Dynamics of a field list of internationally visible journals : a stochastic model

Non Peer-reviewed author version

DYNAMICS OF A FIELD LIST OF INTERNATIONALLY VISIBLE JOURNALS : A STOCHASTIC MODEL.

by

L. Egghe

LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium[1]
and
UIA, Universiteitsplein 1, B-2610 Wilrijk, Belgium

Keywords : visible journal, mil, amart, dynamics

ABSTRACT

A basic model for the dynamics of a field list of internationally visible journals is constructed. We study, in function of time, the remainder of an initially choosen set of source journals. The stochastic process turns out to be a martingale in the limit (mil) or a quasi martingale, according to a.e. or L^1-boundedness conditions on the controlling parameters of the system. Hence by theorems of Mucci of 1976 and of Bellow of 1978, the process converges under these conditions. We do not know of any other "real life" applications of mils and quasi martingales. The application in the field of information science is remarkable.

---

[1]Permanent address

I. Introduction.

The most well-known set of internationally visible journals is the one produced by the Institute for Scientific Information (ISI). This set is called the set of core journals which are used for the citation analysis in their well-known indexes. It is also well-known that this list, although fairly stable, changes from year to year. It suffices to look at the yearly issues of the JCR (Journal Citation Reports).

In addition to these yearly changes one might also ask the following basic question : suppose that, at a certain moment in the past, we would have started with a different list than the one adopted by ISI at their very start. Would we end up with a list that is more or less the same as theirs or could there be significant differences? We, of course, suppose here that the techniques to delete or add a journal from/to the list are the same in both cases. Such a technique (and also the above question) was formulated in detail in Rousseau and Spinak (1996) in which one asks for a dynamic study of these problems. These type of questions are also often formulated by research groups that experience certain barriers (e.g. language, part of the world, etc.), see e.g. Velho (1986, 1987), Gaillard (1989), Vinkler (1986), Garfield (1979), Cronin (1981) and Alvo (1978). For more on these issues we refer to reader to Egghe and Rousseau (1990).

In a model-theoretic way, many problems can be studied. We will refer to them in the closing section of this paper. In this paper we will consider the "easiest" version of the problem described above. Let us consider a universe U of "all" journals. Here we can include all journals that exist (or came into existence) in the time period under study. Of course, in this setting, a journal cannot become a core journal before it actually exists. Let us from now on adopt the short term "core journal" for a member of the (yearly renewable) list of internationally visible journals.

Suppose, in the first year (t=1) we start with an initial "core" list. How this list is constituted is of no importance right now but its "survival", when time passes, will be crucially determined by these criteria that were initially adopted. One could think of an initial core list that consists of the most important journals in a country (including international as well as local ones) and see what happens. For the moment, we can, however, take any nonempty subset $A \subset U$.

Every year (or any other time unit) this list is reconsidered, hereby dropping journals from the present core list and adding new ones, based on scores of the (relative) number of times they are cited in a certain period.

Note that the evolution of the set of core journals that belong to A is not necessarily increasing or decreasing. Indeed, when going from the first year (where A is the core set) to the second year, there is a decrease : some journals of A can be dropped as core journals based on weak citation scores. But already from year 2, and continuing so, journals from A an be added or deleted (the ones added in case they were previously deleted as core journal at least once).

So, the evolution of the "remainder" of A, the problem studied in this paper, does not seem to be a simple process where "expectations" (a term from probability theory) are increasing or decreasing in time, all the time. Such processes were encountered e.g. in Egghe and Rousseau (1995, 1996) and Egghe (1995, 1996) and are called sub- or supermartingales.

It turns out that the process under study is a generalization of martingales, namely a

- martingale in the limit (mil) under a.e.-convergence requirements of the controlling parameters
- quasi martingale under L^1-convergence requirements of the controlling parameters.

We refer the reader to the appendix for a summary (with references) of the theory of stochastic processes, martingales, submartingales, supermartingales, mils and quasi martingales and also to Egghe (1984) and Edgar and Sucheston (1992).

In the next section we develop the model of the "remainder" of A while in the third section its stochastic properties are studied and proved.

The last section is developed to open problems and suggestions for further research.

II. The stochastic process describing the remainder of the initial set of core journals.

II.1 Introduction

Suppose U is the universe of all journals under possible consideration. It is a fixed set but can contain journals that were not existing at the starting time $t=1$. Of course, before their "birth" their probability to become a source journal is zero. We fix a subset $A \subset U$ as our starting set of source journals at $t=1$.

