

This article was downloaded by: [University of Georgia]

On: 17 July 2013, At: 21:00

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Sequential Analysis: Design Methods and Applications

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/lsqa20>

Change-Point Detection in Binomial Thinning Processes, with Applications in Epidemiology

Xian Yu ^a, Michael Baron ^b & Pankaj K. Choudhary ^b

^a Department of Mathematics and Statistics, University of Arkansas, Little Rock, Arkansas, USA

^b Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, Texas, USA

To cite this article: Xian Yu, Michael Baron & Pankaj K. Choudhary (2013) Change-Point Detection in Binomial Thinning Processes, with Applications in Epidemiology, Sequential Analysis: Design Methods and Applications, 32:3, 350-367, DOI: [10.1080/07474946.2013.803821](https://doi.org/10.1080/07474946.2013.803821)

To link to this article: <http://dx.doi.org/10.1080/07474946.2013.803821>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Change-Point Detection in Binomial Thinning Processes, with Applications in Epidemiology

Xian Yu¹, Michael Baron², and Pankaj K. Choudhary²

¹Department of Mathematics and Statistics, University of Arkansas,
Little Rock, Arkansas, USA

²Department of Mathematical Sciences, The University of Texas at Dallas,
Richardson, Texas, USA

Abstract: Most of the classical change-point detection schemes are designed for the sequences of independent and identically distributed (i.i.d.) random variables. In this article, motivated by the outbreak of 2009 H1N1 pandemic influenza, we develop change-point detection procedures for the susceptible–infected–recovered (SIR) epidemic model, where a change-point in the infection rate parameter signifies either the beginning or the end of an epidemic trend.

The considered model falls into a general class of *binomial thinning* processes, which is a Markov chain. The cumulative sum (CUSUM) change-point detection procedure is developed for this class, and its performance is evaluated. Apparently, the CUSUM stopping rule is no longer optimal for this non-i.i.d. case. It can be improved by introducing a non constant adaptive threshold. The resulting modified scheme attains a shorter mean delay and at the same time a longer expected time of a false alarm, given that a false alarm eventually occurs.

Proposed detection procedures are applied to the 2001–2012 influenza data published by the Centers for Disease Control and Prevention.

Keywords: Adaptive threshold; Change-point detection; CUSUM; Markov chain; SIR; Thinning process.

Subject Classifications: 62L10; 62L15; 62M05; 92D30; 94C12.

1. INTRODUCTION

In the current practice, epidemics are formally declared when the associated mortality exceeds the epidemic threshold. However, to attain a threshold before

Received September 12, 2012, Revised May 3, 2013, May 5, 2013, Accepted May 6, 2013
Recommended by Ivan Chang

Address correspondence to M. Baron, Department of Mathematical Sciences, University of Texas at Dallas, FO-35, Richardson, TX 75080, USA; Fax: +1 (972) 883-6622; E-mail: mbaron@utdallas.edu

the end of an epidemic season, mortality has to grow at a rate unusual for the nonepidemic distribution creating a *pre-epidemic trend* (Dwyer and Groves, 2001; Weber et al., 2001). Other deviations from the nonepidemic behavior will also occur; for example, unusually high numbers of infections, doctor visits, hospitalizations, sold medications, and searches of related keywords (Fienberg and Shmueli, 2005; Ginsberg et al., 2009; Goldenberg et al., 2002). That is, a change-point occurs when the observed process starts deviating from its “normal,” nonepidemic distribution.

Timely detection of such a change-point will enable to detect an epidemic early during the epidemic season and possibly predict an epidemic before it becomes official (Baron, 2002, 2004; Dwyer and Groves, 2001; Hashimoto et al., 2000; Rath et al., 2003; Tabnak et al., 2000; Weber et al., 2001; also see Brauer et al., 2008; Chowell et al., 2009; Diekmann and Heesterbeek, 2000; and others for mathematical modeling in epidemiology).

In this article, the susceptible–infected–recovered (SIR) *epidemic model* is considered, introduced in Kermack and McKendrick (1927) and widely used ever since. According to this model, the population consists of three groups, (S)usceptible, (I)nfectious, and (R)ecovered, that form a non-stationary Markov chain. Any individual can move through the states in order,

$$S \rightarrow I \rightarrow R.$$

The first transition, from S to I , occurs when an individual gets infected, and the second transition, from I to R , means recovery from the disease (and presumably, immunity from re-infection until the next epidemic season).

Every time unit (say, every day), members of the susceptible group can become infected with probability θ , the infection rate, and members of the infected group can recover with probability η , the recovery rate. If individuals get infected and recover independently of each other, the number of transitions of each type during each day is conditionally binomial. Then the overall number of susceptible people reduces every day according to a general *binomial thinning model* (Weiß, 2008, 2009; Zhu and Joe, 2006) with a *thinning operator* (Steutel and van Harn, 1979) converting the number of susceptible X_t on day t into $X_{t+1} = \theta \circ X_t$ that has a binomial (X_t, θ) distribution, given X_t .

Besides population dynamics where the number of people or species reduces due to epidemics, migration, and other causes, binomial thinning models are applicable to other areas; for example, the distribution of goods such as tickets or apartments, each being offered for sale/rent every day while it is available.

This article focuses on the optimal change-point detection tools for the binomial thinning processes. For epidemics, a change in the infection rate from θ_0 to $\theta_1 > \theta_0$ marks the beginning of an epidemic trend; for the distribution of goods, it means a change in the market demand.

Performance criteria for the detection procedures include the mean delay, probability of a false alarm, expected time until the false alarms, or a risk function that is a combination of these.

We start with the performance evaluation of the cumulative sum (CUSUM) procedure for the change-point detection, which is the optimal stopping rule under the independent and identically distributed (i.i.d.) assumption according to a number of criteria (Basseville and Nikiforov, 1993; Lai, 2001; Lorden, 1971; Moustakides, 1986; Ritov, 1990). Performance characteristics of the CUSUM rule

are derived for large populations in Section 2. In particular, we show that the CUSUM stopping time for thinning processes has an asymptotically Gumbel distribution under the no-change hypothesis. These results enable one to choose a threshold that guarantees the desired performance of the procedure.

A recent discussion in the *Journal of Sequential Analysis* (Han and Tsung, 2009; Tartakovsky and Moustakides, 2010) questioned whether CUSUM remains optimal in the case of Markov chains. The answer appears *negative* for the binomial thinning models. In fact, the standard CUSUM procedure can be improved by introducing a variable *adaptive threshold* that we derive in Section 3. Having a shorter mean delay and at the same time a longer expected time of a false alarm, it yields a lower risk.

