In "Longevity, Aging and Degradation Models in Reliability, Public Health, Medicine and Biology", edited by V. Antonov, C. Huber, M. Nikulin, and V. Polischook; volume 2, St. Petersburg, 2004; pp. 31–43

Early Detection of Epidemics as a Sequential Change-Point Problem

Michael Baron
Department of Mathematical Sciences, University of Texas at Dallas,
Richardson, Texas, 75083-0688, USA
mbaron@utdallas.edu

An outbreak of an infectious disease becomes *an epidemic* when the associated mortality exceeds an epidemic threshold. It is often possible though to detect an epidemic trend earlier and to predict an epidemic. Such a trend is marked by a rapidly increasing number of deaths, diagnosed patients, or sold medicines. Substantial prior information for the beginning of such an epidemic trend is usually available, and thus, its detection can be stated as a *Bayes sequential change-point detection problem*.

The exact Bayes solution involves rather complicated computation of the payoff function that is feasible only for relatively simple prior distributions or by means of an extensive Monte Carlo study. Instead, we propose *asymptotically pointwise optimal (APO)* stopping rules for change-point detection whose computation is straightforward even under complicated prior distributions that arise in epidemiology.

Recent results on APO rules for change-point detection are extended to discrete-time stochastic processes with nuisance pre-change and post-change parameters. Direct expressions for the APO rules under suitable risk functions are derived. Results are applied to the 1996–2004 influenza mortality data published by Centers for Disease Control and Prevention.

Key words and phrases: asymptotically pointwise optimal, Bayes sequential, change point, epidemic threshold, influenza mortality, optimal stopping, payoff function.

1 Introduction

In the modern world, the timely detection of epidemics has been recognized as an extremely important problem of disease control and biosurveillance.

According to the World Health Organization, influenza epidemics result in three to five million severe cases and between 250,000 and 500,000 deaths each year. Between 5% and 15% of the population develop respiratory tract infections. It is estimated that in the United States, influenza epidemics account for

a 71-167 million dollars loss in the form of health care costs and lost productivity. A 274 million dollar budget has been requested for the fiscal year 2005 for "an integrated biosurveillance initiative" that will include "development of the national data collection and analysis system for identifying possible bioterrorist incidents and other disease outbreaks."

This article concentrates on *statistical methods* for the detection of epidemics. In the current practice, epidemics are detected and reported when mortality exceeds the *epidemic threshold*. Recent results (e.g.,[3], [12], [17]) show the possibility of epidemics detection at early stages, even before the epidemic threshold is exceeded. The main tool in this methodology is detection of "unusual trends" such as unusually high rate of new diagnoses, fatalities, sold medicines, or doctors' visits. In each case, beginning of an unusual trend marks a switch from a "control state" to an "epidemic state". Therefore, a beginning of an epidemic trend is a *change point* whose timely detection will predict occurrence of a new epidemic.

An important feature of this analysis is availability of rich *prior information*. Besides the data obtained by disease surveillance, there are various factors, easily observed, that influence occurrence of epidemics. For instance, one can relate the beginning of pre-epidemic trends to weather conditions, pollution, pollen, ozone, and other factors ([14], [22]). In particular, rapidly changing weather contributes to the possibility of a pre-epidemic trend.

Thus, a realistic prior distribution exists for the change-point parameter, and the statistical problem of early detection of epidemics can now be stated as a *Bayes sequential change-point detection problem*.

Exact Bayes solution to this problem (Bayes stopping time) has been found for the cases of geometric prior ([21]), the least favorable prior ([18]), and a prior driven by a stationary Markov chain ([2]), assuming independent observations. Each algorithm and especially its practical implementation put heavy constraints on the structure of the prior distribution of a change point. As a result, this does not allow to utilize all the available prior information for accurate detection of epidemics.

Alternatively, *asymptotically pointwise optimal (APO)* rules can be used for the detection of epidemics, as proposed in [3]. This extends the concept of an APO stopping rule in classical sequential analysis ([6], [7]) to the problems of change-point detection.

In this paper, the observed data (number of cases, deaths, sold medicines, etc.) are assumed to come from a stationary or nonstationary *time series*, which is standard in epidemic modelling. *Nuisance parameters* before and after the change point are introduced allowing the model to recalibrate itself sequentially as the new data becomes available. Under these general conditions and rather mild assumptions about the prior distribution, we derive closed-form expressions for the APO stopping rule. Thus, implementation of the obtained rule for epidemic detection is straightforward. They combine observed data with the

prior information and and achieve a controlled balance between the mean delay and the rate of false alarms.