In the investigation of the evolution of the source set, when t increases, we came along two problems.

1. If we take the unity of time to be a year, say, then we have that several journals in A might leave as a source journal and also that many other journals become a source journal. This is not easy to model.

   Therefore we assume that the step of increase of one (from t to $t+1$) stands for one change in the set of source journals (of any type : one in, one out). This is not a restriction since the many transactions in this way, encountered in a year, can be subdivided as indicated above. We hence end up with time varying in a slower way but, by taking t high enough (and indeed this is possible since we allow t to go to infinity), we can reach any (high) number of years.

   Dealing with one transaction at a time has the advantage that we can use the same framework (but with different models!) as we did in the modelling of the Success-Breeds-Success phenomenon (SBS) in Egghe and Rousseau (1996). Also there we looked at changes in time being equal to the addition of a new item to be allocated to a (new or old) source.

2. If one wants to follow the evolution of a set A (or of what is left of it as a set of source journals) one is inclined to define the stochastic process to be

$$X_t = \chi_{A_t} , \tag{1}$$

where A_t is the set of source journals at time t. However, studying (1) means that we have processes with values in an infinite dimensional space and it is well-known than convergence of such processes in such spaces is difficult to prove, cf. Egghe (1984). In Egghe (1998) we give an attempt in this direction.

But, also, by using (1), one requires too much in relation to the problem studied here. Studying (1) implies that we want to be informed, at any time t, of the behavior of every individual journal (i.e. by name) in A. This would be nice but is a far too demanding requirement. We repeat that the purpose of this paper is to study "what fraction (or number of journals) of the initial source set A is left after t steps?"

Hence we can suffice with a stochastic process with values in the natural numbers, hence real valued processes that are indeed much easier to handle (see again Egghe (1984)).

II.2 The stochastic process.

Let t be as described in subsection II.1. Define (not a stochastic process yet!)

X_t = the number of journals from A that, after t steps, are source journals

Note that X_t represents a snapshot at time t. The number X_t refers to journals of A that stayed as a source journal all the time as well as to journals of A that left as a source journal but then (before or at t) were again picked up as a source journal.

Note also that $X_1 = \#A$, the number of elements in A.

First we will determine the underlying probability space $(\Omega,\mathscr{F},P)$ on which all X_t are defined and then we will determine the stochastic process by means of its conditional expectations (cf. also Egghe and Rousseau (1996), Egghe (1984) or the appendix which gives a brief account on the stochastic processes that are used in this paper).

II.2.1 The underlying probability space $(\Omega,\mathscr{F},P)$.

Let $\Omega_1 = \{1\}$. To go from $t=1$ to $t=2$ (i.e. at the first change) we have the following possibilities : or (with probability $\alpha(1)$) an element from A (the source set at $t=1$) leaves as a source journal, or (with probability $1-\alpha(1)$) a journal from U\A enters as a source journal. Let us call the new set of source journals A_2 (call $A=A_1$) and

$$X_2 = \#(A_2 \cap A) \tag{2}$$

To go from t to t+1 for $t \geq 2$, we can give the following general algorithm :

- With probability $\alpha(t)$ there will be a journal from $A_t \cap A$ that leaves as a source journal at t+1.

  If this is not the case (probability $1-\alpha(t)$) then there are two possibilities :

- With probability $\beta(t)$ a journal belonging to $A \backslash A_t$ re-enters the set of source journals at t+1

- With probability $1-\beta(t)$ it will hence be a journal from U\A that enters as a source journal at t+1.

Let us call the new set of source journals A_{t+1} and, according to the definition :

$$X_{t+1} = \#(A_{t+1} \cap A) \tag{3}$$

Note that also the case $t=1$ is covered here by putting $\beta(1)=0$.

