

Methodol Comput Appl Probab (2013) 15:821–839
DOI 10.1007/s11009-012-9281-z

Alarm Systems and Catastrophes from a Diverse Point of View

Pasquale Cirillo · Jürg Hüsler · Pietro Muliere

Received: 1 September 2011 / Revised: 29 February 2012 /
Accepted: 12 March 2012 / Published online: 25 March 2012
© Springer Science+Business Media, LLC 2012

Abstract Using a chain of urns, we build a Bayesian nonparametric alarm system to predict catastrophic events, such as epidemics, black outs, etc. Differently from other alarm systems in the literature, our model is constantly updated on the basis of the available information, according to the Bayesian paradigm. The papers contains both theoretical and empirical results. In particular, we test our alarm system on a well-known time series of sunspots.

Keywords Alarm system · Catastrophe · Risk analysis · Forecasting · Exchangeability · Urn model

AMS 2010 Subject Classifications 60G20 · 60G70 · 60G09 · 97Kxx

1 Alarm Systems, an Introduction

In observing a given phenomenon, it is often important to predict if the variable or quantity of interest will or will not reach a predetermined critical level sometime in the near future. This type of analysis becomes fundamental when the studied phenomenon may have a strong impact on human life, as in the case of floods,

This work has been partially supported by the Swiss National Science Foundation.

P. Cirillo (✉) · J. Hüsler
Institute of Mathematical Statistics and Actuarial Science (IMSV), University of Bern,
Sidlerstrasse 5, CH3012 Bern, Switzerland
e-mail: pasquale.cirillo@stat.unibe.ch

J. Hüsler
e-mail: juerg.huesler@stat.unibe.ch

P. Muliere
Department of Decision Sciences, Bocconi University, Milan, Italy
e-mail: pietro.muliere@unibocconi.it

tides, epidemics or large power outages. In these cases the critical level frequently corresponds to some sort of catastrophe and the possibility of giving an alarm in due time is invaluable.

An alarm system is a tool that predicts the occurrence of a catastrophic event at a specified time in the future on the basis of the available information. To be more exact, an alarm system is meant to activate when the probability of a catastrophe in a given time horizon overcomes a certain threshold of tolerance.

In the literature (e.g. Antunes et al. 2003; or Grage et al. 2010), an alarm system is said optimal when, for a set of available data, it possesses the highest probability of correct alarm. This means that when the alarm is given, the probability that the catastrophic event actually manifest itself is the highest among all possible systems.

As outlined in Monteiro et al. (2008), a naive alarm system is represented by the predictor $\hat{X}_{t+h} = E[X_{t+h}|X_s, -\infty < s \leq t, h > 0]$, where an alarm is cast every time the predictor exceeds some risk level. Naturally this system is not optimal at all, because, as shown in de Maré (1980), it does not show good performances in detecting exceedances, in correctly locating them in time and in reducing the number of false alarms.

The principles of optimal prediction in level crossings for continuous-valued processes have been outlined in Lindgren (1975a, b, 1980), de Maré (1980) and Svensson et al. (1996). A first Bayesian approach has been proposed in Amaral-Turkman and Turkman (1990). More recently Antunes et al. (2003) have derived interesting results for autoregressive models of order k .

For the discrete case, results can be found in Monteiro et al. (2008), Zheng et al. (2006) and Brännäs and Nordström (2004). In particular, Monteiro et al. (2008) develop an alarm system that predicts whether a count process will upcross a certain threshold. A nice model using neural networks is presented in Grage et al. (2010).

For a quite complete review of the literature related to alarm systems we refer to Monteiro et al. (2008) and Grage et al. (2010).

Following de Maré (1980) and Antunes et al. (2003), we now give a formal description of some basic theoretical concepts related to alarm systems and catastrophe prediction.

Let $\{X_t\}$ be a stochastic process with parameter space $\theta \in \mathbb{R}^d$, d fixed. For some $q > 0$, we divide the time sequel $\{1, \dots, t-1, t, t+1, \dots\}$ into three intuitive sections: the past $\{1, \dots, t-q\}$, the present $\{t-q+1, \dots, t\}$ and the future $\{t+1, \dots\}$. Hence we have that $\mathbf{X}_1 = \{X_1, \dots, X_{t-1}\}$, $\mathbf{X}_2 = \{X_{t-q+1}, \dots, X_t\}$ and $\mathbf{X}_3 = \{X_{t+1}, \dots\}$ represent respectively the informative, the present and the future experiment at time t .

A catastrophe C_{t+j} is an event in $\sigma(\mathbf{X}_3)$, the σ -algebra generated by \mathbf{X}_3 . Without loss of generality, the catastrophe is assumed to occur in j steps from time t . Even if C_{t+j} can be any event, it is usual to see it as a rare event, i.e. an event whose occurrence in time or whose magnitude is not common. A nice definition of rare event is given in Hüsler (1993).

An alarm (or predictor event) for C_{t+j} is any event in $\sigma(\mathbf{X}_2)$, a typical example being the process entering some risk region or overcoming some risk level.

We say that an alarm is cast at time t for a future event C_{t+j} if the observed value \mathbf{x}_2 of \mathbf{X}_2 belongs to the predictor event. If a catastrophe occurs after that an alarm is given, the catastrophe is detected. If, on the contrary, the catastrophe does not occur we have a false alarm.

From Antunes et al. (2003) we borrow the following definitions.

Definition 1 An event $A_t \in \sigma(\mathbf{X}_2)$ is an alarm for $C_{t+j} \in \sigma(\mathbf{X}_3)$ if, whenever $\mathbf{x}_2 \in A_t$, it is likely that C_{t+j} takes place, i.e. $P(C_{t+j}|A_t) \geq \psi$, where $\psi \in [0, 1]$ is a predetermined value.

Definition 2 The alarm $A_t \in \sigma(\mathbf{X}_2)$ has size α_t if $P(A_t|\mathbf{X}_1) = \alpha_t$, and detection probability γ_t if $P(C_{t+j}|A_t, \mathbf{X}_1) = \gamma_t$. We also say that A_t is optimal of size α_t if

$$P(C_{t+j}|A_t, \mathbf{X}_1) = \sup_{B \in \sigma(\mathbf{X}_2)} P(C_{t+j}|B, \mathbf{X}_1), \tag{1}$$

where the supremum is taken over all the sets $B \in \sigma(\mathbf{X}_2)$ such that $P(B|\mathbf{X}_1) = \alpha_t$.

All in all, as pointed out in de Maré (1980), in predicting catastrophes and developing alarm systems, one is interested in:

- the alarm size $P(A_t|\mathbf{X}_1)$;
- the probability of detecting the catastrophe $P(C_{t+j}|A_t, \mathbf{X}_1)$;
- the probability of correct alarm $P(A_t|C_{t+j}, \mathbf{X}_1)$;
- the probability of false alarm $P(A_t|C_{t+j}^c, \mathbf{X}_1)$;
- the probability of undetected event $P(C_{t+j}|A_t^c, \mathbf{X}_1)$.

These quantities are also called operating characteristics of an alarm system. For further details we refer to Antunes et al. (2003).

2 An Alternative Modeling

Starting from Cirillo et al. (2010) and Cirillo and Hüsler (2011), we aim to propose a Bayesian nonparametric approach to catastrophe prediction using urn processes.