2 Exact Bayes and Bayes-like stopping rules for change-point detection

Assume a rather general model with a change point. Let $X_1, X_2, \ldots$ be an observed sequence of random variables or random vectors, *possibly dependent*, following the distribution law F for $j \le \nu$ and distribution $G(x)$ for $j > \nu$ for some $\nu \ge 0$, with ν being the unknown change point, the parameter of interest. Denote $\boldsymbol{X}^{(k:n)} = (X_k, \ldots, X_n)$, $\boldsymbol{X}^{(n)} = \boldsymbol{X}^{(1:n)}$, assume that for each n, measures $F(\boldsymbol{X}^{(n)})$ and $G(\boldsymbol{X}^{(n)})$ are absolutely continuous with respect to a reference measure μ_n, and consider joint densities $f_{(n)}(\boldsymbol{X}^{(n)})$ and $g_{(n)}(\boldsymbol{X}^{(n)})$, marginal densities $f_k(X_k)$ and $g_k(X_k)$, conditional densities $f_{n|k}(\boldsymbol{X}^{(n)}|\boldsymbol{X}^{(k)})$ and $g_{n|k}(\boldsymbol{X}^{(n)}|\boldsymbol{X}^{(k)})$, etc. In unambiguous cases, subscripts will typically be omitted. Also, let

$$\mathcal{L}_k(\boldsymbol{X}^{(n)} = f(\boldsymbol{X}^{(k)})g(\boldsymbol{X}^{(k+1:n)}|\boldsymbol{X}^{(k)})$$

be the joint density of $\boldsymbol{X}^{(n)}$ under the condition $\nu = k$.

Further, suppose that the distributions F come from a family $\mathcal{F}$, indexed by an unknown (*nuisance*) parameter θ, and distributions G come from a family $\mathcal{G}$ with a nuisance parameter η. This will ensure a wide range of applications where it is impractical to assume completely known distributions. Often a pre-change (say, "in-control") distribution is known, but an after-change ("out-of-control") distribution is not. Families $\mathcal{F}$ and $\mathcal{G}$ may actually coincide, in which case only its parameter changes from θ to η at the time ν.

One can think of two stochastic sequences

$$\{U_n\} \sim F_\theta \in \mathcal{F} \text{ and } \{V_n\} \sim G_\eta \in \mathcal{G},$$

and let

$$X_n = \begin{cases} U_n & \text{for} \quad n \le \nu \\ V_n & \text{for} \quad n > \nu. \end{cases}$$

For problems of epidemic detection, the process $\{X_n\}$ may represent the number of diagnoses, the number or percentage of deaths, the number of sold medicines, or the number of doctor's visits.

The objective is to detect and report a change in distribution "as soon as possible" after it occurs, at the same time not allowing a high frequency of false alarms. A decision-theoretic approach to this sequential detection problem will therefore involve a risk function that balances the mean delay of a stopping time and the probability of a false alarm. Generalization to sequences with more than two change points, with the objective to detect all of them is natural and straightforward ([9],[13]). One possible solution is to sample the data

sequentially and to detect one change point at a time by means of a sequential detection mechanism ([4]).

Next, the random nature of change points and available rich prior information about their possible occurrence justify a Bayesian model where the unknown parameter of interest ν has a rather objective prior distribution

$$\pi_t = \boldsymbol{P}\{\nu = t\}.$$

In epidemiology, a rather realistic prior distribution for the change-point parameter ν can be constructed based on weather, ozone conditions, pollution level, etc.

The nuisance parameters θ and η will have their own prior distributions $\pi_\theta(\theta)$ and $\pi_\eta(\eta)$.

2.1 Shiryaev's Bayes rule

Shiryaev ([20], [21]) formulated and solved the Bayesian change-point detection problem for the case of a geometric prior distribution (with a known and fixed parameter), i.i.d. observations before and after the change point, and the risk function

$$R_1(T) = \lambda \boldsymbol{E}(T-\nu)^+ + \boldsymbol{P}\{T < \nu\} \tag{1}$$

The Bayes stopping rule T in this case can be computed as

$$T = \inf\{n : \Pi_n > \pi^*\}, \tag{2}$$

where

$$\Pi_n = \boldsymbol{P}\left\{\nu \le n | \boldsymbol{X}^{(n)}\right\} = \frac{\sum_{k\le n} \pi_k f(X_1)\cdots f(X_k)g(X_{k+1})\cdots g(X_n)}{\sum_k \pi_k f(X_1)\cdots f(X_k)g(X_{k+1})\cdots g(X_n)}$$

is the posterior probability that a change point has occurred by the time n, and the threshold π^* is computed from the *payoff function*

$$s(\pi) = -\inf_{\{\text{all stopping rules } T\}} R_1(T|\pi_0 = \pi). \tag{3}$$

This prior implies a constant hazard rate function of ν. That is, at any time moment, there is a constant probability p that a change occurs, given that it has not occurred earlier. The Bayes rule in this case signals a change point when the posterior probability of its occurrence exceeds a certain value. Taking a limit as $p \to 0$ leads to a famous Shiryaev-Roberts procedure ([19]). The latter is asymptotically Bayes risk efficient ([16]), also under the geometric prior distribution.

2.2 Cusum algorithm as the Bayes rule

Ritov ([18]) proved decision-theoretic optimality and Bayesian properties of the cusum scheme. Under the risk function

$$R_2(T,\nu) = C_1 \boldsymbol{P}_\nu \{T < \nu\} + C_2 \boldsymbol{E}_\nu (T-\nu)^+ - C_3 \boldsymbol{E}_\nu \min\{T,\nu\} \qquad (4)$$

and a special prior distribution

$$\boldsymbol{P}\left\{\nu = n \mid \nu \geq n, \boldsymbol{X}^{(n)}\right\} = p\left(1 - \frac{f}{g}(X_n)e^{-W_{n-1}}\right)^+,$$

the cusum stopping rule

$$T(h) = \inf\{n : W_n \geq h\}; \quad W_n = \max_{k\in[0;n)} \log \frac{\mathcal{L}(X_1,\ldots,X_n \mid \nu = k)}{\mathcal{L}(X_1,\ldots,X_n \mid \nu = \infty)},$$

is Bayes.