Section 4 contains a simulation study that supports the obtained theoretical results. Developed change-point detection procedures are applied in Section 5 to the seasonal influenza and the 2009 influenza A (H1N1) pandemic data published by the Centers for Disease Control and Prevention (CDC).

2. CUSUM RULE FOR BINOMIAL THINNING PROCESSES

In this section, we develop a CUSUM procedure for change-point detection in the binomial thinning process and evaluate its performance characteristics.

2.1. Epidemic Modeling and Binomial Thinning

Assume that the initial population (say, in the beginning of an epidemic season) consists of a large number N of susceptible individuals. Then, each time unit t , every susceptible gets infected with probability θ , and every infected recovers with probability η . Conditioned on the number of susceptible people X_t at time t , their number X_{t+1} at time $(t + 1)$ has the binomial distribution,

$$X_{t+1} | X_t \sim \text{binomial}(X_t, 1 - \theta). \quad (2.1)$$

A random operator performing the transition from X_t to X_{t+1} is called the *thinning operator* (Steutel and van Harn, 1979) that arises in all of the situations where individuals leave the subpopulation with the constant probability and independently of each other. The resulting discrete valued *thinning process* X_t represents a homogeneous irreversible Markov chain.

While there is no reason for the recovery rate η to change, a jump in the infection rate θ is of particular importance in epidemiology. A sudden increase of the infection parameter from the “normal” value θ_0 to an “epidemic” value θ_1 inevitably causes the number of infections to grow at an unusually high rate that may eventually lead to a formal epidemic. Detection of such a change allows to signal a pre-epidemic trend and predict an epidemic.

2.2. CUSUM Stopping Rule

Most of the classical change-point detection schemes are designed for sequences of i.i.d. observations $X_1, X_2, \dots, X_v \sim f_0(\cdot)$ before the change-point v and i.i.d.

$X_{v+1}, X_{v+2}, \dots \sim f_1(\cdot)$ after the change. An optimal, according to several criteria, sequential detection scheme for this case is the CUSUM procedure that is based on the likelihood ratios of the prechange and postchange densities (with respect to some reference measure) f_0 and f_1 . The CUSUM stopping rule is the first time

$$T = \inf \{t : W_t \geq h\}, \quad (2.2)$$

when the CUSUM process

$$W_t = \max_{k=0, \dots, t} \log \frac{f_1(X_{k+1}, \dots, X_t)}{f_0(X_{k+1}, \dots, X_t)} = \max_{k=0, \dots, t} \sum_{j=k+1}^t \log \frac{f_1(X_j)}{f_0(X_j)}$$

attains the prechosen constant threshold $h > 0$.

Assumption of independent observations is often too restrictive although not many papers go beyond the i.i.d. case (Baron, 2004; Basseville and Nikiforov, 1993, chs. 8–9; Beibel, 2000; Davis et al., 1995; Fuh, 2003, 2004; Lai, 1998; Moustakides, 2004; Tartakovsky, 1995; Yakir, 1994). A natural extension of the CUSUM scheme is based on the conditional likelihood ratios,

$$W_t = \max_{k=0, \dots, t} \log \frac{f_1(X_{k+1}, \dots, X_t)}{f_0(X_{k+1}, \dots, X_t)} = \max_{k=0, \dots, t} \sum_{j=k+1}^t \log \frac{f_1(X_j | X_1, \dots, X_{j-1})}{f_0(X_j | X_1, \dots, X_{j-1})}$$

(Baron, 2004; Basseville and Nikiforov, 1993, Section. 8.3), which simplifies as

$$W_t = \max_{k=0, \dots, t} \sum_{j=k+1}^t \log \frac{f_1(X_j | X_{j-1})}{f_0(X_j | X_{j-1})}$$

for the first-order Markov chains, including the thinning process (2.1).

Optimality of the CUSUM stopping rule (2.2) in terms of the mean delay and the rate of false alarms is shown for the i.i.d. case in Lorden (1971), Moustakides (1986), Ritov (1990), and others. Does it remain optimal for the case of non-i.i.d. observations? Han and Tsung (2009) claimed that the CUSUM stopping time is optimal for change-point detection in Markov processes under the condition of invariant distribution. However, Tartakovsky and Moustakides (2010) pointed out flaws in the proof and questioned the conclusion.

In this article, we evaluate performance of the CUSUM stopping rule (2.2) for change-point detection in thinning processes and propose its improvement by an adaptively chosen threshold. In particular, this shows that the standard CUSUM procedure with a constant threshold h can be improved for Markov chains of type (2.1).

2.3. Performance of the CUSUM Procedure

Theoretical evaluation of the CUSUM scheme is based on the following result that extends Wald's identity to the case of independent but not identically distributed observations.

Theorem 2.1 (Extended Wald's lemma). *Let Z_1, Z_2, \dots be independent random variables with $E(Z_j) = \mu_j$, where $|\mu_j| \leq C$ for some $C > 0$. Let T be a Markov stopping time with respect to $\{Z_j\}$. Then, for $S_t = \sum_{j=1}^t Z_j$ and $g(t) = \sum_{j=1}^t \mu_j$,*

$$E(S_T) = Eg(T). \quad (2.3)$$

For the case of $\mu_j \equiv \mu$, this result reduces to the Wald's identity $E(S_T) = \mu E(T)$. The proof of Theorem 2.1 is given in the Appendix.

We apply Theorem 2.1 to conditional log-likelihood ratios of the binomial thinning process (2.1),

$$Z_k = \log \frac{f_1(X_k|X_{k-1})}{f_0(X_k|X_{k-1})} = (X_{k-1} - X_k) \log \frac{\theta_1}{\theta_0} + X_k \log \frac{1 - \theta_1}{1 - \theta_0}, \quad (2.4)$$

where for $j = 0, 1$,

$$f_j(X_k|X_{k-1}) = \binom{X_{k-1}}{X_k} \theta_j^{X_{k-1}-X_k} (1 - \theta_j)^{X_k},$$

with θ_0 and θ_1 being infection rates before and after the change-point, respectively.

2.3.1. Mean Delay

To find the *mean delay* of the CUSUM procedure for the binomial thinning process, we solve

$$E_1(S_T) = E_1\left(\sum_{j=1}^T Z_j\right) = Eg(T) \approx h \quad (2.5)$$

for the stopping time T .