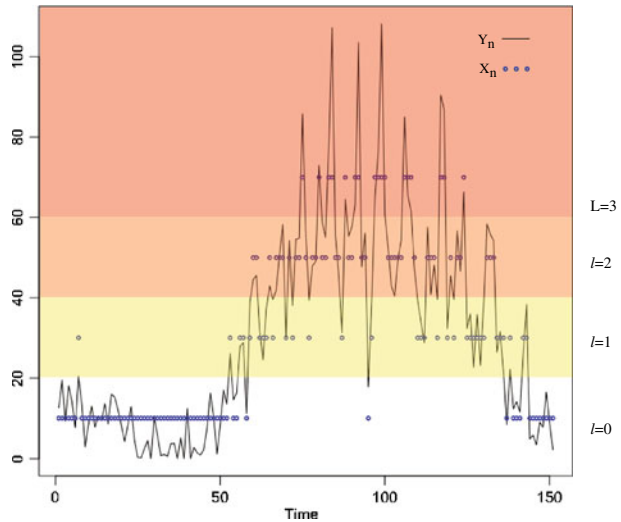
Urn processes form a very large family of probabilistic models in which the probability of certain events is represented in terms of sampling, replacing and adding balls in one or more urns or boxes. They have powerful combinatorial properties and their use in Bayesian nonparametrics is well known, see for example Coppersmith and Diaconis (1986) and Muliere et al. (2000). Good introductions to urn models and their properties are Johnson and Kotz (1977) and Mahmoud (2009).

Our idea is to use a special process that can only assume a finite number of values over time but that, thanks to Polya-like reinforcement mechanism, is able to learn from the past. Every value must be seen as a level of risk, on a scale that goes from 0, no risk, to L, catastrophe. We will assume time to be discrete, in order to have a more intuitive description of the model. Nevertheless, it is possible to extend our alarm system to continuous time using the results of Muliere et al. (2003).

The use of a finite-valued process may seem reductive in many cases where the observed phenomenon assumes a continuous range of values. In reality, if $\{X_n\}$ is our process, it can be seen as the “simplified” version of another underlying process $\{Y_n\}$, characterized by a (much) larger space of states. In Fig. 1 we give a stylized idea of what we mean. For the underlying process $\{Y_n\}$ we define four different regions of risk $([0, 20], (20, 40], (40, 60], (60, +\infty))$ so that when $\{Y_n\}$ is in lowest region, $[0, 20]$, $\{X_n\}$ assumes the value 0, meaning no risk. Whether $\{Y_n\}$ enters the $(20, 40]$ region, $\{X_n\}$ becomes equal to 1 and so on. For $\{X_n\}$ equal to 3 we have that $\{Y_n\}$ has entered the catastrophe region, i.e. $(60, +\infty)$.

Categorizing data surely implies some loss of information. The precise amount of this loss is still an open question, since it depends on several different factors.

Fig. 1 Stylized idea of a finite-value process $\{X_n\}$ as approximation of a more complex one $\{Y_n\}$



Heuristically, we have that, the larger the number of categories, the smaller the amount of information we lose. Similarly, the smoother the underlying process $\{Y_n\}$, the smaller is the number of categories necessary to satisfactorily follow it, since sudden jumps are generally ruled out. We will come back to this point later on in Section 3.

The construction we propose belongs to the wider class of reinforced urn processes (RUP). RUP have been introduced in Muliere et al. (2000) as reinforced random walks on a state space of urns. They have then been further extended in Muliere et al. (2003). In the last years several generalizations and applications have been presented in the field of clinical trials (Mezzetti et al. 2007), survival analysis (Bulla 2005), risk modeling (Amerio et al. 2004; Cirillo et al. 2010) and shock models (Cirillo and Hüsler 2011).

2.1 The Urn-based Alarm System

We take into account the following:

- a state space $V = \mathbb{N}_0^+ \times \{0, 1, 2, \dots, L\}$, whose elements (n, l) represent levels of risk l at time instant n . The last level of risk L corresponds to a catastrophe.
In the rest of the paper we indicate with $v \in V$ a general state in V . With the notation (n, l) , on the contrary, we refer to a specific couple $(n, l) \in V$.
- every state $v \in V$ is endowed with a Polya urn $U(v)$, i.e. an urn that is sampled with replacement and reinforcement. In other words, every time we sample a ball of a given color, we replace that ball, also adding $s(v) > 0$ extra balls of the same color.
- every urn $U(v)$ is characterized by a set of 4 colors $C = (c_1, c_2, c_3, c_4)$ (think for example c_1 is white, c_2 is red, c_3 is green and c_4 is black) and a reinforcement $s(v)$. With $m_v(c)$ we indicate the number of balls of color c in urn $U(v)$. Notice that $m_v(c)$ may be equal to 0 for some c .

- the composition of each urn $U(v)$, $v \in V$, is given by $Q(v) = \sum_{i=1}^4 m_v(c_i)$, i.e. the sum of all the balls in the urn.
- a function $d : V \times C \rightarrow V$ that represents a rule of motion.

Now, fix an initial state $(0, 0)$ and recursively define a reinforced random walk $\{X_n\}$ on V starting in $(0, 0)$, such that: $X_0 = (0, 0)$. For all $n \geq 1$, if $X_{n-1} = v \in V$, a ball is sampled from $U(v)$, its color c is registered and the ball is returned into the urn together with $s(v)$ balls of the same color. At this point, according to the rule of motion d , we have $X_n = d(v, c)$.

The rule of motion d is one of the most important elements of the entire construction, and we need to better specify its behavior in order to characterize the alarm system. In particular, since we have 4 possible colors in every urn, we set the following moves

$$d((n, l), c) = \begin{cases} (n + 1, l) & \text{if } c = c_1, \\ (n + 1, l + 1) & \text{if } c = c_2, \\ (n + 1, l - 1) & \text{if } c = c_3, \\ (n + 1, L) & \text{if } c = c_4. \end{cases} \tag{2}$$

Remembering that the couple (n, l) describes a level of risk l at time n , the rule of motion d says that, when we are in $X_{n-1} = (n, l) \in V$ and we sample a ball from $U((n, l))$, then

- if the color is c_1 , $X_n = (n + 1, l)$, i.e. time passes but the process remains at the same level of risk l ,
- if c_2 , the level of risk in the next time period increases by 1,
- if c_3 , the level of risk decreases by 1,
- if c_4 , the process jumps to the highest level of risk L and we have a catastrophic event.

Since the levels of risk are between 0 and L , we need to impose some conditions on the urns compositions to avoid some limiting cases. First, we set $m_{(n,0)}(c_3) = 0$ for all n . In this way, every time the process $\{X_n\}$ is at the minimum level of risk 0, the level of risk cannot further decrease, since it is not possible to sample a c_3 ball. In a similar way, $m_{(n,L)}(c_2) = 0$, so that it is not possible to reach a level of risk higher than L .

Figure 2 contains a graphical example of our urn-based alarm system. At time $n = 0$ we are in $(0, 0)$, hence $X_0 = (0, 0)$. We sample urn $U((0, 0))$. Imagine we get a c_1 (white) ball. According to the rule of motion $X_1 = d((0, 0), c_1) = (1, 0)$, that is to say we move to the next time instant but we remain at the same level of risk. In $(1, 0)$ we sample $U((1, 0))$ and we extract c_2 (red) ball, so that $X_2 = (2, 1)$, where the level of risk increases. Then we sample a c_3 green ball from $U((2, 1))$, the level of risk decreases and $X_3 = (3, 0)$. Finally in $(3, 0)$ we sample a c_4 (black) ball and the process jumps to $(4, 3)$, meaning that a catastrophe takes place.

