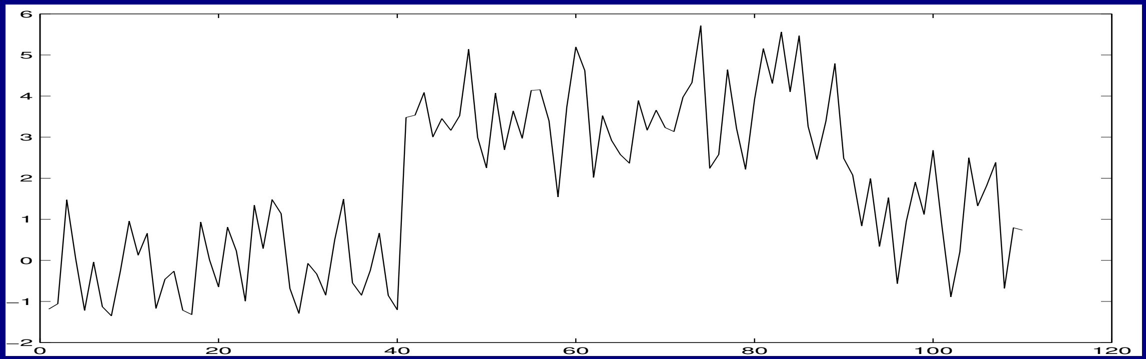
Sequential change-point methods for the early detection of epidemics



by

Michael Barón

University of Texas at Dallas
and
IBM Research Division

mbaron@utdallas.edu

DIMACS Tutorial on Statistical and Other Analytic Health Surveillance Methods

Plan

- Intro to Change-point analysis, multistate processes, and hidden Markov chains

- Detection of epidemics as a change-point problem

- Why Bayes?
  - What loss functions?
  - What priors?

- When the solution has been found

- Alternatives to Bayes rules

- Examples

Classical Change-Point Problem

Observe

$$\mathbf{X} = (X_1, \ldots, X_\nu, X_{\nu+1}, \ldots, X_n) = (\mathbf{X}_1, \mathbf{X}_2)$$

$$\begin{array}{rll} X_1, \ldots, X_\nu & \text{from} & f(x) \\ X_{\nu+1}, \ldots, X_n & \text{from} & g(x) \end{array}$$