It can be seen that the expectation of the thinning process decreases at an exponential rate,

$$E(X_t|v) = N(1 - \theta_0)^{\min\{t,v\}}(1 - \theta_1)^{(t-v)^+},$$

where $N = X_0$ is the initial state that is reduced every time unit by an expected fraction of θ_0 before the change-point v and by θ_1 after v . Then, applying Theorem 2.1,

$$\begin{aligned} h &\approx E(W_T) \approx E(S_T - S_v)^+ \approx E_{\theta_1} \sum_{j=1}^{(T-v)^+} Z_j \\ &= N(1 - \theta_0)^v E \sum_{k=1}^{(T-v)^+} \left(\{(1 - \theta_1)^{k-1} - (1 - \theta_1)^k\} \log \frac{\theta_1}{\theta_0} + (1 - \theta_1)^k \log \frac{1 - \theta_1}{1 - \theta_0} \right) \\ &= N(1 - \theta_0)^v \left(\theta_1 \log \frac{\theta_1}{\theta_0} + (1 - \theta_1) \log \frac{1 - \theta_1}{1 - \theta_0} \right) \frac{1 - E(1 - \theta_1)^{(T-v)^+}}{\theta_1} \\ &\approx N(1 - \theta_0)^v K_1 \frac{1 - (1 - \theta_1)^{E(T-v)^+}}{\theta_1}, \end{aligned} \quad (2.6)$$

where $K_1 = K(\theta_1, \theta_0)$ is the Kullback-Leibler information number for Bernoulli distributions with parameters θ_1 and θ_0 , and the last approximation is justified for

small θ_1 , when function $(1 - \theta_1)^x$ is approximately linear, up to terms of order θ_1^2 . Several other approximations are used here, all ignoring the terms of order $o(h)$ as $h \rightarrow \infty$. Firstly, the overshoot $(W_T - h)$ is ignored, as in Wald's approximation (e.g., Basseville and Nikiforov, 1993). Secondly, we let $W_t \approx S_t - S_v$ for $t > v$, assuming $\min\{S_k, k = 1, \dots, T\} \approx S_v$, because the random walk S_t has a negative drift for $t \leq v$ and a positive drift for $t > v$. Thirdly, applying Theorem 2.1, we assume that Z_k are nearly independent since for small θ_0, θ_1 (which are much less than 1% in epidemics), the size of susceptible group changes very slowly. Finally, we let $T \approx E(T)$ when solving equation (2.5) for T . The accuracy of these approximations is assessed by simulations in Section 4.

"Solving" equation (2.6) for $E(T - v)^+$, which is the mean delay for the stopping rule T , we obtain the asymptotic expression for the mean delay given a change-point at time v ,

$$MD(v) = E_v(T - v)^+ \approx \frac{\log\left(1 - \frac{h\theta_1}{NK_1(1-\theta_0)^v}\right)}{\log(1 - \theta_1)} \approx \frac{h}{NK_1(1 - \theta_0)^v}, \quad (2.7)$$

where N is the initial population size, θ_0 and θ_1 are (small) binomial parameters before and after the change, and h is the threshold that does not increase faster than N .

It can be seen from (2.7) that in order to satisfy a constraint on the mean delay, the threshold h should be chosen to be of the same order as the population size N . In particular, to attain $MD(v) \leq d$, choose $h \approx NdK_1(1 - \theta_0)^v$.

2.3.2. Probability of a False Alarm (PFA)

Another performance characteristic of the CUSUM procedure, counter balancing the mean delay, is the probability of a false alarm.

We assume that the observation period is finite even though the change-point detection problem is sequential in nature. That is, should a change occur in the distribution of observed data, it has to take place before a certain known time t_{\max} . In epidemiology, t_{\max} is duration of the epidemic season; in tickets distribution, it is the time until the start of the event. Then, the probability of a false alarm is defined as

$$PFA = P(T \leq t_{\max} \mid \text{no change, or } v = \infty).$$

To evaluate PFA for the CUSUM scheme, we first consider the probability $p_{x,h}$, under θ_0 , that the random walk S_n that starts from zero when the population size is x , exceeds the threshold h before it crosses zero again,

$$p_{x,h} = P_{\theta_0}\{S_n \geq h \text{ before } S_n \leq 0 \mid X_0 = x\}.$$

Lemma 2.1. *As the initial population size $N \rightarrow \infty$ proportionally to the current size $x = cN$ ($c \leq 1$) and threshold $h = dK_1N$ satisfying the mean delay condition $MD(0) \leq d$ ($d \geq 1$),*

$$p_{x,h} \approx e^{-N(dK_1 + cK_0 - cV/2)}(1 + o(1)), \quad (2.8)$$

for sufficiently small θ_0 and $\theta_1 = r\theta_0$ with $1 < r < 6$, Kullback information numbers $K_0 = K(\theta_0, \theta_1)$ and $K_1 = K(\theta_1, \theta_0)$, and $V = \theta_0(1 - \theta_0) \log^2 \frac{(1-\theta_1)\theta_0}{(1-\theta_0)\theta_1}$, where the approximation is up to θ_0^2 terms.

In other words,

$$\lim_{N \rightarrow \infty} (-N^{-1} \log p_{x,h}) = dK_1 + cK_0 - V/2 + O(\theta_0^2), \quad \theta_0^2 \downarrow 0.$$

The proof is given in the Appendix. In particular, it shows that $K_0 - V/2 > 0$ under the given assumptions, so our approximation of the probability $p_{x,h}$ is clearly between 0 and 1.

To use Lemma 2.1, let $\tau = \sum_{i=1}^{t_{\max}} \mathbf{1}_{S_i=0}$ be the number of times random walk visits 0, and let X_1, \dots, X_τ be the population size at these moments. Also, let A_j be the event that after the j th visit to 0, S_t returns to 0 before crossing threshold h .

According to Lemma 2.1, the probability of a false alarm before time t_{\max} can be bounded as

$$\begin{aligned} PFA &= p_{N,h} + (1 - p_{N,h})E(p_{X_1,h} \mid A_1) \\ &\quad + (1 - p_{N,h})(1 - E(p_{X_1,h} \mid A_1))E(p_{X_2,h} \mid A_1 \cap A_2) + \dots \\ &\sim E(p_{X_\tau,h} \mid A_1 \cap \dots \cap A_{\tau-1}) \\ &\leq \exp \{-N[dK_1 + (1 - \theta_0)^{t_{\max}}(K_0 - V/2)]\} (1 + o(1)). \end{aligned} \quad (2.9)$$

Ignoring for small θ_0 the change in the population size between the last visit to 0 and t_{\max} and recalling that $h = dK_1 N$, we obtain the approximation

$$PFA \approx \exp \{-h - CN(1 - \theta_0)^{t_{\max}}\}, \quad (2.10)$$

where $C = K_0 - V/2 > 0$.