Consider a finite sequence $\psi = ((0, 0), \dots, (n, l))$ of elements of the space V . Such a sequence is said admissible (Diaconis and Freedman 1980) for the process $\{X_n\}$ if its element can be effectively visited by $\{X_n\}$ in the order given by ψ . In other words, for every two adjacent elements of ψ , say (n, l) and $(n + 1, h)$, such that $l, h \in L$, we have that urn $U((n, l))$ contains at least one ball of color c_r , $r = 1, \dots, 4$, so that $d((n, l), c_r) = (n + 1, h)$. Considering the example of Fig. 2, we have that $((0, 0), (1, 0), (2, 1), (3, 1), (4, 0), (5, 2))$ is admissible, while

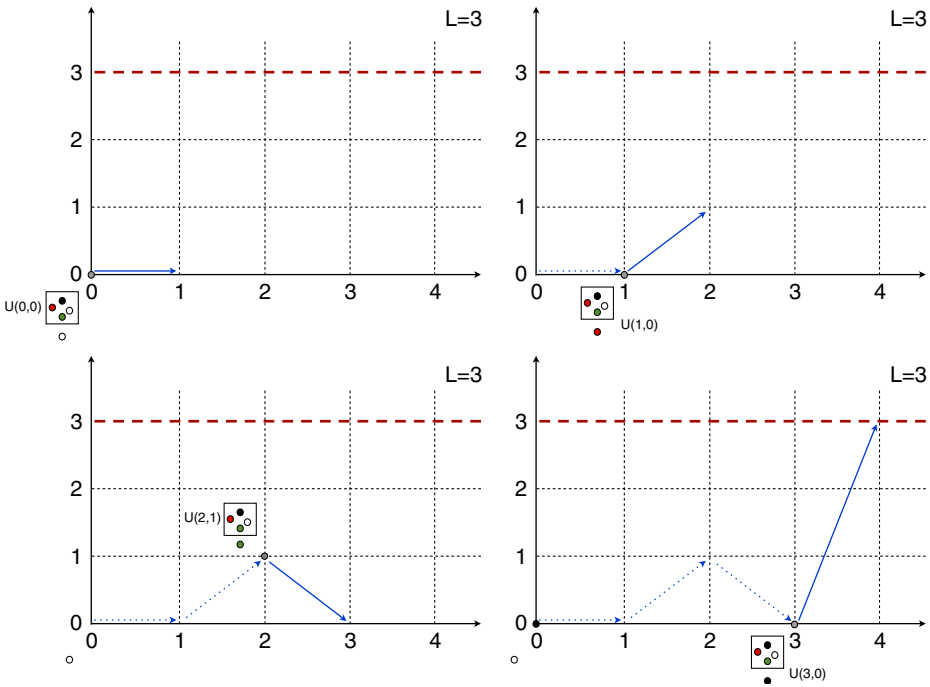


Fig. 2 Example of a realization of the urn-based alarm system with maximum level of risk $L = 3$

$((0, 0), (1, 0), (2, 0), (3, 2), (4, 0), (5, 1))$ is not. This is due to the fact that, given our construction, it is not possible to move in one step from level of risk 0 to level 2, i.e. without passing through level 1.

Again, if two finite sequences ψ_1 and ψ_2 of states of V start with the same element, say v , and, for every $u, w \in V$, the number of transitions $t(u, w)$ from state u to state w is the same for both ψ_1 and ψ_2 , then ψ_1 and ψ_2 are said equivalent (Diaconis and Freedman 1980).

In what follows, for any admissible finite sequence ψ of elements of V , set $h_v(c_i) = t(v, d(v, c_i))$, $i = 1, \dots, 4$, to be the number of transitions from state $v \in V$ to state $d(v, c_i)$, i.e. the state reachable after sampling color c_i from $U(v)$. Moreover, let $H(v) = \sum_{u \in V} t(v, u)$ be the total number of transitions from v to all the other states $u \in V$ present in ψ .

With a little abuse in notation, for $\psi = (v_{0,0}, \dots, v_{n,l})$, set $P[\{X_n\} = \psi] = P[X_0 = (0, 0), \dots, X_n = (n, l)]$. Moreover set L^* to be the set of risk levels actually visited by $\{X_n\}$ in the exact order in which they are visited. We can then explicitly obtain the joint probability of $P[\{X_n\} = \psi]$.

Theorem 1 For $n \geq 0$ and all finite sequences $\psi = ((0, 0), \dots, (n, l))$ of elements of V , we have $P[\{X_n\} = \psi] = 0$ if ψ is not admissible, otherwise

$$P[\{X_n\} = \psi] = \prod_{v \in V} \frac{\prod_{i=1}^4 \prod_{j=0}^{h_v(c_i)-1} (m_v(c_i) + js(v))}{\prod_{g=0}^{H(v)-1} (gs(v) + \sum_{i=1}^4 m_v(c_i))}, \tag{3}$$

with $\prod_0^{-1} = 1$.

Proof The proof comes directly from the construction of the process.

If ψ is not admissible, then, by definition, the process $\{X_n\}$ cannot visit the sequence of states given in ψ . For this reason $P[X_0 = (0, 0), \dots, X_n = (n, l)] = 0$.

For the second part, for ψ admissible, set $c(v, w)$ to be the color that, if sampled, brings from v to w . Hence, $\forall j, l \in L^*$,

$$P[\{X_n\} = \psi] = \prod_{i=0}^{n-1} \frac{m_{(i,j)}(c((i, j), (i + 1, l))) + q_{(i,j)}(c((i, j), (i + 1, l)))}{\sum_{i=1}^4 (m_{(i,j)}(c_i) + q_{(i,j)}(c_i))}, \tag{4}$$

where $q_{(0,0)} = 0$ and $q_{(i,j)}(c) = s((i, j)) \sum_{k=0}^{i-1} 1_{\{(k,j)=(i,j), c((k,j), (k+1,l))=c\}}$, $c \in C$.

Now let v to be an element in ψ and consider the ordered sequence $(w_1, \dots, w_{H(v)})$ of elements following v in ψ . We have that, relatively to v , Eq. 4 becomes

$$\frac{\prod_{i=1}^4 \prod_{j=0}^{h_v(c_i)-1} (m_v(c_i) + js(v))}{\prod_{g=0}^{H(v)-1} (gs(v) + \sum_{i=1}^4 m_v(c_i))} = \frac{m_v(c(v, w_1))}{\sum_{i=1}^4 m_v(c_i)} \times \frac{m_v(c(v, w_2)) + s(v)1_{\{w_1=w_2\}}}{s(v) + \sum_{i=1}^4 m_v(c_i)} \times \dots$$

$$\times \frac{m_v(c(v, w_{H(v)})) + s(v) \sum_{i=1}^{H(v)-1} 1_{\{w_i=w_{H(v)-1}\}}}{s(v)(H(v) - 1) + \sum_{i=1}^4 m_v(c_i)}.$$

and Eq. 3 follows. □

It is interesting to notice that the joint probability expressed in Eq. 3 essentially depends on the number of transitions from state u to state w in ψ . As a consequence of this, we have that, if ρ is another sequence of elements of V , equivalent to ψ , then $P[\{X_n\} = \psi] = P[\{X_n\} = \rho]$. This fact is worth noticing, because it essentially means that the process $\{X_n\}$ is partially exchangeable in the sense of Diaconis and Freedman (1980). In Aldous (1985) a very interesting discussion about partial exchangeability and its implications is given.

2.2 Learning from the Past

The process we have defined is characterized by the Polya reinforcements of the urns it walks on. However, up till now, we have not been exploiting the mechanism of reinforcement, as a way for learning from past observations. In order to use our process for Bayesian prediction, we need to set a so-called condition of recursiveness.

