Detection of Transient Changes

Michael Baron[1] & Sergey V. Malov[2]

[1] American University, Washington DC, USA, baron@american.edu
supported by NSF grant 1737960

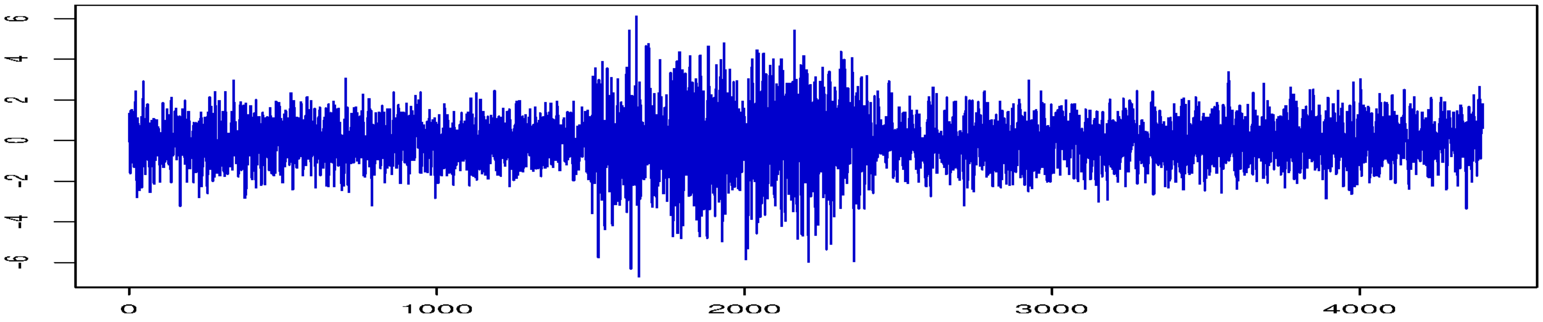
[2] St. Petersburg State University, St. Petersburg, Russia
supported by RSF grant 20-14-00072

IMS Annual Meeting in Probability and Statistics
London, 27–30 June 2022

Introduction: Transient changes

The distribution eventually returns to the initial form,

$$\begin{cases} X_1, \ldots, X_a & \sim F \\ X_{a+1}, \ldots, X_b & \sim G \\ X_{b+1}, \ldots, X_n & \sim F \end{cases}$$



Goals: Detect the change; estimate a and b .

Tartakovsky (1987), Repin (1991), Guépié et al (2012), Noonan and Zhigljavsky (2020), Tartakovsky et al (2021)

Transient changes may reappear at unknown moments,

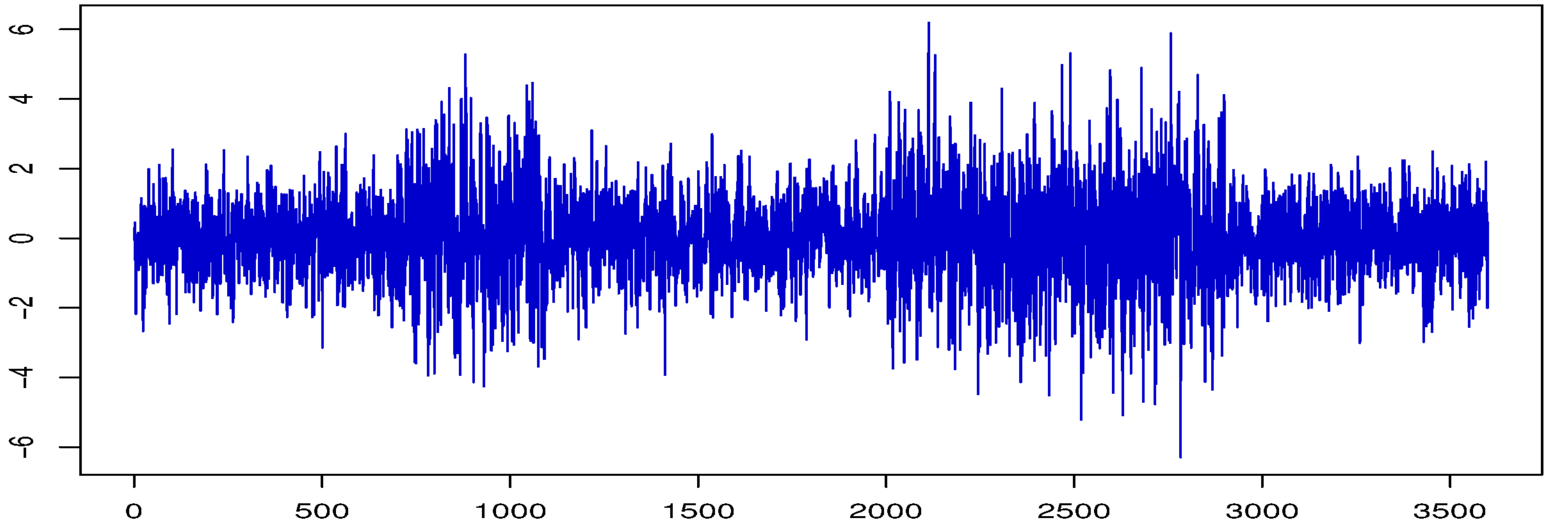
$$\begin{cases} \mathbb{X}_{0:a_1} & = & X_1, \ldots, X_{a_1} & \sim & F \\ \mathbb{X}_{a_1:b_1} & = & X_{a_1+1}, \ldots, X_{b_1} & \sim & G \\ \mathbb{X}_{b_1:a_2} & = & X_{b_1+1}, \ldots, X_{a_2} & \sim & F \\ \mathbb{X}_{a_2:b_2} & = & X_{a_2+1}, \ldots, X_{b_2} & \sim & G \\ \cdots & & \cdots & & \cdots \\ \mathbb{X}_{b_K:n} & = & X_{b_K+1}, \ldots, X_n & \sim & F \end{cases}$$

Goals:

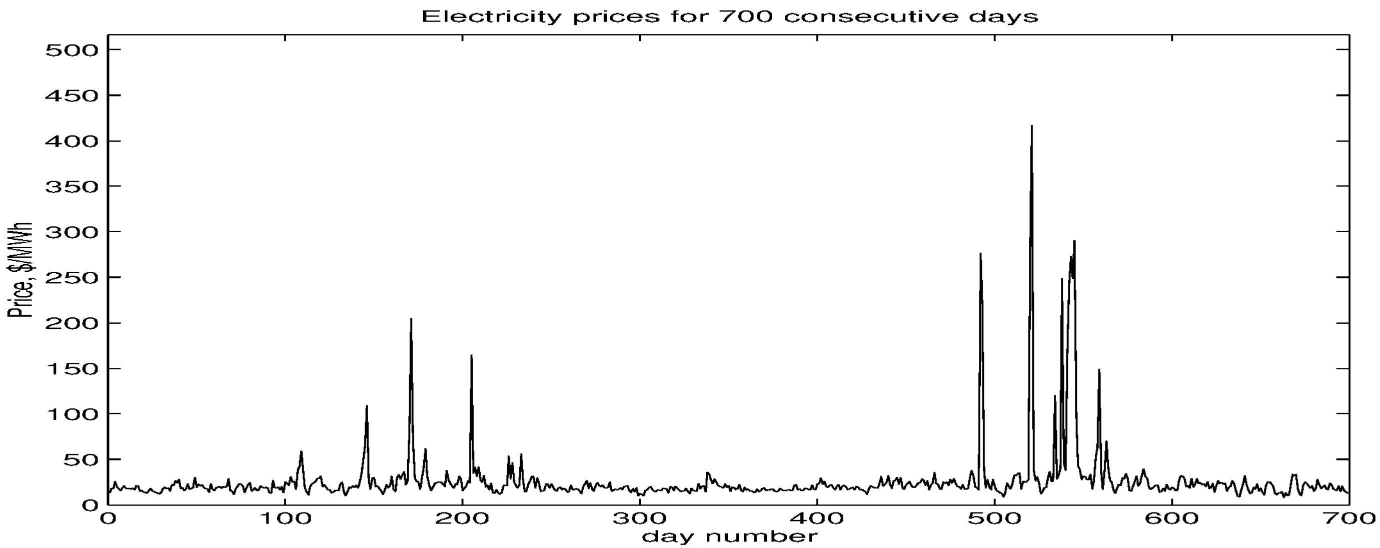
– Detect all changes

– estimate all a_k and b_k

– control familywise false alarm rates



Applications: Deregulated Energy Markets

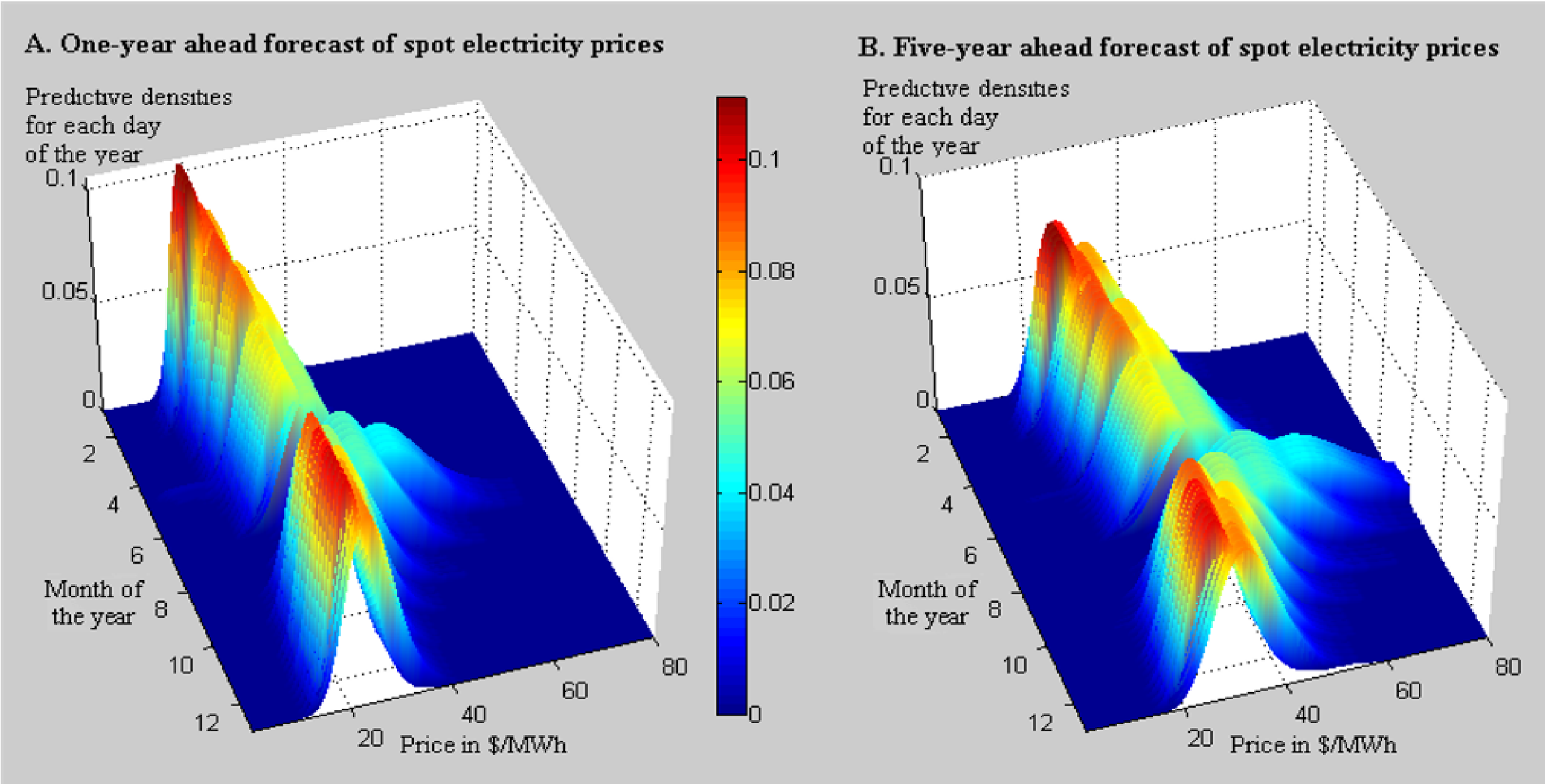


Goals:

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

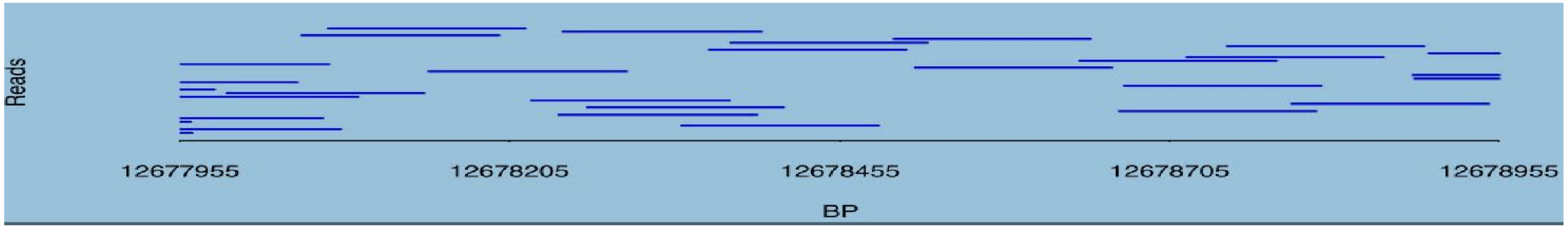
(b) Forecast; predictive distribution of electricity prices for any given day. Baron et al (2001)

Applications: Deregulated Energy Markets

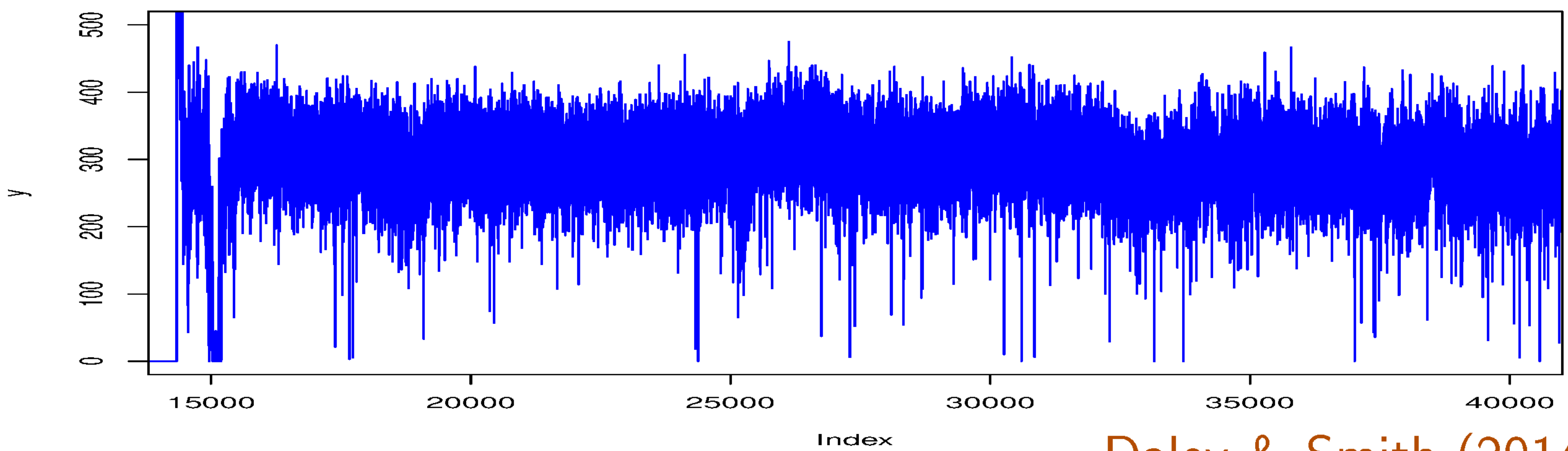


Applications: Genome coverage process

Reads attach to a chromosome at random locations.



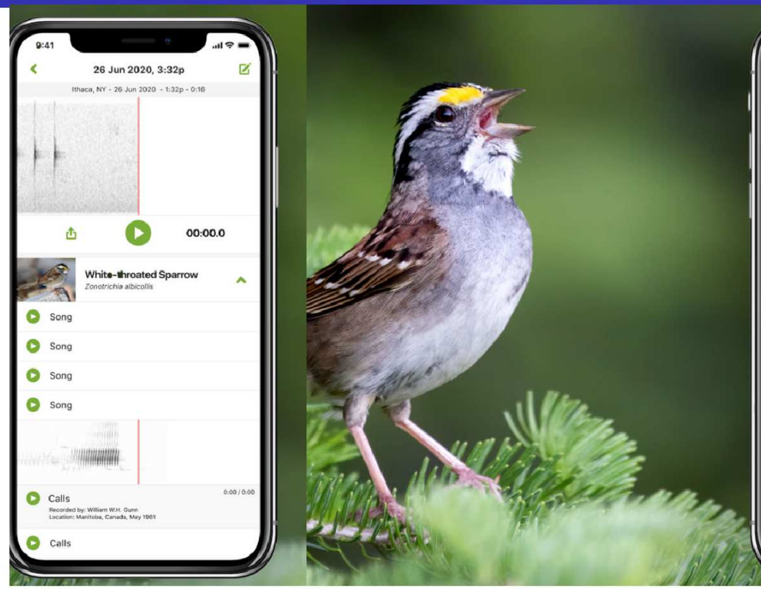
Shifts occur in the coverage depth.

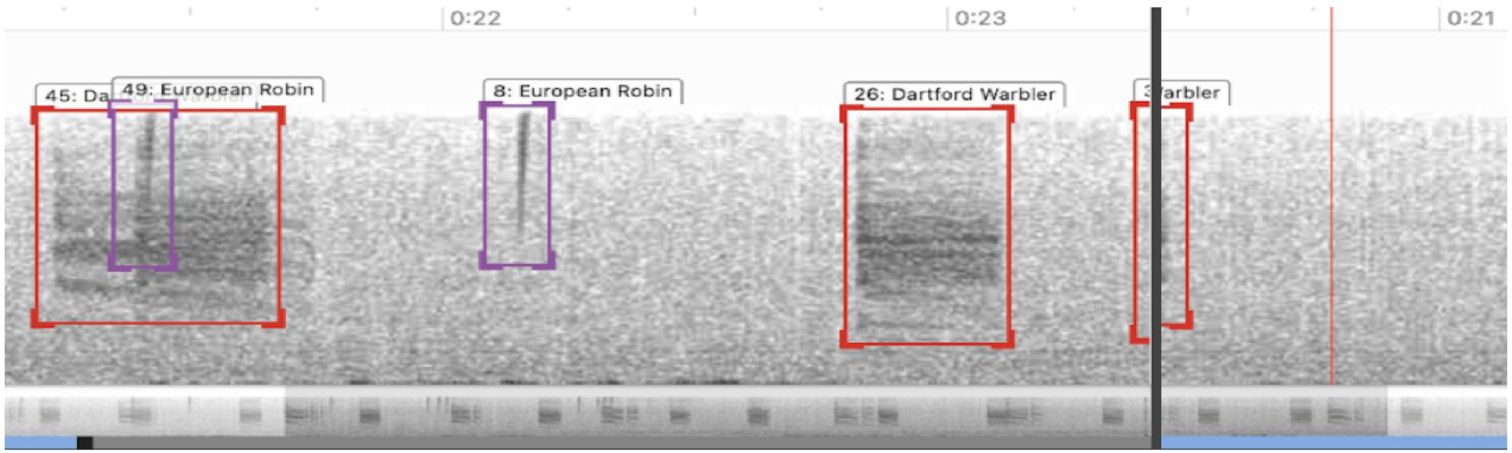


Daley & Smith (2014)

Other Applications

- Industrial process control
- Signal processing
- Image processing
- Target tracking
- Bird song recognition





One transient change: maximum likelihood estimation

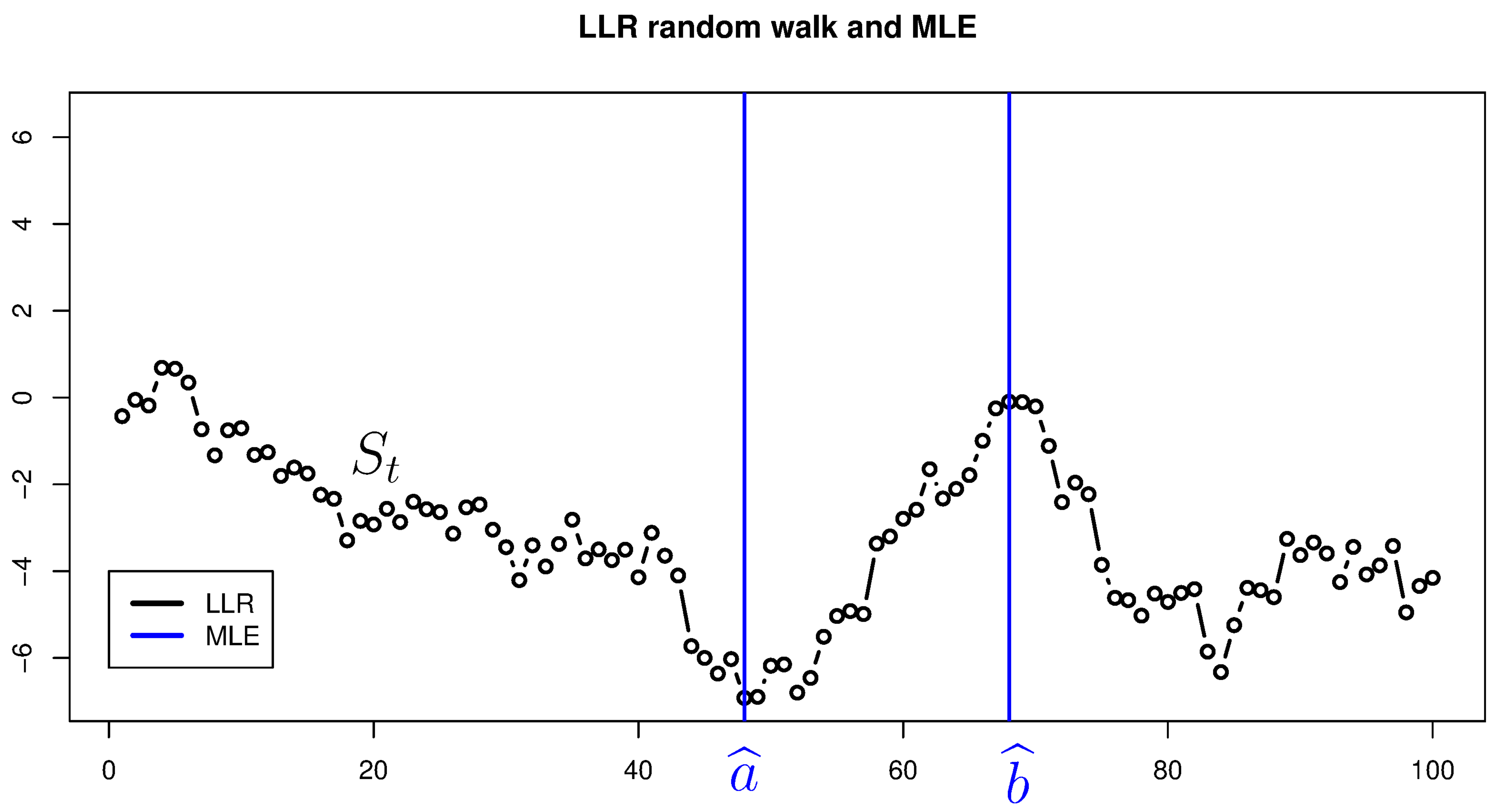
For $\theta = (a, b)$, the log-likelihood is

$$L(X;\theta) = \sum_{i=1}^{a} \log f(X_i) + \sum_{i=a+1}^{b} \log g(X_i) + \sum_{i=b+1}^{n} \log f(X_i)$$
$$\cong \sum_{i=a+1}^{b} \log \frac{g(X_i)}{f(X_i)}$$