2.3.3. Distribution of the Stopping Time

In fact, formula (2.10) gives the probability of crossing threshold h by an arbitrary time t . Therefore, we have approximated the whole cumulative distribution function of the stopping time T ,

$$P(T \leq t) \approx \exp \{-h - CN(1 - \theta_0)^t\} = \exp \left\{ -h - \exp \left(-\frac{t - \mu}{\beta} \right) \right\}, \quad t > 0, \quad (2.11)$$

where $\mu = \log(CN)/|\log(1 - \theta_0)|$ and $\beta = |\log^{-1}(1 - \theta_0)|$. Noticeably, $\lim_{t \rightarrow \infty} F(t) = e^{-h} < 1$. Hence, if hypothetically, the observation period can be extended unboundedly, the stopping time T becomes *improper*. With probability $1 - e^{-h} = P(T = \infty)$, the entire “susceptible” population becomes “infected” before raising a false alarm.

Equation (2.11) also shows that conditionally on $\{T < \infty\}$, the stopping time T has the Gumbel distribution with location parameter (mode) μ and scale parameter β ,

$$P(T \leq t \mid T < \infty) = \exp \left\{ -\exp \left(-\frac{t - \mu}{\beta} \right) \right\}, \quad t > 0.$$

Thus, if there is no change in distribution, the most likely time period for a false alarm is around the mode of the Gumbel distribution, which is μ .

2.3.4. Expected Time until a False Alarm (ETFA)

From the Gumbel distribution of T , should a false alarm occur, it has the (conditional) expected time

$$ETFA = E_{\theta_0}(T \mid T < \infty) \approx \mu + \beta\gamma = \frac{\log(CN) + \gamma}{|\log(1 - \theta_0)|},$$

where $\gamma \approx 0.5772$ is the Euler constant. It is a (slowly) increasing function of the initial population size N because the threshold h increases proportionally to it.

The following result summarizes the performance characteristics of the CUSUM stopping rule under the binomial thinning model.

Theorem 2.2. *As the initial population size $N \rightarrow \infty$, the CUSUM stopping rule T has the following performance characteristics for sufficiently small binomial parameters θ_0 and $\theta_1 = r\theta_0$,*

1. *The mean delay given a change point at v is*

$$MD(v) = E_v(T - v)^+ \approx \frac{h}{NK_1(1 - \theta_0)^v} (1 + o(1)), \quad (2.12)$$

if $h = O(N)$.

Further results are for the threshold proportional to the population size $h = dK_1N(1 + o(1))$, $d \geq 1$, that satisfies the constraint $MD \leq d$ when $v = 0$, and for $r = \theta_1/\theta_0 \in (1, 6]$.

2. *The expected time until a false alarm, conditional on the occurrence of a false alarm, is*

$$ETFA = E_{\theta_0}(T \mid T < \infty) \approx \frac{\log(CN) + \gamma}{|\log(1 - \theta_0)|} (1 + o(1)) \quad (2.13)$$

3. *Probability of a false alarm by the time t is*

$$PFA(t) = P_{\theta_0}(T \leq t) \approx \exp\{-h - CN(1 - \theta_0)^t\} \quad (2.14)$$

4. *Conditioned on a false alarm, the stopping time T has approximately Gumbel distribution under θ_0 with parameters $\mu = \log(CN)/|\log(1 - \theta_0)|$ and $\beta = |\log^{-1}(1 - \theta_0)|$*

where $C = K_0 - V/2 > 0$; $V = \theta_0(1 - \theta_0) \log^2[(1 - \theta_1)\theta_0]/[(1 - \theta_0)\theta_1]$; $K_0 = K(\theta_0, \theta_1) = \theta_0 \log[\theta_0/\theta_1] + (1 - \theta_0) \log[(1 - \theta_0)/(1 - \theta_1)]$ and $K_1 = K(\theta_1, \theta_0) = \theta_1 \log[\theta_1/\theta_0] + (1 - \theta_1) \log[(1 - \theta_1)/(1 - \theta_0)]$ are Kullback-Leibler information numbers for the corresponding Bernoulli distributions; $\gamma \approx 0.5772$ is the Euler constant; and the approximations are up to the terms of order θ_0^2 .

Theorem 2.2 can be used to design change-point detection procedures satisfying certain desired performance characteristics by choosing a suitable threshold h as we do in Section 5. Alternatively, a threshold can be chosen to minimize the risk function that combines several performance characteristics, such as $R(v, h) = PFA + \lambda MD(v)$ (e.g., Ritov, 1990; Shiryaev, 1978).

3. CUSUM WITH ADAPTIVE THRESHOLD

According to the binomial thinning model, the population size X_t steadily reduces over time. This questions the use of CUSUM procedures with constant thresholds. Indeed, Theorem 2.2 shows that with constant h , it takes longer to reach the threshold under either parameter θ_0 or θ_1 , if the change occurs late during the observation period. As a result, the mean detection delay increases exponentially, $MD(v) \sim (1 - \theta_0)^{-v}$. Although the prechange rate θ_0 is assumed small, this mean delay may be considerable if a change-point occurs late during the observed period. On the other hand, if no detection occurred by the time t , the probability of a false alarm decreases with t , and it is even possible that the population reduces to zero before having a false alarm. This may be an overwhelming protection against false alarms at the expense of progressively longer mean delays.

In non-i.i.d. situations, when future observations can be predicted from the past—for example, in Markov chains—it seems natural to introduce *adaptive thresholds* that are constructed based on the current state of the observed process. Indeed, in sequential problems, it is not necessary to calculate the threshold for the entire procedure in advance. At any time, only the current threshold is needed in order to decide on the stopping time. Thus, it is perfectly fine to make the threshold dependent on the previous observations.

An adaptive threshold can be used to stabilize the mean delay. From (2.7), in order to have the mean delay $MD(v) \leq d$, threshold h should be proportional to the population size at v . This suggests to choose a threshold at any time t proportional to the current value of X_t ,

$$\tilde{h}_t = dK_1 X_t.$$

To simplify the supporting theory, we modify this threshold making it a function of the population size at the last visit to zero, a renewal point of the CUSUM process. That is, we introduce an adaptive threshold

$$h_t = dK_1 X_{j(t)}, \quad (3.1)$$

where $j(t) = \max\{k \leq t : W_k = 0\}$ is the time of the last visit to zero before time t (Figure 1).