Set $\tau_0 = 0$ and, for $n \geq 1$, define $\tau_n = \inf\{j > \tau_{n-1} : X_j = (0, 0)\}$. In this way, $\{\tau_n\}$ represents the sequence of times elapsed between two consecutive visits of state $(0, 0)$ by process $\{X_n\}$. We say that the reinforced urn process $\{X_n\}$ is recurrent if $P[\cap_{n=0}^{+\infty} \{\tau_n < +\infty\}] = 1$.

In particular, we can assume that every time the process $\{X_n\}$ reaches a level of risk equal to L , i.e. a catastrophe happens, then the process is reset to $(0, 0)$. Restarting the process after each catastrophe allows us to exploit past information through the updated urns. In fact, at every cycle, if a state v has already been visited in the past, the composition of urn $U(v)$ has been changed through reinforcement and we keep memory of that to make our future forecast. If a state w is on the contrary visited

for the first time, then the urn $U(w)$ is unchanged and it only contains the prior information related to its initial composition.

Given our modeling, the process $\{X_n\}$ can reach the maximum risk level L in 2 ways:

- through the extraction of a c_4 (black) ball at every time;
- through the repeated sampling of c_2 (red balls).

Hence, it is easy to verify that a simple condition for recursiveness is

$$\lim_{n \rightarrow +\infty} \prod_{i=0}^n \frac{m_{(i,j)}(c_1) + m_{(i,j)}(c_3)}{\sum_{k=1}^4 m_{(i,j)}(c_k)} = 0, \quad \forall j \in L. \tag{5}$$

If the process is recurrent and we reset it to $(0, 0)$ every time it reaches the maximum level of risk, then we can define the so-called 0-blocks, i.e. blocks of states of V visited by $\{X_n\}$, that start with $(0, 0)$ and terminate with (n, L) , $n \geq 1$. Assume $L = 3$, an example can be: $((0, 0), (1, 0), (2, 1), (3, 1), (4, 2), (5, 3), (0, 0), (1, 1), (2, 2), (3, 3), (0, 0), (1, 0), (2, 0), (3, 3), (0, 0), (1, 1), (2, 2), (3, 2), (4, 2), (5, 3), \dots)$. Then we have the following grouping in terms of 0-blocks (we just consider the levels of risk, omitting the time index):

$$\underbrace{0, 0, 1, 1, 2, 3}_{\text{Block 1}} \mid \overbrace{0, 1, 2, 3}^{\text{Block 2}} \mid \underbrace{0, 0, 0, 3}_{\text{Block 3}} \mid \overbrace{0, 1, 2, 2, 2, 3}^{\text{Block 4}} \mid \dots$$

We have already seen that $\{X_n\}$ is partially exchangeable. If $\{X_n\}$ is also recurrent then, as showed in Diaconis and Freedman (1980), $\{X_n\}$ can be expressed as mixture of Markov Chains. Intuitively this means that every 0-block within $\{X_n\}$ is a Markov chain, an immediate consequence of our construction, and that the 0-block are exchangeable, i.e. their joint distribution (Eq. 3) is immune to permutations.

Following Muliere et al. (2000) let $R^{(p)}$ be the states that a reinforced urn process $\{X_n\}$, with initial state $(0, 0)$, can visit in p time steps with positive probability. Also set $R = \cup_{p=0}^{+\infty} R^{(p)}$. Theorem 1 assures that with probability 1 the states visited by $\{X_n\}$ are elements of R , i.e. $P[X_0 \in R, \dots, X_n \in R] = 1$.

For Diaconis and Freedman (1980), the fact that $\{X_n\}$ is a mixture of Markov chains means that there exists a probability distribution μ on the set M of stochastic matrices on $R \times R$ such that, for $n \geq 1$ and $h, k \in L$,

$$P[X_0 = (0, 0), \dots, X_n = (r, l)] = \int_M \prod_{i=0}^{r-1} \pi((i, h), (i + 1, k)) \mu(d\pi). \tag{6}$$

Now, let Π a random element of M with probability μ . For every $v \in R$, let $\Pi(v)$ be the v -th row of Π .

A simple adaptation of Theorem 2.16 in Muliere et al. (2000) tells us that, for $l \in L$, $\Pi((0, 0)), \dots, \Pi((n, l))$ are mutually independent random distributions on R , and that the law of the generic $\Pi(v)$ is that of a Dirichlet process $Dirichlet\left(\frac{m_v(c_i)}{s(v)} : i = 1, \dots, 4\right)$.

All this implies that the vector of probability measures $(\Pi((0, 0)), \dots, \Pi((n, j)))$, $j \in L$, is a Product Dirichlet process F , as introduced in Cifarelli and Regazzini (1978) and further analyzed in Giudici et al. (2003). This also suggests that the

process F is the de Finetti measure, i.e. the mixing measure, of the 0-blocks, and its parameters depend on the initial compositions of the urns.

In other words, for $(n + 1, l), l \in L$, the random mass assigned by F to the subset $\{(0, 0), \dots, (n + 1, l)\}$ is equal to

$$1 - \prod_{p=1}^{n+1} (1 - Y_{p,l}), \quad l \in L^* \tag{7}$$

where $Y_v, v \in V$, are independent random variables such that Y_v is distributed as a

$$\text{Dirichlet} \left(\frac{m_v(c_i)}{s(v)} : i = 1, \dots, 4 \right). \tag{8}$$

Product Dirichlet processes are a special kind of neutral to the right processes (Doksum 1974).

Theorem 2 *If $\{X_n\}$ is recurrent, the sequence $\{\tau_k\}$ is exchangeable and its de Finetti measure is the Product Dirichlet process defined in Eq. 8.*

Proof The stopping time τ_i is a measurable function of the i -th 0-block. Since 0-blocks are exchangeable, given the partial exchangeability of $\{X_n\}$, we have that the sequence $\{\tau_k\}$ is exchangeable too and its de Finetti measure is the one associated to the 0-blocks. □

As shown in Muliere et al. (2000), reinforced urn processes are conjugated. This means that if $\{X_n\}$ is a RUP characterized by a space state V , a set of colors C , a set of urns $U = \{U(v) : v \in V\}$ and a law of motion d , then if $\{X_n\}$ is also recurrent, given $X_0 = (0, 0), \dots, X_n = (r, l)$, the process (X_n, X_{n+1}, \dots) is a $RUP(V, C, \tilde{U}, d)$, where \tilde{U} is the set of updated urns. The conjugacy of RUP is strictly connected with the conjugacy of Product Dirichlet processes, as special types of neutral to the right processes.

Being a special case of RUP, our urn-based alarm system is also conjugated. This means that, thanks to recurrence and Eq. 3, we are able to produce nonparametric Bayesian prediction. In fact, when we initialize the process for the first time, we can use our prior knowledge of the phenomenon under investigation by acting on the initial composition the urns. For example, if we have no particular knowledge, we can initialize all the urns in the same way, with exactly the same number of colors. If, on the contrary, we know that the risk of catastrophe increases over time, i.e. there is a trend, then we can put a larger number of c_4 balls in $U((n, l))$ for n growing. Infinite possibilities are available. We will come back to this topic in the applied section of the paper.

2.3 Alarm: the Next Big One

Imagine we are in the first cycle of our urn-based alarm process and that we are visiting state $(n, l), l < L$. This means that we have not observed any catastrophic event yet. The probability of observing a catastrophe at time $n + k$ is $P(\tau_1 = n + k)$.

An intuitive (and - we will see - effective) alarm mechanism is the following: whenever $P(\tau_1 = n + k) \geq \gamma$, with $\gamma \in [0, 1]$ a predetermined quantity, an alarm is cast at time n .