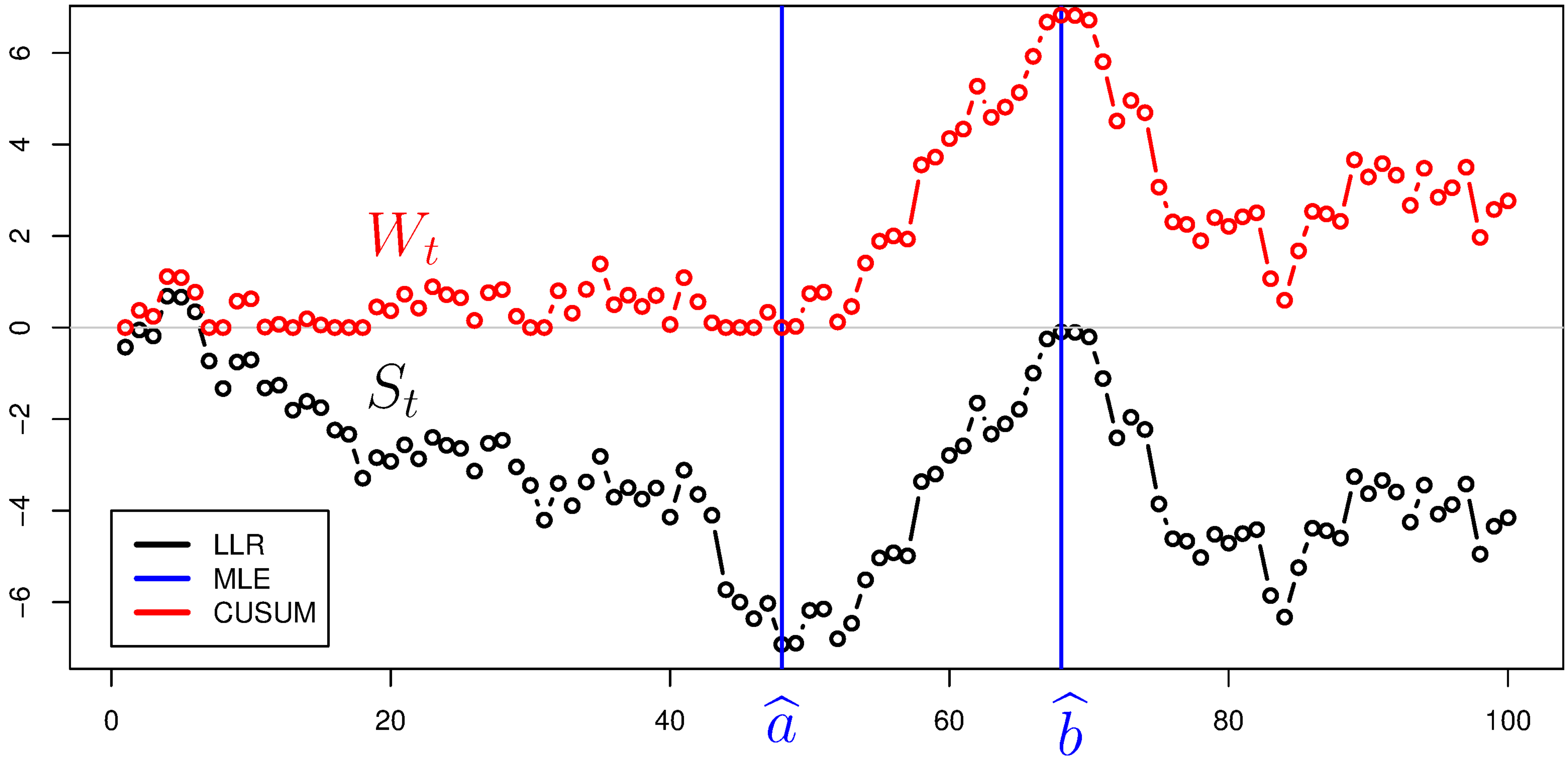
Hence, the MLE of θ is $\widehat{\theta} = (\widehat{a}, \widehat{b}) = \arg\max_{a \le b}(S_b - S_a)$, where $S_t = \sum_{i=1}^{t} \log \frac{g}{f}(X_i)$

In terms of the CUSUM process $W_t = S_t - \min_{i \le t} S_i$ with $\text{Ker}(W) = \{t : W_t = 0\}$, the MLE is

$$\widehat{b} = \arg\max W_t, \quad \widehat{a} = \max\left\{\text{Ker}(W) \cap [0, \widehat{b})\right\}$$


LLR random walk and MLE
6
4
2
0
−2
−4
−6
0
20
40
60
80
100
$\widehat{a}$
$\widehat{b}$
S_t
LLR
MLE

LLR random walk, CUSUM process, and MLE

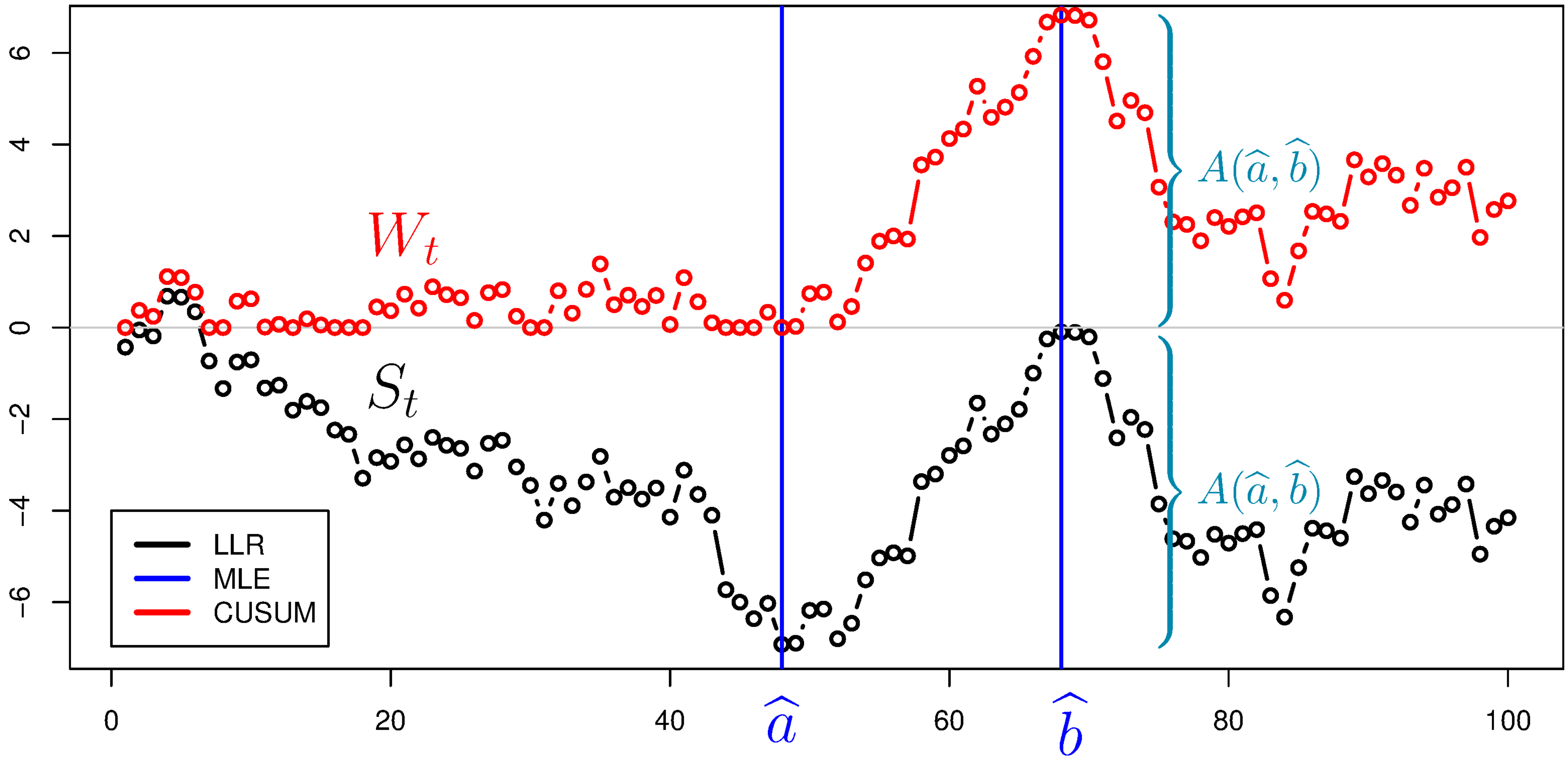


Control of the false alarm rate

Decide between $K = 0$ (no change) and $K = 1$ (one change)?

- Testing
$$H_0 : \begin{array}{c} K = 0 \\ \text{all } \mathbb{X}_{0:n} \sim F \end{array} \quad \text{vs} \quad H_1 : \begin{array}{c} K = 1 \\ \mathbb{X}_{a:b} \sim G \text{ for some } a, b \end{array}$$
where $\mathbb{X}_{k:m} := (X_{k+1}, \ldots, X_m)$.
- The log-likelihood ratio test statistic is
$$\Lambda = \log \frac{\max_{a<b} f(\mathbb{X}_{0:a}) g(\mathbb{X}_{a,b}) f(\mathbb{X}_{b:n})}{f(\mathbb{X}_{0:n})} = A(\hat{a}, \hat{b})$$
- Reject H_0 in favor of H_1 if $\Lambda \geq h$ for some *threshold* h, which controls the balance between the sensitivity and the rate of false alarms.

LLR random walk, CUSUM process, and MLE



Control of the false alarm rate

- By the Doob's maximal inequality,

$$\mathbb{P}_{H_0}\{\max_{0\le t\le n} W_t \ge h\} = \mathbb{P}_F\{\max_{0\le t\le n} e^{W_t} \ge e^h\} \le e^{-h}\mathbb{E}_F(e^{W_n})$$

- RESULT: the threshold $h = -\log \frac{\alpha}{\mathbb{E}_F(e^{W_n})}$ for the increment

$$A(\widehat{a},\widehat{b}) = S_{\widehat{b}} - S_{\widehat{a}} = \max_{0\le t\le n} W_t$$

controls the false alarm rate at level α,

$$\mathbb{P}\{\text{false alarm}\} = \mathbb{P}\{\text{Type I error}\} = \mathbb{P}_F\{A(\widehat{a},\widehat{b}) \ge h\} \le \alpha.$$

- Report a change-point if $A(\widehat{a},\widehat{b}) \ge h$.