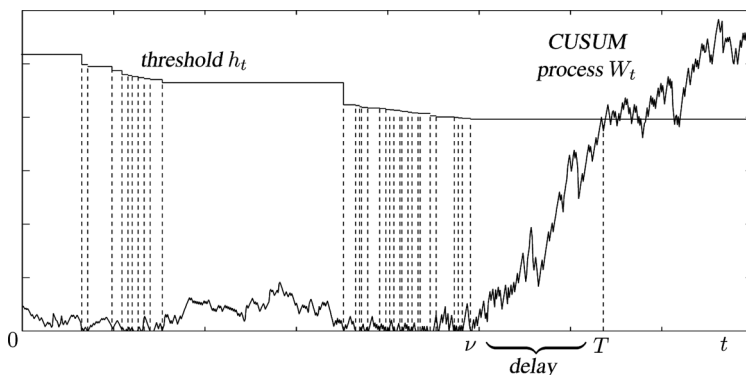


Figure 1. CUSUM process with an adaptive threshold. Threshold changes at renewal points.

3.1. Performance Characteristics

How does the adaptive threshold (3.1) improve performance of the CUSUM procedure?

First of all, the *mean delay* is shortened because the threshold is reduced,

$$h_t = dK_1 X_{j(t)} \leq dK_1 N,$$

and a lower threshold is attained faster. Also, the mean delay stabilizes and becomes practically independent of the change-point location,

$$MD(v) \approx \frac{h_{j(v)}}{NK_1(1 - \theta_0)^v} (1 + o(1)) \approx \frac{dNK_1(1 - \theta_0)^v}{NK_1(1 - \theta_0)^v} (1 + o(1)) \rightarrow d,$$

as $N \rightarrow \infty$. The derivation of this expression follows along the lines of Section 2.3.1. The mean delay is reduced approximately by a factor of $(1 - \theta_0)^v$.

Second, since the threshold is reset every time the CUSUM process W_t returns to zero, the probability of crossing it before returning to zero again follows from Lemma 2.1. For example, with $d = 1$,

$$p_{x,h} = p_{x,dK_1x} \approx e^{-x(dK_1+cK_0-cV/2)} (1 + o(1)) = e^{-Kx} (1 + o(1)),$$

where $K = dK_1 + cK_0 - cV/2$. Then, the probability of a false alarm by the time t , similar to (2.10), is

$$PFA \sim E(p_{x_{\tau,h}} \mid A_1 \cap \dots \cap A_{\tau-1}) \approx \exp \{-KN(1 - \theta_0)^t\}.$$

According to this formula, the stopping time T has approximately *Gumbel distribution* with parameters $\tilde{\mu} = \log(KN)/|\log(1 - \theta_0)|$ and $\tilde{\beta} = |\log^{-1}(1 - \theta_0)|$. Comparing with the distribution of the constant-threshold stopping time in Section 2.3.3, $\tilde{\mu} = \mu + \log(1 + dK_1/C)/|\log(1 + \theta_0)|$ and $\tilde{\beta} = \beta$. Hence, the mode of the time T until false alarm ($\tilde{\mu}$) increased by a positive amount of order θ_0^{-1} , and so did the mean

$$ETFA \approx \frac{\log(KN) + \gamma}{|\log(1 - \theta_0)|}.$$

This expected time until a false alarm is also understood as conditional expectation given that a false alarm occurs eventually, if hypothetically, the observation period is extended to infinity. There is actually a positive probability that $T = \infty$ and there is no false alarm at all, for example, if the entire population gets infected during the first time unit. The theoretical probability of no false alarm until $t = \infty$ is practically zero, lower than for the constant-threshold procedure.

The following result summarizes the performance characteristics of the CUSUM stopping rule with adaptive threshold.

Theorem 3.1. *The CUSUM stopping rule T with adaptive threshold (3.1) has the following performance characteristics, as $N \rightarrow \infty$, for sufficiently small θ_0 and θ_1 ,*

1. *The mean delay is*

$$MD \approx d + o(1) \tag{3.2}$$

2. The expected time until a false alarm, conditional on the occurrence of a false alarm, is

$$ETFA \approx \frac{\log(KN) + \gamma}{|\log(1 - \theta_0)|} (1 + o(1)) \quad (3.3)$$

3. Probability of a false alarm by the time t is

$$PFA(t) \approx \exp \{-KN(1 - \theta_0)^t\} \quad (3.4)$$

4. Conditioned on a false alarm, the stopping time T has approximately Gumbel distribution under θ_0 with parameters $\tilde{\mu} = \log(KN)/|\log(1 - \theta_0)|$ and $\beta = |\log^{-1}(1 - \theta_0)|$,

where $K = dK_1 + K_0 - V/2 > 0$, and the other constants are defined in Theorem 2.2.

Comparing with the classical CUSUM, our adaptive threshold yields a *shorter mean delay* and a *longer expected time until a false alarm*, given that a false alarm eventually occurs. However, the probability of a false alarm during an unbounded time period is higher.

4. MONTE CARLO STUDY

To evaluate performance of the proposed detection procedures and compare CUSUM schemes with the constant and adaptive thresholds, random samples are generated from the model (2.1), with the initial population size $N = 3 \cdot 10^6$, change-point $v = 0$, an epidemic season consisting of $t_{\max} = 30$ weeks, and different values of the prechange parameter θ_0 and the ratio $r = \theta_1/\theta_0$. Results are based on 100,000 Monte Carlo runs.

Table 1 shows the estimated mean delay of the detection procedure with the desired mean delay set at $d = 3$. The actual mean delay is always slightly below the desired value, which agrees with (2.9).

False alarm rates and the expected times until false alarms are shown in Table 2. Thresholds are also selected to attain the mean delay of $d = 3$. It can be seen that although the adaptive correction of threshold h is relatively small for low values of θ_0 , it always results in a significant increase of the expected time until a false alarm, essentially without sacrificing the probability of a false alarm or a mean delay. This

Table 1. Estimated mean delay under constant and adaptive thresholds

Ratio $r = \theta_1/\theta_0$	MD (Constant h)	MD (Adaptive h)
1.03	2.616	2.572
1.05	2.531	2.484
1.1	2.499	2.418
1.3	2.503	2.243
1.5	2.497	2.107
1.7	2.500	2.032
2.0	2.501	2.003

