Detection of Multiple Transient Changes

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

[1] American University, Washington DC, USA, baron@american.edu
supported by NSF grant 1737960 and DARPA grant HR0011-18-C-0051

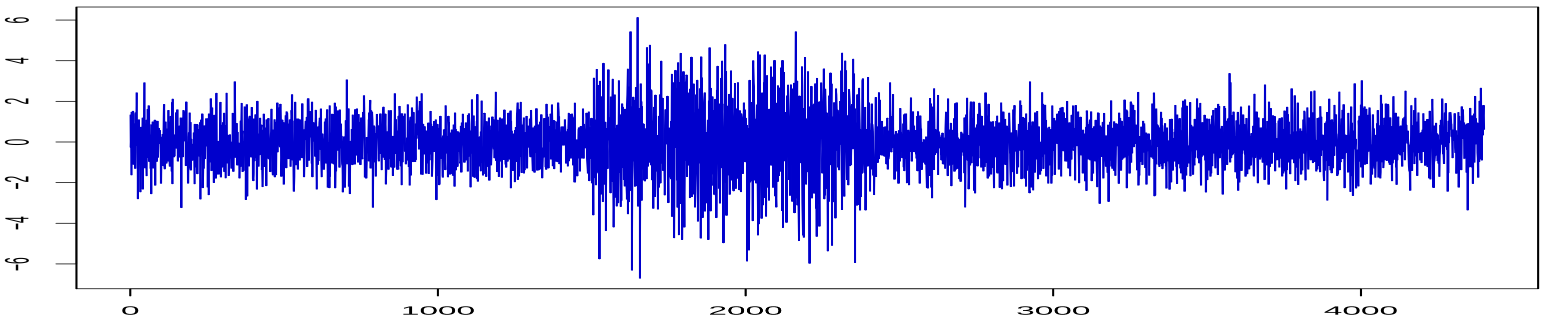
[2] St. Petersburg State University, St. Petersburg, Russia
supported by RSF grant 20-14-00072

The International Chinese Statistical Association
ICSA 2021 Applied Statistics Symposium
September 12-15, 2021

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 .

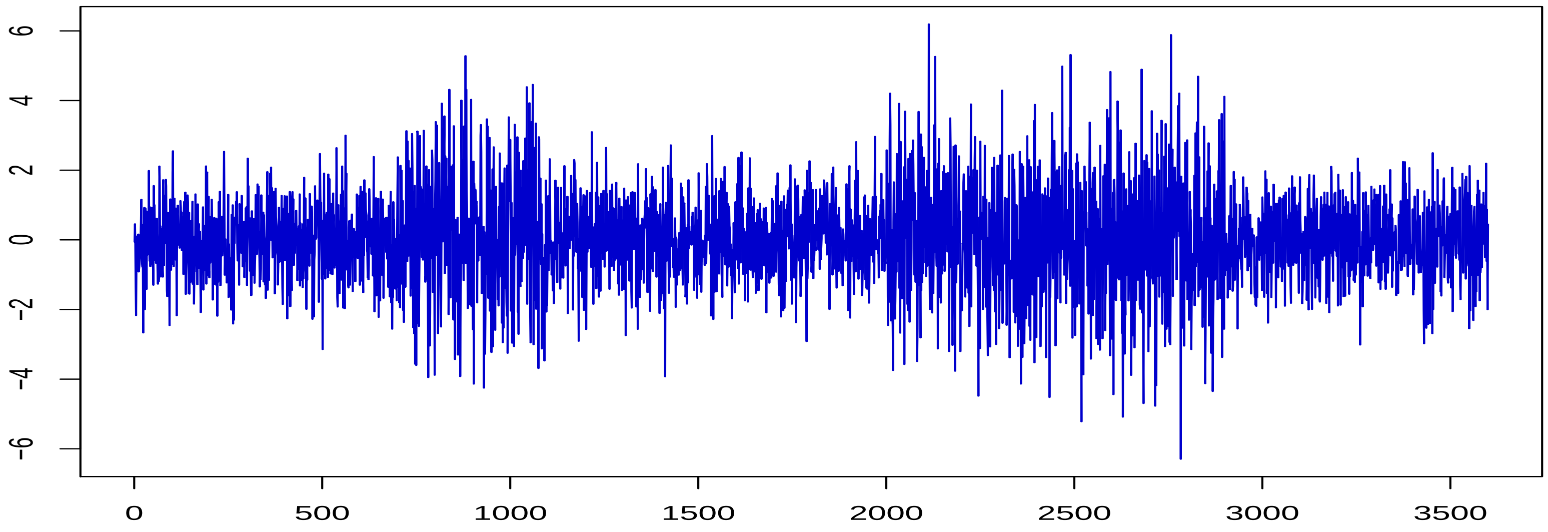
Tartakovsky (1987), Repin (1991), Guépié et al (2012), Noonan and Zhigljavsky (2020), Tartakovsky et al (2021)