Multiple transient changes

Oscillation between distributions F and G, switching at unknown times,

$$\begin{cases} \mathbb{X}_{0:a_1} \sim & X_1, \ldots, X_{a_1} & \sim F \\ \mathbb{X}_{a_1:b_1} \sim & X_{a_1+1}, \ldots, X_{b_1} & \sim G \\ \mathbb{X}_{b_1:a_2} \sim & X_{b_1+1}, \ldots, X_{a_2} & \sim F \\ \mathbb{X}_{a_2:b_2} \sim & X_{a_2+1}, \ldots, X_{b_2} & \sim G \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & & \\ \mathbb{X}_{b_K:n} \sim & X_{b_K+1}, \ldots, X_n & \sim F \end{cases}$$

Estimate a $(2K)$-dimensional change-point parameter

$$\boldsymbol{\theta} = \{a_k, b_k\}_{k=1}^{k=K} = \{a_1, b_1; \ldots; a_K, b_K\}$$

by a $2\widehat{K}$-dimensional estimator $\widehat{\boldsymbol{\theta}} = \left\{\widehat{a}_k, \widehat{b}_k\right\}_{k=1}^{k=\widehat{K}}$.

Distinguish two cases

Known number of changes: maximum likelihood estimation

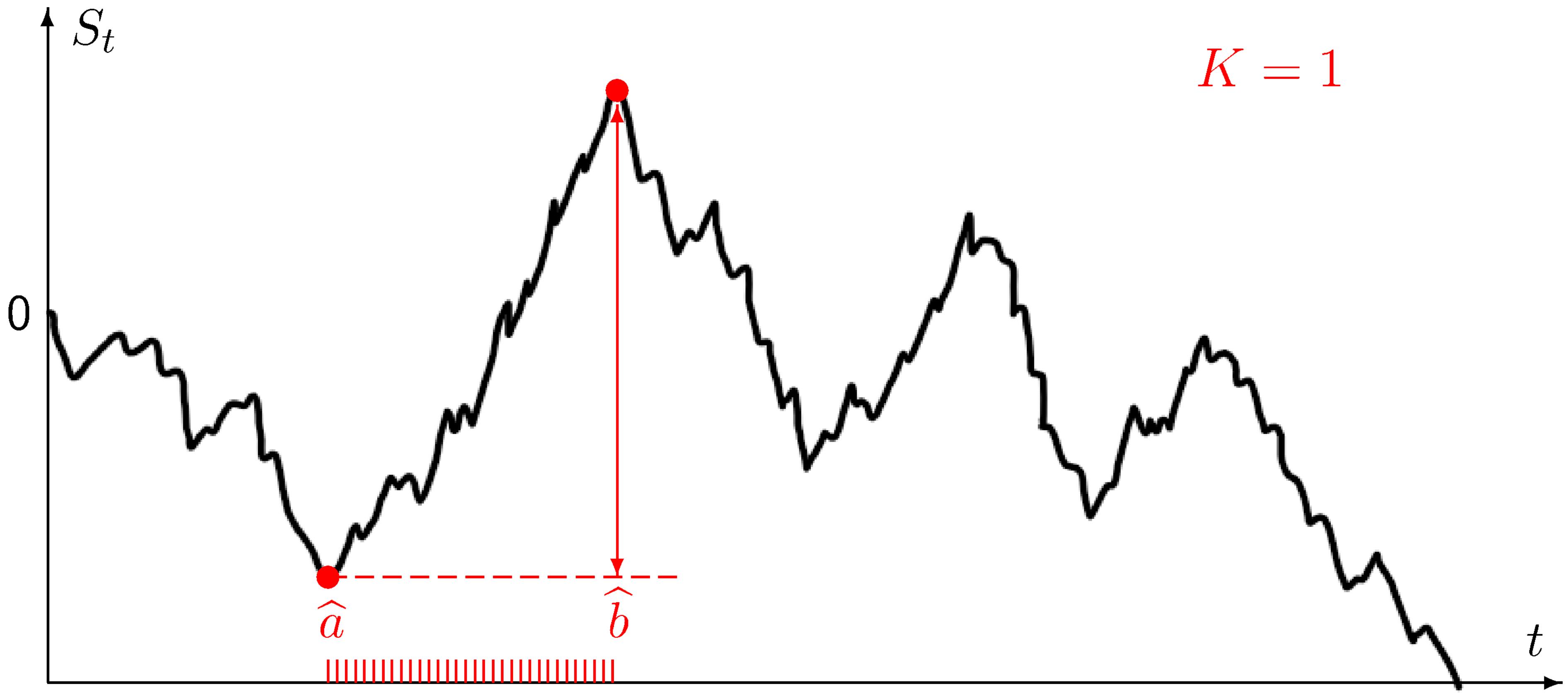
For $\boldsymbol{\theta} = \{(a_k, b_k),\ k = 1, \ldots, K\}$, K known, the log-likelihood is

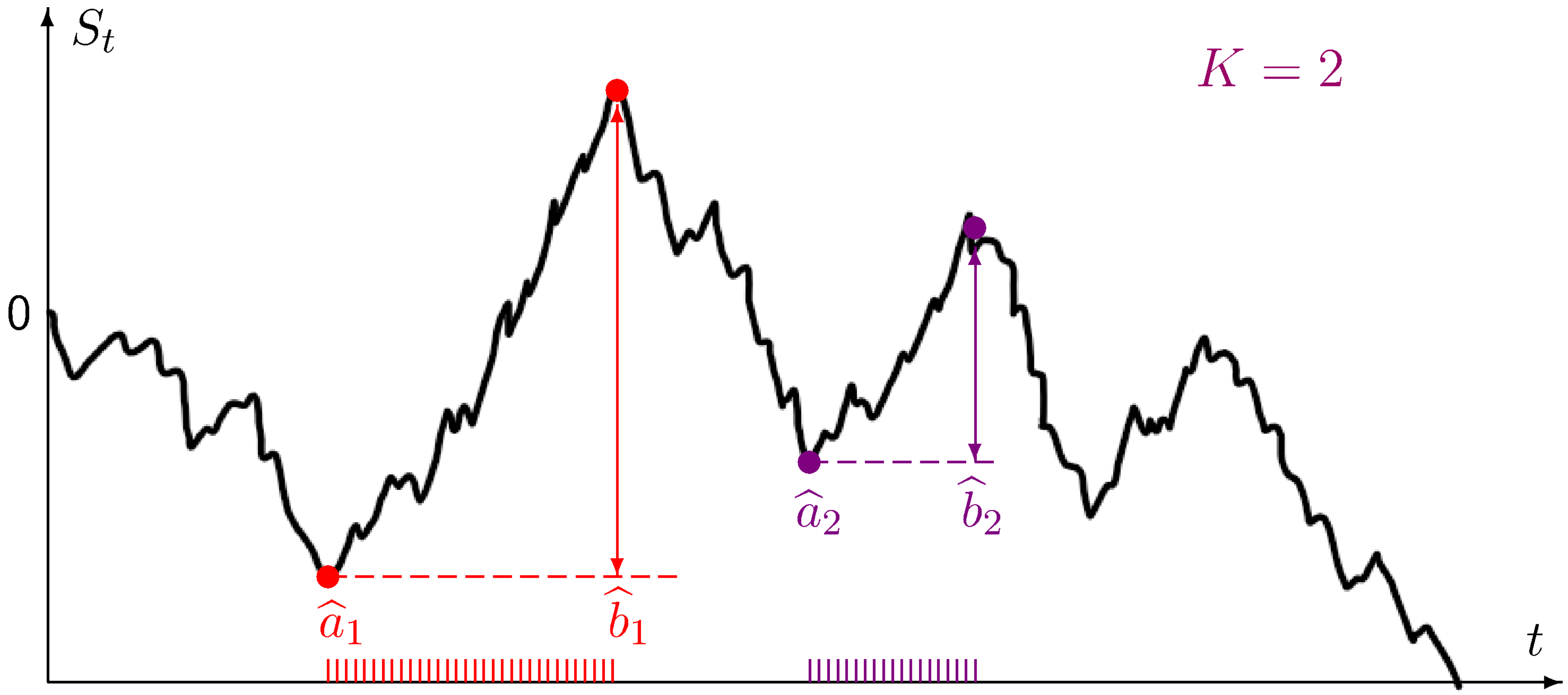
$$L(X;\theta) = \sum_{k=1}^{K} \sum_{i=a_k+1}^{b_k} \log \frac{g(X_i)}{f(X_i)}$$