2.3 Hierarchical Bayes analysis

Usually the available prior information yields to a prior distribution different from the ones mentioned above. Results of Shiryaev and Ritov about Bayes stopping rules were generalized in [2] to allow for a wider class of prior distributions. Values of the prior discrete hazard function of a change point $\phi_n = \boldsymbol{P}\{\nu = n+1 \mid \nu > n\}$ were assumed to follow a homogeneous Markov chain. Clearly, this included geometric priors as a special case, but also allowed the prior probabilities to depend on another stochastic process. This model can be used in situations when occurrence of a change point is driven by another time series or a random phenomenon. The latter is observed. By the time t, one observes the states a_j, $j \leq t$ of a Markov chain and computes ϕ_j for $j \leq t$. Then, the posterior probability (for the case of independent observations) that a change has occurred by the time t can be computed as

$$\Pi_t = \boldsymbol{P}\left\{\nu \leq t | \boldsymbol{X}^{(t)}, \phi_0, \ldots, \phi_t\right\} = \frac{\sum_{k\leq t} \pi_k \rho(X_{k+1}) \cdots \rho(X_t)}{\sum_{k<t} \pi_k \rho(X_{k+1}) \cdots \rho(X_t) + 1 - \sum_{k<t} \pi_k},$$

where $\rho = g/f$ and $\pi_k = \phi_0 \cdots \phi_{k-1}(1-\phi_k)$. Derivation of the Bayes stopping rule is essentially based on the theory of optimal stopping ([21]), and likewise, it requires computation of the *payoff function* (3). Finally, it also has the form (2), under the risk function (1) or (4).

2.4 Sequence of Bayes tests

The previous three subsections outlined the situations when the exact Bayes stopping rule for the change-point detection can be constructed. Each case

places certain tight conditions on the prior distribution. This is a major disadvantage because in many applications, especially including epidemiology, rich prior information is available, and it is desirable to utilize it for timely and accurate detection of change points (epidemic trends).

The last case allows a wide class of priors, but the construction of the Bayes stopping rule is based on the payoff function. The latter can be computed numerically, through a sequence of iterations, by solving a certain fixed-point functional equation. Carrying out this computation seems feasible for fairly simple types of prior distributions.

In order to overcome these difficulties and to allow richer and more practical classes of prior distributions, especially of a nonstationary nature, we extended our search to "Bayes-like" stopping rules. The following sequential scheme, generated by a sequence of Bayes tests, is proposed mainly in order to extend the class of considered prior distributions and allow richer and more complicated models.

We begin by noticing that the popular cusum stopping rule was derived from Wald's sequential probability ratio test ([15]), that is, from a sequence of likelihood ratio tests. Occurrence of a change point is reported as soon as the no-change null hypothesis is rejected.

In presence of a prior distribution of ν, we repeat this derivation replacing the likelihood ratio tests by Bayes tests. For $n = 1, 2, \ldots$, we test the null hypothesis $H_o : \nu > n$ that a change point has not occurred yet against the alternative hypothesis $H_1 : \nu \leq n$ that a change has already occurred by the time n. For a chosen significance level α, the Bayes test rejects the null hypothesis if $\Pi_n > 1 - \alpha$. Thus, we define a stopping rule

$$T(\alpha) = \inf \left\{ n \mid H_o \text{ is rejected based on } \boldsymbol{X}^{(n)} \right\} = \inf \left\{ n \mid \Pi_n > 1 - \alpha \right\}.$$

Bayes stopping rules in subsections 1-3 have the form $T(\alpha)$ for some α ([2], Theorem 1). However, no constructive method is available to relate α with risk functions (1) and (4), namely, with coefficients λ and $C_{1,2}$, besides Monte Carlo simulations. Thus, computing $T(\alpha)$ does not yet achieve an optimal balance between the mean delay and the mean time between false alarms. It only explains, intuitively, why optimal change-point detection rules report occurrence of a change point when a high value of Π_n is observed.

2.5 Asymptotically pointwise optimal rules for dependent variables

A viable alternative to Bayesian procedures are *asymptotically pointwise optimal (APO)* stopping rules. Unlike the Bayes stopping rule, it has a closed form expression that can be computed for virtually any arbitrary prior distribution.

Bickel and Yahav ([6], [7]) define an APO rule $T = T_c$ for a loss $L(\delta, \theta))$

function and a risk

$$\boldsymbol{E}\left\{L(\delta,\theta)+cT\right\} \tag{5}$$

as a stopping rule satisfying the inequality

$$\limsup_{c\downarrow 0}\frac{\boldsymbol{E}_X L(\delta(T;X_1,\ldots,X_T),\theta)+cT}{\boldsymbol{E}_X L(\delta(U;X_1,\ldots,X_U),\theta)+cU}\le 1 \tag{6}$$

for any stopping rule U, where the subscript X means that the expectations are taken with respect to the posterior distribution of ν, given all the observations X_j until the stopping time T or U, respectively. Clearly, the Bayes stopping rule is APO, but there are other APO rules that almost reach the minimum posterior risk for small values of λ.