We take $\Omega_2=\{1,2\}$ and, for every $t\geq 3$, $\Omega_t=\{1,2,3\}$ with $P_t(1)=\alpha(t)$, $P_t(2)=(1-\alpha(t))\beta(t)$, $P_t(3)=(1-\alpha(t))(1-\beta(t))$. Each time we take $\mathscr{F}_t=\mathfrak{P}(\Omega_t)$ ($\mathfrak{P}$ denotes the set of all subsets of Ω_t) as our σ-algebra. Finally, define $(\Omega,\mathscr{F},P)$ to be the product probability space of the spaces $(\Omega_t,\mathscr{F}_t,P_t)$, $t=1,2,3,\ldots$. Define $\mathcal{G}_t$ to be the σ-algebra generated by $\mathscr{F}_1,\ldots,\mathscr{F}_t$ via the sets $\mathrm{Proj}_t^{-1}(x)$, where x is any element in $\prod_{i=1}^{t}\Omega_i$ and where

$$\mathrm{Proj}_t : \Omega\to\{1,2,3\}^t$$
$$(x_i)_{i\in N}\to x=(x_1,\ldots,x_t)$$

denotes the projection on the first t coordinates. Each $\mathcal{G}_t$, clearly, is a sub-σ-algebra of $\mathscr{F}$, and, furthermore, the sequence $(\mathcal{G}_t)$ increases.

II.2.2 The stochastic process $(X_t,\mathcal{G}_t,P)$.

It is clear that, by construction and by definition of the σ-algebras $\mathcal{G}_t$, that every X_t is $\mathcal{G}_t$-measurable. We hence have a stochastic process $(X_t,\mathcal{G}_t,P)$ (also called an adapted sequence - cf. Egghe (1984). The conditional expectations, relating X_t to X_{t+1} are as follows ($\omega\in\Omega$) :

$$X_1(\omega) = \#A$$

$$E^{\mathcal{G}_1}X_2(\omega) = \alpha(1)(\omega)(X_1(\omega)-1) + (1-\alpha(1)(\omega))X_1(\omega) \quad (4)$$

and for $t\geq 2$

$$E^{\mathcal{G}_t} X_{t+1}(\omega) = \alpha(t)(\omega)(X_t(\omega)-1) + (1-\alpha(t)(\omega))[\beta(t)(\omega)(X_t(\omega)+1)+(1-\beta(t)(\omega))X_t(\omega)] \quad (5)$$

Here we allowed the α(t)s and β(t)s to be dependent on ω as well but in this case we assume them to be $\mathcal{G}_t$-measurable (a logical assumption : the α(t)s and β(t)s act on the situation at time t ; everything is $\mathcal{G}_t$-measurable at that time!). In order not to overload the notation we will drop the ω-dependence, whenever there cannot be any confusion.

Equation (5) boils down to

$$E^{\mathcal{G}_t}X_{t+1} = X_t+\beta(t)-\alpha(t)-\alpha(t)\beta(t) \tag{6}$$

which can also be used in case t=1, assuming β(1)=0.

From (6) it is immediately clear that, $\forall t\in\mathbb{N}$

$$E(X_{t+1}) = E(X_t)+E(\beta(t))-E(\alpha(t))-E(\alpha(t)\beta(t)) \tag{7}$$

Hence, by induction on t (and since $X_1=E(X_1)=\#A$),

$$E(X_t) = \#A - E(\alpha(1))+\sum_{i=2}^{t-1} [E(\beta(i))-E(\alpha(i))-E(\alpha(i)\beta(i))] \tag{8}$$

$\forall t\in\mathbb{N}$, $t\geq 2$ (for t=2 we use $\sum_{i=2}^{1} =0$).
If the limit exists, we also have

$$\lim_{t\to\infty} E(X_t) = \#A - E(\alpha(1)) + \sum_{i=2}^{\infty} [E(\beta(i))-E(\alpha(i))-E(\alpha(i)\beta(i))]. \tag{9}$$

III. Stochastic properties of the process $(X_t,\mathcal{G}_t,P)$.

Obviously, the properties of the stochastic process $(X_t,\mathcal{G}_t,P)$ depend on the interrelations of the probabilities α(i) and β(i), $i\in\mathbb{N}$.

Since $X_1 = \#A$, a relation of the form

$$E^{\mathcal{G}_t}X_{t+1} > X_t \tag{10}$$

can only be true if one has the opposite inequality for lower t. Indeed, otherwise one would have $E(X_{t_0}) > \#A$ for some t_0, a contradiction. Hence, since also equality cannot be true [based on (6) this would require $\beta(t) = \alpha(t)+\alpha(t)\beta(t)$ for all t, which is not so since $\beta(1)=0$ and $\alpha(1)\neq 0$, by the very definition of the process] we can definitely say that $(X_t,\mathcal{G}_t,P)$ is not a submartingale (see the appendix for an introduction on these processes). An equality of the form

$$E^{\mathcal{G}_t}X_{t+1} \leq X_t \tag{11}$$

is possible, even for all t, although very unlikely in practise. This would mean that, on the average and at each step, the set A looses source journals all the time.