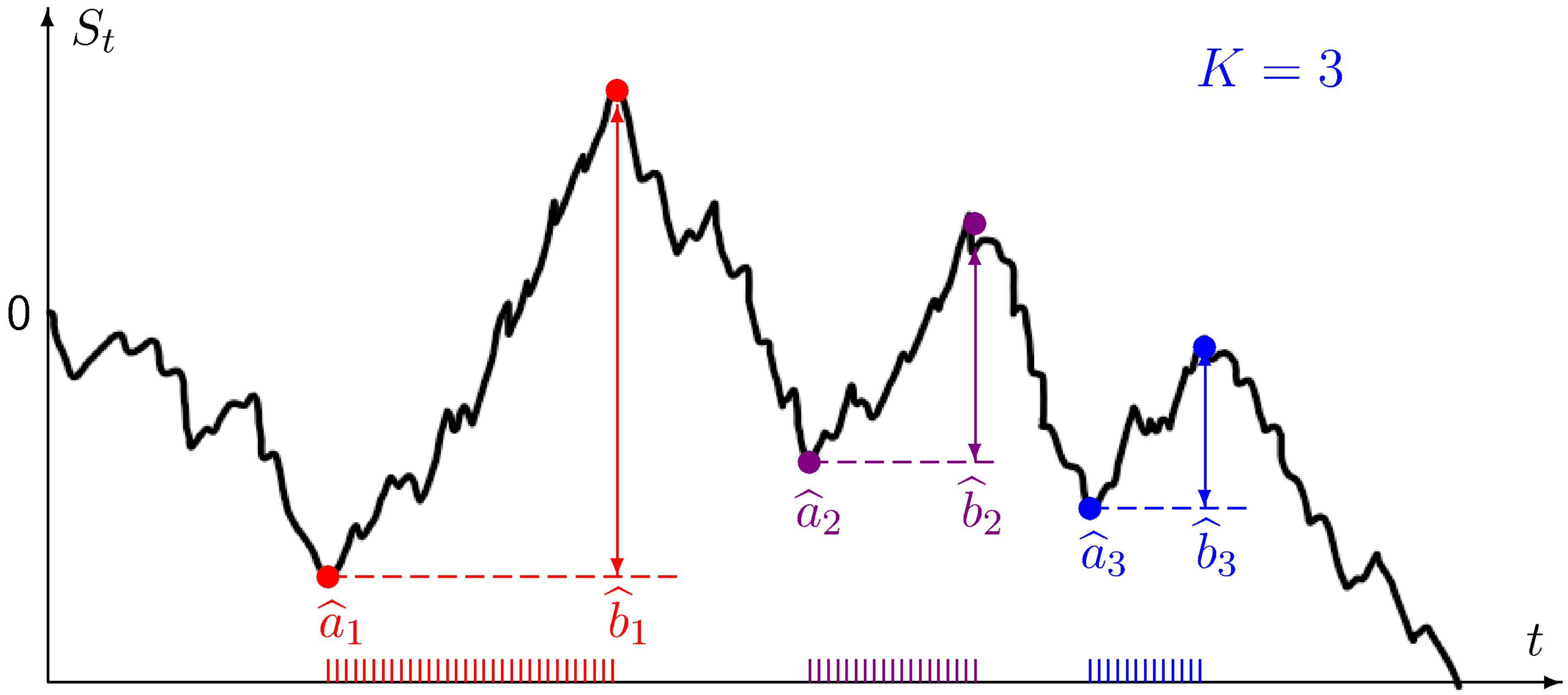
Hence, the MLE of θ is

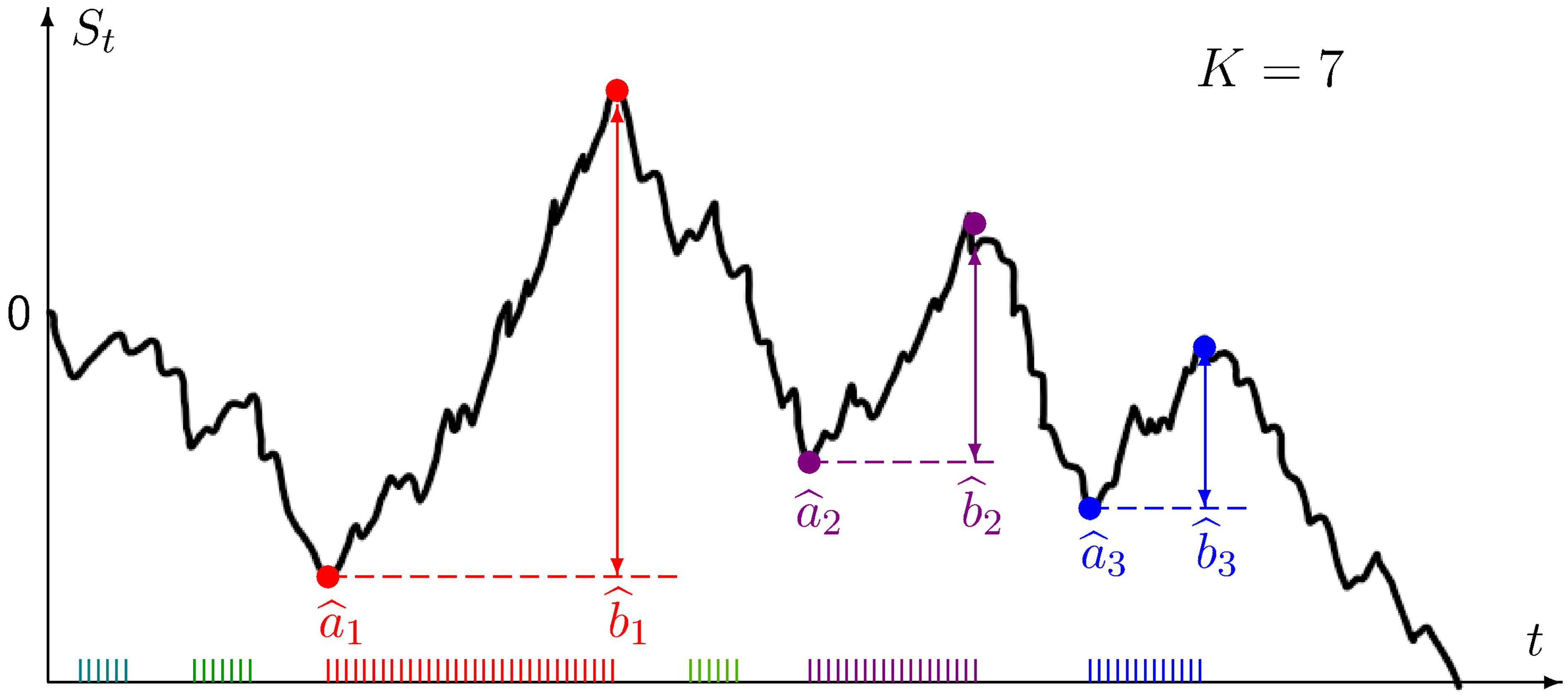
$$\widehat{\boldsymbol{\theta}} = \left\{\widehat{\theta}_k\right\}_{k=1}^{k=K} = \left\{(\widehat{a}_k, \widehat{b}_k)\right\}_{k=1}^{k=K} = \underset{a_1<b_1<\ldots<a_k<b_K}{\arg\max} \sum_{k=1}^{K} (S_{b_k} - S_{a_k}),$$

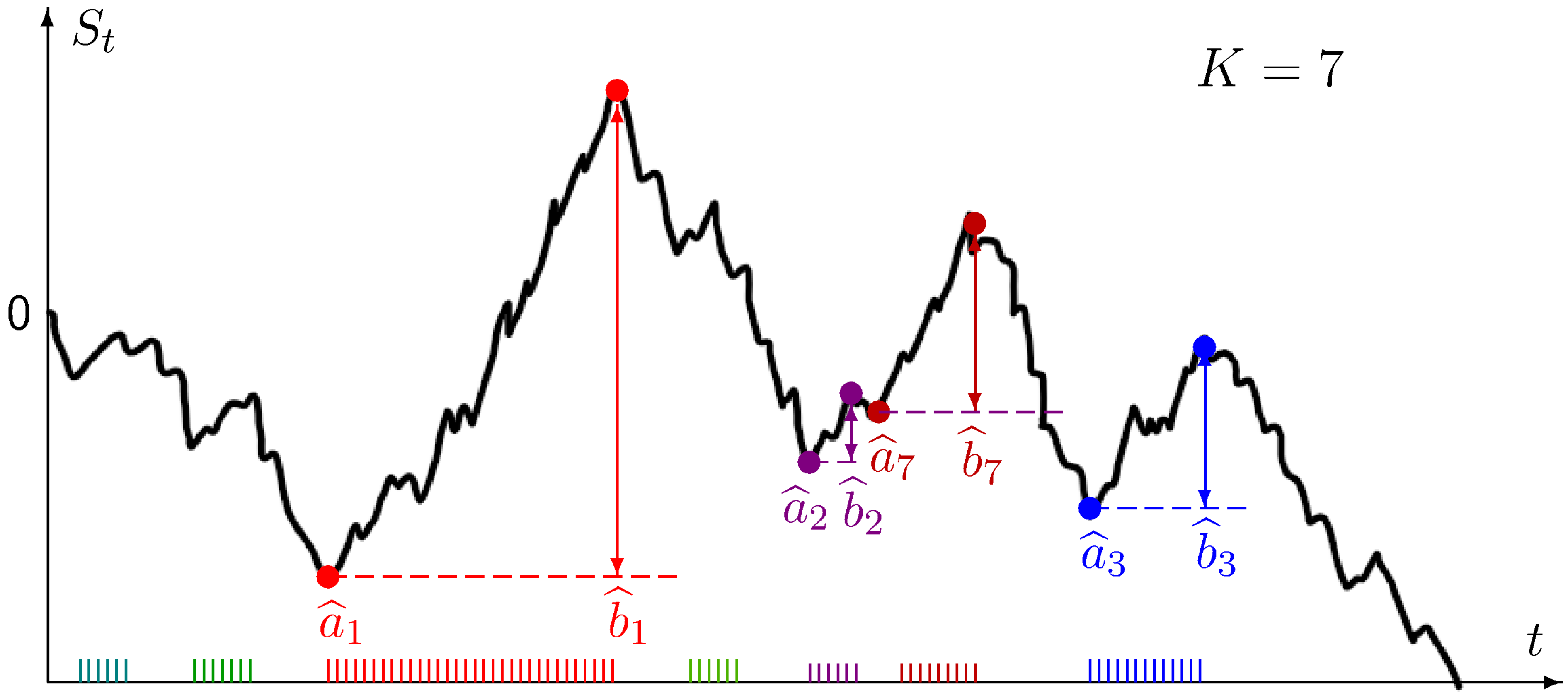
which are K intervals of the biggest growth of S_t.

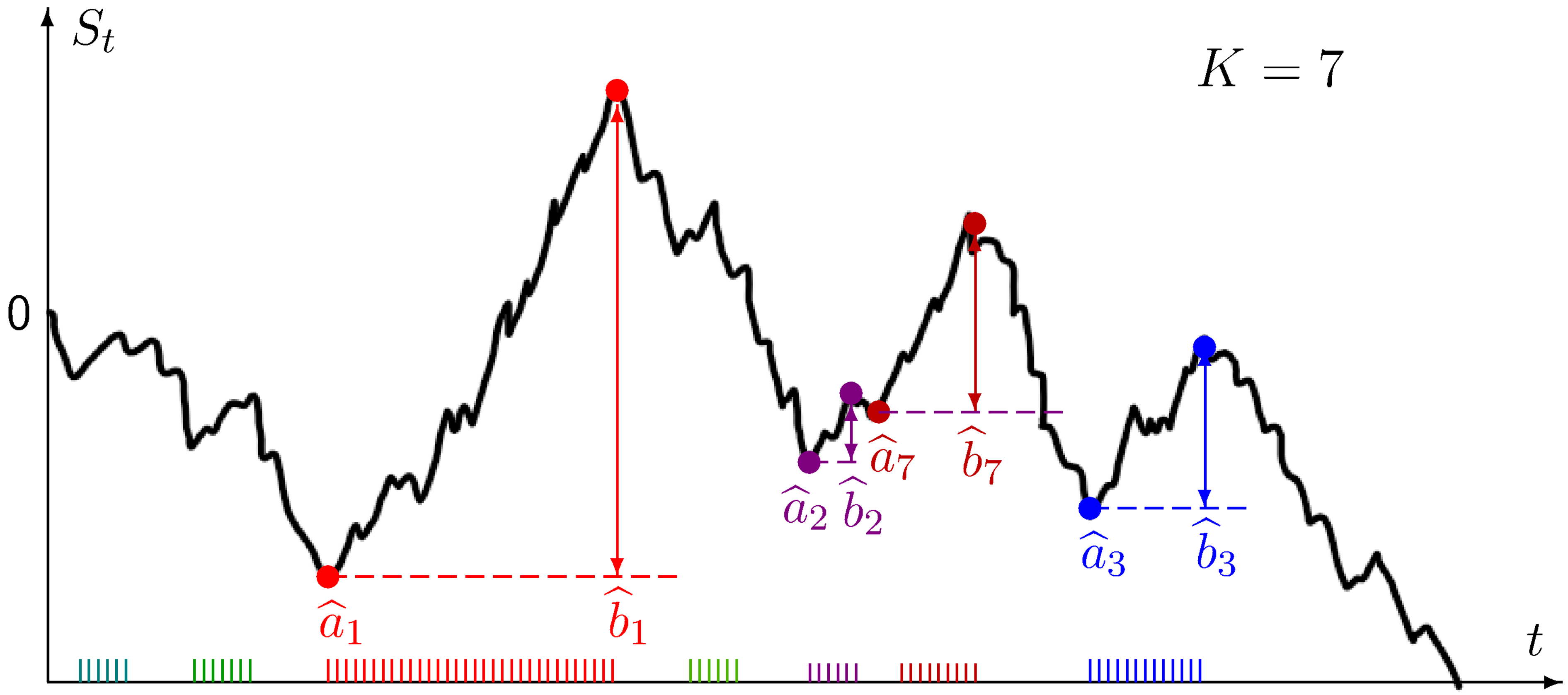

S_t
$K = 1$
0
$\widehat{a}$
$\widehat{b}$
t


S_t
$K = 2$
0
$\widehat{a}_1$
$\widehat{b}_1$
$\widehat{a}_2$
$\widehat{b}_2$
t


S_t
$K = 3$
0
$\widehat{a}_1$
$\widehat{b}_1$
$\widehat{a}_2$
$\widehat{b}_2$
$\widehat{a}_3$
$\widehat{b}_3$
t


S_t
$K = 7$
0
$\widehat{a}_1$
$\widehat{b}_1$
$\widehat{a}_2$
$\widehat{b}_2$
$\widehat{a}_3$
$\widehat{b}_3$
t


S_t
$K = 7$
0
$\widehat{a}_1$
$\widehat{b}_1$
$\widehat{a}_2$
$\widehat{b}_2$
$\widehat{a}_7$
$\widehat{b}_7$
$\widehat{a}_3$
$\widehat{b}_3$
t



Some detected change-points may be false alarms.
Or false adjustments.