Bickel and Yahav also prove that the stopping rule

$$\tilde{T}=\min\left\{n\mid n^{-1}\boldsymbol{E}_X\left\{L(\delta,\theta)+cn\right\}\le\frac{c}{\beta}\right\} \tag{7}$$

is APO if

$$n^{\beta}\,\boldsymbol{E}_X L(\delta(n;X_1,\ldots,X_n),\theta) \tag{8}$$

converges to a positive limit a.s., as $n\to\infty$, for some $\beta>0$.

Risk functions of the form (5), with a linear cost cT are standard in sequential analysis. However, they are inappropriate in sequential change-point problems. Indeed, there should be no penalty for collecting observations (and not terminating the process) before a change point occurs. After the change point, a price is to be paid for each observation collected. Thus, the cost term is replaced by the mean delay $\boldsymbol{E}(T-\nu)^+$, and risk functions (1) and (4) are justified. The role of the loss function in (5) is played by the probability of a false alarm.

However, (8) is not satisfied by risk functions (1) and (4). Moreover, it is desirable to find an optimal balance between the mean delay and the *logarithm* of the probability of a false alarms. This is motivated by an asymptotically linear relation between the mean delay and the logarithm of the mean time between false alarms for cusum-type schemes ([5], sect. 5.2.1, [1]).

Therefore, we consider a different risk function

$$R(T)=\lambda\boldsymbol{E}(T-\nu)^+-\log^{-1}\boldsymbol{P}\left\{T<\nu\right\} \tag{9}$$

that is also an increasing function of a mean delay and the probability of a false alarm. Similarly to (6), a stopping rule T will be called *asymptotically pointwise optimal for change-point detection* if

$$\limsup_{\lambda\downarrow 0}\frac{\lambda\boldsymbol{E}_X(T-\nu)^+-\log^{-1}\boldsymbol{P}_X\left\{T<\nu\right\}}{\lambda\boldsymbol{E}_X(U-\nu)^+-\log^{-1}\boldsymbol{P}_X\left\{U<\nu\right\}}\le 1 \tag{10}$$

a.s. for any stopping rule U.

Theorems 1 and 2 establish a class of APO rules for the sequential change-point detection under very mild conditions about the prior distribution of ν. These APO rules and the conditions on the prior distribution $\pi(\nu)$ are stated in terms of the *prior survival function of a change point*

$$S(t) = \sum_{j=t+1}^{\infty} \pi_j = \boldsymbol{P}^{\pi}(\nu > t)$$

and the *posterior survival function* of a change point

$$S_X(t) = 1 - \Pi_t = \boldsymbol{P}^{\pi}(\nu > t \,|\, \boldsymbol{X}^{(t)}). \tag{11}$$

Introducing the likelihood ratios

$$\rho_1 = \frac{g(X_1)}{f(X_1)}, \quad \rho_{n+1} = \frac{g(X_{n+1}|\boldsymbol{X}^{(n)})}{f(X_{n+1}|\boldsymbol{X}^{(n)})},$$

we have

$$S_X(t) = \frac{\sum_{k>t} \pi_k f(\boldsymbol{X}^{(t)})}{\sum_k \pi_k f(\boldsymbol{X}^{(k)}) g(\boldsymbol{X}^{(k+1:t)}|\boldsymbol{X}^{(k)})} = \frac{S(t)}{\sum_{k\le t} \pi_k \rho_{k+1} \cdots \rho_t + S(t)} \tag{12}$$

The role of the posterior expected loss term in (7) is now played by

$$r_t = -\log^{-1} S_X(t) = \log^{-1}\left(1 + \frac{\sum_{k\le t} \pi_k \rho_{k+1} \cdots \rho_t + S(t)}{S(t)}\right). \tag{13}$$

The next Theorem establishes asymptotics of r_t for large t.

Theorem 1 *Suppose that the strong law of large numbers holds for log-likelihood ratios* ρ_k,

$$t^{-1} \sum_{k=1}^{t} \log \rho_k \to K > 0, \ \textit{as } t \to \infty, \tag{14}$$

with G-probability one. Then there exists an a.s. limit

(i)	$\lim_{t\to\infty}(tr_t) = 1/K$	*if*	$t^{-1}\lvert\log S(t)\rvert \to 0$ *as* $t \to \infty$
(ii)	$\lim_{t\to\infty}(tr_t) = 1/\{K+L\}$	*if*	$t^{-1}\lvert\log S(t)\rvert \to L > 0$ *as* $t \to \infty$
(iii)	$\lim_{t\to\infty}(t^{\beta} r_t) = 1/L$	*if*	$t^{-\beta}\lvert\log S(t)\rvert \to L > 0$ *as* $t \to \infty$ *for some* $\beta > 1$.