Transient changes may reappear at unknown moments,

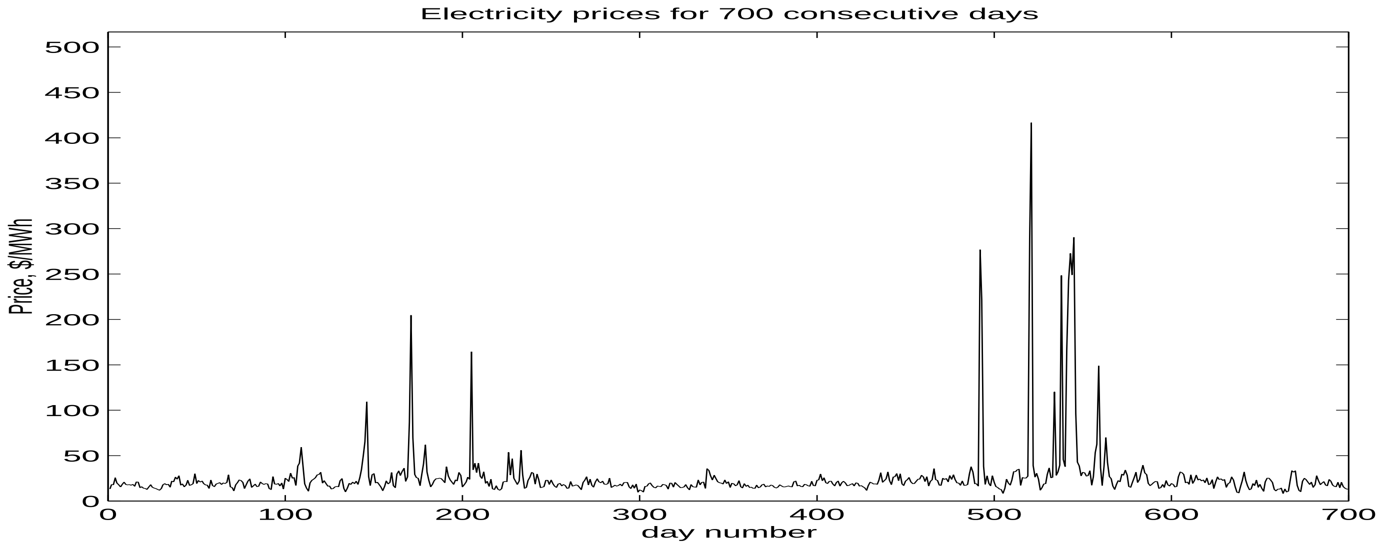
$$\begin{cases} \boldsymbol{X}_{0:a_1} = X_1, \ldots, X_{a_1} \sim F \\ \boldsymbol{X}_{a_1:b_1} = X_{a_1+1}, \ldots, X_{b_1} \sim G \\ \boldsymbol{X}_{b_1:a_2} = X_{b_1+1}, \ldots, X_{a_2} \sim F \\ \boldsymbol{X}_{a_2:b_2} = X_{a_2+1}, \ldots, X_{b_2} \sim G \\ \ldots \qquad \ldots \qquad \ldots \\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

Goals:

- Detect all changes
- estimate all a_k and b_k
- control familywise false alarm rates



Applications: Deregulated Energy Markets



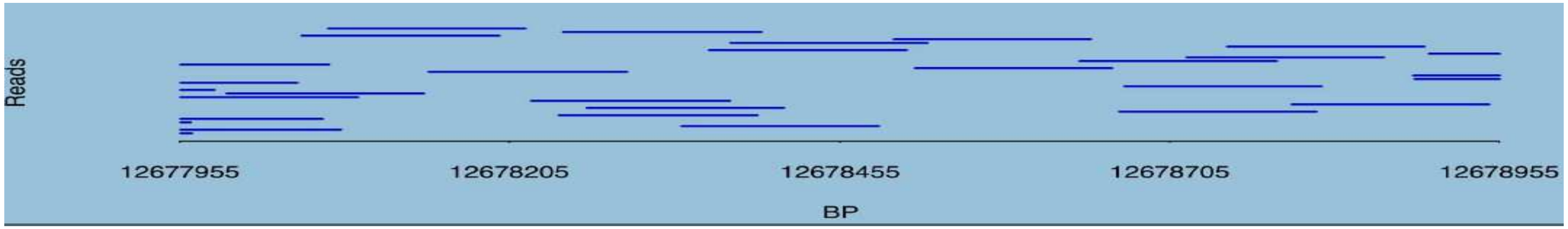
Goals:

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

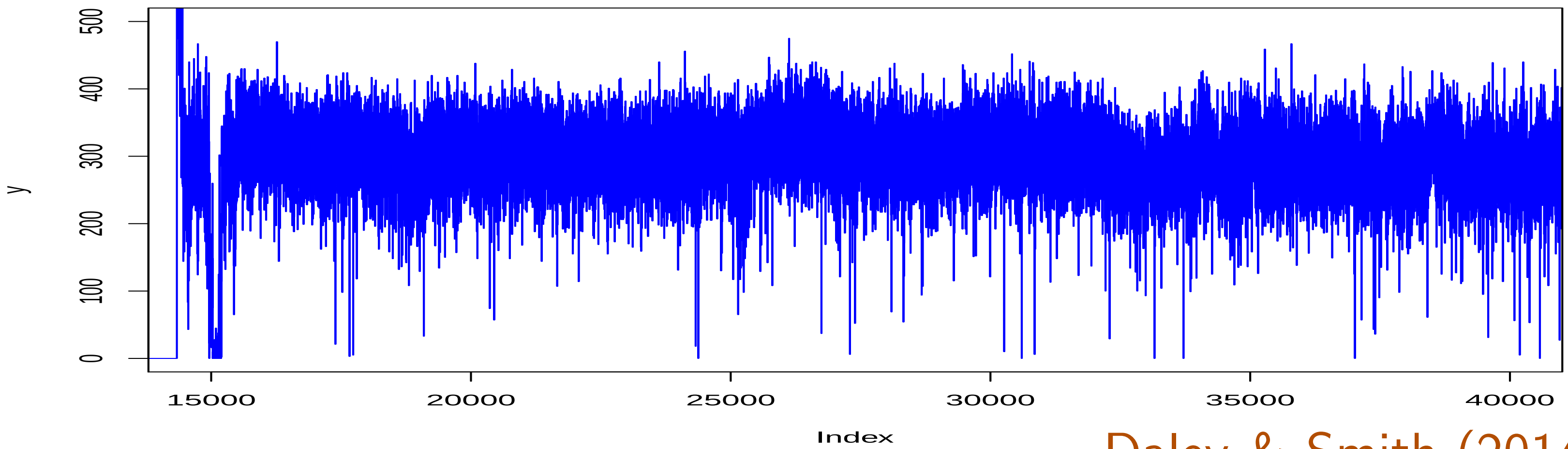
(b) Forecast; predictive distribution of electricity prices for any given day. Baron et al (2001)

Applications: 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

One transient change: maximum likelihood estimation

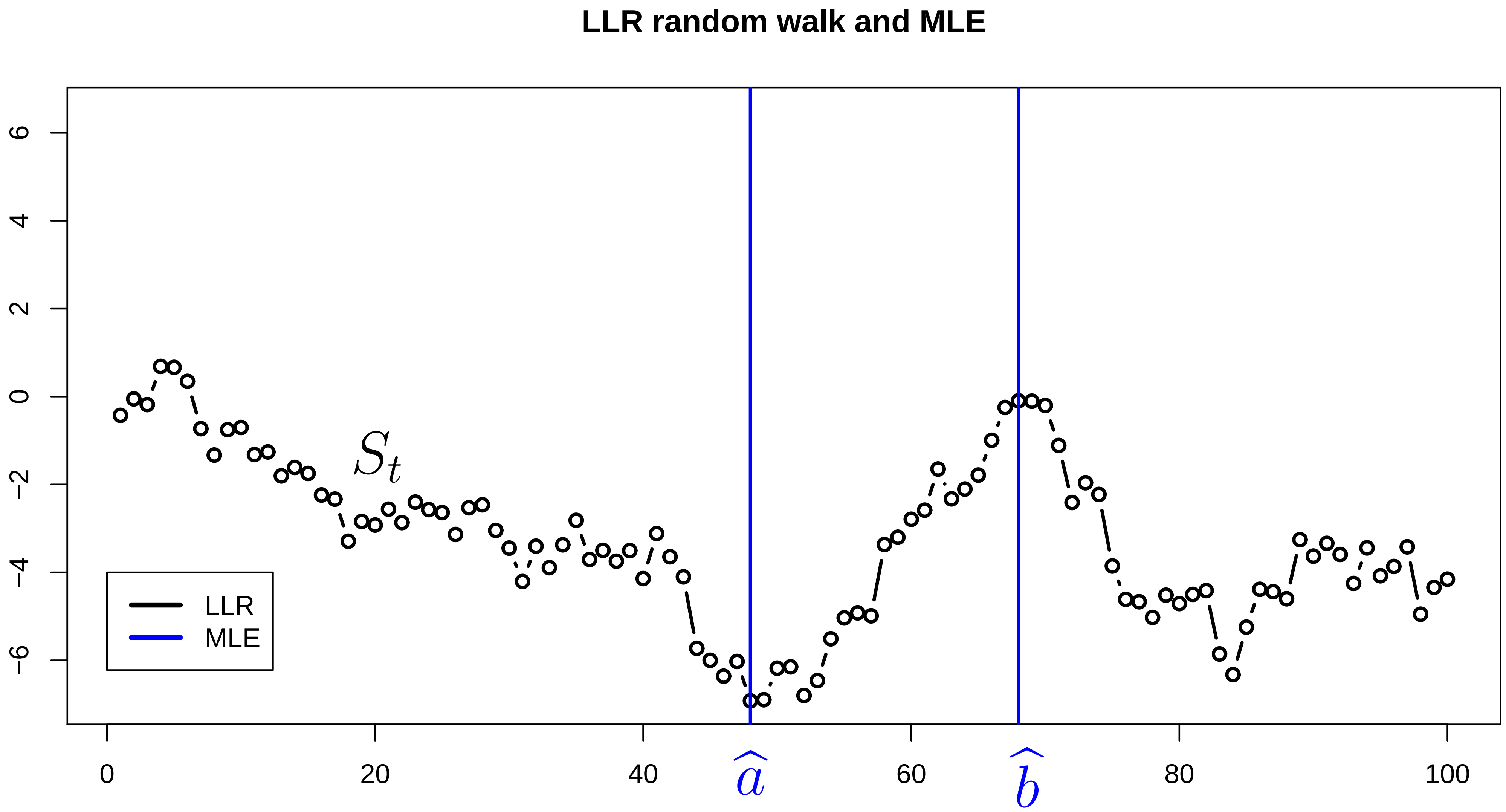
For $\theta = (a, b)$, the log-likelihood is