The very nature of a process as determined by (6) is of going up and down respectively by $\beta(t)$ and $\alpha(t)(1+\beta(t)))$ and we can wonder if a stable limit situation exists. This is certainly not so in all cases. Indeed, according to (6) we can have many cases in which heavy fluctuations occur. Let us just give some examples :

(1) The only limitation on the $\alpha(t)$s and $\beta(t)$s is that, $\forall t \in \mathbb{N}$:

$$-1 \leq \beta(t)-\alpha(t)-\alpha(t)\beta(t) \leq 1 \tag{12}$$

If one applies this to its maximal possibility and from $t=1$ on we even have that, at certain times, we have the whole of A as source journals and at other times we have nothing of A left as source journals. Hence we have here a divergent process.

(2) But even more "moderate" behavior of the $\alpha(t)$s and $\beta(t)$s does not always lead to a convergent process. Indeed, f.i. only requiring that

$$\lim_{t\to\infty} [\beta(t)-\alpha(t)-\alpha(t)\beta(t)] = 0, \text{ a.e.} \tag{13}$$

leads to processes with the property

$$\lim_{t\to\infty} |E^{\mathcal{G}_t} X_{t+1} - X_t| = 0, \text{ a.e.} \tag{14}$$

and it is well-known that examples of such processes exist (and reproducable here) for which $(X_t)_{t\in N}$ does not converge (cf. Egghe (1984) and Edgar and Sucheston (1992)).

If (13) is not sufficient and if we keep our idea that fluctuations must diminish at high t, equation (6) indicates that we must require that "something like"

$$\sum_{i=1}^{\infty} (\beta(i)-\alpha(i)-\alpha(i)\beta(i)) \tag{15}$$

converges (in some sense) must be true. We will investigate two cases

(i) the case that

$$\sum_{i=1}^{\infty} \alpha(i) < \infty, \text{ a.e.} \tag{16}$$

$$\sum_{i=1}^{\infty} \beta(i) < \infty, \text{ a.e.} \tag{17}$$

implying (15) in the a.e. sense

(ii) the case that

$$\sum_{i=1}^{\infty} E(\alpha(i)) < \infty \tag{18}$$

$$\sum_{i=1}^{\infty} E(\beta(i)) < \infty \tag{19}$$

implying (15) in the L^1-sense.

It will turn out that these requirements lead us to generalizations of martingales that have been studied only 20 years ago. This link with the advanced theory of stochastic processes is remarkable. To a certain extent, the generalizations are also optimal : further relaxation of the conditions on α and β lead to diverging processes.

III.1 First case : martingales in the limit.

Let us continue to work with the process $(X_t,\mathcal{G}_t,P)$ for which (16) and (17) are valid. In the literature (cf. Egghe (1984), Edgar and Sucheston (1992)) the following definition is known : the process $(X_t,\mathcal{G}_t,P)$ is called a martingale in the limit (abbreviated mil) if

$$\lim_{t\to\infty} \sup_{t'\geq t} |E^{\mathcal{G}_t} X_{t'} - X_t| = 0, \text{ a.e.} \tag{20}$$

If we can show that our process is a mil then we can use the following theorem of Mucci (1976) :

Theorem III.1 (Mucci). Let $(X_t,\mathcal{G}_t,P)$ be an L^1-bounded mil. Then there exists an integrable function X_∞ such that

$$\lim_{t\to\infty} X_t = X_\infty, \text{ a.e.} \tag{21}$$

The proof of this theorem (see e.g. also Egghe (1984), theorem VII.2.12, p.258) is non trivial and extends the classical downcrossing argument as given in the proof of the convergence of martingales (see e.g. Neveu (1975)).

We have the following result :

Theorem III.2 : Let $(X_t,\mathcal{G}_t,P)$ be as above for which (16) and (17) are valid. Then $(X_t,\mathcal{G}_t,P)$ is a mil and converges a.e. to an integrable function.