Proof: From (13),

$$\frac{1}{r_t} = -\log(1 - \Pi_t) = -\log S(t) + \log \sum_{k=0}^{\infty} \pi_k \rho(X_{k+1}) \cdots \rho(X_t) = A_t + B_t \text{ (say)}, \tag{15}$$

where $\rho(X_{k+1} \cdots \rho(X_t) = 1$ if $k \geq t$. Consider the terms A_t and B_t. In cases (i) and (ii), $A_t/t \to L \geq 0$, as $t \to \infty$, and in case (iii), $A_t/t \to \infty$ and $A_t/t^\beta \to L > 0$. Also, for any t,

$$B_t \leq \log\left(\sum_{k=0}^{\infty} \pi_k\right) \max_{0 \leq k \leq t} \rho(X_{k+1}) \cdots \rho(X_t) = \max_{0 \leq k \leq t} \sum_{i=k+1}^{t} \log \rho_i = W_t,$$

where W_t is the cusum process based on log-likelihood ratios. There exists an a.s. limit (under G)

$$\lim_{t \to \infty} t^{-1} W_t = \lim_{t \to \infty} t^{-1} \sum_{k=1}^{t} \log \rho_k = K,$$

because

$$W_t = \max_{0 \leq k \leq t} \sum_{i=k+1}^{t} \log \rho_i = \sum_{i=1}^{t} \log \rho_i - \min_{0 \leq k \leq t} \sum_{i=1}^{k} \log \rho_i,$$

where $\sum_1^t \log \rho_i \sim Kt$ and $\min_{[0,t]} \sum_1^k \log \rho_i = O(1)$, as $t \to \infty$. The latter is uniformly bounded for all t because $\sum_1^k \log \rho_i \to \infty$, $k \to \infty$, G-a.s., in view of (14). Thus, we have the inequality

$$\limsup_{t \to \infty} \frac{B_t}{t} \leq K. \tag{16}$$

Also,

$$B_t \geq \log\left(\pi_m \rho(X_{m+1}) \cdots \rho(X_t)\right) = C_m + D_t,$$

where $C_m = \log \pi_m - \sum_1^m \log \rho(X_k)$, $D_t = \sum_1^t \log \rho(X_k)$, and $m = \min\{k \mid \pi_k > 0\}$ is the essential infimum of ν under $\pi(\nu)$. Since $C_m/t \to 0$ and $D_t/t \to K$, as $t \to \infty$, we have

$$\liminf_{t \to \infty} \frac{B_t}{t} \geq K. \tag{17}$$

From (15), (16) and (17), we have, as $t \to \infty$,

$$tr_t = \frac{1}{A_t/t + B_t/t} \to \frac{1}{-\lim_t (A_t/t) + K},$$

which leads to statements (i) and (ii). In case (iii),

$$t^\beta r_t = \frac{1}{A_t/t^\beta + B_t/t^\beta} \to \frac{1}{L + 0},$$

which proves statement (iii). □

Theorem 1 establishes a convergence property of the posterior expected loss that is similar to (8).

We notice that conditions $(i) - (iii)$ of Theorem 1 include many common distributions of ν, however, ν is not allowed to be a bounded variable. The case

of bounded ν is trivial because only stopping rules $\tilde{T}$ that equal $\operatorname{ess\,sup}(\nu)$ with probability tending to 1 can be APO in this case.

In the classical situation of F-i.i.d. variables before the change point and G-i.i.d. variables after it, condition (14) holds with K being the Kullback-Leibler information, $K = K(G, F)$. Certainly, (14) is true in a much wider range of processes: L_p-mixingales, L_p-NED (near-epoch-dependent in $L-p$ norm) sequences, stationary invertible ARMA ([8], [10]).

Theorem 2 *Under conditions (i) or (ii) of Theorem 1, the stopping rule*

$$\tilde{T} = \inf\left\{n \mid -n \log S_X(n) \geq \lambda^{-1}\right\} \tag{18}$$

is asymptotically pointwise optimal under the risk function (9). Under condition (iii) of Theorem 1, the stopping rule

$$\tilde{T} = \inf\left\{n \mid -n \log S_X(n) \geq \beta\lambda^{-1}\right\} \tag{19}$$

is asymptotically pointwise optimal under the risk function (9).

Proof: In order to prove (10), we compare the posterior risk of $\tilde{T}$ with that of an arbitrary stopping rule U. Without loss of generality, we can consider only such stopping rules $U = U(\lambda)$ that tend to ∞ a.s., as $\lambda \downarrow 0$.

Indeed, if $U \not\to \infty$ with a positive probability, then with the same probability $S_X(U) = \boldsymbol{P}\left\{\nu > U \mid \boldsymbol{X}^{(U)}\right\} \not\to 0$ which precludes the posterior risk $\lambda \boldsymbol{E}_X(U-\nu)^+ - \log^{-1} S_X(U)$ from tending to 0.

At the same time, there exist stopping rules that have a posterior risk tending to 0 a.s. A trivial example is $U \equiv [\lambda^{-1/2}]$ for any $X_1, X_2, \ldots$. For this rule, $\lambda \boldsymbol{E}_X(U-\nu)^+ \leq \lambda[\lambda^{-1/2}] \to 0$, as $\lambda \to 0$, and $S_X(U) \to 0$ according to Theorem 1. Hence, the posterior risk of this stopping rule converges to 0 a.s.

Thus, we consider only the stopping rules U that tend to infinity a.s., as $\lambda \downarrow 0$, and satisfy $S_X(U) \to 0$. Also, $\tilde{T} \to \infty$ a.s. For these rules, there exists a positive limit $\lim_{\lambda \downarrow 0} U^\beta r_U = V$, given by Theorem 1, where we set $\beta = 1$ in cases (i) and (ii). Since $(\tilde{T}-1)$ also tends to ∞ as $\lambda \downarrow 0$, we have $U^\beta r_U \sim (\tilde{T}-1)^\beta r_{\tilde{T}-1}$.