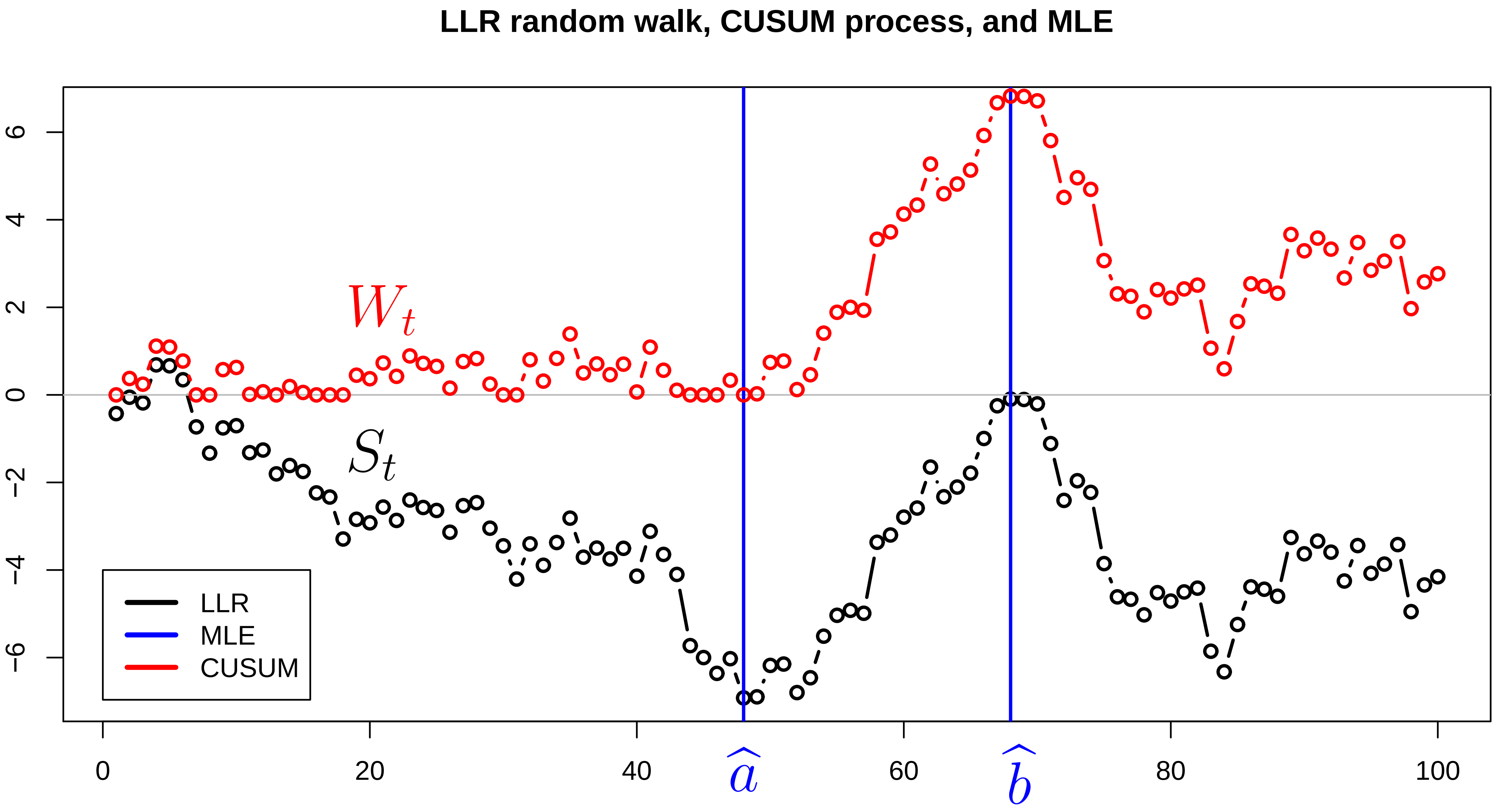
$$\begin{aligned} L(X;\theta) &= \sum_{i=1}^{a} \log f(X_i) + \sum_{i=a+1}^{b} \log g(X_i) + \sum_{i=b+1}^{n} \log f(X_i) \\ &\cong \sum_{i=a+1}^{b} \log \frac{g(X_i)}{f(X_i)} \end{aligned}$$