Controlling familywise error rates

The *familywise false alarm rate* will be defined as the probability of at least one false alarm,

$$\mathsf{FAR} = \mathbb{P}\left\{\cup_k \left([\widehat{a}_k, \widehat{b}_k] \cap (\cup_j [a_j, b_j]) = \varnothing\right)\right\}.$$

A *false readjustment* occurs when the estimated "in control" interval $[\widehat{b}_k, \widehat{a}_{k+1}]$ does not contain any in-control observations,

$$\mathsf{FRR} = \mathbb{P}\left\{\cup_k \left([\widehat{b}_k, \widehat{a}_{k+1}] \cap (\cup_j [b_j, a_{j+1}]) = \varnothing\right)\right\}.$$

Control the familywise rates at levels α and β, respectively,

$$\mathsf{FAR} \leq \alpha, \quad \text{and} \quad \mathsf{FRR} \leq \beta.$$

Detection scheme with an unknown number of changes

Simultaneous detection of disorders and adjustments

- $W_{\tau,t}$ = CUSUM based on $S_{\tau+t}$, renewed at τ
  $\widetilde{W}_{\tau,t}$ = CUSUM based on $(-S_{\tau+t})$, renewed at τ
- Detection times...

$$\tau_0 = 0 \ ,$$

$$\tau_k = \inf\{t > \tau_{k-1} : W_{\tau_{k-1}, t-\tau_{k-1}} \geq h_\alpha\} \wedge n \ , \text{ for odd } k,$$

$$h_\alpha = -\log(\alpha \mathbb{E}_F^{-1}(e^{W_n}));$$

$$\tau_k = \inf\{t > \tau_{k-1} : \widetilde{W}_{\tau_{k-1}, t-\tau_{k-1}} \geq \widetilde{h}_\beta\} \wedge n \ , \text{ for even } k,$$

$$\widetilde{h}_\beta = -\log(\beta \mathbb{E}_G^{-1}(e^{\widetilde{W}_n})).$$

- Restarted and grounded CUSUM process $W_0^{(h)} = 0$,

$$W_t^{(h)} = \begin{cases} W_{\tau_{k-1}, t-\tau_{k-1}} & \text{if } k \le 2K \text{ is odd,} \\ \widetilde{W}_{\tau_{k-1}, t-\tau_{k-1}} & \text{if } k \le 2K \text{ is even} \end{cases} \quad \text{for } t \in (\tau_{k-1}, \tau_k]$$

For the last stopping time τ^* before n,

$$W_t^{(h)} = \begin{cases} W_{\tau^*, t-\tau^*} & \text{if } \tau^* \text{ is even,} \\ \widetilde{W}_{\tau^*, t-\tau^*} & \text{if } \tau^* \text{ is odd} \end{cases} \quad \text{for } t \in (\tau^*, n]$$

- $\nu_k = \sup\left(\text{Ker}(W_t^{(h)}) \cap [0, \tau_k)\right)$ = last zero of $W_t^{(h)}$ before τ_k
- $\theta_k = (a_k, b_k)$ is estimated by (ν_{2k-1}, ν_{2k}) for $k = 1, \ldots, 2K$.

RESULT: *The algorithm detects disorders with familywise FAR $\le \alpha$ and readjustments with familywise FRR $\le \beta$ for $\forall K$.*

No Bonferroni or Holm type correction is needed!

Power Analysis

Considered scenarios – a change from the Standard Normal base distribution to:

- the Normal distribution with mean μ and unit variance (change in the mean);
- the Normal distribution with mean 0 and variance σ^2 (change in the variance);
- the Laplace distribution with mean 0 and variance 1 (change neither in the mean nor in the variance).

Results suggest to estimate threshold h as $\widehat{h}_\alpha = \log\left(\frac{\widehat{\mathbb{E}_F(e^{W_n})}}{\alpha}\right)$.

Very high variance. Alternatively, $\widehat{h}_\alpha =$ the 95-th empirical percentile of the distribution of $\Lambda = \max_{0 \le t \le n} (W_t) \;\Rightarrow\;$ FAR $= \alpha$.

Location changes

Disturbed distribution		Threshold	Detection probability	Accuracy of Estimation			
μ	σ	h	$\mathbb{P}_{a,b}(\Lambda \geq h)$	$\mathbb{E}(\widehat{a})$	Std$(\widehat{a})$	$\mathbb{E}(\widehat{b})$	Std$(\widehat{b})$
0.05	1	2.65	0.109	351.2	238.0	694.5	238.5
0.10	1	4.16	0.212	404.0	208.9	690.9	209.9
0.15	1	5.02	0.394	444.5	171.2	691.0	174.8
0.20	1	5.60	0.618	472.1	127.8	695.3	133.8
0.25	1	6.03	0.804	487.5	92.0	697.7	97.6
0.30	1	6.35	0.915	495.6	64.6	699.6	68.0
0.35	1	6.62	0.969	498.2	45.8	700.4	46.9
0.40	1	6.84	0.991	499.8	32.7	700.4	32.9
0.60	1	7.45	1	500.0	14.1	700.1	14.1
0.80	1	7.80	1	499.9	7.9	700.0	7.9
1.00	1	8.00	1	500.0	5.1	700.0	5.0

$N = 50,000$ MC runs; h_α is estimated from $N_h = 200,000$ MC runs.

Scale changes

Disturbed distribution		Threshold	Detection probability	Accuracy of Estimation			
μ	σ	h	$\mathbb{P}_{a,b}(\Lambda \geq h)$	$\mathbb{E}(\widehat{a})$	Std$(\widehat{a})$	$\mathbb{E}(\widehat{b})$	Std$(\widehat{b})$
0	0.50	8.20	1	498.4	5.4	701.6	5.5
0	0.75	6.92	0.989	494.7	32.6	704.8	34.1
0	0.90	5.01	0.355	430.2	171.8	701.2	175.4
0	0.95	3.41	0.137	361.2	222.6	703.6	223.9
0	1.05	3.33	0.146	385.7	230.7	678.7	232.4
0	1.10	4.73	0.362	447.8	181.3	681.2	185.5
0	1.25	6.22	0.950	502.9	55.1	693.9	58.2
0	1.50	6.95	1	503.4	15.7	696.9	15.8
0	2.00	7.25	1	501.6	5.6	698.4	5.6
Laplace$(0,\ 1/\sqrt{2})$		6.4	0.975	499.9	45.7	698.1	47.4

Change duration

Change		Duration of the transient period Δ									
	μ	50	100	150	200	250	300	350	400	450	500
From N(0,1) to N(μ,1)	0.1	0.068	0.102	0.152	0.213	0.281	0.346	0.394	0.465	0.532	0.571
	0.2	0.113	0.259	0.457	0.611	0.740	0.828	0.889	0.924	0.949	0.968
	0.3	0.216	0.567	0.808	0.916	0.968	0.983	0.992	0.997	0.999	1
	0.4	0.421	0.839	0.964	0.992	0.998	0.999	1	1	1	1
	0.5	0.659	0.957	0.995	1	1	1	1	1	1	1
	0.6	0.836	0.993	1	1	1	1	1	1	1	1
	0.7	0.937	0.999	1	1	1	1	1	1	1	1
	0.8	0.978	1	1	1	1	1	1	1	1	1
	0.9	0.994	1	1	1	1	1	1	1	1	1
	1.0	0.998	1	1	1	1	1	1	1	1	1

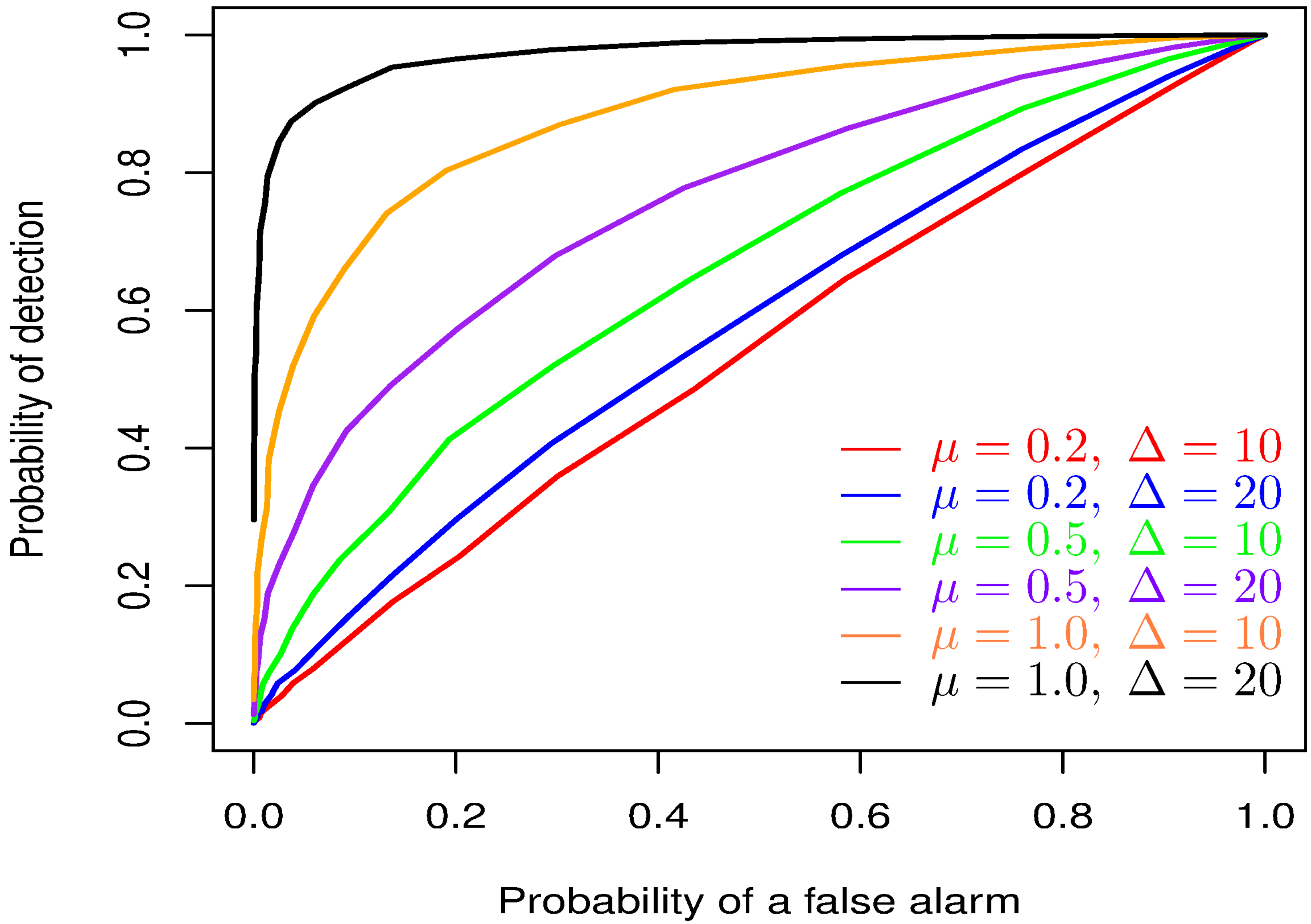
Detection probabilities as functions of magnitude and duration of a transient change