where ν = change-point parameter

Features – No consistent estimator

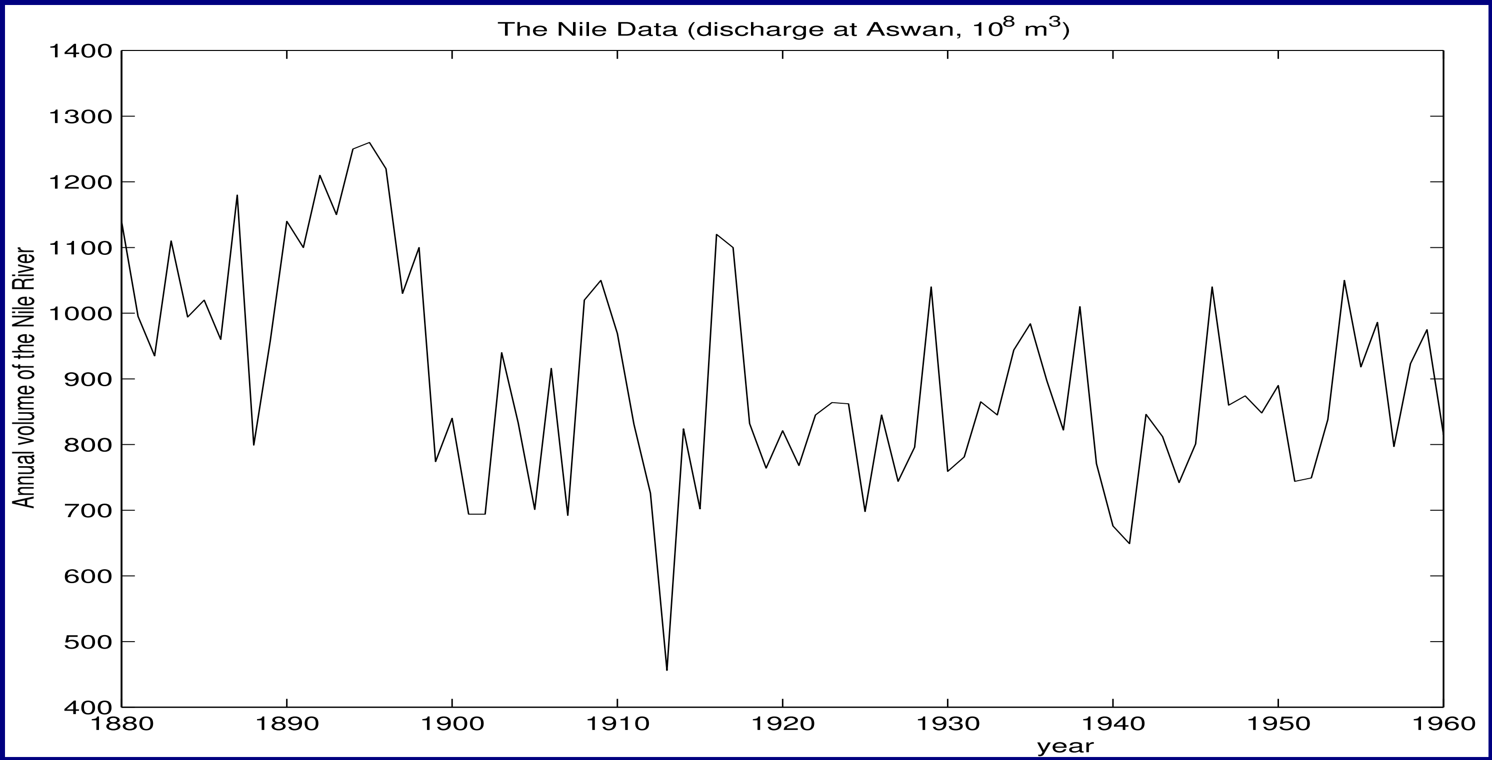
Large n does not help

No data reduction

No regularity

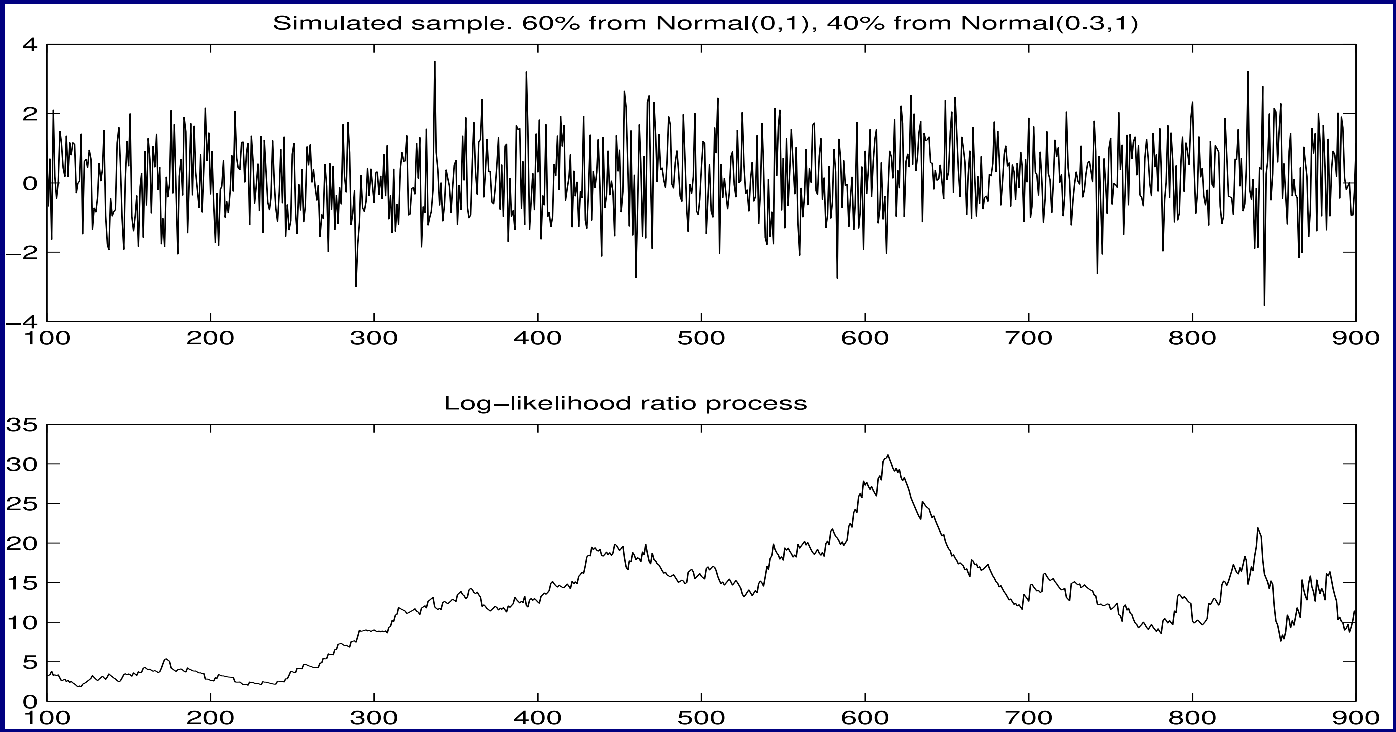
No asymptotic normality

Problem of the Nile



Sensitivity. Detecting small changes

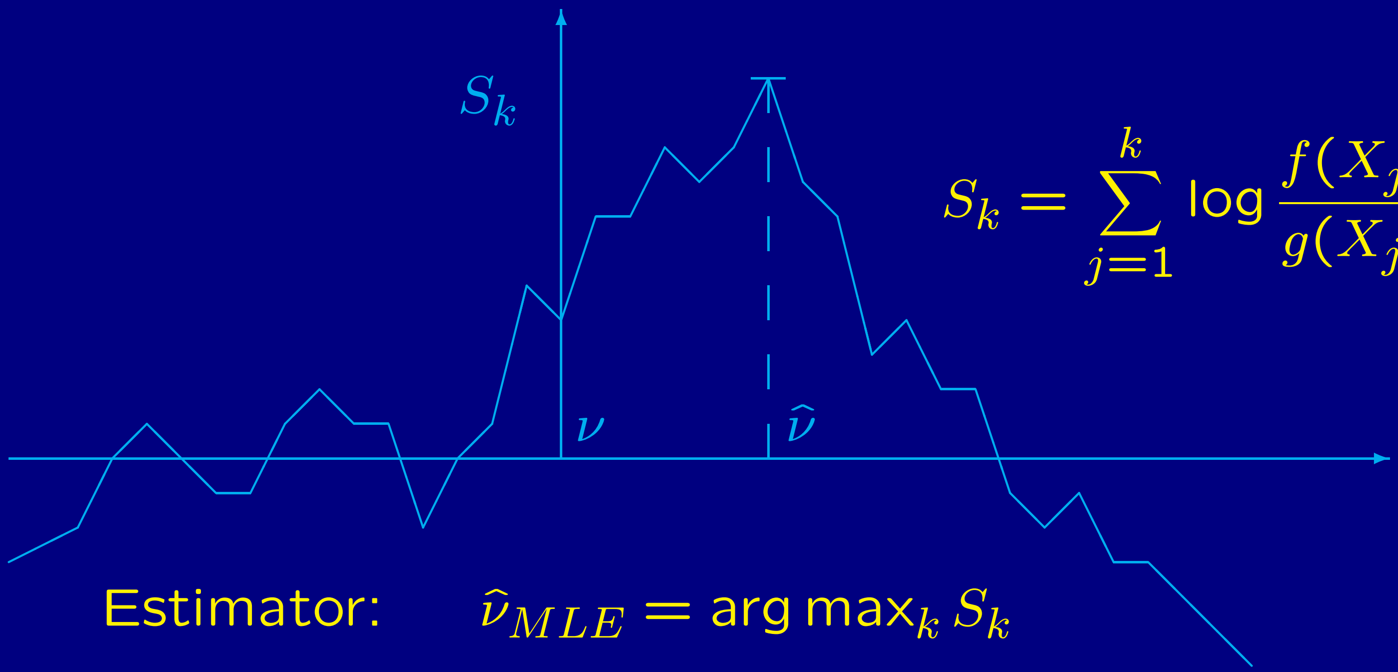
$$X_1 \ldots X_{600} \sim \textit{Normal } (0,\ 1)$$
$$X_{601} \ldots X_{1000} \sim \textit{Normal } (0.3,\ 1)$$



.

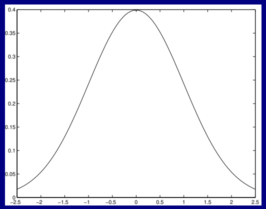
Maximum Likelihood Estimation

MLE is based on a log-likelihood ratio random walk

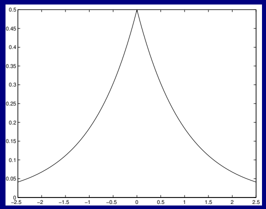


$$S_k = \sum_{j=1}^{k} \log \frac{f(X_j)}{g(X_j)}$$

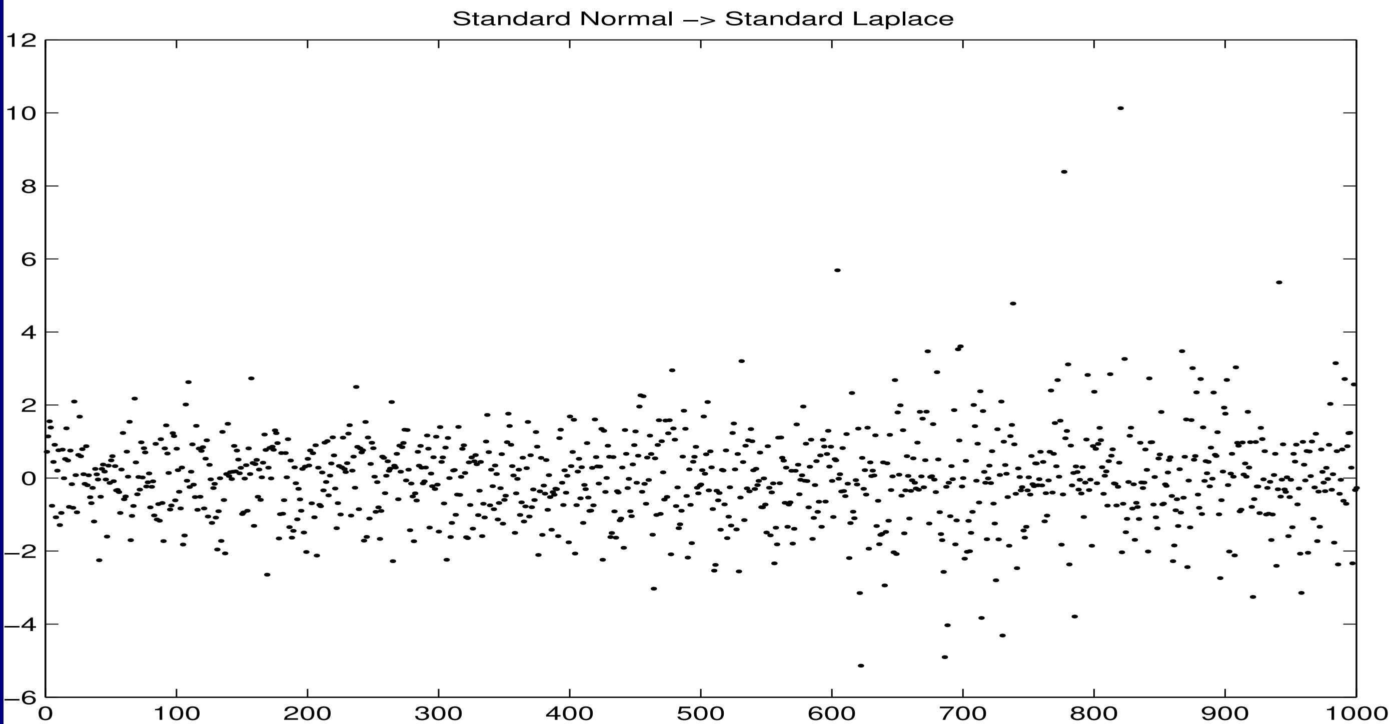
Estimator: $\hat{\nu}_{MLE} = \arg\max_k S_k$