Table 2. Comparison of the false alarm rates

Prechange parameter θ_0	Ratio $r = \theta_1/\theta_0$	Constant threshold				Adaptive threshold			
		Threshold h	PFA bounded by formula	PFA estimated by simulation	ETFA	Final threshold $h(t_{\max})$	PFA bounded by formula	PFA estimated by simulation	ETFA
0.001	1.03	1.74	0.1273	0.0135	214.1	1.69	0.134	0.0220	2,056
0.002	1.03	3.49	0.0164	0.0003	453.7	3.29	0.020	0.0006	1,374
0.01	1.03	17.6	$2 \cdot 10^{-9}$	$<10^{-5}$	251.3	13.0	$2 \cdot 10^{-7}$	0.019	434.7
0.02	1.03	35.5	$9 \cdot 10^{-18}$	$<10^{-5}$	159.9	19.4	10^{-10}	0.0003	251.1
0.001	1.05	4.81	0.0034	$<10^{-5}$	1,227	4.67	0.0039	$<10^{-5}$	3,070
0.002	1.05	9.63	10^{-5}	$<10^{-5}$	960.0	9.07	$2 \cdot 10^{-5}$	$<10^{-5}$	1,881
0.01	1.05	48.6	$9 \cdot 10^{-25}$	$<10^{-5}$	352.2	35.9	$2 \cdot 10^{-19}$	$<10^{-5}$	535.7
0.02	1.05	98.1	10^{-47}	$<10^{-5}$	210.0	53.5	$2 \cdot 10^{-28}$	$<10^{-5}$	301.3
0.001	1.1	18.9	$2 \cdot 10^{-10}$	$<10^{-5}$	2,593	18.4	$3 \cdot 10^{-10}$	$<10^{-5}$	4,440
0.002	1.1	37.9	$4 \cdot 10^{-20}$	$<10^{-5}$	1,642	35.7	$4 \cdot 10^{-19}$	$<10^{-5}$	2,565
0.01	1.1	191	$3 \cdot 10^{-95}$	$<10^{-5}$	488.2	141	10^{-73}	$<10^{-5}$	672.0
0.02	1.1	386	10^{-185}	$<10^{-5}$	277.7	211	$2 \cdot 10^{-109}$	$<10^{-5}$	369.1

is where the adaptive threshold offers an improvement of the conventional CUSUM stopping rule.

For some pairs of parameters θ_0 and r , a false alarm has not occurred in the course of 100,000 simulated processes; therefore, its probability is formally estimated to be below 10^{-5} . The exact order of the false alarm probability is still unknown when PFA is extremely low, and it is not practically important. As we know from (2.9), formulas (2.14) and (3.4) give upper bounds for the probabilities of a false alarm, and they are accurate for small prechange parameter θ_0 .

We can also see that since the change with a larger ratio r is easier to detect, the same mean delay $d = 3$ can be satisfied with a higher threshold and, therefore, lower probability of a false alarm and longer time until a false alarm.

5. APPLICATION TO INFLUENZA ACTIVITY

In Sections 2 and 3, two CUSUM-type procedures are developed for the sequential detection of a change-point in a sequence of dependent random variables that forms a binomial thinning process. This section shows the application of the proposed algorithms to the detection of epidemic trends.

According to the SIR epidemic model, the number of susceptible members of the population forms a binomial thinning process. Therefore, we are applying the tools developed in this article. In fact, the possibility of a statistical detection of epidemics was the main motivation of this work.

Table 3 shows the detection results for 2001–2012 influenza data published as weekly reports by the CDC. In most seasons, the CUSUM algorithm with an adaptive threshold detects influenza epidemics before it was officially announced. During severe epidemic seasons of 2001–02, 2003–04, 2004–05, 2007–08, and 2010–11, the extended CUSUM process detects epidemics 7, 3, 6, 1, and 6 weeks earlier than the first upcrossing of the epidemic threshold by the pneumonia and influenza (P&I) mortality rate.

In this study, the threshold is chosen to detect an epidemic trend, on the average, no longer than three time units after the change point; that is, $MD \leq 3$. Under this constraint, the test has a probability of a false alarm no greater than

Table 3. Detection of epidemic trends

Season	Week of detection by CDC threshold	Week of detection by CUSUM
2001–2002	22nd	15th
2002–2003	Not detected	14th
2003–2004	11th	8th
2004–2005	20th	14th
2005–2006	Not detected	12th
2006–2007	Not detected	12th
2007–2008	16th	15th
2008–2009	28th	17th
2009 (H1N1)	22nd–23rd (06/11/09)	23rd (06/13/09)
2010–2011	17th	11th
2011–2012	16th	20th

0.00032 provided that the epidemic never started, and this probability is significantly lower for years with stronger epidemics; that is, higher postchange infection rate θ_1 .

APPENDIX

Proof of Theorem 2.1 (Extended Wald's Lemma)

The proof essentially follows the lines of the proof of Wald's identity in (Siegmund, 1985, ch. 2), translating it to the case of non identical distributions. Suppose $Z_j \geq 0$. The indicator $\{\mathbf{1}_{T \geq j}\} = \{\mathbf{1}_{T > j-1}\}$ is a function of Z_1, \dots, Z_{j-1} ; therefore, it is independent of Z_j, Z_{j+1}, \dots for any $j = 1, 2, \dots$. Then

$$\begin{aligned} E(S_T) &= E\left(\sum_{j=1}^T Z_j\right) = E\left(\sum_{j=1}^{\infty} Z_j \cdot \mathbf{1}_{T \geq j}\right) = \sum_{j=1}^{\infty} E(Z_j \cdot \mathbf{1}_{T \geq j}) = \sum_{j=1}^{\infty} E(Z_j \cdot \mathbf{1}_{T > j-1}) \\ &= \sum_{j=1}^{\infty} E(Z_j) \cdot P(T \geq j) = \sum_{j=1}^{\infty} \mu_j \sum_{k=j}^{\infty} P(T = k) \\ &= \sum_{k=1}^{\infty} P(T = k) \sum_{j=1}^k \mu_j = \sum_{k=1}^{\infty} g(k) P(T = k) = E g(T). \end{aligned}$$

Proof of Lemma 2.1

From (2.4), given $X_0 = x$, the first increment of random walk S_t ,

$$S_1 = Z_1 = x \log \frac{\theta_1}{\theta_0} + X_1 \log \frac{(1 - \theta_1)\theta_0}{(1 - \theta_0)\theta_1}$$

is a linear function of a binomial($x, 1 - \theta_0$) variable X_1 . Conditioning on X_1 , obtain

$$\begin{aligned} p_{x,h} &= P(S_t \geq h \text{ before } S_t \leq 0) \\ &= P(S_1 \geq h) + E P(S_t \geq h \text{ before } S_t \leq 0 \mid S_1) \mathbf{1}_{0 < S_1 < h} \end{aligned} \quad (\text{A.1})$$

$$\approx P(S_1 \geq h) + E \frac{e^{S_1} - 1}{e^h - 1} \mathbf{1}_{0 < S_1 < h} \quad (\text{A.2})$$

$$= \frac{e^h}{e^h - 1} P(S_1 \geq h) - \frac{1}{e^h - 1} P(S_1 > 0) + \frac{1}{e^h - 1} E e^{S_1} \mathbf{1}_{0 < S_1 < h}. \quad (\text{A.3})$$

Wald's formula for the probability of Type I error of SPRT—that is, crossing the upper boundary h before the lower boundary 0 starting from S_1 (e.g., Basseville and Nikiforov, 1993, p. 171, or Govindarajulu, 2004, Corollary 2.2.1)—was used to obtain (A.2) from (A.1). It is approximate because of Wald's approximation; that is, ignoring the overshoot.