The rest of the proof is rather similar to Theorem 5.4.1 of [11]. According to (18) and (19), $r_{\tilde{T}} = -1/\log S_X(\tilde{T}) \leq \lambda\tilde{T}/\beta$, but $r_{\tilde{T}-1} > \lambda(\tilde{T}-1)/\beta$. Hence, as $\lambda \downarrow 0$,

$$\begin{aligned} H(\tilde{T}, U) &= \frac{\lambda \boldsymbol{E}_X(\tilde{T}-\nu)^+ - \log^{-1} S_X(\tilde{T})}{\lambda \boldsymbol{E}_X(U-\nu)^+ - \log^{-1} S_X(U)} \leq \frac{\lambda\tilde{T} + \lambda\tilde{T}/\beta}{\lambda \boldsymbol{E}_X(U-\nu)^+ + r_U} \\ &\sim \frac{\lambda\tilde{T}(1+1/\beta)}{\lambda U + \left(\frac{\tilde{T}-1}{U}\right)^\beta r_{\tilde{T}-1}} < \frac{\lambda\tilde{T}(1+1/\beta)}{\lambda U + \lambda(\tilde{T}/U)^\beta(\tilde{T}-1)/\beta} \\ &\sim \frac{1+1/\beta}{(U/T) + (\tilde{T}/U)^\beta/\beta} \leq \sup_{y>0} \frac{1+1/\beta}{y + y^{-\beta}/\beta} = 1. \end{aligned} \tag{20}$$

Hence $\limsup_{\lambda\downarrow 0} H(\tilde{T}, U) \le 1$, and $\tilde{T}$ is APO.

In (20), we used the equality

$$\lim_{\lambda\downarrow 0} \frac{\boldsymbol{E}_X(U-\nu)^+}{U} = 1 \text{ a.s.} \tag{21}$$

It holds because

$$\frac{\boldsymbol{E}_X(U-\nu)^+}{U} = \boldsymbol{E}_X\left(1-\frac{\nu}{U}\right) I_{\nu\le U} = 1 - S_X(U) - \frac{1}{U}\boldsymbol{E}_X \nu I_{\nu\le U}.$$

As explained above, without loss of generality we consider only such stopping rules U that satisfy $S_X(U) \to 0$, as $\lambda \downarrow 0$. Also,

$$0 \le \limsup_{\lambda\downarrow 0} \frac{1}{U}\boldsymbol{E}_X \nu I_{\nu\le U} \le \lim_{U\to\infty} \frac{1}{U}\sum_{t=0}^{U} S_X(t),$$

where the right-hand side is the Cesaro limit of $S_X(t)$, $t \to \infty$, which equals $\lim_t S_X(t) = 0$. This proves (21). □

We notice that contrary to the computation of Bayes stopping rules, the implementation of the APO scheme (18) is fairly straightforward. It is based on the posterior survival function of ν only, which can be computed directly, from (11) or (12).

2.6 Introducing nuisance parameters

The APO stopping rules proposed by Theorem 2 can only be computed in the case of completely known pre- and post-change distributions F and G. Here, we assume that only families of distributions $\mathcal{F}$ and $\mathcal{G}$ are known, indexed by nuisance parameters $\theta \in \Theta$ and $\eta \in H$, so that $F = F_\theta \in \mathcal{F}$ and $G = G_\eta \in \mathcal{G}$.

Parameters θ and η are considered unknown, having prior distributions π_θ and π_η independently of the change point ν. Asymptotically pointwise optimal stopping rules for change-point detection will now be derived for this case.

Let

$$f^*(\boldsymbol{X}^{(t)}) = \int f_\theta(\boldsymbol{X}^{(t)}) d\pi_\theta(\theta), \quad f^*(X_{t+1}|\boldsymbol{X}^{(t)}) = \frac{f^*(\boldsymbol{X}^{(t+1)})}{f^*(\boldsymbol{X}^{(t)})},$$

$$g^*(\boldsymbol{X}^{(t)}) = \int g_\theta(\boldsymbol{X}^{(t)}) d\pi_\theta(\theta), \quad g^*(X_{t+1}|\boldsymbol{X}^{(t)}) = \frac{g^*(\boldsymbol{X}^{(t+1)})}{g^*(\boldsymbol{X}^{(t)})}$$

be the marginal and conditional densities before and after the change point, respectively. Then the problem becomes to detect a change from the distribution F^* to the distribution G^*, and we define the likelihood ratios

$$\rho_1^* = \frac{g^*(X_1)}{f^*(X_1)}, \quad \rho_{t+1}^* = \frac{g^*(X_{t+1}|\boldsymbol{X}^{(t)})}{f^*(X_{t+1}|\boldsymbol{X}^{(t)})}.$$

Theorem 3 *Suppose that the strong law of large numbers holds for the loglikelihood ratios ρ_t^*, i.e.,*

$$t^{-1}\log\rho_t^* \to K > 0, \ \ as\ t \to \infty, \tag{22}$$

G_η-a.s. for all $\eta \in H$. Then, for change-point detection in presence of nuisance parameters, the stopping rules

(1) the stopping rule $\tilde{T}^* = \inf\left\{t \mid -t\log S_X^*(t) \geq \lambda^{-1}\right\}$ is APO under conditions (i) or (ii) of Theorem 1;

(2) the stopping rule $\tilde{T}^* = \inf\left\{t \mid -t\log S_X^*(t) \geq \beta\lambda^{-1}\right\}$ is APO under condition (iii) of Theorem 1,

where $S_X^*(t) = \boldsymbol{P}\left\{\nu > t|\boldsymbol{X}^{(t)}\right\}$ is the marginal (parameter-free) posterior survival function of the change point.