Detecting all kinds of changes



Normal(0,1) $\longrightarrow$ *Laplace(0,1)*

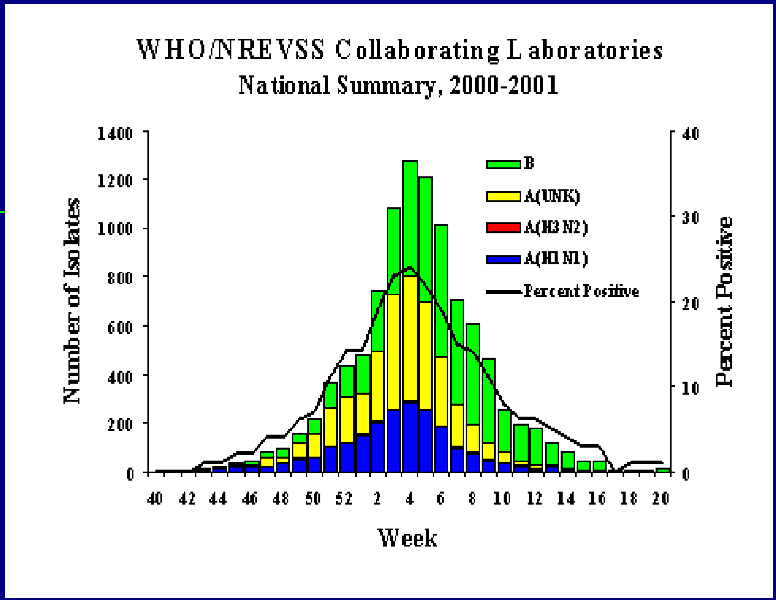


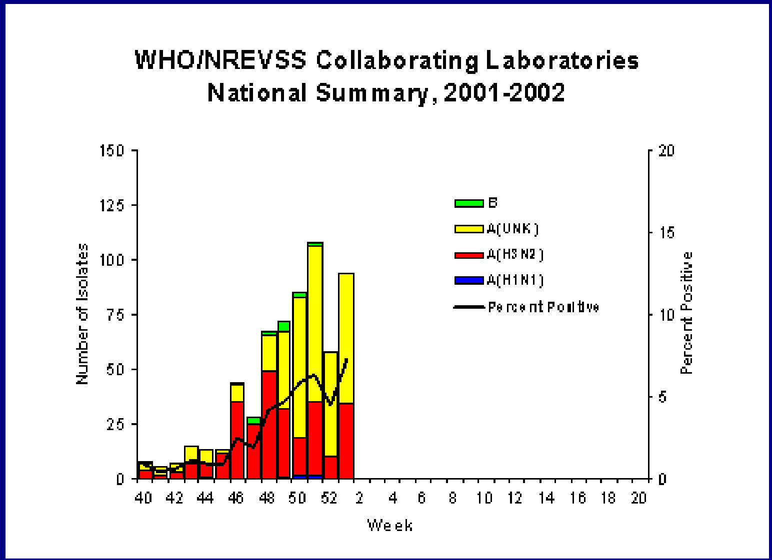
Example 1: Influenza in U.S.

More morbidity and mortality in the U.S. than AIDS

Each year:

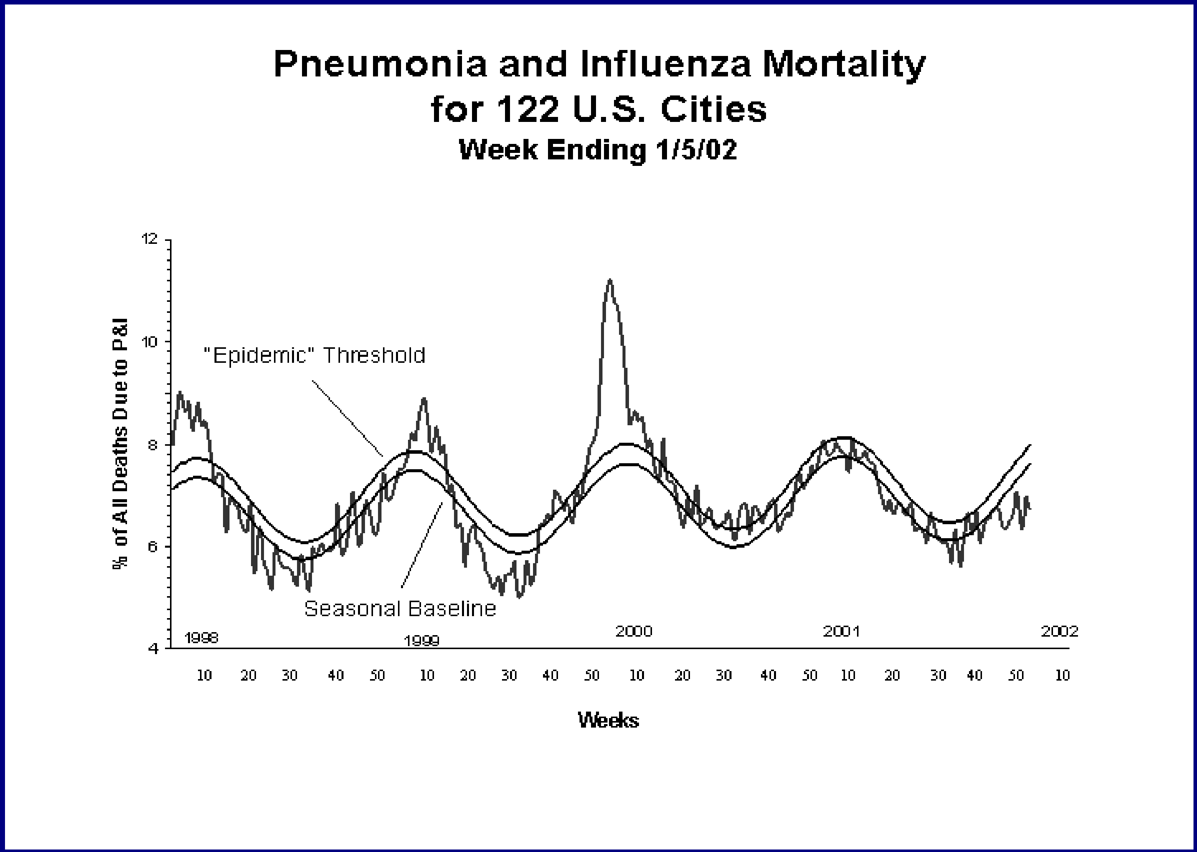
- 10,000 - 40,000 fatalities each year
- 200,000 hospitalizations
- 70 million work-loss days
- associated annual cost of $12 billion





Weeks 2-8 of 2001 = epidemic?

What about 2001/2002 season?



Epidemics = when epidemic threshold is exceeded

Goal: Early detection of epidemic trends

$\Downarrow$

optimal stopping

sequential change-point detection

Sequential change-point detection

Observe $\mathbf{X} = (X_1, \ldots, X_n, \ldots\ \ldots)$ sequentially.

$$\begin{array}{rll} X_1, \ldots, X_\nu & \text{from} & f(x) \\ X_{\nu+1}, \ldots, X_T & \text{from} & g(x) \end{array}$$

where T is a *stopping time*, at which the process is terminated or corrected. *Find T, which*