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 $\text{Ker}(W) = \{t \,:\, W_t = 0\}$,

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


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

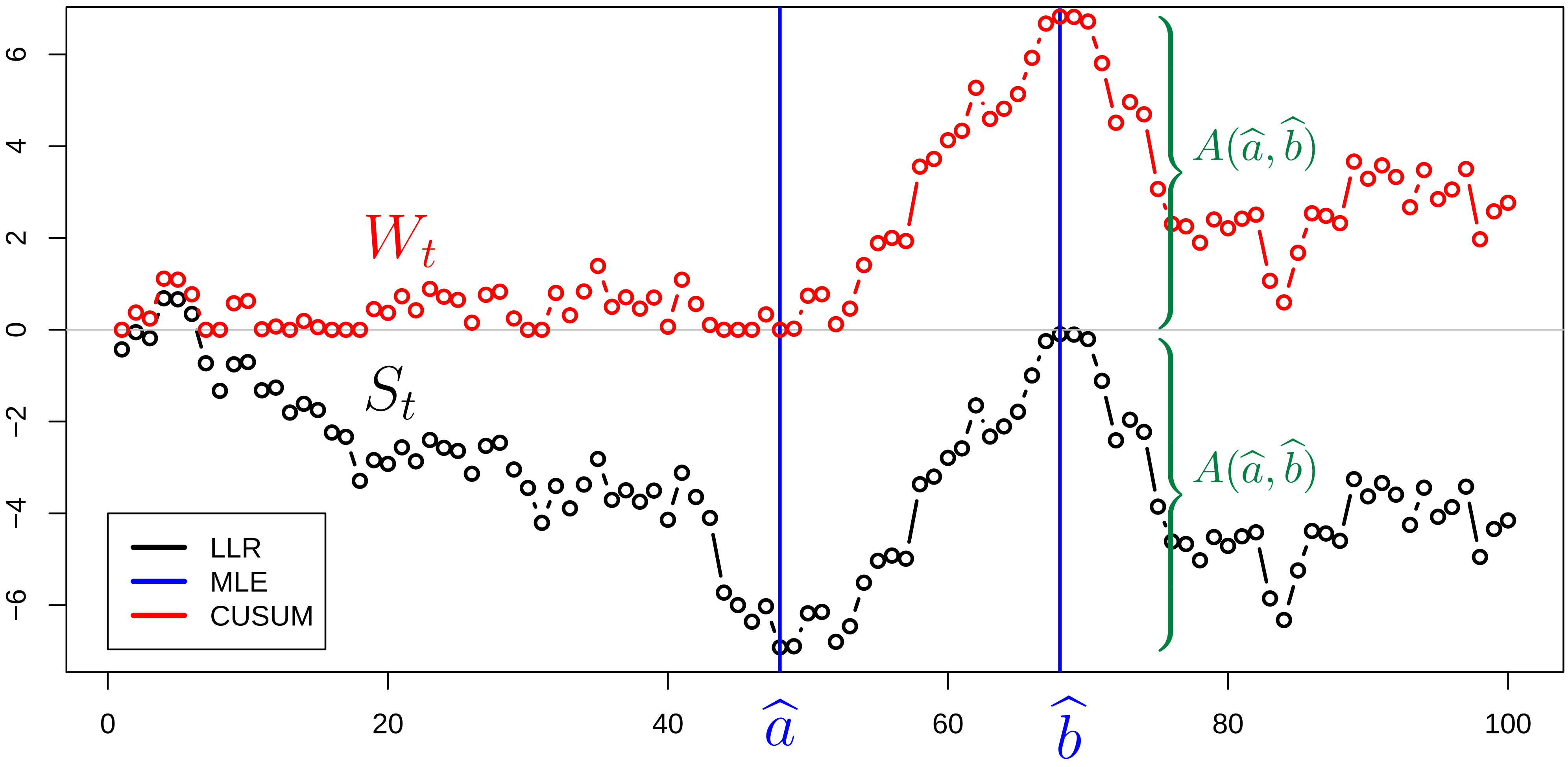

LLR random walk, CUSUM process, and MLE
W_t
S_t
LLR
MLE
CUSUM
6
4
2
0
−2
−4
−6
0
20
40
$\widehat{a}$
60
$\widehat{b}$
80
100

Control of the false alarm rate

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

- Testing
$$H_0 : \begin{array}{c} K = 0 \\ \text{all } \boldsymbol{X}_{0:n} \sim F \end{array} \quad \text{vs} \quad H_1 : \begin{array}{c} K = 1 \\ \boldsymbol{X}_{a:b} \sim G \text{ for some } a, b \end{array}$$
where $\boldsymbol{X}_{k:m} := (X_{k+1}, \ldots, X_m)$.
- The log-likelihood ratio test statistic is
$$\Lambda = \log \frac{\max_{a<b} f(\boldsymbol{X}_{0:a}) g(\boldsymbol{X}_{a,b}) f(\boldsymbol{X}_{b:n})}{f(\boldsymbol{X}_{0:n})} = A(\hat{a}, \hat{b})$$
- 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,

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

- Hence, 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 $A(\widehat{a},\widehat{b}) \ge h$.

