω', Σ, P') is the product space of the spaces Ω_t = a set of $t+1$ points on which every singleton is measurable and with probability of the singletons $\alpha(t)$, respectively, $(1-\alpha(t))\, x(t,n)$ $(n = 1, \ldots, t)$. By construction, $X_t(n)$ is Σ-measurable, but more is true; as in Section 2, since $X_t(n)$ depends only on $t' = 1, \ldots, t$, we have that $X_t(n)$ is Σ_t-measurable for every $n \in \mathbb{N}_0$ and $t \geq n$, where Σ_t is the σ-algebra generated by the sets $\text{Proj}_t^{-1}\, x$, where $x \in \text{Proj}_t(\Omega')$ arbitrarily (again Proj_t denotes the projection of Ω' onto the first t coordinates).

We have the following result.

PROPOSITION 9. *For any $n, t \in \mathbb{N}_0$ we have for $n = 1$*

$$E^{\Sigma_t}(X_{t+1}(1)) = X_t(1) + \alpha(t) - x(t,1)(1-\alpha(t)), \tag{19}$$

for $n = 2, \ldots, t$

$$E^{\Sigma_t}(X_{t+1}(n)) = X_t(n) + (1-\alpha(t))(x(t,n-1) - x(t,n)), \tag{20}$$

and for $n = t+1$

$$E^{\Sigma_t}(X_{t+1}(t+1)) = (1-\alpha(t))x(t,t). \tag{21}$$

PROOF. By the definition of SBS we have, for $n = 1$,

$$\begin{aligned} E^{\Sigma_t}(X_{t+1}(1)) &= \alpha(t)(X_t(1)+1) + (1-\alpha(t))((X_t(1)-1)x(t,1) + (1-x(t,1))X_t(1)) \\ &= X_t(1) + \alpha(t) - x(t,1)(1-\alpha(t)), \end{aligned}$$

and, for $n = 2, \ldots, t$,

$$\begin{aligned} E^{\Sigma_t}(X_{t+1}(n)) &= \alpha(t)X_t(n) + (1-\alpha(t)) \\ &\quad \times [x(t,n)(X_t(n)-1) + x(t,n-1)(X_t(n)+1) + (1-x(t,n)-x(t,n-1))X_t(n)] \\ &= X_t(n) + (1-\alpha(t))(x(t,n-1) - x(t,n)). \end{aligned}$$

For $n = t+1$, one has clearly (21). ∎

NOTE. One can take $\alpha(t)$ as a random variable as in the previous sections (cf. formula (3)), but one can take it even more general: a random variable with respect to the Σ_t, i.e., $\alpha(t)$ is Σ_t-measurable (i.e., depends on the variation in the $X_t(n)$). The same is true for random variables $x(t,n)$, $t \geq n$; $n \in \mathbb{N}_0$.

COROLLARY 10. *For any $n \in \mathbb{N}_0$, $n \geq 2$, the process $(X_t(n), \Sigma_t)_{t\geq n}$ is a supermartingale, (respectively, submartingale, or a martingale) if and only if*

$$x(t,n-1) \leq x(t,n), \tag{22}$$

for every $t \geq n$, (respectively, $\geq$, or $=$). For $n = 1$, the process $(X_t(1), \Sigma_t)_{t\in\mathbb{N}_0}$ is a supermartingale, (respectively, a submartingale, or a martingale) if and only if

$$\alpha(t) \leq \frac{x(t,1)}{1+x(t,1)} \tag{23}$$

(respectively, $\geq$ or $=$).

PROOF. For $n \geq 2$, the condition

$$E^{\Sigma_t}(X_{t+1}(n)) \leq X_t(n),$$

for all $t \geq n$ is equivalent with (22) and for $n = 1$,

$$E^{\Sigma_t}(X_{t+1}(1)) \leq X_t(1),$$

if and only if

$$\alpha(t) \leq x(t,1)(1 - \alpha(t)),$$

for all $t \in \mathbb{N}_0$, and hence, equivalently, (23).

The convergence properties of (super), (sub)martingales (cf. [11]) give us the following important stability result.

PROPOSITION 11.

1. *Let*

$$\alpha(t) \leq \frac{x(t,1)}{1 + x(t,1)}, \tag{24}$$

for all $t \in \mathbb{N}_0$ *or*

$$\alpha(t) = (1 - \alpha(t))x(t,1) + \varphi(t), \tag{25}$$

where $\varphi \geq 0$ *is such that*

$$\sum_{t=1}^{\infty} \int_{\Omega'} \varphi(t)(\omega)\, dP'(\omega) < \infty. \tag{26}$$

Then there exists an integrable function $X_\infty(1) \in L^1(\Omega', \Sigma, P')$ *such that*

$$\lim_{t\to\infty} X_t(1) = X_\infty(1), \qquad P'\text{-a.e.}$$

In case (24) is valid, we also have that

$$E^{\Sigma_t}(X_\infty(1)) \leq X_t(1), \tag{27}$$

for all $t \in \mathbb{N}_0$.