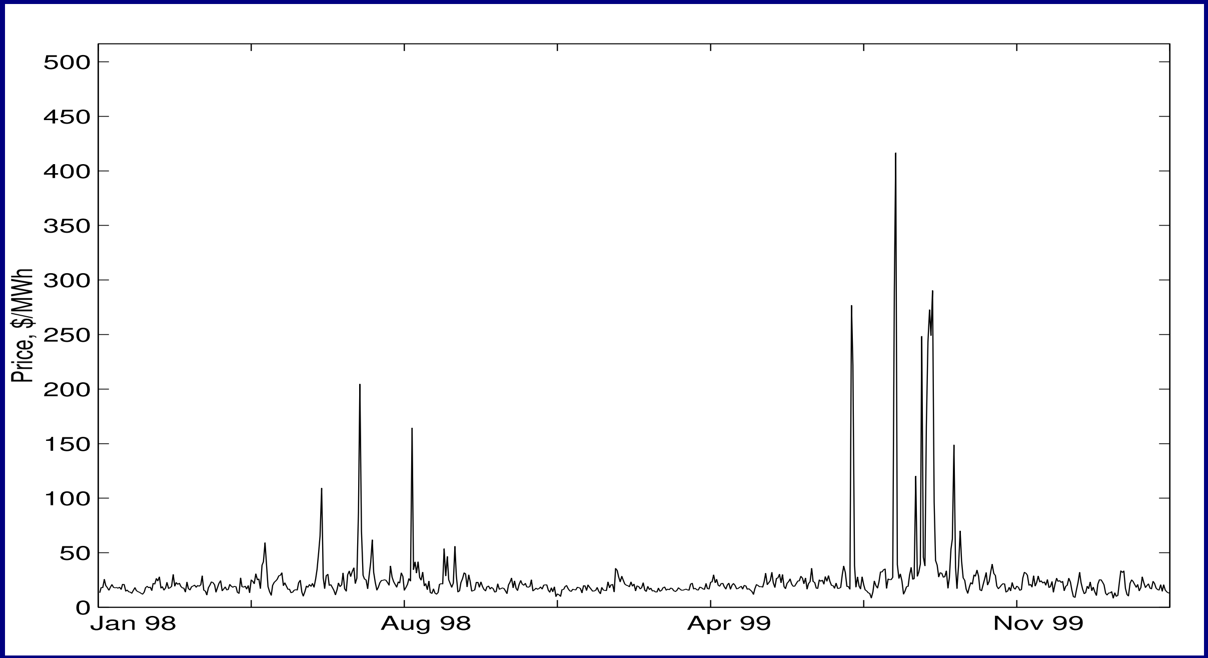
- *minimizes the mean delay*

$$\mathit{MD} = \mathbf{E}_\nu (T - \nu)^+$$

- *maximizes the mean time between false alarms*

$$\mathit{MTBFA} = \mathbf{E}\{T | no\ change\} = \mathbf{E}_{\nu=\infty} T$$

Example 2: Price of Electricity



Need:

(a) working stochastic model
$\Rightarrow$ Monte Carlo simulation study
$\Rightarrow$ valuation of energy derivatives

(b) forecast; predictive distribution of electricity prices for any given day

Systems with multiple change points

Observe $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_\lambda)$

where $$\begin{cases} \mathbf{X}_1 = (X_1, \ldots, X_{\nu_1}) & \sim \; f_1 \\ \mathbf{X}_2 = (X_{\nu_1+1}, \ldots, X_{\nu_2}) & \sim \; f_2 \\ \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ \mathbf{X}_\lambda = (X_{\nu_{\lambda-1}+1}, \ldots, X_{\nu_\lambda}) & \sim \; f_\lambda \end{cases}$$

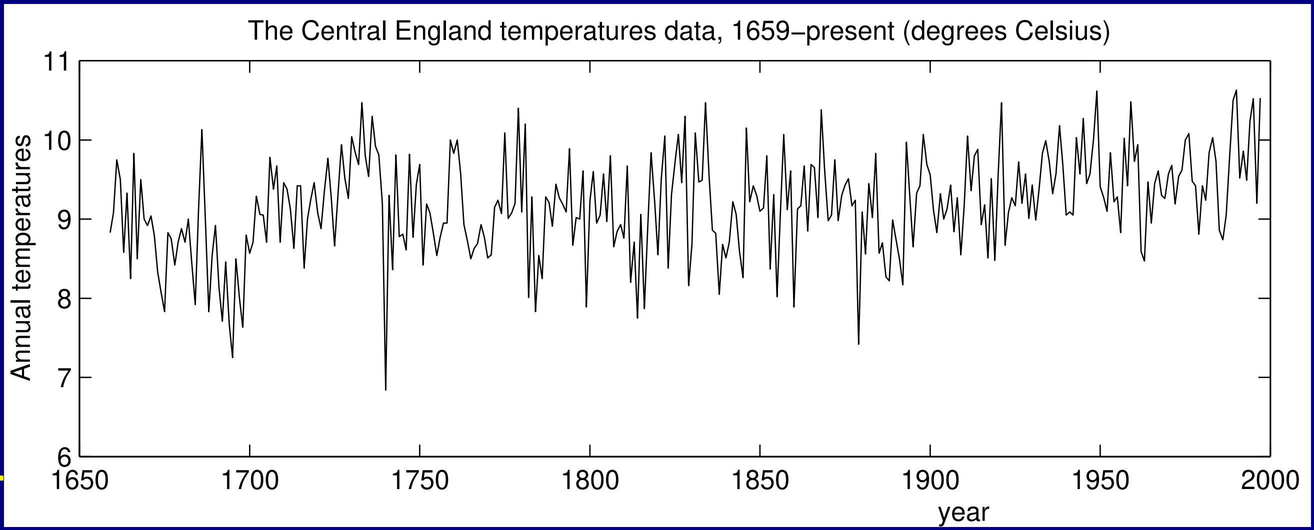
Hidden Markov chains:

States $= \{f_1, f_2, \ldots\}$ are hidden

Transition probabilities $\mathcal{P}_{jk}(t) = \mathbf{P}\left\{X_{t+1} \sim f_k \mid X_t \sim f_j\right\}$

Example 3: Climate

Central England temperatures data 1659 – present



Bayes sequential change-point detection (theory)

§1. Shiryaev (1978)

Under the loss function $L_1(T,\nu) = \lambda(T-\nu)^+ + I_{\{T<\nu\}}$ and the *Geometric* prior $\pi_n = \mathbf{P}\{\nu = n\} = pq^n$,

$$T^* = \inf\{n \geq 0 \mid \Pi_n \geq \pi^*\}$$

is *Bayes*, where

$$\begin{aligned} \Pi_n &= \mathbf{P}\{\nu \leq n \mid X_1, \ldots, X_n\}, \\ \pi^* &= \arg\min_{\pi} s(\pi), \end{aligned}$$

and

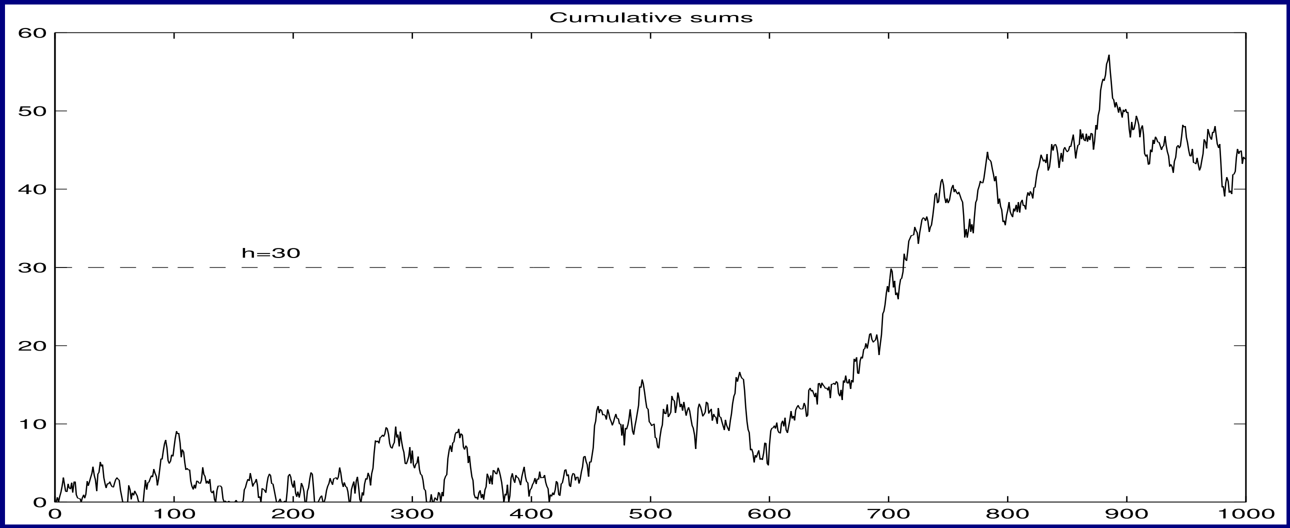
$$s(\pi) = \sup_{\{T\}} \left(-\mathbf{E}^{\mathbf{X},\nu}\{L_1(T,\nu) \mid \Pi_0 = \pi\} \right)$$

is *the payoff function*.