Known number of changes: maximum likelihood estimation

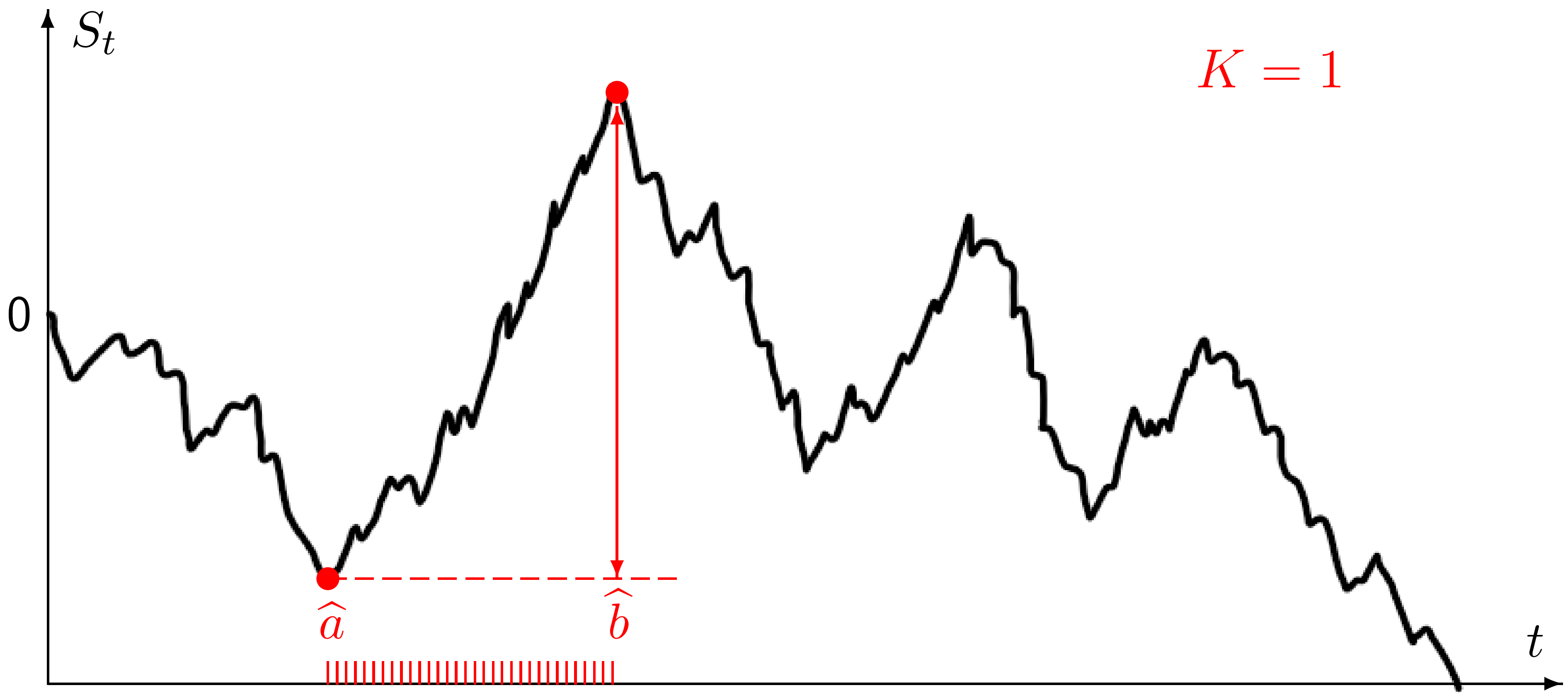
For $\boldsymbol{\theta} = \{(a_k, b_k),\ k = 1, \ldots, K\}$, K known, the log-likelihood is