Change duration

Change		Duration of the transient period Δ									
	σ	50	100	150	200	250	300	350	400	450	500
From N(0,1) to N(0,σ)	0.50	0.993	1	1	1	1	1	1	1	1	1
	0.75	0.305	0.797	0.952	0.989	0.997	0.999	1	1	1	1
	0.90	0.082	0.144	0.241	0.361	0.465	0.576	0.664	0.738	0.796	0.837
	0.95	0.062	0.084	0.108	0.142	0.171	0.211	0.254	0.290	0.330	0.357
	1.05	0.064	0.086	0.115	0.143	0.181	0.212	0.256	0.289	0.325	0.366
	1.10	0.086	0.162	0.256	0.365	0.462	0.559	0.637	0.703	0.756	0.801
	1.25	0.325	0.693	0.871	0.950	0.979	0.991	0.997	0.999	1	1
	1.50	0.865	0.992	0.999	1	1	1	1	1	1	1
	2.00	1	1	1	1	1	1	1	1	1	1
Normal to Laplace		0.323	0.731	0.905	0.973	0.991	0.997	0.999	1	1	1

Detection probabilities as functions of magnitude and duration of a transient change

ROC curves for detecting a change in the mean



Detection of multiple changes

Shift μ	Threshold h	FAR	FRR	Probability of detecting exactly k intervals				
				$k=0$	$k=1$	$k=2$	$k=3$	$k>4$
0.1	4.14	0	0	0.77	0.23	0	0	0
0.2	5.60	0.001	0	0.44	0.49	0.07	0	0
0.3	6.36	0.002	0	0.09	0.37	0.41	0.13	0
0.4	6.84	0.007	0	0	0.07	0.36	0.56	0
0.5	7.18	0.013	0	0	0	0.11	0.87	0.01
0.6	7.44	0.020	0.002	0	0	0.02	0.96	0.02
0.7	7.64	0.024	0.003	0	0	0	0.97	0.03
0.8	7.79	0.028	0.006	0	0	0	0.97	0.03
0.9	7.94	0.027	0.008	0	0	0	0.97	0.03
1.0	8.01	0.030	0.010	0	0	0	0.96	0.04

Detection of multiple changes

Shift	Means and standard deviations of change-point estimators											
μ	$\mathbb{E}(\widehat{a}_1)$	$\mathbb{E}(\widehat{b}_1)$	$\mathbb{E}(\widehat{a}_2)$	$\mathbb{E}(\widehat{b}_2)$	$\mathbb{E}(\widehat{a}_3)$	$\mathbb{E}(\widehat{b}_3)$	$\sigma(\widehat{a}_1)$	$\sigma(\widehat{b}_1)$	$\sigma(\widehat{a}_2)$	$\sigma(\widehat{b}_2)$	$\sigma(\widehat{a}_3)$	$\sigma(\widehat{b}_3)$
0.2	114.0	272.2	426.4	551.0	734.4	870.8	43.9	33.5	35.7	20.0	39.9	21.1
0.3	135.3	260.0	441.4	560.0	739.9	853.9	36.1	30.5	30.9	30.4	30.9	25.6
0.4	145.4	253.9	446.1	553.2	745.7	852.7	26.0	24.6	26.9	25.2	25.4	22.5
0.5	148.8	251.2	448.8	551.2	748.6	850.8	18.8	18.6	20.0	19.2	18.9	18.3
0.6	149.7	250.4	449.8	550.3	749.8	850.3	13.7	13.7	13.9	13.7	13.4	13.3
0.7	149.9	250.1	449.9	550.1	749.9	850.0	10.4	10.2	10.1	10.4	10.1	10.3
0.8	150.0	250.1	450.1	550.1	749.9	850.0	8.2	7.6	7.9	8.0	7.9	7.8
0.9	150.0	250.0	450.1	550.0	750.1	850.1	6.2	6.2	6.4	6.3	6.3	6.3
1.0	149.9	250.0	450.0	550.0	750.0	850.0	5.0	5.0	5.2	4.9	5.1	5.0

Actual intervals of change:

$$[a_1, b_1] = [150, 250], \quad [a_2, b_2] = [450, 550], \quad \text{and} \quad [a_3, b_3] = [750, 850].$$

Threshold simplified

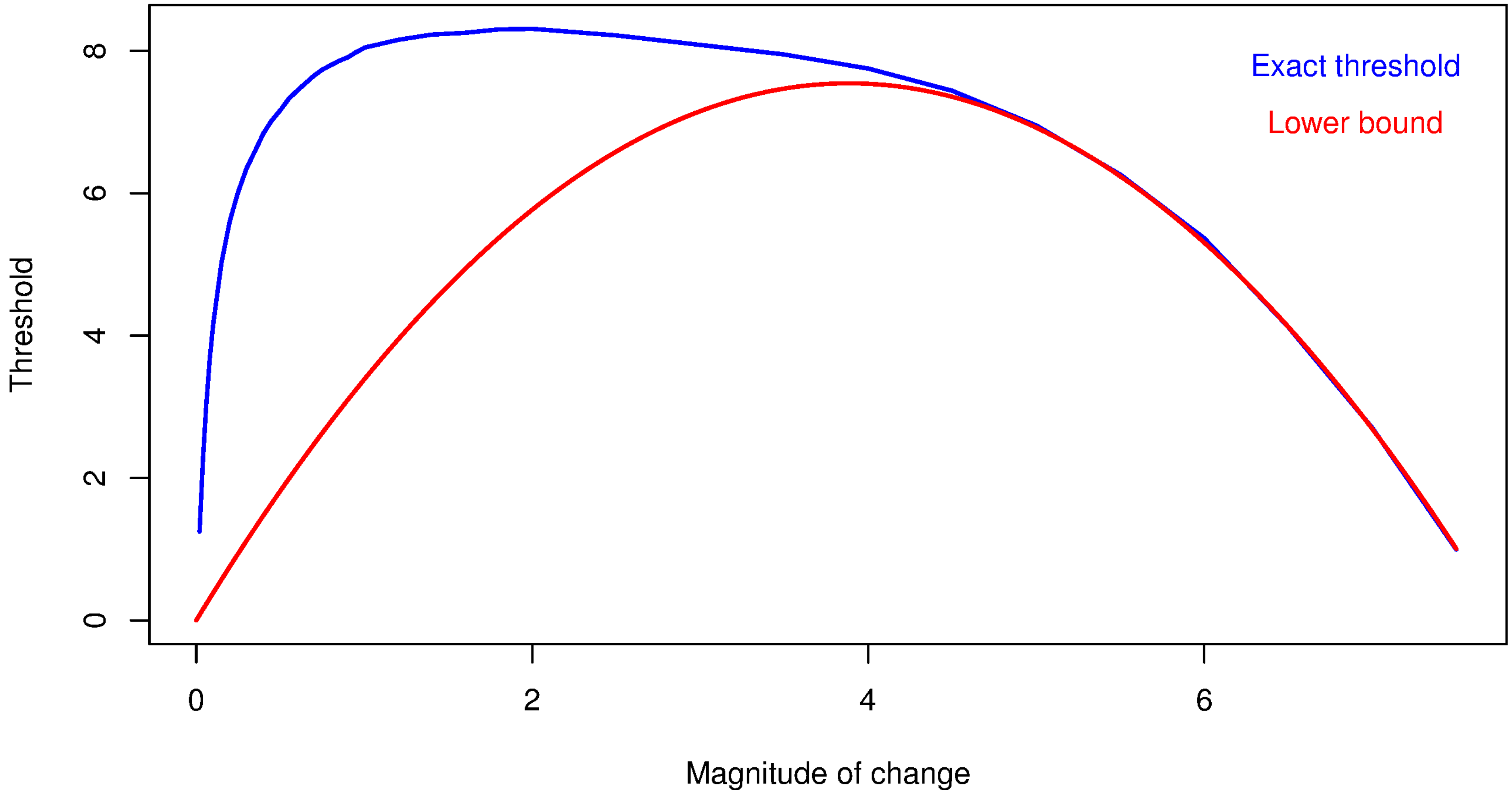
Based on one increment:

$$\text{FAR} \geq \mathbb{P}\left\{\bigcup_{i=1}^{n} z_i \geq h\right\} = 1 - \Phi^n\left(\frac{h + \mu^2/2}{\mu}\right),$$

for a change from N(0,1) to N(μ,1), with $z_i \sim N(-\mu^2/2, \mu^2)$.

$$\text{FAR} \leq \alpha \;\Rightarrow h \geq \mu\Phi^{-1}\left((1-\alpha)^{1/n}\right) - \frac{\mu^2}{2}$$

– a lower bound, appears accurate for large changes.



Conclusion: in a problem of detecting a change between substantially different distributions, a false alarm is likely to be cause by one extreme observation.

Conclusions:

- Temporary changes in the distribution of data
- Even small transient changes can be detected, if they last long
- Detection power reduces with smaller magnitudes, shorter durations
- Sensitivity depends on the selected threshold, which can be chosen to satisfy familywise error rates

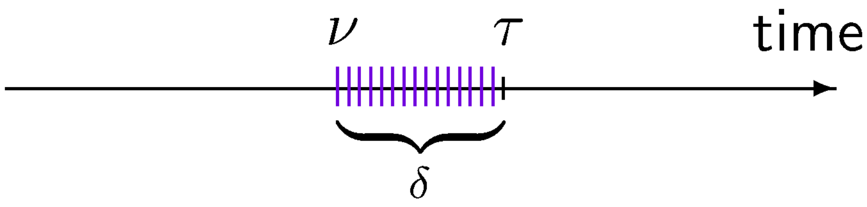
Extensions:

- Nuisance parameters
- Correlated time series data
- Multiple dimensions
- Bayesian approach
- Applications

Bayes sequential problem

$$\begin{cases} \nu = \text{change-point time} \sim \text{prior } \pi_k \\ \delta = \text{disturbance time} \sim \text{prior } p_d \end{cases}$$

$\delta = b_i - a_i$ is independent of $\nu = a_i$



- Detect ν “quickly” after the change-point
- Detect $\tau = \nu + \delta$ “quickly” after adjustment

Look for stopping times $T^{(\nu)}$ and $T^{(\tau)}$.