§2. *Ritov (1990)*

CUSUM stopping rule $\quad T(h) = \inf\{n : W_n \geq h\}$,

$$W_n = \max_{k \in [0;n)} \log \frac{\mathcal{L}(X_1, \ldots, X_n \mid \text{change at } \nu = k)}{\mathcal{L}(X_1, \ldots, X_n \mid \text{no change so far})},$$



is *Bayes* under the loss function

$$L_2(T, \nu) = C_1 I_{\{T<\nu\}} + C_2 (T - \nu)^+ - C_3 \min\{T, \nu\}$$

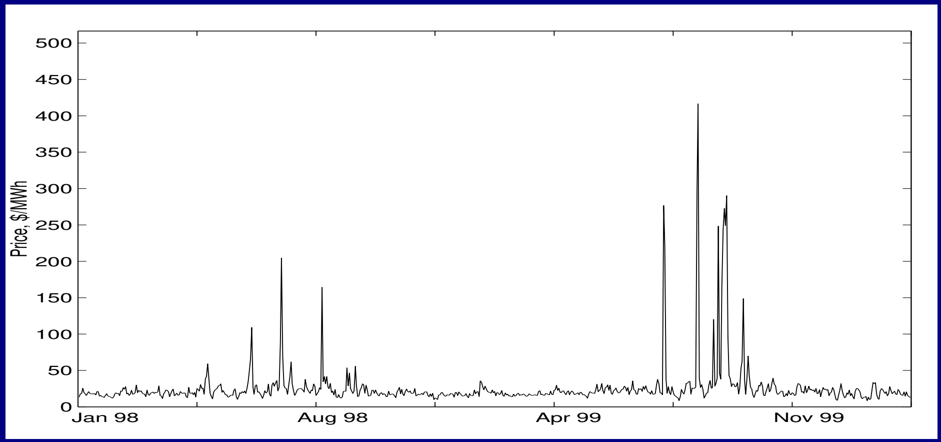
and the prior

$$\mathbf{P}\left\{\nu = n \mid \nu \geq n, X_1, \ldots, X_n\right\} = p\left(1 - \frac{f}{g}(X_n)e^{-W_{n-1}}\right)$$

Bayes sequential change-point detection (practice)

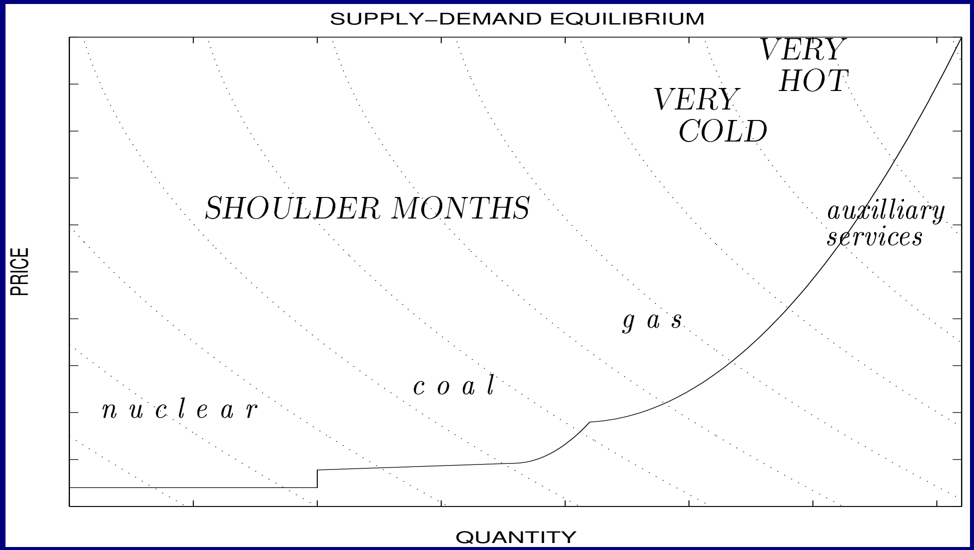
Seeking Mr. Right Prior: what causes change points?

Electricity prices:



.

Electricity prices:

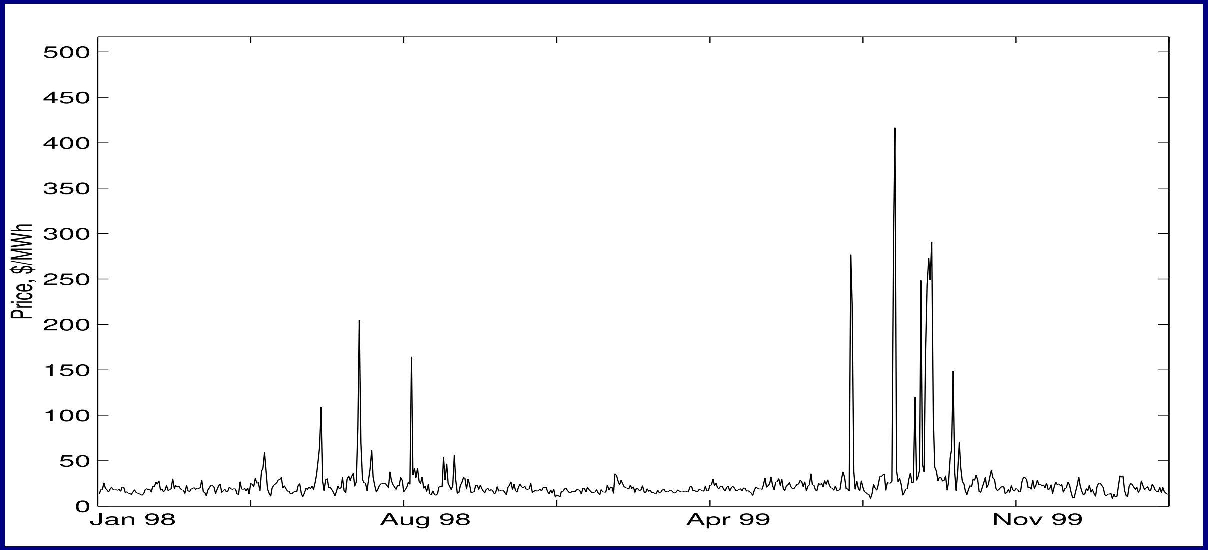


Spikes are likely to start:

- during a season of peak demand
- during a day
- during a weekday
- during unusually hot weather
- during closure or maintanence of a power plant

$\Rightarrow$ prior π_n

Detrended log-prices form a *hidden Markov chain* with 2 states: "regular" and "spikes".



$$\begin{array}{c}\textit{Prior}\\ \textit{distribution}\end{array} \Leftrightarrow \begin{array}{c}\textit{Transition}\\ \textit{probability}\\ \textit{matrix}\end{array} \begin{pmatrix} 1-p(t) & p(t) \\ q & 1-q \end{pmatrix},$$

$p(t) = \mathbf{P}\{$ spike @ $t \mid$ no spike @ $t-1\} = p($*season, weekday, time of the day*$)$,

$q = \mathbf{P}\{$ no spike @ $t \mid$ spike @ $t-1\} = 1/\mathbf{E}$(duration of a spike)

Epidemic trend

Significant factors:

- weather
- air pollution
- pollen
- ozone level
- others

Examples

cold front followed by high pressure

$\Downarrow$

likelihood of epidemic $\uparrow$

- -

high humidity, vigorous air movement

$\Downarrow$

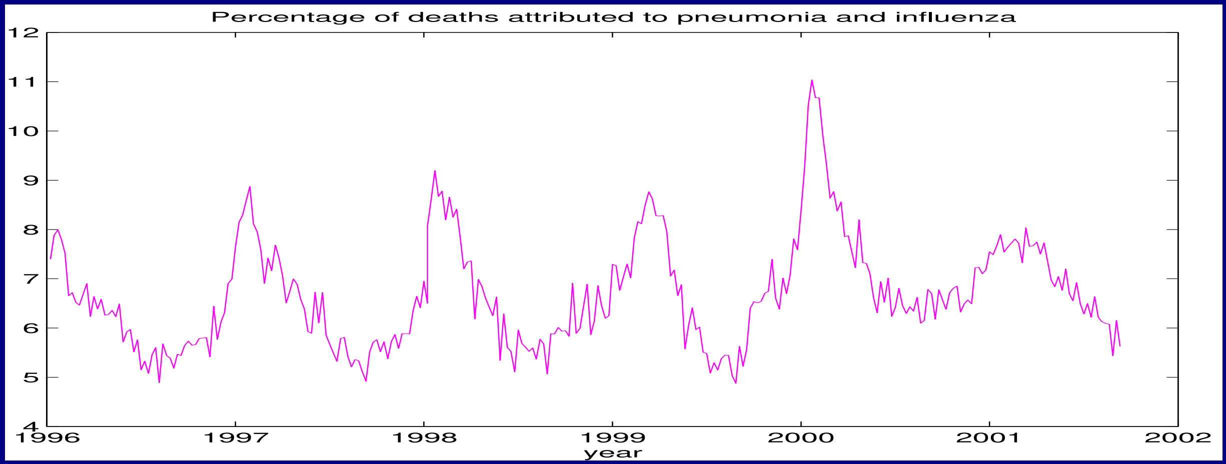
virus dies rapidly

$\Downarrow$

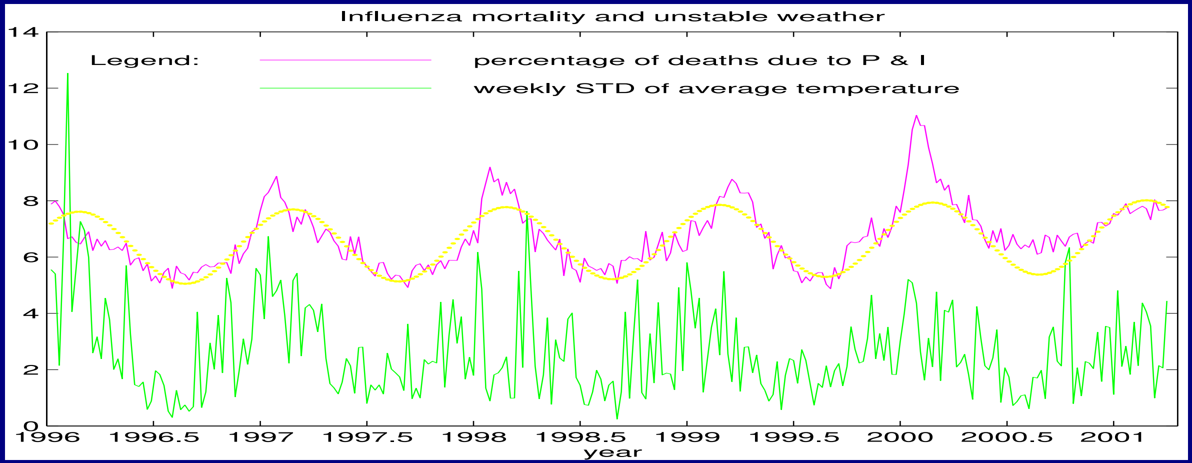
likelihood of epidemic $\downarrow$

Right Prior

Influenza mortality...



... and weather conditions



.

Prior distribution of ν $\quad \phi_t = \mathbf{P}\{\nu = t \mid \nu \geq t\}$

depends on weather, etc.
At time N, only $\{\phi_t,\ t \leq N\}$ are known.

Hyperprior distribution

Based on factors, there are "*favorable*" and "*unfavorable*" weeks for the beginning of an influenza epidemic period

$$\phi_t = \begin{cases} p_0 & \text{on unfavorable weeks} \\ p_1 & \text{on favorable weeks},\ \ p_1 > p_0 \end{cases}$$

Favorable $\Leftrightarrow$ Unstable weather

So, $\{\phi_t\}$ = *homogeneous Markov chain*

$$\mathcal{P} = \begin{pmatrix} 1-q_0 & q_0 \\ q_1 & 1-q_1 \end{pmatrix}$$

A favorable week follows an unfavorable week w/prob q_0

An unfavorable week follows a favorable week w/prob q_1

Bayes stopping rules

Assumption:

$\phi_n = \mathbf{P}\{\nu = n \mid \nu \geq n\} =$ homogeneous Markov chain

(includes the cases of Shiryaev and Ritov)

Method (theory of optimal stopping):

If $\exists$ functions $\eta(\cdot)$ and $\zeta(\cdot)$ and a Markov sequence $\{U_n\}_{n=0}^{\infty}$,

$$\mathbf{E}^{\mathbf{X},\nu} L(T,\nu) = \mathbf{E}\left\{\sum_{n<T} \eta(U_n) - \zeta(U_T)\right\},$$

then $T^* = \inf\{n \mid s(U_n) = \zeta(U_n)\}$ is Bayes,

where $s(u) = \sup_{\{T\}}(-Risk \mid U_0 = u) =$ payoff function

Implementation

Need suitable $\eta, \zeta, \{U_n\}$.

For $L_1(T,\nu)$, take $\begin{cases} \eta(\pi,\cdot) &= \lambda\pi \\ \zeta(\pi,\cdot) &= \pi - 1 \end{cases}$

For $L_2(T,\nu)$, take $\begin{cases} \eta(\pi,\cdot) = (C_2 + C_3)\pi - C_3 \\ \zeta(\pi,\cdot) = C_1(\pi - 1) \end{cases}$

and $U_n = \left(\Pi_n, \phi_{n+1}\right)$,

$$\Pi_n = \mathbf{P}\left\{\nu \le n \,\middle|\, \begin{matrix} X_1,\ldots,X_n \\ \phi_0,\ldots,\phi_n \end{matrix}\right\} = \frac{\sum_{k\le n} \phi_k \prod_{j<k}(1-\phi_j)\rho(k,n)}{\sum_{k<n} \phi_k \prod_{j<k}(1-\phi_j)\rho(k,n) + 1 - \prod_{j<n}(1-\phi_j)},$$

and $\rho(k,n) = \dfrac{\mathcal{L}(X_1,\ldots,X_n \mid \text{change at } \nu = k)}{\mathcal{L}(X_1,\ldots,X_n \mid \text{no change so far})}$

Computing the payoff function

$s(u)$ solves the equation

$$s(u) = \max \left\{ \zeta(u),\ \mathbf{E}\left\{ s(U_{n+1}) \mid U_n = u \right\} - \eta(u) \right\}$$

---

(1) Let $Tw(\pi, \phi) = \mathbf{E}\left\{ w(U_{n+1}) \mid U_n = (\pi, \phi) \right\}$.

For $\phi \in \{p_0, p_1\}$, compute

$$\begin{aligned} Tw(\pi, p_0) = \mathbf{E} \Bigg\{ & (1 - q_0) w \left(\frac{\pi\rho(0,1) + (1-\pi)p_0}{\pi\rho(0,1) + 1 - \pi}, p_0 \right) \\ & + q_0 w \left(\frac{\pi\rho(0,1) + (1-\pi)p_0}{\pi\rho(0,1) + 1 - \pi}, p_1 \right) \Bigg\} (\pi\rho(0,1) + 1 - \pi) \end{aligned}$$

and

$$Tw(\pi, p_1) = \mathbf{E}\left\{(1-q_1)w\left(\frac{\pi\rho(0,1)+(1-\pi)p_1}{\pi\rho(0,1)+1-\pi}, p_1\right)\right.$$
$$\left. + q_1 w\left(\frac{\pi\rho(0,1)+(1-\pi)p_1}{\pi\rho(0,1)+1-\pi}, p_0\right)\right\}(\pi\rho(0,1)+1-\pi)$$

---

(2) *Iterations*. Let

$$Qw(u) = \max\left\{w(u), Tw(\pi,\phi) - \eta(u)\right\}.$$

Compute
$$\begin{array}{rcl} Q^1\zeta(\pi,\phi) & = & Q\zeta(\pi,\phi), \\ Q^2\zeta(\pi,\phi) & = & QQ\zeta(\pi,\phi), \\ \cdots\cdots\cdots & & \cdots\cdots \end{array}$$

Fact:
$$s(u) = \lim_{N\to\infty} Q^N\zeta(u)$$

Large number of iterations $\Rightarrow s(\pi,\phi)$.