2. *For every* $n \geq 2$, *let*

$$x(t, n-1) \leq x(t,n), \tag{28}$$

for all $t \geq n$ *or*

$$(1 - \alpha(t))(x(t, n-1) - x(t,n))) = \psi_n(t), \tag{29}$$

where $\psi_n \geq 0$ *is such that*

$$\sum_{t=1}^{\infty} \int_{\Omega'} \psi_n(t)(\omega)\, dP'(\omega) < \infty. \tag{30}$$

Then there exists an integrable function $X_\infty(n) \in L^1(\Omega', \Sigma, P')$ *such that*

$$\lim_{t\to\infty} X_t(n) = X_\infty(n), \qquad P'\text{-a.e.}$$

In case (28) is valid, we also have that

$$E^{\Sigma_t}(X_\infty(n)) \leq X_t(n), \tag{31}$$

for all $t \geq n$.

PROOF. Conditions (24) and (28) imply (by Corollary 10) that the processes $(X_t(n), \Sigma_t)_{t \geq n}$ $(n \in \mathbb{N})$ are positive supermartingales. Hence, using [11], we have the asserted convergence and corresponding inequalities (27) and (31).

Condition (25) shows that

$$\alpha(t) \geq \frac{x(t,1)}{1 + x(t,1)},$$

and hence, $(X_t(1), \Sigma_t)_{t \in \mathbb{N}_0}$ is a submartingale, by Corollary 10. But (25) and (19) imply

$$\int_{\Omega'} X_{t+1}(1)(\omega)\, dP'(\omega) - \int_{\Omega'} X_t(1)(\omega)\, dP'(\omega) = \int_{\Omega'} \varphi(t)(\omega)\, dP'(\omega),$$

so that, by (26),

$$\sup_{t \in \mathbb{N}_0} \int_{\Omega'} X_t(1)(\omega)\, dP'(\omega) < \infty.$$

This, together with the fact that $(X_t(1), \Sigma_t)_{t \in \mathbb{N}_0}$ is a submartingale again shows the asserted convergence (the inequality (27) is not valid here) (Doob's theorem, see [11]).

The same arguments, using (28),(20),(29) and (30) now show the same for the processes $(X_t(n), \Sigma_t)_{t \geq n}$ for all $n \in \mathbb{N}_0$. ∎

NOTE. Note that (24) and (25) combine to the single condition

$$\alpha(t) \leq \frac{x(t,1)}{1 + x(t,1)} + \varphi(t), \tag{32}$$

and (28) and (29) to

$$(1 - \alpha(t))(x(t, n-1) - x(t,n)) \leq \psi_n(t). \tag{33}$$

These conditions are very clear restrictions on α (for (32)) and the $x(t,n)$ (for (33)) in order to have stable distributions $X_t(n)$, $t \geq n$ for large t.

REFERENCES

1. L. Egghe and R. Rousseau, A general success-breeds-success principle, leading to time-dependent informetric distributions, *Journal of the American Society for Information Science* (1995) (to appear).
2. H.A. Simon, On a class of skew distribution functions, *Biometrika* **52**, 425–440 (1955).
3. D. De Solla Price, A general theory of bibliometric and other cumulative advantage processes, *Journal of the American Society for Information Science* **27**, 292–306 (1976).
4. Y. Ijiri and H.A. Simon, *Skew Distributions and the Sizes of Business Firms*, North-Holland, Amsterdam, (1977).
5. Y.S. Chen, Analysis of Lotka's law: The Simon-Yule approach, *Information Processing and Management* **25**, 527–544 (1989).
6. L. Egghe, The duality of informetric systems with applications to the empirical laws, Ph.D. Thesis, The City University, London, (1989).
7. L. Egghe, The duality of informetric systems with applications to the empirical laws, *Journal of Information Science* **16**, 17–27 (1990).
8. L. Egghe, Bridging the gaps—Conceptual discussions on informetrics, *Proceedings of the 4th International Conference on Bibliometrics, Scientometrics and Informetrics*, Berlin, 1993, *Scientometrics* **30**, 35–47 (1994).
9. L. Egghe and R. Rousseau, *Introduction to Informetrics*, Elsevier, Amsterdam, (1990).
10. L. Egghe, Stopping time techniques for analysts and probabilists, *London Mathematical Society Lecture Notes Series, 100*, Cambridge University Press, (1984).
11. J. Neveu, *Discrete-Parameter Martingales*, North-Holland, Amsterdam, (1975).
12. G.R. Grimmett and D.R. Stirzaher, *Probability and Random Processes*, Clarendon Press, Oxford, (1985).
13. Y.S. Chen, Zipf's laws in text modeling, *International Journal of General Systems* **15**, 233–252 (1989).
14. Y.S. Chen, Zipf-Halstead's theory of software metrication, *International Journal of Computer Mathematics* **412**, 125–138 (1992).
15. L. Egghe, Extension of the general success-breeds-success principle to the case that items can have multiple sources, In *Proceedings of the Fifth International Conference on Scientometrics and Informetrics*, River Forest, IL, U.S.A., 1995 (to appear).
16. P.R. Halmos, *Measure Theory*, Graduate Texts in Mathematics 18, Springer-Verlag, New York, (1974).
17. Y.S. Chen, P.P. Chong and Y. Tong, The Simon-Yule approach to bibliometric modelling (to appear).