Proof: It suffices to notice that distributions F^* and G^* satisfy conditions of the previous subsection. The posterior survival function $S_X^*(t)$ is computed as

$$\begin{aligned}
S_X^*(t) &= \frac{\sum_{k>t}\pi_k\mathcal{L}(\boldsymbol{X}^{(t)}|\nu=k)}{\sum_k\pi_k\mathcal{L}(\boldsymbol{X}^{(t)}|\nu=k)} \\
&= \frac{\sum_{k>t}\pi_k\int\int\mathcal{L}(\boldsymbol{X}^{(t)}|\nu=k,\theta,\eta)d\pi_\theta(\theta)d\pi_\eta(\eta)}{\sum_k\pi_k\int\int\mathcal{L}(\boldsymbol{X}^{(t)}|\nu=k,\theta,\eta)d\pi_\theta(\theta)d\pi_\eta(\eta)} \\
&= \frac{\sum_{k>t}\pi_k f^*(\boldsymbol{X}^{(t)})}{\sum_k\pi_k f^*(\boldsymbol{X}^{(k)})g^*(\boldsymbol{X}^{(t)}|\boldsymbol{X}^{(k)}) + \sum_{k>t}\pi_k f^*(\boldsymbol{X}^{(t)})} \\
&= \frac{S(t)}{\sum_{k\leq t}\pi_k\rho_k^* + S(t)}
\end{aligned}$$

This coincides with (12) with ρ_k replaced by ρ_k^*. After this, the proof goes along the lines of Theorems 1 and 2. □

3 Application - detecting the beginning of an epidemic trend

Influenza epidemics are declared when mortality (the percentage of deaths attributed to influenza-like illnesses) exceeds the *epidemic threshold* (Figure 1). However, a change-point may be detected a few weeks before the threshold is exceeded, signaling the beginning of an epidemic trend and allowing to predict epidemics.

In this study, the observed process X_t represents the percentage of patients tested positive for influenza-like illnesses, computed weekly (Figure 2). The dotted curve represents the general seasonal trend, estimated from the historical data. Clear deviations from the general trend are seen during 1999-2000 and

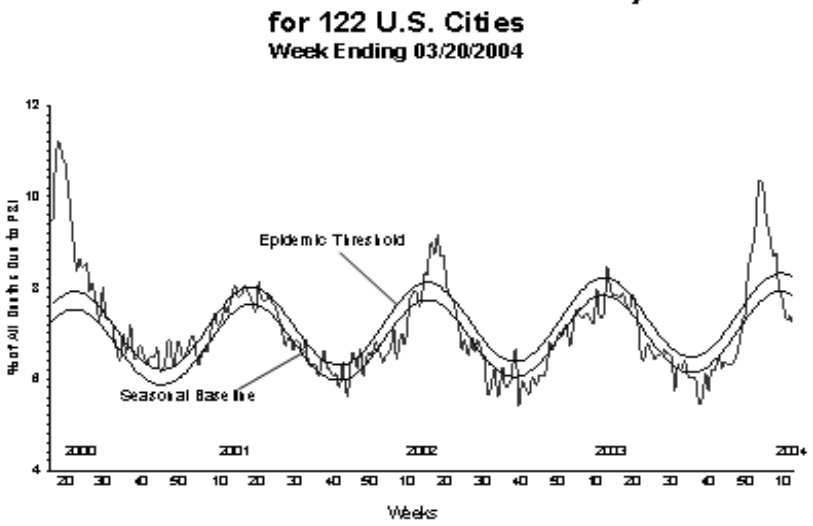

Figure 1: Influenza mortality and epidemic threshold

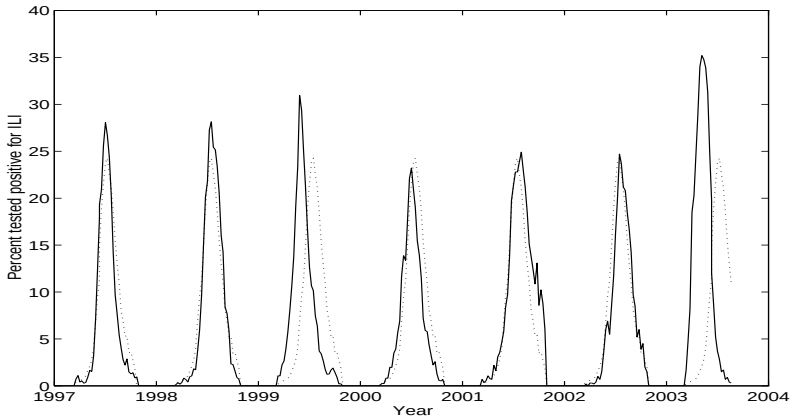

Figure 2: Percentage of patients tested positive for influenza-like illnesses

2003-2004 epidemic seasons. But, the goal is to detect such deviations as soon as possible.