Computing the critical probability

$$\pi^*(p_i) = \min\{\pi \mid s(\pi, p_i) = \zeta(\pi, p_i)\}, \quad i = 0, 1$$

Computing the stopping time

$$T^* = \inf\{n \geq 0 \mid \Pi_n \geq \pi^*\}$$

Application - influenza epidemics

Model

Trend + seasonal variation

$$Y_t = m_t + s_t + Z_t$$

$$\begin{aligned} Y_t &= \text{percentage of deaths attributed to flu} \\ Z_t &= \text{deviation from "normal"}, \ \mathbf{E}(Z_t) = 0 \end{aligned}$$

Differencing

$$X_t = \nabla Z_t = Z_t - Z_{t-1}$$

Start of epidemics $\Leftrightarrow$ "unusually" fast increase of mortality

Two phases:

$$\begin{array}{lcl} \text{"in-control"} & \Leftrightarrow & Z_t \sim N(0, \sigma^2) \\ \text{"epidemic"} & \Leftrightarrow & Z_t \sim N(\mu, \tau^2), \;\; \mu > 0 \end{array}$$

Change point

$\nu =$ first week of epidemic trend

$$Z_1, \ldots, Z_\nu \sim N(0, \sigma^2), \quad Z_{\nu+1}, \ldots \sim N(\mu, \tau^2)$$

Parameter estimates from historical data

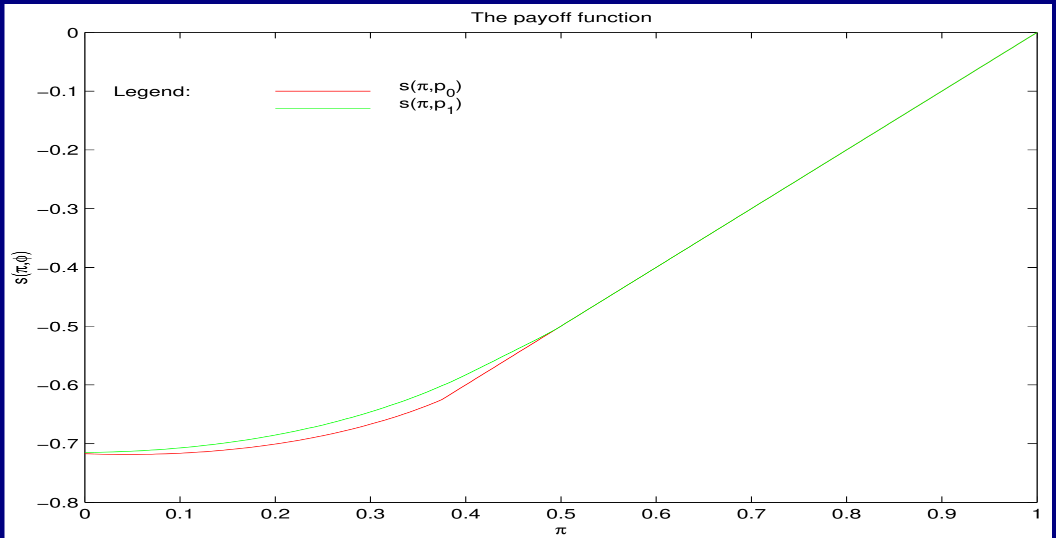
$\mu = 0.0654, \sigma = 0.4278, \tau = 0.4893$

Rates of mortality $p_0 = 0.01,\ p_1 = 0.04$

Transition probabilities $q_0 = 0.2286, q_1 = 0.75$

Loss $\quad L(T, \nu) = 0.03(T - \nu)^+ + I_{\{T<\nu\}}$

Payoff function

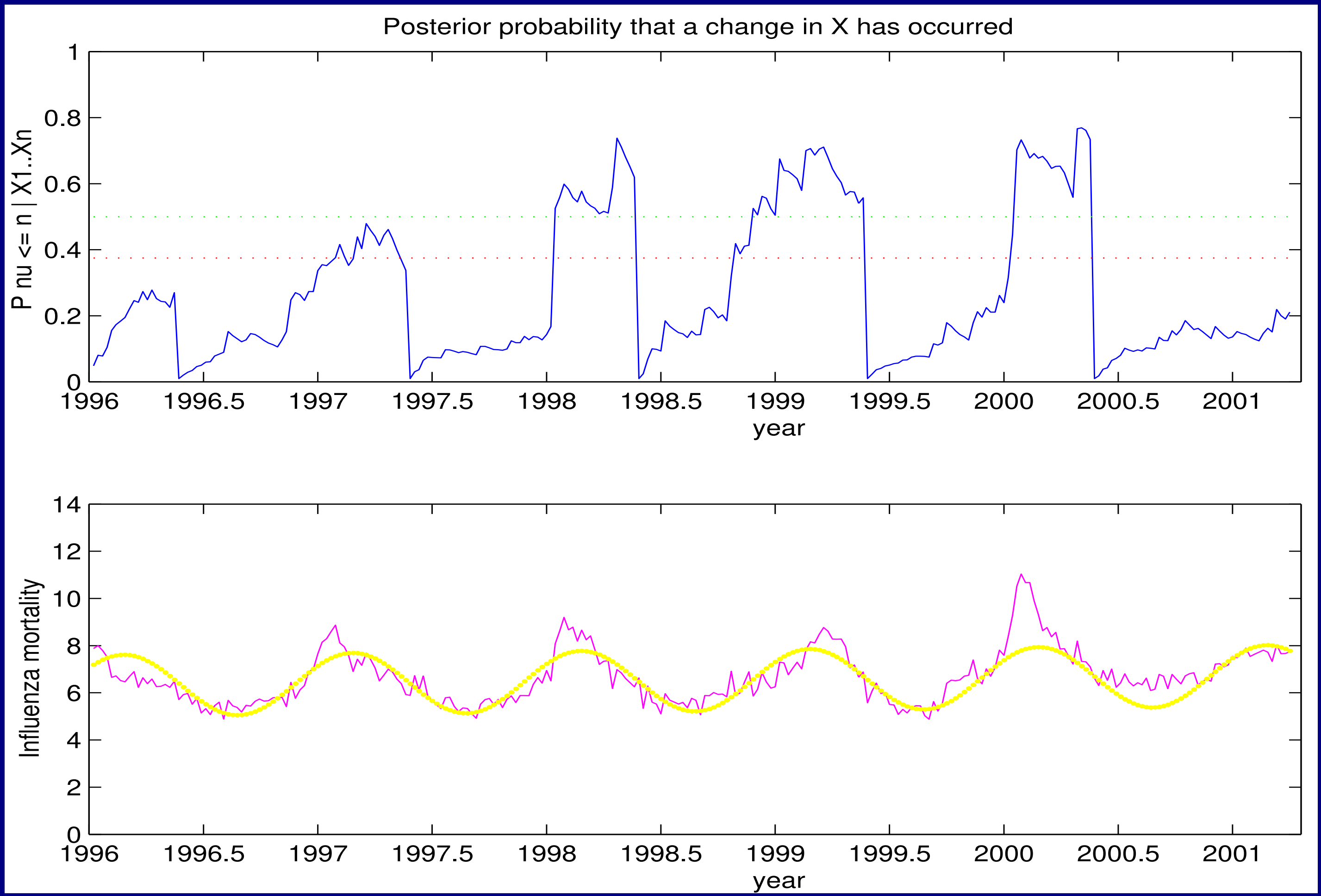


And finally ...

$$\pi^*(p_i) = \min\{\pi \mid s(\pi, p_i) = \zeta(\pi, p_i)\}, \quad i = 0, 1$$

$$T^* = \inf\{n \geq 0 \mid \Pi_n \geq \pi^*\}$$

Posterior distribution



Days of detection:
12/28/1996, 01/10/1998, 10/24/1998, 01/01/2000

Disadvantages of the method

Computation is complicated.

No general algorithm of computing $s(\pi, \phi)$ accurately.

The prior ϕ_t must be very simple.

Alternatives to Bayes sequential rules

§1. *Sequential testing*

For $n = 1, 2, 3, 4, \ldots\ldots,$ test

$$H_0 : \underset{\text{(no change)}}{\nu > n} \quad \text{vs} \quad H_A : \underset{(\exists \text{ change})}{\nu \leq n}$$

where $\quad T(\alpha) = \inf\{n \ : \ H_A \text{ is rejected } \}$

- Sequential probability ratio test $\Rightarrow$ CUSUM

$$\begin{cases} W_n \leq h & \Rightarrow \text{ accept } H_0 \text{ and take } X_{n+1} \\ W_n \geq h & \Rightarrow \text{ reject } H_0 \text{ and stop sampling} \end{cases}$$

- Bayes test $\Rightarrow$?