Proof :

$$E^{\mathcal{G}_t} X_{t'} - X_t$$

$$= E^{\mathcal{G}_t} \left(\sum_{i=t}^{t'-1} (\beta(i) - \alpha(i) - \alpha(i)\beta(i)) \right) \tag{22}$$

as follows readily from (6) by induction. Now, since $\beta(i) \geq 0$,

$$\sum_{i=t}^{t'-1} \beta(i) \leq \sum_{i=t}^{\infty} \beta(i).$$

Conditional expectations are order preserving. So

$$E^{\mathcal{G}_t} \left(\sum_{i=t}^{t'-1} \beta(i) \right) \leq E^{\mathcal{G}_t} \left(\sum_{i=t}^{\infty} \beta(i) \right) \tag{23}$$

Now

$$E^{\mathcal{G}_t} \left(\sum_{i=t}^{\infty} \beta(i) \right) = E^{\mathcal{G}_t} \left(\sum_{i=t}^{\infty} \beta(i) \right) - E^{\mathcal{G}_t} \left(\sum_{i=1}^{t-1} \beta(i) \right) \tag{24}$$

Now $(E^{\mathcal{G}_t}(\sum_{i=1}^{\infty} \beta(i)), \mathcal{G}_t, P)$ is an elementary martingale which converges to $\sum_{i=1}^{\infty} \beta(i)$, a.e., cf. Egghe (1984), theorem II.1.6, p.24. Furthermore

$$E^{\mathcal{G}_t} \left(\sum_{i=t}^{t-1} \beta(i) \right)$$

$$= \sum_{i=1}^{t-1} \beta(i)$$

since each $\beta(i)$ is $\mathcal{G}_i$-measurable and since $i < t$. Also this converges to $\sum_{i=1}^{\infty} \beta(i)$, a.e., by assumption. In conclusion, using (23) and (24) we have that

$$\lim_{t\to\infty} \sup_{t'\geq t} E^{\mathcal{G}_t} \left(\sum_{i=t}^{t'-1} \beta(i) \right) = 0, \text{ a.e.} \tag{25}$$

The same argument yields

$$\lim_{t\to\infty} \sup_{t'\geq t} E^{\mathcal{G}_t} \left(\sum_{i=t}^{t'-1} \alpha(i) \right) = 0, \text{ a.e.} \tag{26}$$

Now, since $\sum_{i=1}^{\infty} \alpha(i)$ and $\sum_{i=1}^{\infty} \beta(i)$ converge a.e., the same is true for $\sum_{i=1}^{\infty} \alpha(i)\beta(i)$ by the comparison test for the convergence of series and since all $\alpha(i),\beta(i)$ are positive and inferior to 1. Hence again we have that

$$\lim_{t\to\infty} \sup_{t'\geq t} E^{\mathcal{G}_t} \left(\sum_{i=t}^{t'-1} \alpha(i)\beta(i) \right) = 0, \text{ a.e.} \tag{27}$$

(25), (26) and (27) yield that (by(22)) and the triangle inequality,

$$\lim_{t\to\infty} \sup_{t'\geq t} |E^{\mathcal{G}_t} X_{t'} - X_t| = 0, \text{ a.e. }, \tag{28}$$

hence $(X_t, \mathcal{G}_t, P)$ is a mil. Since, by construction, every X_t is bounded by #A, the process is uniformly bounded, hence L^1-bounded. Mucci's theorem now yields the stated convergence property. □

<u>Note</u> : From the proof it is easy to see that conditions (16) and (17) can be relaxed to the single condition that

$$E^{\mathcal{G}_t} \left(\sum_{i=1}^{\infty} (\beta(i) - \alpha(i) - \alpha(i)\beta(i)) \right) \tag{29}$$

converges a.e., for all $t \in \mathbb{N}$

III.2 Second case : quasi martingales.

Let us first introduce the notion of quasi martingales. Suppose $(X_t,\mathcal{G}_t,P)$ is our process. It is called a quasi martingale if

$$\sum_{t=1}^{\infty} E(|E^{\mathcal{G}_t}X_{t+1} - X_t|) < \infty \tag{30}$$

We will use the following result :

Theorem III.3 (A. Bellow (1978)).

Let $(X_t,\mathcal{G}_t,P)$ be an L^1-bounded quasi martingale. Then it converges a.e. to an integrable function.

In fact, the result of Bellow is more generally valid : it turns out to be true for uniform amarts (extending the notion of quasi martingales - see Egghe (1984) p.122-125) and even vector-valued in certain Banach spaces ; we do not go into this here.

The following result is now easy :

Theorem III.4 : Let $(X_t,\mathcal{G}_t,P)$ be as above for which (18) and (19) are valid. Then $(X_t,\mathcal{G}_t,P)$ is a quasi martingale and converges a.e. to an integrable function.