At every time t, test

$$H_0^{(\nu)} : \nu > t \quad \text{vs.} \quad H_1^{(\nu)} : \nu \le t$$
$$H_0^{(\tau)} : \tau > t \quad \text{vs.} \quad H_1^{(\tau)} : \tau \le t$$

Bayesian problem

After $\mathbb{X}_{0:t} = (X_1, \ldots, X_t)$, the posterior survival functions are

$$\begin{cases} S_X^{(\nu)}(t) = \mathbb{P}(H_0^{(\nu)} \mid \mathbb{X}_{0:t}) = \dfrac{S^{(\nu)}(t)}{S^{(\nu)}(t) + \sum_{k=1}^{t} \pi_k \left(\sum_{d=1}^{t-k-1} p_d \rho_{k:k+d} + \sum_{d=t-k}^{\infty} p_d \rho_{k:t} \right)} \\ S_X^{(\tau)}(t) = \mathbb{P}(H_0^{(\tau)} \mid \mathbb{X}_{0:t}) = \dfrac{S^{(\nu)}(t) + \sum_{k=0}^{t-1} \pi_k \sum_{d=1}^{t-k-1} p_d \rho_{k:k+d}}{S^{(\nu)}(t) + \sum_{k=1}^{t} \pi_k \left(\sum_{d=1}^{t-k-1} p_d \rho_{k:k+d} + \sum_{d=t-k}^{\infty} p_d \rho_{k:t} \right)} \end{cases}$$

where

$S^{(\nu)}(t) = \mathbb{P}(\nu > t)$ and $S^{(\tau)}(t) = \mathbb{P}(\tau > t)$ are prior survival functions;

$\rho_{k:t} = \rho_{k+1} \cdot \ldots \cdot \rho_t$ are likelihood ratios; $\rho_i = g(X_i)/f(X_i)$.

Bayesian problems

Reject $H_0^{(\nu)} : \nu > t$ in favor of $H_1^{(\nu)} : \nu \leq t$ when $S_X^{(\nu)} < \alpha$

Reject $H_0^{(\tau)} : \tau > t$ in favor of $H_1^{(\tau)} : \tau \leq t$ when $S_X^{(\tau)} < \beta$

$\Rightarrow$ stopping rules $T^{(\nu)}$, $T^{(\tau)}$.

Risk functions:

$$\begin{aligned} R_\nu(T,\nu) &= \lambda_\nu \mathbb{E}(T-\nu)^+ - \log^{-1} \mathbb{P}(T<\nu) \\ R_\tau(T,\tau) &= \lambda_\tau \mathbb{E}(T-\tau)^+ - \log^{-1} \mathbb{P}(T<\tau) \\ R_\nu(T,\nu,\tau) &= R_\nu(T,\nu) + cR_\tau(T,\tau) \end{aligned}$$

Asymptotically pointwise optimal rules

For the risk

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

a stopping rule T is APO if

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

a.s. for any stopping rule S.

Bickel, Yahav 1967, 1968
Ghosh, Mukhopadhyay, Sen 1997 [sec. 5.4]

Asymptotically pointwise optimal rules

Theorem (Bickel, Yahav)

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

$$T = \inf\left\{ n \;\middle|\; \frac{\mathbb{E}\left\{L(n,\theta,\delta)\mid \mathbb{X}_{1:n}\right\}}{n} \le \frac{c}{\beta} \right\}$$

is APO.

Asymptotically pointwise optimal rules

In change-point problems:

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

Let a stopping time T be APO if

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

In the single change-point problem, there is a closed-form formula

M.B. 2014

For transient changes, it depends on the rate of $\delta = \delta(\lambda)$.

Theorem

Let $r_t = -\log^{-1} S_X(t)$, "posterior expected loss",
$\rho_1 = \frac{g(X_1)}{f(X_1)}$, $\rho_{t+1} = \frac{g(X_{t+1}|\mathbb{X}_{1:t})}{f(X_{t+1}|\mathbb{X}_{1:t})}$, log-likelihood ratios.

Assume a strong law of large numbers

$$t^{-1}\sum_{k=1}^{t} \log \rho_k \to K > 0, \text{ as } t \to \infty, \; G\text{-a.s.}$$

This condition holds for the i.i.d. case with $K = K(G, F)$, L_p-mixingales, L_p-NED (near-epoch-dependent in L_p norm) sequences, invertible ARMA time series, etc.

For the prior distributions of ν and δ, assume that

$-t^{-\beta} \log S^{\nu}(t) \to L \in [0, \infty)$, $\beta \in [1, \infty)$,

where $S^{\nu}(t)$ is the prior survival function of ν;

and $\dfrac{\mathbb{E}(\delta \wedge t)}{t} \to C \in [0, \infty)$

Then there exists an a.s. limit

(a) If $\beta > 1$ then $\lim\limits_{t\to\infty} (t^{\beta} r_t) = \dfrac{1}{L}$

(b) If $\beta = 1$ and $C = 0$ then $\lim\limits_{t\to\infty} (t r_t) = \dfrac{1}{L}$

(c) If $\beta = 1$ and $C > 0$ then $\lim\limits_{t\to\infty} (t^{\beta} r_t) = \dfrac{1}{L + CK}$

(d) If $\beta = 1$, $C > 0$, and $L = 0$ then $\lim\limits_{t\to\infty} (t r_t) = \dfrac{1}{CK}$

For the prior distributions of ν and δ, assume that

$$-t^{-\beta}\log S^{\nu}(t) \to L \in [0,\infty),\ \beta \in [1,\infty),$$

where $S^{\nu}(t)$ is the prior survival function of ν;

and $\dfrac{\mathbb{E}(\delta \wedge t)}{t} \to C \in [0,\infty)$

Then there exists an a.s. limit

(a) If $\beta > 1$ then $\lim_{t\to\infty}(t^{\beta}r_t) = \dfrac{1}{L}$

(b) If $\beta = 1$ and $C = 0$ then $\lim_{t\to\infty}(tr_t) = \dfrac{1}{L}$

(c) If $\beta = 1$ and $C > 0$ then $\lim_{t\to\infty}(t^{\beta}r_t) = \dfrac{1}{L + CK}$

(d) If $\beta = 1$, $C > 0$, and $L = 0$ then $\lim_{t\to\infty}(tr_t) = \dfrac{1}{CK}$

Detection is dominated by the prior

Detection is dominated by the data

Case (b) $\dfrac{\mathbb{E}(\delta \wedge t)}{t} \to 0$ includes the case of constant or bounded transient period duration δ.

Example: $\delta \sim$ Geometric(p).

Then $\mathbb{E}(\delta \wedge t) = \dfrac{1-(1-p)^t}{p} \to C \in [0,\infty)$ if $\log(1-p) = O(t^{-1})$

Our stopping rule will be of order $T_{APO} \sim \lambda^{-\frac{1}{\beta+1}}$, so in (b-d), we'll need $\log(1-p) = O(\lambda^{\frac{1}{\beta+1}})$, including $o(\lambda^{\frac{1}{\beta+1}})$ in (b), when the data is dominated by the prior.

Theorem (the form of APO stopping rules for ν)
Under condition (a), the stopping rule

$$\tilde{T} = \inf\left\{ t \;\middle|\; -t \log S_X^{(\nu)}(t) \geq \frac{\beta}{\lambda} \right\}$$

is APO.

Under conditions (b-d), the stopping rule

$$\tilde{T} = \inf\left\{ t \;\middle|\; -t \log S_X^{(\nu)}(t) \geq \frac{1}{\lambda} \right\}$$

is APO.

Recursive formula for the posterior survival function

APO stopping rules are in terms of $S_X^{(\nu)}(t) = \mathbb{P}(\nu > t \,|\, \mathbb{X}_{0:t})$

$$= \frac{S^{(\nu)}(t)}{S^{(\nu)}(t) + \sum_{k=1}^{t} \pi_k \left(\sum_{d=1}^{t-k-1} p_d \rho_{k:k+d} + \sum_{d=t-k}^{\infty} p_d \rho_{k:t} \right)} = \frac{S^{(\nu)}(t)}{S^{(\nu)}(t) + U_t + V_t}$$

Recursive computation:

$$U_{t+1} = \left(\frac{U_t}{1 - S^{(\nu)}(t)} + p_{t-k+1} \rho_{k:t+1} \right) \left(1 - S^{(\nu)}(t+1) \right)$$

$$V_{t+1} = V_t \rho_{t+1} \frac{1 - S^{(\nu)}(t+1)}{1 - S^{(\nu)}(t)} \frac{S^{(\delta)}(t-k+1)}{S^{(\delta)}(t-k)}$$

So, the number of operations is $O(n)$.

The case of nuisance parameters

Assume $\theta_1, \ldots, \theta_d \in \Theta$ and $\eta_1, \ldots, \eta_d \in H$, unknown nuisance parameters, so that $F = F_\theta \in \mathcal{F}$ and $G = G_\eta \in \mathcal{G}$;
Priors: $\theta \sim \pi_\theta$, $\eta \sim \pi_\eta$, independently of ν and δ.
Define marginal and conditional densities

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

$$g^*(\mathbb{X}_{1:t}) = \int g_\eta(\mathbb{X}_{1:t} \mid \eta) d\pi_\eta(\eta), \quad g^*(X_{t+1} | \mathbb{X}_{1:t}) = \frac{g^*(\mathbb{X}_{1:t+1})}{g^*(\mathbb{X}_{1:t})}.$$

Now detect transient changes between F^* and G^*

Consider $\rho^*_{1j} = \frac{g^*(X_{1j})}{f^*(X_{1j})}, \quad \rho^*_{t+1} = \frac{g^*(X_{t+1}|\mathbb{X}_{1:t})}{f^*(X_{t+1}|\mathbb{X}_{1:t})}$

Theorem (Bayesian approach, with nuisance parameters)
Assume SLLN

$$t^{-1} \log \rho^*_{tj} \to K > 0, \text{as } t \to \infty, \; G_\eta\text{-a.s., for all } \eta \in H.$$

Then
(1) the stopping rule $\tilde{T}^* = \inf\{t \,|\, -t \log S^*_X(t) \geq \beta/\lambda\}$ is APO under condition (a);

(2) the stopping rule $\tilde{T}^* = \inf\{t \,|\, -t \log S^*_X(t) \geq 1/\lambda\}$ is APO under conditions (b-d),

where $S^*_X(t) = \mathbb{P}\{\min \nu > t|\mathbb{X}_{1:t}\}$ is the *marginal* (parameter-free) posterior survival function of ν.

Baron, M. (2014). Asymptotically pointwise optimal change detection in multiple channels. *Sequential Analysis 33*(4), 440–457.

Baron, M., M. Rosenberg, and N. Sidorenko (2001). Electricity pricing: modeling and prediction with automatic spike detection. *Energy, Power, and Risk Management October 2001*, 36–39.

Bickel, P. J. and J. A. Yahav (1967). Asymptotically pointwise optimal procedures in sequential analysis. In *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, Volume 1, Univ. California Press, Berkeley, pp. 401–413.

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

Daley, T. and A. D. Smith (2014). Modeling genome coverage in single-cell sequencing. *Bioinformatics 30*(22), 3159–3165.

Ghosh, M., N. Mukhopadhyay, and P. K. Sen (1997). *Sequential Estimation*. New York: Wiley.

Guépié, B. K., L. Fillatre, and I. V. Nikiforov (2012). Sequential detection of transient changes. *Sequential Analysis 31*(4), 528–547.

Noonan, J. and A. Zhigljavsky (2020). Power of the mosum test for online detection of a transient change in mean. *Sequential Analysis 39*(2), 269–293.

Repin, V. G. (1991). Detection of a signal with unknown moments of appearance and disappearance. *Problemy Peredachi Informatsii 27*(1), 61–72.

Tartakovskii, A. G. (1987). Optimal detection of random-length signals. *Problemy Peredachi Informatsii 23*(3), 39–47.

Tartakovsky, A. G., N. R. Berenkov, A. E. Kolessa, and I. V. Nikiforov (2021). Optimal sequential detection of signals with unknown appearance and disappearance points in time. *IEEE Transactions on Signal Processing 69*, 2653–2662.

Thank you!