Suppose now that we have observed $r - 1$ catastrophes so far, and hence we are in the r -th cycle of our urn process. Moreover, consider we are visiting state (n, l) , $l < L$, and we want to compute the probability of having a catastrophe (reaching level of risk L) in k steps, i.e. $P(\tau_r = n + k)$.

Figure 3 contains three simple examples of the functioning of the urn-based alarm system.

During the r -th cycle, we are in state (n, l) , $l < L - 1$, as outlined in Fig. 3a. What is the probability of having a catastrophic event in $n + 1$, i.e. $P(\tau_r = n + 1)$? As clearly visible in the Figure, the only possibility is to sample a c_4 (black) ball from urn $U((n, l))$. Black balls make the process jump to the maximum level of risk. Calculating $P(\tau_r = n + 1)$ is hence trivial.

Consider now Fig. 3b. The process is in $(n, L - 1)$. In this case the probability of a catastrophe at time $n + 1$ is given by the probability of extracting a black ball, as before, but we also have to consider the probability of extracting a c_2 (red) ball, that makes the level of risk increase by a factor of 1.

Finally a little more general case is sketched in Fig. 3c. We are in $(n, L - 2)$ and we are interested in $P(\tau_r = n + 2)$. The possible paths to catastrophe are:

- extraction of a red ball in $(n, L - 2)$ and a red ball in $(n + 1, L - 1)$, i.e. $\{c_2, c_2\}$;
- extraction of a red ball in $(n, L - 2)$ and a black ball in $(n + 1, L - 1)$, i.e. $\{c_2, c_4\}$;

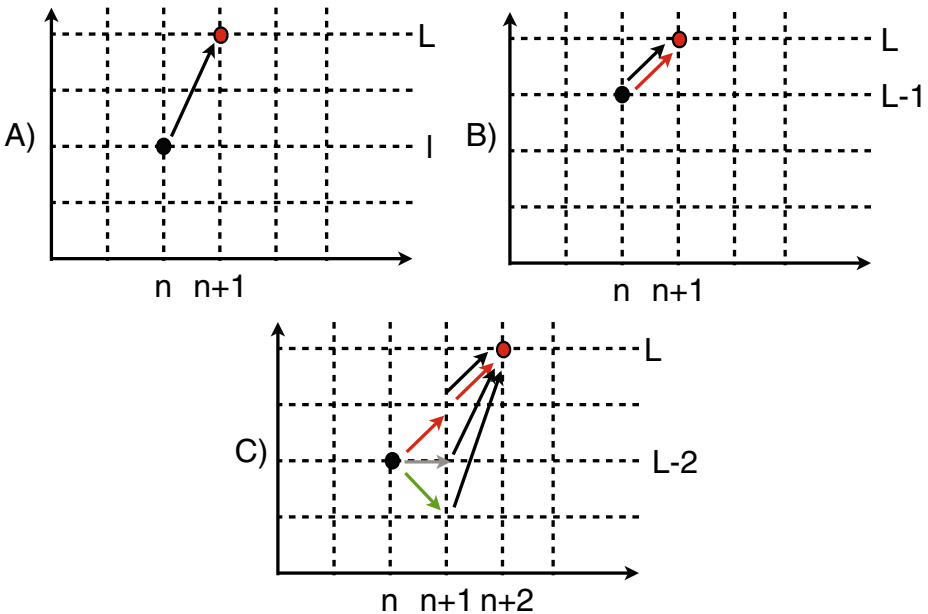


Fig. 3 Graphical examples of the functioning of the urn-based alarm system. The color of the arrows indicates the possible move after the extraction of a ball of the same color

- extraction of a white ball in $(n, L - 2)$ and a black ball in $(n + 1, L - 2)$, i.e. $\{c_1, c_4\}$;
- extraction of a green ball in $(n, L - 2)$ and a black ball in $(n + 1, L - 3)$, i.e. $\{c_3, c_4\}$.

Thus we have to consider all these probabilities together in order to compute $P(\tau_r = n + 2)$, and if $P(\tau_r = n + 2) \geq \gamma$ an alarm is given. Notice that if we want to analyze the event “catastrophe within $n + 2$ ”, we also have to add the case “black ball in $(n, L - 2)$ ”.

It is easy to understand that, in general, computing $P(\tau_r = n + k)$ involves taking into account a lot of different possibilities. With simple combinatorial reasoning, we have that for the probability of having a catastrophe at time $n + k$ when we are in $(n, L - k)$, there are a maximum of $3^{k-1} + 1$ possible paths to consider. Base 3 comes from observing that we have 4 possible colors in the urns, but that we do not want to extract black balls. Notice then that the number of actual paths could be smaller. A typical example is given by the extraction of $k - 1$ successive green balls: this could be impossible because we know that, whenever we reach risk level 0, there is no green ball in urn $U((r, 0))$, $\forall r$.

If the level of risk of the state visited by $\{X_n\}$ is less than $L - k$, then there are at most 3^{k-1} paths, i.e. all the possible extractions of not black balls up to $(n + k - 2, l)$, $\forall l$, and then a black ball in $(n + k - 1, l)$. Whether we consider the probability of a catastrophe within k steps, we obviously have to consider all the intermediate possibilities $P(\tau_r = n + k - 1)$, $P(\tau_r = n + k - 2)$, etc.

Theorem 3 *Let $p_i^r((n, l), (n + k, L))$ be the probability of the i -th feasible path from (n, l) to $(n + k, L)$ for process $\{X_n\}$ during the r -th cycle, i.e. after that $r - 1$ catastrophes have occurred. Set M to be the total number of feasible paths between (n, l) and $(n + k, L)$, hence*

$$P(\tau_r = n + k | \tau_1, \dots, \tau_{r-1}) = \sum_{i=1}^M p_i^r((n, l), (n + k, L)), \tag{9}$$

where p_i^r is computed using Eq. 3.

Proof Once again the proof is based on the construction of the process.

Every time a catastrophe happens, the process $\{X_n\}$ is reset to $(0, 0)$ and it starts a new cycle. Given the urn construction, the probabilities associated to every feasible path p_i^r during cycle r clearly depend on the updates of the different urns so far, according to a naive Bayesian principle.

Finally, the probability of reaching state v starting from state u is simply given by the sum of all the possible paths with an associated positive probability. \square

Despite its theoretical simplicity, the practical computation of the different probabilities p_i^r can be a cumbersome procedure. We have indeed to consider all the possible paths, eliminate the non-feasible ones for $\{X_n\}$ and then calculate the probability of every single path. Luckily it is possible to use computers to speed up. Nevertheless, even in this case it can be a very slow process to predict the probability of a catastrophe in k steps when $k \geq L/2$. In any case, to foster computations, it is

possible to use algorithms as the Floyd–Warshall (Cormen et al. 1990), to individuate the feasible paths.

3 An Application to Solar Sunspots

We now propose a simple application of the urn-based alarm system using freely available data about sunspots. Sunspots are temporary phenomena on the photosphere of the sun (Chattopadhyay 2000) that appear visibly as dark spots compared to surrounding regions. They are caused by very intense magnetic activity, which inhibits convection, forming areas of reduced surface temperature. By counting the frequency and the placement of sunspots visible on the sun, it is possible to analyze the so-called solar cycle, which is the main source of the ≈ 11 -year (the actual estimate is 10.7) periodic solar variation (Grandpierre 2004). Studying sunspots and solar variation is essential for determining space weather and, to some degree, atmospheric weather and climate change. Sunspots are also somehow correlated with solar storms, hence their study can be very useful for telecommunications' safety.