Proof : Using (6) yields

$$E(|E^{\mathcal{G}_t}X_{t+1} - X_t|)$$

$$= E(|\beta(t)-\alpha(t)-\alpha(t)\beta(t)|)$$

$$\leq E(\beta(t))+E(\alpha(t))+E(\alpha(t)\beta(t)). \quad (31)$$

(18) and (19) imply that the series consisting of the first two terms in the above addition converge. From this it also follows that

$$\sum_{t=1}^{\infty} E(\alpha(t)\beta(t)) < \infty$$

by the comparison test for series and since all $\alpha(t),\beta(t)$ are positive and inferior to 1. This concludes the proof of the theorem, by involving theorem III.3, again because the process (X_t) is uniformly bounded, hence L^1-bounded. □

Note : From the proof it is easy to see that conditions (18) and (19) can be relaxed to the single condition

$$\sum_{t=1}^{\infty} E(|\beta(t)-\alpha(t)-\alpha(t)\beta(t)|) < \infty \quad (32)$$

IV. Problems and suggestions for further research.

IV.1 Work out the stochastic process X_t with X_t as in (1). Other σ-algebras (than the $\mathfrak{F}_t$ s) wil be involved and we have here a process with values in a Banach space, e.g. the Hilbert space L^2 (to take the easiest one). See also Egghe (1998) for a first attempt.

IV.2 This paper only studies one of the simplest problems, namely : what is left from A as a source journal after t steps. The more intricate problem consists of describing the stable limit set after t steps (cf. the above problem). So, here we do not focus so much on A but on where we are going to. Also, if we start from two different sets (say A and B), do we end up with a fixed "limit" set of source journals, for high t ? (cf. the suggestions of Rousseau and Spinak (1996)). Here also values X_t as in IV.1 will be encountered, hence the values are in infinite dimensional Banach spaces.

If we do not end up with one stable limit set, how then does the limit set (found when we start with A) relates to the limit set (found when we start with B) and what is the relation with $A \cap B$ or $\#(A \cap B)$?

IV.3 Model the total number of source journals that ever existed from $t=1$ on (from A or from U in total).

IV.4 How can rankings (e.g. based on impact factors) be involved in these models? What is the relation of a ranking (at a certain time t) of a source journal with the number of times the journal was a source journal ?

IV.5 Is it possible to apply fixed point theory (cf. Smart (1974), Istratescu (1981)) to these type of problems ?

IV.6 What is the stability of the results obtained here (and of the other problems raised here) i.e. if we change the criteria to become a source journal a little bit, will the set of source journals, when followed for t increasing, experience dramatic changes or not ?

IV.7 It is our feeling that these models can also be applied to other domains in information science : evolution of retrieved sets over time (e.g. w.r.t. a fixed query), evolutions of bibliographies, of research groups, etc.

IV.8 What is the involvement of stochastic differential equations in this (cf. Gard (1988), Gardiner (1997)) ?

References

Austin, D.G., Edgar, G.A. and Ionescu Tulcea, A. (1974). Pointwise convergence in terms of expectations. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 30, 17-26.

Bellow, A. (1978). Uniform amarts : a class of asymptotic martingales for which almost sure convergence obtains. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 41, 177-191.

Cronin, B. (1981). Transatlantic citation patterns in educational psychology. Education Libraries Bulletin 24, 48-51.

Edgar, G.A. and Sucheston, L. (1992). Stopping times and directed processes. Cambridge University Press, Cambridge, UK.

Egghe, L. (1984). Stopping time techniques for analysts and probabilists. London Mathematical Society Lecture Notes Series 100, Cambridge University Press, Cambridge, UK.

Egghe, L. (1995). Extension of the general "success breeds success" principle to the case that items can have multiple sources. Proceedings of the fifth biennial Conference of the international Society for Scientometrics and Informetrics, Rosary College, River Forest, Ill., USA, 1995. Learned Information, Medford, NJ.

Egghe, L. (1996). Source-item production laws for the case that items have multiple sources with fractional counting of credits. Journal of the American Society for Information Science 47(10), 730-748.

Egghe, L. (1998). The evolution of core collections can be described via Banach space valued stochastic processes. To appear in Mathematical and Computer Modelling.

Egghe, L. and Rousseau, R. (1990). Introduction to informetrics. Quantitative methods in library, documentation and information science. Elsevier, Amsterdam.

Egghe, L. and Rousseau, R. (1995). Generalized success-breeds-success principle leading to time dependent informetric distributions. Journal of the American Society for Information Science 46, 426-445.