- Bayes test $\Rightarrow$

  Reject H_0 if $\mathbf{P}\left\{\nu > n | X_1, \ldots, X_n\right\} \leq \alpha$, or

  $\Pi_n \geq 1 - \alpha$ (Bayes rule: $\Pi_n \geq \pi^*$)

$$T(\alpha) = \inf \left\{ n : \sum_{k<n} \frac{\phi_k \rho(k,n)}{\prod_k^n (1-\phi_i)} > \frac{1-\alpha}{\alpha} \right\}$$

Application - energy pricing

P_t = price at time t

Trends

$$\log(P_t) = \alpha + \beta t + \gamma(m_t) + \delta(w_t) + X_t,$$

where m_t =month, w_t = day of the week

Regular mode

X_t is an autoregressive process

$$X_t = \phi X_{t-1} + \sigma \varepsilon_t,$$

$\mathbf{X}_{1:k} \sim N_k(\mathbf{0}, \Sigma)$, with $\Sigma_{ij} = \sigma^2 \phi^{|i-j|}/(1 - \phi^2)$

Spikes

$X_t \sim N(\mu(spike), \tau)$, where $\mu \sim N(\theta, \eta)$

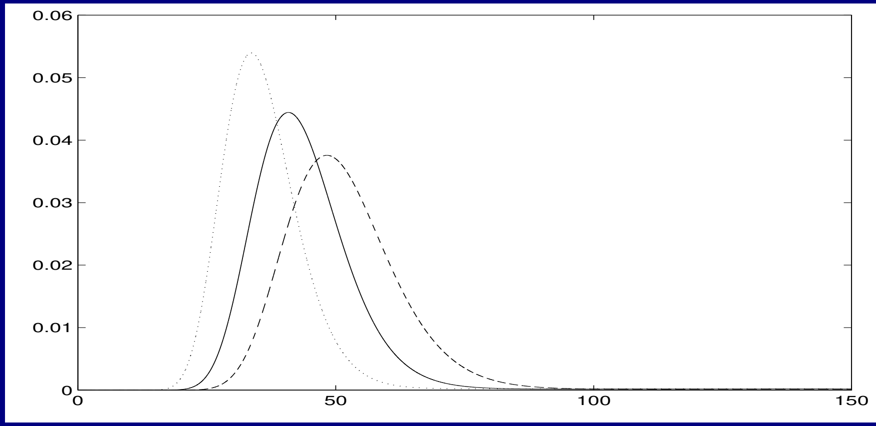
Results

Spikes detected on 05/19-21, 06/25-27, 07/21-23, 08/25, 09/15, and 09/22 in 1998, and on 06/08-09, 07/06-07, 07/24, 07/27-08/01, and 08/12-14 in 1999.

	Total	1998	1999
Mean spike duration ($1/q$)	2.3636	2	2.8
Mean interspike period (during the peak season)	18.7778	23	13.5
mean spike effect on log-prices θ	1.5473	1.1611	2.0107
within-spike variance τ	0.3730	0.1313	0.2693
between-spike variance η	0.0765	0.0648	0.1079
Transition probabilities			
$P_{peak}\{spike \rightarrow control\}$	0.4231	0.5000	0.3571
$P_{peak}\{control \rightarrow spike\}$	0.0533	0.0435	0.0740
$P_{off-peak}\{control \rightarrow control\}$	1	1	1

Forecast

Predictive densities



§2. *Asymptotically pointwise optimal rules*

- For the risk

$$R(T,\theta) = \mathbf{E}\left\{L(T,\theta,\delta) + cT\right\} = \mathbf{E}\left\{loss + cost\right\},$$

a stopping rule T is APO if

$$\limsup_{c\downarrow 0} \frac{\inf_{\delta} \mathbf{E}\left\{L(T,\theta,\delta) \mid \mathbf{X}_{1:T}\right\} + cT}{\inf_{\delta} \mathbf{E}\left\{L(S,\theta,\delta) \mid \mathbf{X}_{1:S}\right\} + cS} \leq 1$$

a.s. for $\forall$ stopping rule S. (Bickel & Yahav, 1968)

Theorem (Bickel, Yahav)

If $N^\beta \mathsf{E}\{L(N,\theta,\delta) \mid \mathbf{X}_{1:N}\} \to V$ a.s. for some $\beta, V > 0$, then

$$T = \text{inf}\left\{n \mid \frac{\mathsf{E}\{L(n,\theta,\delta) \mid \mathbf{X}_{1:n}\}}{n} \le \frac{c}{\beta}\right\}$$

is APO.

In change-point problems,

- Replace cT by the *delay term*
- Add a term penalizing for *false alarms*

Let

$$R(T,\nu) = \lambda \mathbf{E}(T-\nu)^{+} - \log^{-1} \mathbf{P}\left\{T < \nu\right\}.$$

Let a stopping time T be APO if

$$\limsup_{\lambda \downarrow 0} \frac{\lambda \mathbf{E}_X(T-\nu)^{+} - \log^{-1} \mathbf{P}_X\left\{T < \nu\right\}}{\lambda \mathbf{E}_X(S-\nu)^{+} - \log^{-1} \mathbf{P}_X\left\{S < \nu\right\}} \leq 1$$

Theorem

Let $r_t = -\log^{-1} \mathbf{P}^\pi(\nu > t \mid X_1, \ldots, X_t)$
("posterior expected loss").
Then there exists an a.s. limit

$$\lim_{t\to\infty} (t r_t) = \frac{1}{K(G,F)} \qquad \text{if } t^{-1}|\log S(t)| \to 0$$

$$\lim_{t\to\infty} (t r_t) = \frac{1}{K(G,F)+L} \qquad \text{if } t^{-1}|\log S(t)| \to L > 0$$

$$\lim_{t\to\infty} (t^\beta r_t) = \frac{1}{L} \qquad \text{if } t^{-\beta}|\log S(t)| \to L > 0$$

for some $\beta > 1$, where $K(G,F) = \mathbf{E}_G \log \frac{g(X)}{f(X)}$
and $S(t) = 1 - \Pi(t) = \mathbf{P}\{\nu > t \mid X_1, \ldots, X_t\}$.

Theorem

Under (a) or (b), the stopping rule

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

is APO. Under condition (c), the stopping rule

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

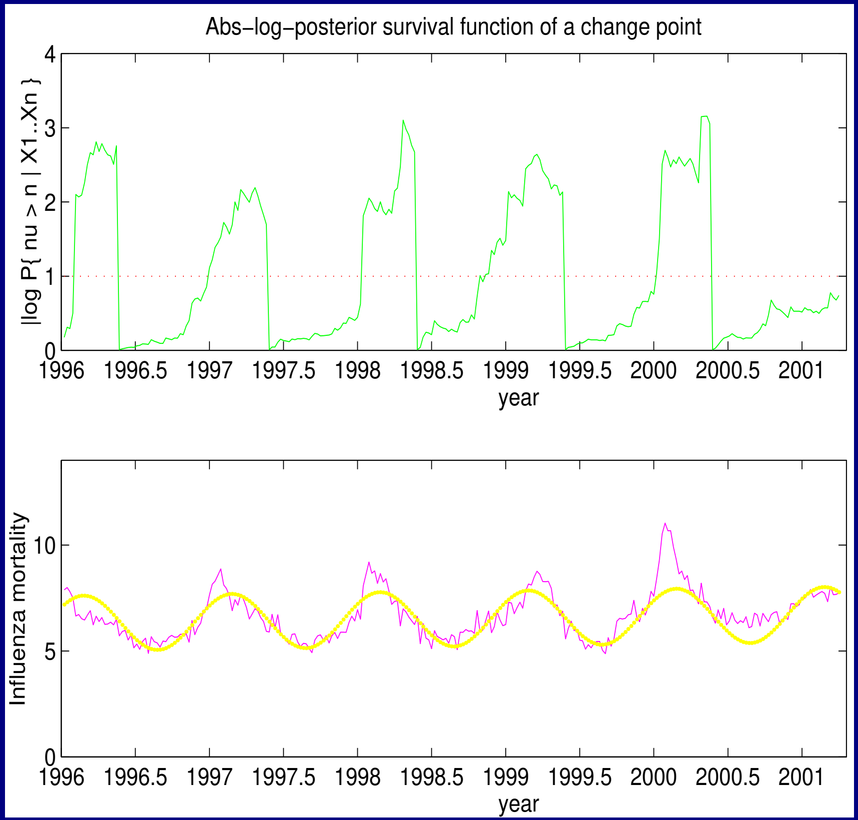
is APO.

APO rules

Application - influenza epidemics

Prior distribution

$$\phi_n = P\{\nu = n \mid \nu \geq n\} \propto Var(Temperature)$$



$-n \log(1 - \Pi_n)$

P&I mortality

Result -

Days of detection			
01/25/97	01/10/98	11/07/98	01/08/00

On the average, this is 3.25 weeks before P&I mortality exceeded the epidemic threshold

---

Any questions?