Next, as $x \rightarrow \infty$ (proportionally to the initial population size N and threshold h), the distribution of S_1 is asymptotically normal, with

$$E(S_1) = x \left(\log \frac{\theta_1}{\theta_0} + (1 - \theta_0) \log \frac{(1 - \theta_1)\theta_0}{(1 - \theta_0)\theta_1} \right) = -K_0 x$$

and

$$\text{Var}(S_1) = x\theta_0(1 - \theta_0) \log^2 \frac{(1 - \theta_1)\theta_0}{(1 - \theta_0)\theta_1} = Vx.$$

The three terms in (A.3) are evaluated using the asymptotic normality of S_k . This results in a linear combination of exponential functions of N and, thus, the final asymptotic expression will be determined by the dominating, highest degree term. Comparing the exponents, we notice that as θ_0 and $\theta_1 = r\theta_0$ tend to zero,

$$K_1 = K(\theta_1, \theta_0) = \theta_0(r \log r - r + 1) + O(\theta_0^2), \quad (\text{A.4})$$

$$K_0 = K(\theta_0, \theta_1) = \theta_0(r - 1 - \log r) + O(\theta_0^2), \quad \text{and} \quad (\text{A.5})$$

$$V = \theta_0 \log^2 r + O(\theta_0^2). \quad (\text{A.6})$$

Notice that up to θ_0^2 terms, $K_0 > V/2$ and for $r \leq 6$, $K_0 < V$. In what follows, ϕ and Φ denote the standard normal probability density function and cumulative distribution function. It is easy to verify by the L'Hôpital Rule that as $z \rightarrow +\infty$,

$$1 - \Phi(z) \sim \frac{\phi(z)}{z} \quad \text{or} \quad \lim_{z \rightarrow +\infty} \frac{1 - \Phi(z)}{\phi(z)/z} = 1. \quad (\text{A.7})$$

As $h = dK_1N \rightarrow \infty$ and $x = cN \rightarrow \infty$, the first term in (A.3) is asymptotically equivalent to

$$\begin{aligned} P(S_1 \geq h) &\sim 1 - \Phi\left(\frac{h + K_0x}{\sqrt{Vx}}\right) \sim \frac{\phi(\sqrt{N} \frac{dK_1 + cK_0}{\sqrt{cV}})}{\sqrt{N}(dK_1 + cK_0)/\sqrt{cV}} \\ &= \frac{\sqrt{cV}}{\sqrt{2\pi N}(dK_1 + cK_0)} e^{-N \frac{(dK_1 + cK_0)^2}{2cV}}. \end{aligned} \quad (\text{A.8})$$

The second term is asymptotically equivalent to

$$e^{-h} P(S_1 > 0) \sim \frac{e^{-dK_1N} \phi(\sqrt{N} K_0 \sqrt{\frac{c}{V}})}{\sqrt{N} K_0 \sqrt{c/V}} = \frac{\sqrt{V}}{\sqrt{2\pi Nc} K_0} e^{-N(dK_1 + \frac{K_0^2 c}{2V})}. \quad (\text{A.9})$$

Denoting the standardized value of S_1 by $U = (S_1 + K_0x)/\sqrt{Vx}$ and using the formula

$$Ee^{\lambda U} \mathbf{1}_{a < U < b} = e^{\lambda^2/2} (\Phi(b - \lambda) - \Phi(a - \lambda))$$

for the standard normal variable U , the third term is asymptotically equivalent to

$$\begin{aligned} &e^{-h} E e^{-K_0x + U\sqrt{Vx}} \mathbf{1}_{K_0\sqrt{x/V} < U < (h + K_0x)/\sqrt{Vx}} \\ &= e^{-N(dK_1 + cK_0 - cV/2)} \left\{ \Phi\left(\sqrt{N} \frac{dK_1 + cK_0 - cV}{\sqrt{cV}}\right) - \Phi\left(\sqrt{Nc} \frac{K_0 - V}{V}\right) \right\} \\ &\sim e^{-N(dK_1 + cK_0 - cV/2)}. \end{aligned} \quad (\text{A.10})$$

This is because from (A.4)–(A.6), we have $V - dK_1 < K_0 < V$ for sufficiently small θ_0 , $r \in (1, 6]$, $c \leq 1$, and, $d \geq 1$, and, therefore,

$$\Phi\left(\sqrt{N} \frac{dK_1 + cK_0 - cV}{\sqrt{cV}}\right) \rightarrow 1 \quad \text{and} \quad \Phi\left(\sqrt{Nc} \frac{K_0 - V}{V}\right) \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

Moreover, comparing exponents in (A.8) and (A.10), we find that

$$\frac{(dK_1 + cK_0)^2}{2cV} - (dK_1 + cK_0 - cV/2) = \frac{(dK_1 + cK_0 - cV)^2}{2cV} > 0,$$

because $dK_1 + cK_0 - cV > 0$, and comparing exponents in (A.9) and (A.10),

$$\left(dK_1 + \frac{K_0^2 c}{2V}\right) - \left(dK_1 + cK_0 - \frac{cV}{2}\right) = \frac{c(K_0 - V)^2}{2V} > 0,$$

because $K_0 < V$. Hence, term (A.10) is dominating in (A.3) as $N \rightarrow \infty$, and (2.8) follows.

ACKNOWLEDGMENTS

The authors are thankful to the Editor and the Associate Editor of *Sequential Analysis* for their helpful comments. Research of M. Baron was funded by the National Science Foundation and the National Security Agency. This support is greatly appreciated.