Egghe, L. and Rousseau, R. (1996). Stochastic processes determined by a general success-breeds-success principle. Mathematical and Computer Modelling 23(4), 93-104.

Gaillard, J. (1989). La science du tiers monde est-elle visible ? La Recherche 20, 636-640.

Gard, T.C. (1988). Introduction to stochastic differential equations. Marcel Dekker, New York.

Gardiner, C.W. (1997). Handbook of stochastic methods for physics, chemistry and the natural sciences. Springer, Berlin.

Garfield, E. (1979). Journal citation studies 32. Canadian journals, Part 2 : an analysis of Canadian research published at home and abroad. Current Contents, August 20. Reprinted in : Essays of an Information Scientist, 4 (ISI, Philadelphia), 249-253, 1981.

Halmos, P.R. (1974). Measure theory. Graduate Texts in Mathematics 18, Springer Verlag, New York.

Inhaber, H. and Alvo, M. (1978). World science as an input-output system. Scientometrics 1, 43-64.

Istratescu, V.I. (1981). Fixed point theory. An introduction. Mathematics and its Applications 7. D. Reidel Publishing Company, Dordrecht.

Mucci, A.G. (1976). Another martingale convergence theorem. Pacific Journal of Mathematics 62(2), 539-541.

Neveu, J. (1975). Discrete-parameter martingales. North-Holland Publishing Company, Amsterdam.

Rousseau, R. and Spinak, E. (1996). Do a field list of internationally visible journals and their journal impact factors depend on the initial set of journals ? A research proposal. Journal of Documentation 52(4), 449-456.

Smart, D.R. (1974). Fixed point theorems. Cambridge University Press, Cambridge, UK.

Velho, L. (1986). The "meaning" of citation in the context of a scientific peripheral country. Scientometrcs 9, 71-89.

Velho, L. (1987). The author and the beholder : how paradigm commitments can influence the interpretation of research results. Scientometrics 11, 59-70.

Vinkler, P. (1986). Evaluation of some methods for the relative assessment of scientific publications. Scientometrics 10, 157-177.

Appendix.

For elementary notions on measure theory we refer to Egghe (1984) or to Halmos (1974).

Let $(\Omega,\mathcal{G},P)$ be a probability space, where $\mathcal{G}$ is a σ-algebra. Let $t\in\mathbb{N}$ and $\mathcal{G}_t\subset\mathcal{G}$ be a sub-σ-algebra of $\mathcal{G}$ such that $(\mathcal{G}_t)_{t\in N}$ increases. Let X_t be an integrable function. We say that the process $(X_t,\mathcal{G}_t,P)$ is a stochastic process (or adapted sequence) if each X_t is $\mathcal{G}_t$-measurable (cf. Egghe (1984)).

Let $\mathscr{F}\subset\mathcal{G}$ be an arbitrary sub-σ-algebra of $\mathcal{G}$. For every integrable function f (integrable w.r.t. $(\Omega,\mathcal{G},P)$) there exists a a.e. unique integrable

$$E^{\mathscr{F}}(f) \qquad \text{(A1)}$$

such that

$$\int_A E^{\mathscr{F}}(f) = \int_A f \qquad \text{(A2)}$$

$\forall A\in\mathscr{F}$ and such that $E^{\mathscr{F}}(f)$ is $\mathscr{F}$-measurable. $E^{\mathscr{F}}(f)$ is called the conditional expectation of f w.r.t. $\mathscr{F}$ (cf. Egghe (1984)).

Let $(X_t,\mathcal{G}_t,P)$ be a stochastic process. It is called a martingale if, for every $t\in\mathbb{N}$,

$$E^{\mathcal{G}_t}(X_{t+1}) = X_t, \text{ a.e.} \qquad \text{(A3)}$$

We also denote $E^{\mathcal{G}_t}(X_{t+1})$ by $E^{\mathcal{G}_t}X_{t+1}$ if no confusion can arise. The process is called a submartingale if

$$E^{\mathcal{G}_t}X_{t+1} \geq X_t, \text{ a.e.} \qquad \text{(A4)}$$

and a supermartingale if

$$E^{\mathcal{G}_t}X_{t+1} \leq X_t, \text{ a.e.} \tag{A5}$$

Positive supermartingales and L^1-bounded submartingales converge a.e. to an integrable function X_∞ (cf. Neveu (1975)).