The data set `sunspot.month` is available in R and it contains the monthly number of sunspots as collected by the Royal Observatory of Belgium from January 1749–December 1997. For our analysis we use a subsample of the original time series, focusing our attention on the period 01/1900–12/1990. Figure 4 shows the data. The ≈ 11 -year cycle is clearly evident. In particular we use the data going from 01/1900–12/1974 as running-in for Bayesian learning. We then perform catastrophe prediction in the period 1975–1990. Figure 5 focuses on this second period.

As shown in Fig. 5, we have chosen three threshold values in order to determine the levels of risk of our alarm system. These are: 70, 140, and 180. Starting from the data, we have that our process $\{X_n\}$ is at risk level 0 for a monthly number of sunspots in between 0 and 69, risk 1 for $[70, 139]$, risk 2 for $[140, 179]$ and risk 3 (what

Fig. 4 Monthly sunspots time series between January 1900 and December 1990

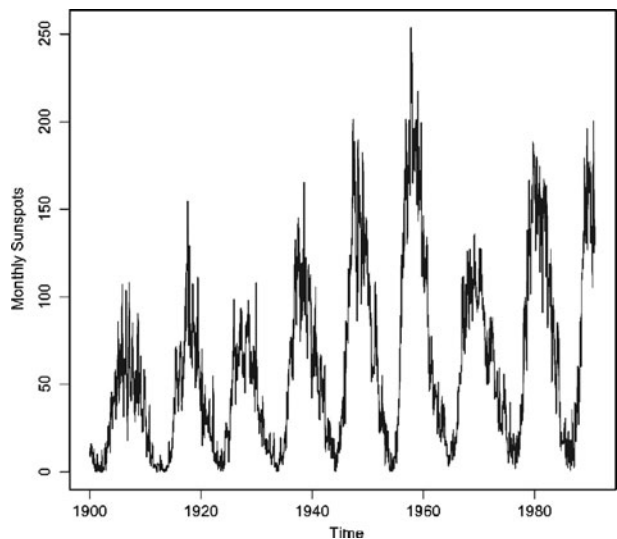
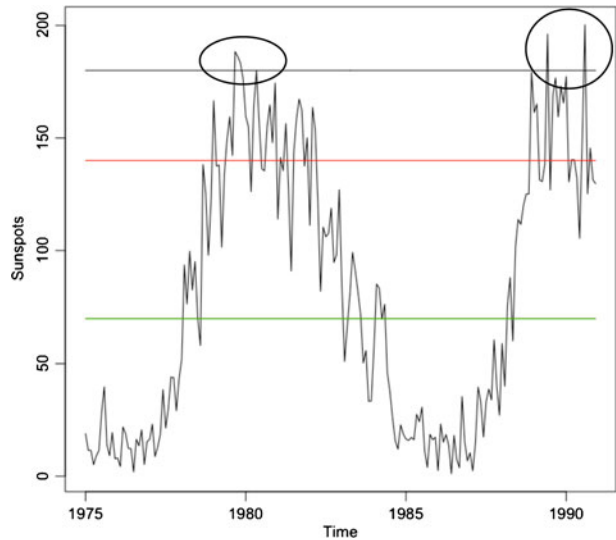


Fig. 5 Sunspots time series with thresholds of risk and catastrophes



we can call a “catastrophe”) for a number of sunspots greater than or equal to 180. In the period 1900–1990 we can observe a total of 28 catastrophes (5 in the window 1975–1990). Notice that, even if we have more or less regular peaks in the number of sunspots every 11 years, these peaks may not be catastrophic events.

Hence for our process the state space is $V = \mathbb{N}_0^+ \times \{0, 1, 2, 3\}$, and each time step represents one month.

In order to initialize the alarm system we have to explicit the compositions of all the urns $U(v)$, $v \in V$ and their reinforcement $s(v)$.

For what concerns reinforcement, we can simply start by setting $s(v) = 1$ for all $v \in V$. This choice is arbitrary and may not be optimal. For this reason, we will also discuss an alternative based on an optimality criterion, i.e. the minimization of the number of false alarms.

We know that sunspots follow a cycle of about 11 years (132 months), with a peak after about 70 months. This information could be embedded in the alarm systems by inputting a different number of balls in the urns. For example we could think about an increasing number of red c_2 balls (that increase risk) for the first half of every cycle and then an increasing number of green c_3 balls (that make risk levels decrease) for the second half. Moreover we could assume that the urns whose time index is around the 70th month have a larger number of black c_4 balls and so on.

In reality, having plenty of data, we prefer to make the system learn from them, without embedding too much prior knowledge, also considering that none of us is an expert astronomer. Our idea is then to initialize all the urns with the same starting composition, so that at the beginning of the first cycle, in January 1900, they all contain 45 c_1 , 32 c_2 , 22 c_3 and 1 c_4 balls. All the urns hence have a total of 100 balls. This means that the composition of every Polya urn follows a Dirichlet distribution $Dirichlet(45, 32, 22, 1)$. The probability of instant catastrophe is therefore 1%. This initialization is not non-informative, since we are assuming different numbers of balls for the different colors. Anyway, it does not seem absurd to think that the probability of having a catastrophe is sensibly smaller than the one of not having it.

Table 1 Results of the alarm system for the period 1975–1990

1975–90	False alarms	Correct alarms	Actual catastrophes
$k = 1$	16	4	5
$k = 2$	15	4	5
$k = 5$	21	2	5

An alarm is given in n when the probability of catastrophe in $n + k$ is equal or greater to 0.05

At this point we can start using Bayesian learning. Assume we observe the following blocks (as usual we remove the time index):

$$\underbrace{0, 0, 0, 1, 1, 2, 1, 3}_{\text{Block 1}} \mid \overbrace{0, 1, 2, 3}^{\text{Block 2}} \mid \underbrace{0, 0, 3}_{\text{Block 3}} \mid \dots$$

In other words the first catastrophe is observed after 7 time steps, so that $\tau_1 = 7$. Then we have $\tau_2 = 3$ and $\tau_3 = 2$.

Using the Polya updating rule of every urn, and view that $s(v) = 1$ for every $v \in V$, we have that, at the beginning of the 4-th cycle of the alarm system: $m_{(0,0)}(c_1) = 45 + 2$, $m_{(0,0)}(c_2) = 32 + 1$, $m_{(0,0)}(c_4) = 1, \dots$, $m_{(5,2)}(c_3) = 22 + 1$ and so on. Repeating this updating over and over using the information available in the data we can train the model, making it learn the various patterns present in the data: trends, seasons, etc.

Tables 1 and 2 contain the results of the alarm system for the period 1975–1990 after being trained over the 1900–1974 data.

We have considered 3 possible situations: $k = 1$, $k = 2$ and $k = 5$. In other words, at every time point n (i.e. a month) we try to predict the probability of having a catastrophe in $n + k$, exploiting the information gained so far. As thresholds γ for an alarm to be cast we have chosen 0.05 (Table 1) and 0.2 (Table 2).

In Table 1 we see that the alarm system is able to correctly predict 4 catastrophes in one time step with $\gamma = 0.05$. Anyway we also have to consider that 16 false alarms have been cast. For $k = 2$ the percentage of correctly detected catastrophes is still 80% and we have 15 false alarms.¹ For $k = 5$, we predict 2 catastrophes, with 21 false alarms.

In order to decrease the number of false alarms, we may increase the alarm threshold γ . As shown in Table 2 the growth of γ has an impact on the number of correctly predicted catastrophes as well. In particular, for $k = 1, 2$ we succeed in predicting 3 catastrophes, 1 for $k = 5$. It is hence evident that the choice of γ is fundamental for the alarm system to work properly. In Fig. 6 we show the number of false alarms as a function of γ with $k = 1$. Naturally also the choice of k has an important role.