REFERENCES

- Baron, M. (2002). Bayes and Asymptotically Pointwise Optimal Stopping Rules for the Detection of Influenza Epidemics, in *Case Studies in Bayesian Statistics*, vol. 6, C. Gatsonis, R. E. Kass, A. Carriquiry, A. Gelman, D. Higdon, D. K. Pauler, and I. Verdinelli, eds., pp. 153–163, New York: Springer.
- Baron, M. (2004). Detection of Epidemics as a Sequential Change-Point Problem, in *Longevity, Aging and Degradation Models in Reliability, Public Health, Medicine and Biology*, vol. 2, V. Antonov, C. Huber, M. Nikulin, and V. Polischuk, eds., pp. 31–43, St. Petersburg: St. Petersburg State Polytechnical University.
- Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Application*, Englewood Cliffs: Prentice-Hall.
- Beibel, M. (2000). Sequential Detection of Signals with Known Shape and Unknown Magnitude, *Statistica Sinica* 10: 715–729.
- Brauer, F., Driessche, P., and Wu, J. (2008). *Mathematical Epidemiology*, New York: Springer.
- Chowell, G., Hayman, J. M., Bettencourt, L. M. A., and Castillo-Chavez, C. (2009). *Mathematical and Statistical Estimation Approaches in Epidemiology*, New York: Springer.
- Davis, R. A., Huang, D., and Yao, Y.-C. (1995). Testing for a Change in the Parameter Values and Order of an Autoregressive Model, *Annals of Statistics* 23: 282–304.
- Diekmann, O. and Heesterbeek, J. A. P. (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, New York: Wiley.

- Dwyer, D. M. and Groves, C. (2001). Outbreak Epidemiology, in *Infectious Disease, Epidemiology: Theory and Practice*, K. E. Nelson and C. F. Williams, eds., pp. 119–148, Gaithersburg: Aspen Publishers.
- Fienberg, S. E. and Shmueli, G. (2005). Statistical Issues and Challenges Associated with Rapid Detection of Bio-Terrorist Attacks, *Statistics in Medicine* 24: 513–529.
- Fuh, C. D. (2003). SPRT and CUSUM in Hidden Markov Models, *Annals of Statistics* 31: 942–977.
- Fuh, C. D. (2004). Asymptotic Operating Characteristics of an Optimal Change Point Detection in Hidden Markov Models, *Annals of Statistics* 32: 2305–2339.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. (2009). Detecting Influenza Epidemics Using Search Engine Query Data, *Nature* 457: 1012–1015.
- Goldenberg, A., Shmueli, G., Caruana, R. A., and Fienberg, S. E. (2002). Early Statistical Detection of Anthrax Outbreaks by Tracking Over-the-Counter Medication Sales, *Proceedings of National Academy of Sciences of United States of America* 99: 5237–5240.
- Govindarajulu, Z. (2004). *Sequential Statistics*, Singapore: World Scientific Publishing.
- Han, D. and Tsung, F. (2009). Optimal Stopping Time for Detecting Changes in Discrete Time Markov Processes, *Sequential Analysis* 28: 115–135.
- Hashimoto, S., Murakami, Y., Taniguchi, K., and Nagai, M. (2000). Detection of Epidemics in Their Early Stage through Infectious Disease Surveillance, *International Journal of Epidemiology* 29: 905–910.
- Kermack, W. O. and McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics, *Proceedings of Royal Society of London, Series A* 115: 700–721.
- Lai, T. L. (1998). Information Bounds and Quick Detection of Parameter Changes in Stochastic Systems, *IEEE Transactions on Information Theory* 44: 2917–2929.
- Lai, T. L. (2001). Sequential Analysis: Some Classical Problems and New Challenges, *Statistica Sinica* 11: 303–408.
- Lorden, G. (1971). Procedures for Reacting to a Change in Distribution, *Annals of Mathematical Statistics* 42: 1897–1908.
- Moustakides, G. V. (1986). Optimal Stopping Times for Detecting a Change in Distribution, *Annals of Statistics* 14: 1379–1388.
- Moustakides, G. V. (2004). Optimality of the CUSUM Procedure in Continuous Time, *Annals of Statistics* 32: 302–315.
- Rath, T. M., Carreras, M., and Sebastiani, P. (2003). Automated Detection of Influenza Epidemics Using Hidden Markov Models, in *Advances in Intelligent Data Analysis*, M. R. Berthold, H.-J. Lenz, E. Bradley, R. Kruse, and C. Borgelt, eds., pp. 521–532, Berlin: Springer.
- Ritov, Y. (1990). Decision Theoretic Optimality of CUSUM Procedure, *Annals of Statistics* 18: 1464–1469.
- Shiryaev, A. N. (1978). *Optimal Stopping Rules*, New York: Springer.
- Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*, New York: Springer.
- Steutel, F. and van Harn, K. (1979). Discrete Analogues of Self-Decomposability and Stability, *Annals of Probability* 7: 893–899.
- Tabnak, F., Muller, H.-G., Wang, J.-L., Chiou, J.-M., and Sun, R. K. P. (2000). A Change-Point Model for Reporting Delays under Change of AIDS Case Definition, *European Journal of Epidemiology* 16: 1135–1141.
- Tartakovsky, A. G. (1995). Asymptotic Properties of CUSUM and Shiryaev's Procedures for Detecting a Change in a Nonhomogeneous Gaussian Process, *Mathematical Methods of Statistics* 4: 389–404.
- Tartakovsky, A. G. and Moustakides, G. V. (2010). A Note on the “Optimal Stopping Time for Detecting Changes in Discrete Time Markov Processes” by Han and Tsung, *Sequential Analysis* 29: 483–486.

- Weber, D. J., Menajovsky, L. B., and Wenzel, R. (2001). Investigation of Outbreaks, in *Epidemiologic Methods for the Study of Infectious Diseases*, D. Weber and J. Thomas, eds., pp. 291–310, New York: Oxford University.
- Weiß, C. H. (2008). Thinning Operations for Modeling Time Series of Counts—A Survey, *Advances in Statistical Analysis* 92: 319–341.
- Weiß, C. H. (2009). A New Class of Autoregressive Models for Time Series of Binomial Counts, *Communications in Statistics – Theory and Methods* 38: 447–460.
- Yakir, B. (1994). Optimal Detection of a Change in Distribution When the Observations Form a Markov Chain with a Finite State Space, in *Change-Point Problems*, E. Carlstein, H. Muller, and D. Siegmund, eds., pp. 346–358, Hayward: Institute of Mathematical Statistics.
- Zhu, R. and Joe, H. (2006). Modelling Count Data Time Series with Markov Processes Based on Binomial Thinning, *Journal of Time Series Analysis* 27: 725–738.