Analysis of detrended series (residuals) shows significant autocorrelation (between 1997 and 2004, the first sample autocorrelation coefficient is $r_1 = 0.958$). Thus, we assume that before a change point, or during an epidemic season without an epidemic, the detrended observed process is weakly stationary, and estimate its autocorrelation coefficients from data. A fully Bayesian approach is certainly possible, with prior distributions on autocorrelation coefficients, as in Subsection 2.6.

Next, one can relate the beginning of pre-epidemic trends to weather conditions, pollution, pollen, ozone, and other factors ([14], [22]). In particular, rapidly changing whether contributes to the possibility of a pre-epidemic trend. We model the prior distribution of the change-point letting its hazard rate to be inversely proportional to the daily standard deviation of temperatures,

$$\phi_n = \frac{\pi_n}{1 - \sum_1^{n-1} \pi_k} = P\{\nu = n \,|\, \nu \geq n\} \propto \sqrt{Var(\text{Temperature})}.$$

An APO stopping rule is then computed according to Theorem 1, detecting epidemic trends in 1999 and 2003 before the epidemic threshold for influenza mortality was crossed.

References

[1] M. Baron. Nonparametric adaptive change-point estimation and on-line detection. *Sequential Analysis*, 19:1–23, 2000.

[2] M. Baron. Bayes stopping rules in a change-point model with a random hazard rate. *Sequential Analysis*, 20:147–163, 2001.

[3] M. Baron. Bayes and asymptotically pointwise optimal stopping rules for the detection of influenza epidemics. In *C. Gatsonis, R. E. Kass, A. Carriquiry, A. Gelman, D. Higdon, D. K. Pauler and I. Verdinelli, Eds., Case Studies of Bayesian Statistics*, volume 6, New York, 2002. Springer-Verlag.

[4] M. Baron. Sequential methods for multistate processes. real-world examples with data analysis. In *N. Mukhopadhyay, S. Datta and S. Chattopadhyay, Eds., Applications of Sequential Methodologies*, New York, 2004. Marcel Dekker.

[5] M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. PTR Prentice-Hall, Inc., 1993.

[6] P. J. Bickel and J. A. Yahav. Asymptotically pointwise optimal procedures in sequential analysis. In *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, volume I: Statistics, pages 401–413, Univ. California Press, Berkeley, Calif., 1967.

[7] P. J. Bickel and J. A. Yahav. Asymptotically optimal Bayes and minimax procedures in sequential estimation. *Ann. Math. Stat.*, 39:442–456, 1968.

[8] J. Davidson. *Stochastic Limit Theory*. Oxford University Press, 1994.

[9] Y.-X Fu, R. N. Curnow. Maximum likelihood estimation of multiple change points. *Biometrika*, 77:563–573, 1990.

[10] A. R. Gallant and H. White. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, New York, 1988.

[11] M. Ghosh, N. Mukhopadhyay and P. K. Sen. *Sequential Estimation*. Wiley, New York, 1997.

[12] S. Hashimoto, Y. Murakami, K. Taniguchi, and Nagai M. Detection of epidemics in their early stage through infectious disease surveillance. *Int. J. Epidemiol.*, 29:905–910, 2000.

[13] C.-B. Lee. Estimating the number of change points in a sequence of independent normal random variables. *Statist. Probab. Letters*, 25:241–248, 1995.

[14] S. Medina, A. Le Tertre, P. Quénel, Y. Le Moullec, P. Lameloise, J. C. Guzzo, B. Festy, R. Ferry, W. Dab. Air pollution and doctors' house calls: Results from the ERPURS system for monitoring the effects of air pollution on public health in Greater Paris, France, 1991-1995. *Environmental Research*, 75:73–84, 1997.

[15] E. S. Page. Continuous inspection schemes. *Biomterika*, 41:100–115, 1954.

[16] M. Pollak. Optimal detection of a change in distribution. *Ann. Statist.*, 13:206–227, 1985.

[17] T. M. Rath, M. Carreras and P. Sebastiani. Automated detection of influenza epidemics using hidden markov models. In *Proc. of the 5th International Symposium on Intelligent Data Analysis*, Berlin, Germany, 2003.

[18] Y. Ritov. Decision theoretic optimality of CUSUM procedure. *Ann. Statist.*, 18:1464–1469, 1990.

[19] S. W. Roberts. A comparison of some control chart procedures. *Technometrics*, 8:411–430, 1966.

[20] A. N. Shiryaev. On optimal method in earliest detection problems. *Theory Probab. Appl.*, 8:26–51, 1963.

[21] A. N. Shiryaev. *Optimal Stopping Rules*. Springer-Verlag, New York, 1978.

[22] M. R. White and I. Hertz-Picciotto. Human health: Analysis of climate related to health. In *Characterization of information requirements for studies of CO2 effects: Water resources, agriculture, fisheries, forests and human health, ed. M.R. White*, pages 172–206, Washington, D.C.: Department of Energy, 1985.