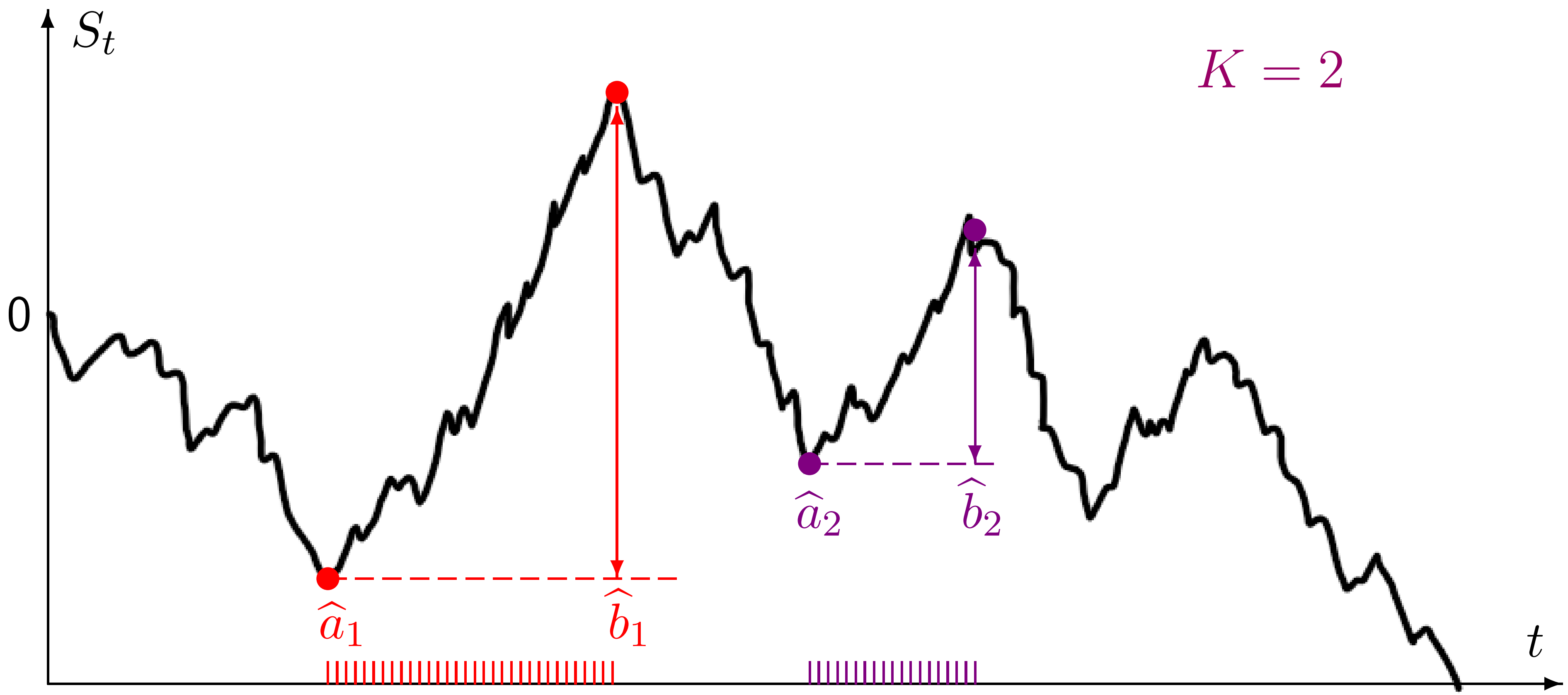
$$L(X; \boldsymbol{\theta}) = \sum_{k=1}^{K} \sum_{i=a_k+1}^{b_k} \log \frac{g(X_i)}{f(X_i)}$$