¹Interestingly the number of false alarms decreases, when we pass from $k = 1$ to $k = 2$, and this may seem counterintuitive. An explanation is due to the presence, in the training set, of several couples of catastrophes (i.e. exceedances for 2 consecutive periods). This makes the alarm system believe (from what it has learnt so far) that the probability of observing a catastrophe in $t + 1$, when there is a catastrophe in t is not at all negligible, and an alarm is thus cast. Setting $k = 2$ solves this problem. It is hence clear the importance of correctly dealing with clusters of catastrophic events. A possible solution is discussed in the next section. Also notice that increasing the threshold γ corrects this peculiarity.

Table 2 Results of the alarm system for the period 1975–1990

1975–90	False alarms	Correct alarms	Actual catastrophes
$k = 1$	8	3	5
$k = 2$	13	3	5
$k = 5$	17	1	5

An alarm is given in n when the probability of catastrophe in $n + k$ is equal or greater to 0.2

The choice of the optimal threshold is definitely a very important step in the definition of a good urn-based alarm system. Using the information in the training set, one could for example choose the value of γ that maximizes the number of correctly predicted alarms. Nevertheless we believe that the choice of the alarm threshold is strictly connected to the phenomenon under study, and should be based on expert evaluations.

Another solution, to ameliorate the predictive power of the urn-based alarm system is to act on the reinforcement $s(v)$, i.e. on the strength of the updates according to the Bayesian paradigm. As anticipated, a simple possibility is to choose the values of $s(v)$ that minimize the number of false alarms. There are a lot of alternative paths that can be pursued. A very basic one is the following heuristic procedure, nothing more than a grid search:

- To speed up calculations we have simply assumed that $s(v) = s$, for all $v \in V$.
- Letting s vary in the interval $[0, 10]$ with steps equal to 0.1, we have generated and trained a total of 101 alarm systems over the period 1900–1975. We have chosen this interval, since values greater than 10 would represent an unrealistically strong reinforcement, if we consider that all the urns are initialized with 100 balls.
- We have then looked for the value of s that minimizes the number of false alarms for $k = 5$, i.e. the value of k for which we get the poorer results as per Tables 1 and 2.

The minimum value we have found is $s = 2.2$ and the corresponding performances (definitely improved), when $\gamma = 0.05$, are given in Table 3. Figure 7 shows the number of false alarms for $k = 5$ as a function of s : interestingly 2.2 seems to be

Fig. 6 Number of false alarms as a function of γ for $k = 1$

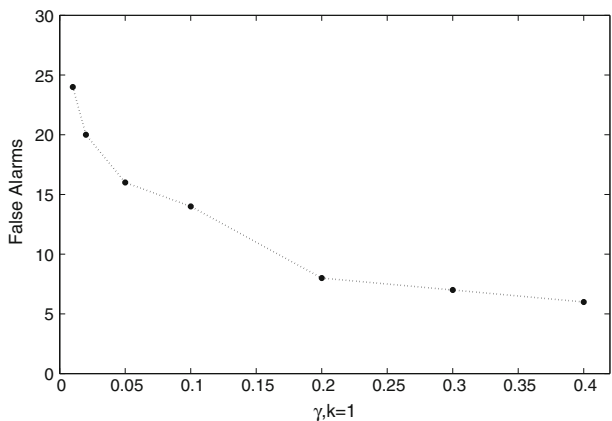


Table 3 Results of the alarm system for the period 1975–1990 using the values of $s(v)$ (i.e. 2.2), which minimize the number of false alarms

1975–90	False alarms	Correct alarms	Actual catastrophes
$k = 1$	7	4	5
$k = 2$	10	4	5
$k = 5$	14	3	5

An alarm is given in n when the probability of catastrophe in $n + k$ is equal or greater to 0.05

the only minimum in the interval $[0, 10]$. For readability, only the values for $s \in [0, 5]$ are given.

Naturally the calibration of $s(v)$ can also be performed with more sophisticated techniques. A possibility could be the use of the generalized expectation maximization algorithm (see McLachlan and Krisnan 1997) for Polya urns, as outlined in Cirillo (2008).

Finally, in order to fully exploit the Bayesian mechanism behind the urn construction, one should be able to express some prior belief about the behavior of the underlying process, embedding this information into the composition of the urns. Think for example of a Gaussian process $\{Y_n\}$, that we discretise as RUP $\{X_n\}$. If we know the behavior of the first derivative of $\{Y_n\}$, we have information about its slope, i.e. we may have an idea of how likely the process can experience large oscillations and jumps. This knowledge can be used to decide an optimal number of risk levels, but also to initialize an appropriate number of say c_4 (black) balls, those reproducing the event of a sudden unexpected catastrophe.

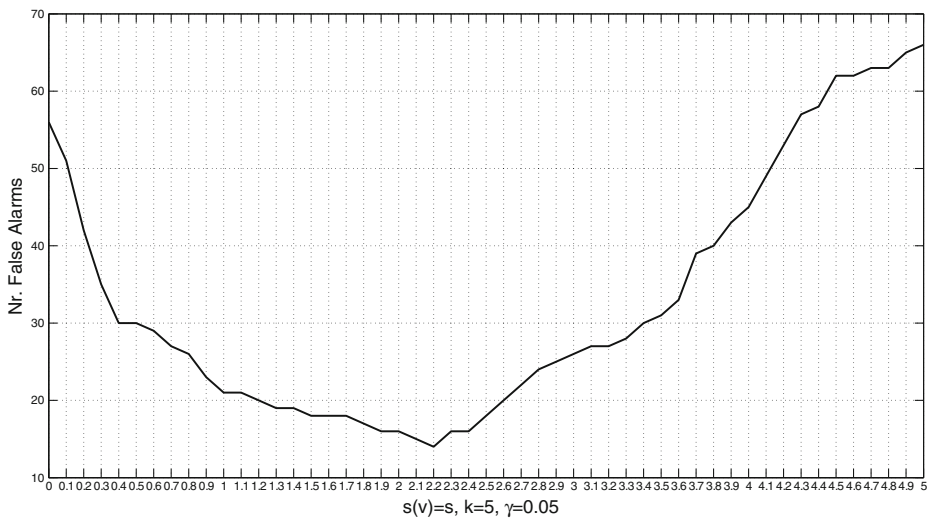


Fig. 7 Number of false alarms as a function of s for $k = 5$ and $\gamma = 0.05$. Only the values up to $s = 5$ are given to increase readability. For $s > 5$ the number of false alarms slightly increases up to a maximum of 86 for $s = 10$

4 Conclusions and Possible Extensions

Using a special reinforced urn process, we have built a Bayesian nonparametric alarm system to predict catastrophic events, such as epidemics, black outs, etc. Differently from other alarm systems in the literature, our model is constantly updated on the basis of the available information, according to the Bayesian paradigm.

The process we have presented is rather flexible: acting on its parameters it is possible to obtaining a large spectrum of results. For example, working with the initial compositions of the urns, it is possible to embed our prior knowledge of the phenomenon in the analysis. In the same way, increasing or decreasing the reinforcement $s(v)$ we can give a larger or a smaller weight to actual data in computing the posterior distribution.