A classical interpretation of (A3), (A4) and (A5) goes in the direction of gambling : if X_t denotes the gambler's fortune at time t, a martingale is a process where the gambler can expect to keep his fortune (on the average) ; he/she will win (on the average) in case of a submartingale and will loose (on the average) in case of a supermartingale. The latter case is the most likely one since casinos must make money to pay for salaries and infrastructure.

One can think of more "irregular" processes for which conditional expectations do not behave in a "monotonic" way as described by (A4) or (A5). These extensions have been formulated and studied about two decades ago (hence relatively recent). One such an extension is easy to formulate but leads to a non-trivial stochastic process : the so-called martingale in the limit :

A stochastic process $(X_t, \mathcal{G}_t, P)$ is called a martingale in the limit (shortly mil) if

$$\lim_{t\to\infty} \sup_{t'\geq t} |E^{\mathcal{G}_t}X_{t'} - X_t| = 0, \text{ a.e.} \tag{A6}$$

We refer again to theorem III.1 of Mucci (1976) stating that L^1-bounded mils converge a.e. to an integrable function X_∞.

This is a very good result and, in a way, optimal. Indeed, relaxing (A6) to

$$\lim_{t\to\infty} |E^{\mathcal{G}_t}X_{t+1} - X_t| = 0, \text{ a.e.} \tag{A7}$$

leads to possibly divergent processes (X_t).

However, requiring that

$$\sum_{t=1}^{\infty} E(|E^{\mathcal{G}_t} X_{t+1} - X_t|) < \infty \tag{A8}$$

leads to convergent processes (see below). A process that satisfies (A8) is called a quasi martingale. It is a special case of another generalization of martingales, namely the asymptotic martingales (amarts for short). We can introduce amarts as follows. Let $(X_t, \mathcal{G}_t, P)$ be a stochastic process on Ω. A stopping time τ is a function

$$\tau : \Omega \rightarrow \mathbb{N} \tag{A9}$$

such that, for every $t \in \mathbb{N}$

$$\{\omega \in \Omega \| \tau(\omega) = t\} \in \mathcal{G}_t. \tag{A10}$$

We shortly, denote

$$\{\tau = t\} = \{\omega \in \Omega \| \tau(\omega) = t\} \tag{A11}$$

Let T denote the set of all bounded stopping times. Hence every $\tau \in T$ can only have a finite range in $\mathbb{N}$. For $\tau \in T$, denote X_τ by

$$(X_\tau)(\omega) = X_{\tau(\omega)}(\omega) \tag{A12}$$

$$= \left(\sum_{t=\min \tau}^{\max \tau} X_t \, \chi_{\{\tau = t\}} \right)(\omega),$$

where $\chi_{\{\tau=t\}}$ denotes the characteristic function of $\{\tau = t\}$, i.e. the function which is 1 if $\tau(\omega) = t$ and 0 if $\tau(\omega) \neq t$.

The process $(X_t, \mathcal{G}_t, P)$ is called an amart if the net

$$\int_\Omega X_\tau = \int_\Omega X_\tau(\omega)dP(\omega) \tag{A13}$$

converges. Here we use the natural order $\leq$ on T : $\sigma \leq \tau$ iff $\sigma(\omega) \leq \tau(\omega)$ for every $\omega \in \Omega$.

There is another definition, equivalent with the notion of amarts for real valued processes as is the case here, which is called "uniform amart". A process $(X_t, \mathcal{G}_t, P)$ is called a uniform amart if

$$\lim_{\sigma \in T} \sup_{\substack{\tau \geq \sigma \\ \tau \in T}} E(|E^{\mathcal{G}_\sigma} X_\tau - X_\sigma|) = 0 \tag{A14}$$

Here, logically, we use the definition

$$\mathcal{G}_\sigma = \{S \in \mathcal{G} \| S \cap \{\sigma = t\} \in \mathcal{G}_t, \text{ for each } t \in \mathbb{N}\} \tag{A15}$$

Quasi martingales are uniform amarts, hence amarts (see Egghe (1984), p.123-124). There is the following convergence result for (uniform) amarts, hence for quasi martingales : every L^1-bounded (uniform) amart converges a.e. to an integrable function. The proof that amarts converge was given in Austin, Edgar and Ionescu Tulcea (1974). The proof that uniform amarts converge (in the vector space context) was given in Bellow (1978).

Problem : Find a condition on our process $(X_t, \mathcal{G}_t, P)$ such that it becomes an amart (and not necessarily a quasi martingale).