Our alarm system could be generalized by substituting the simple Polya urns with more general ones. An idea would be to use the randomly reinforced urns of Muliere et al. (2006) in order to add extra randomness. Otherwise one could use one of the many time-varying Polya like urns described for example in Pemantle (2007). Time varying urns could be a way of integrating trends and seasonality in the model. Anyway, as we have shown in the paper, it is possible to incorporate simple trends in the model: 1) essentially acting on the compositions of its Polya urns, without entering more complicated elements; 2) using Bayesian learning. In fact, it is pretty obvious that in case of randomly reinforced or time varying urns, most of the results we have given (e.g. Product Dirichlet processes) would not hold.

A possible problem connected with our alarm system, at least in its present version, is the treatment of clusters of catastrophic events. In our construction, every time a catastrophe happens the process is reset to the initial state $(0, 0)$. In this way, if the actual phenomenon shows a cluster of catastrophes, we reset the process several subsequent times and this may be inappropriate. A possible solution can be the following: whether the process reaches risk level L in (n, L) , it is reset to $(0, 0)$ only when a non-black ball is sampled in the successive sampling. For example the process could be reset in $(n + 3, L)$, if 3 black balls are consecutively sampled. This apparently simple modification has however important implications for the model itself, since it modifies the basic recurrence conditions. As a consequence, we would probably move from a mixture of Markov chains, as in the present framework, to a mixture of semi-Markov processes.

When a catastrophe takes place, it may also be interesting to consider its magnitude. Such an eventuality can be modeled by splitting c_4 (black) balls into balls of more colors, say c'_j , $j = 1, \dots, J$ (a sort of grey scale, for example), each associated with a different magnitude. We can then choose not to change the stopping rule: every time a c'_j ball is selected the process is reset to the initial state. However, it is also possible to combine this modeling with the one of clusters.

As suggested by one of the referees, it could also be worth to import some ideas from discrete time Markov decision processes, trying to embed them in the urn construction. In particular, one could assume that each decision taken at time t has a certain cost at time $t + k$. Hence one would be interested in minimizing the expected long term costs, introducing some optimality criteria. These extensions will be subject of future research.

In the last part of the paper we have proposed a first application to certain sunspot data that are freely available on the R platform. The results are promising and foster

the idea of better developing the urn-based alarm system, also testing it on other data sets.

Acknowledgements The authors are grateful to the Editor and the anonymous referees for their comments and suggestions.

References

- Aldous D (1985) *Lecture notes in mathematics: exchangeability and related topics*. Springer Verlag, New York
- Amaral-Turkman MA, Turkman KF (1990) Optimal alarm systems for autoregressive process; a Bayesian approach. *Comput Stat Data An* 19:307–314
- Amerio E, Muliere P, Secchi P (2004) Reinforced urn processes for modeling credit default distributions. *Int J Theor Appl Financ* 7:407–423
- Antunes M, Amaral-Turkman MA, Turkman FK (2003) A Bayesian approach to event prediction. *J Time Ser An* 24:631–646
- Blackwell D, MacQueen JB (1973) Ferguson distributions via Polya-urn schemes. *Ann Stat* 1:353–355
- Brännäs K, Nordström J (2004) An integer-valued time series model for hotels that accounts for constrained capacity. *Stud Nonlinear Dyn E* 8:97–105
- Bulla P (2005) Application of reinforced urn processes to survival analysis. PhD Thesis Bocconi University
- Chattopadhyay R (2000) Covariation of critical frequency of F_2 -layer and relative sunspot number. *Bulletin Astron Soc India*
- Cifarelli DM, Regazzini E (1978) Problemi statistici non parametrici in condizioni di scambiabilità parziale. *Impiego di medie associative*. IMF University of Turin scientific report 3, 12. English translation available online: [http://www.unibocconi.it/wps/allegatiCTP/CR-Scamb-parz\[1\].20080528.135739.pdf](http://www.unibocconi.it/wps/allegatiCTP/CR-Scamb-parz[1].20080528.135739.pdf)
- Cirillo P (2008) New urn approach to shock and default models. PhD Thesis Bocconi University
- Cirillo P, Hüsler J (2009) An urn-based approach to generalized extreme shock models. *Stat Probab Lett* 79:969–976
- Cirillo P, Hüsler J (2011) Extreme shock models: an alternative approach. *Stat Probab Lett* 81:25–30
- Cirillo P, Hüsler J, Muliere P (2010) A nonparametric approach to interacting failing systems with an application to credit risk modeling. *Int J Theor Appl Financ* 13:1–18
- Coppersmith D, Diaconis P (1986) Random walk with reinforcement. Unpublished Manuscript
- Cormen TH, Leiserson CE, Rivest RL (1990) *Introduction to algorithms*. MIT Press and McGraw-Hill
- de Finetti B (1975) *Theory of probability II*. Wiley, New York
- de Maré J (1980) Optimal prediction of catastrophes with application to Gaussian process. *Ann Probab* 8:841–850
- Diaconis P, Freedman D (1980) de Finetti's theorem for Markov chains. *Ann Probab* 8:115–130
- Doksum K (1974) Tailfree and neutral random probabilities and their posterior distributions. *Ann Stat* 2:183–201
- Eggenberger F, Polya G (1923) Über die Statistik verketteter Vorgänge. *Zeitschrift für Angewandte Mathematik and Mechanik* 1:279–289
- Giudici P, Mezzetti M, Muliere P (2003) Mixtures of products of Dirichlet processes for variable selection in survival analysis. *J Stat Plan Infer* 111:101–115
- Grage H, Holst J, Lindgren G, Saklák M (2010) Level crossing prediction with neural networks. *Methodol Comput Appl Probab* 12:623–645
- Grandpierre A (2004) On the origin of solar cycle periodicity. *Astrophys Space Sci* 243:393–400
- Hüsler J (1993) A note on exceedances and rare events of non-stationary sequences. *J Appl Probab* 30:877–888
- Johnson NL, Kotz S (1977). *Urn models and their applications*. Wiley, New York
- Lindgren G (1975a) Prediction for a random time point. *Ann Probab* 3:412–433
- Lindgren G (1975b). Prediction of catastrophes and high level crossings. *Bulletin Int Stat Institut* 46:225–240

- Lindgren G (1980) Model process in non-linear prediction, with application to detection and alarm. *Ann Probab* 8:775–792
- Mahmoud HM (2009) *Polya urn models*. CRC Press, New York
- Marshall AW, Olkin I (1993) Bivariate life distributions from Polya's urn model for contagion. *J Appl Probab* 30:497–508
- Mclachlan GJ, Krishnan T (1997) *The EM algorithm and extensions*. Wiley, New York
- Mezzetti M, Muliere P, Bulla P (2007) An application of reinforced urn processes to determining maximum tolerated dose. *Stat Probab Lett* 77:740–747
- Monteiro M, Pereira I, Scotto MG (2008) Optimal alarm systems for count process. *Commun Stat-Theor M* 37:3054–3076
- Muliere P, Secchi P, Walker SG (2000) Urn schemes and reinforced random walks. *Stoch Proc Appl* 88:59–78
- Muliere P, Secchi P, Walker SG (2003) Reinforced random processes in continuous time. *Stoch Proc Appl* 104:117–130
- Muliere P, Paganoni AM, Secchi P (2006) A randomly reinforced urn. *J Stat Plan Infer* 136:1853–1874
- Pemantle R (2007) A survey of random processes with reinforcement. *Probab Surv* 4:1–79
- Svensson A, Lindquist R, Lindgren G (1996) Optimal prediction of catastrophes in autoregressive moving average processes. *J Time Ser An* 17:511–531
- Zheng HT, Basawa IV, Datta S (2006) Inference for the p -th order random coefficient integer-valued autoregressive processes. *J Time Ser An* 27:411–440