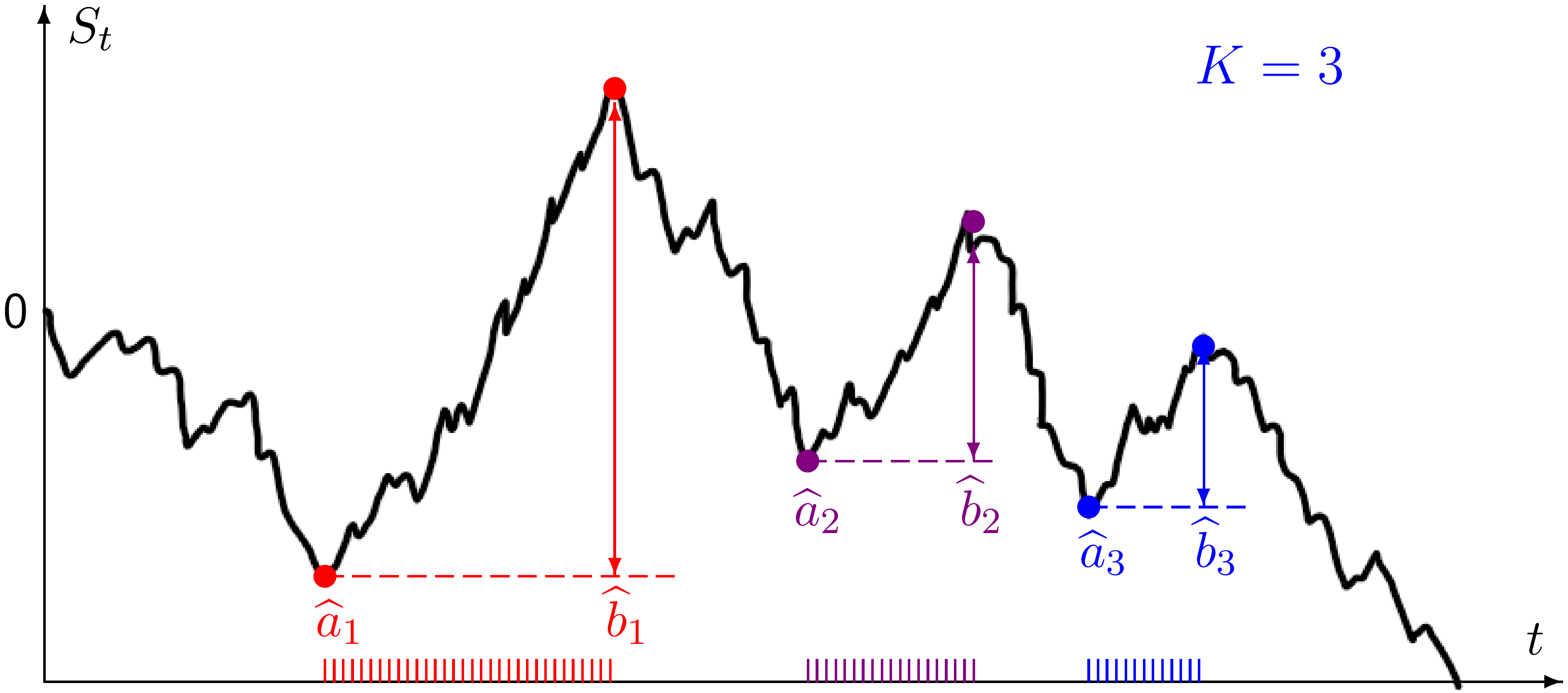
Hence, the MLE of θ is

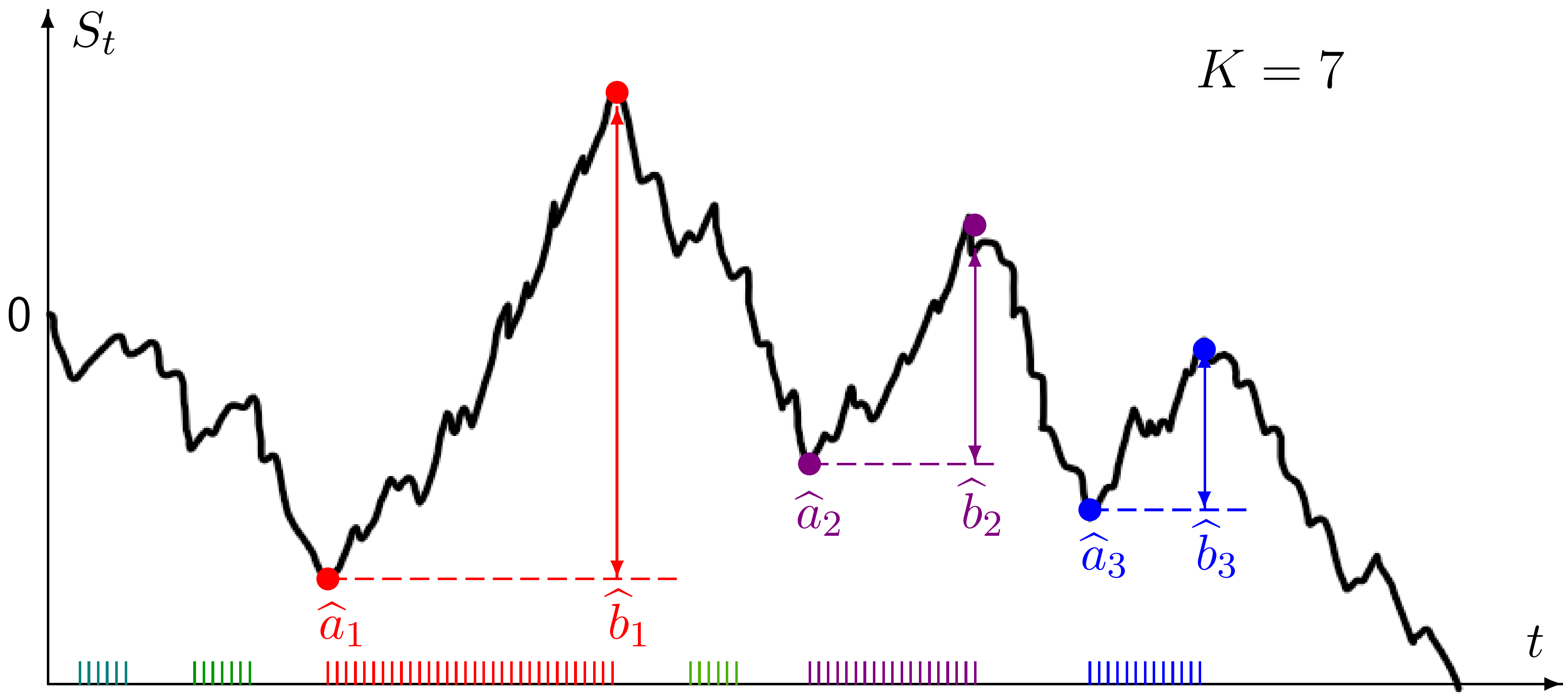
$$\widehat{\boldsymbol{\theta}} = \left\{ \widehat{\theta}_k \right\}_{k=1}^{k=K} = \left\{ (\widehat{a}_k, \widehat{b}_k) \right\}_{k=1}^{k=K} = \underset{a_1 < b_1 < \ldots < a_k < b_K}{\arg\max} \sum_{k=1}^{K} (S_{b_k} - S_{a_k}),$$

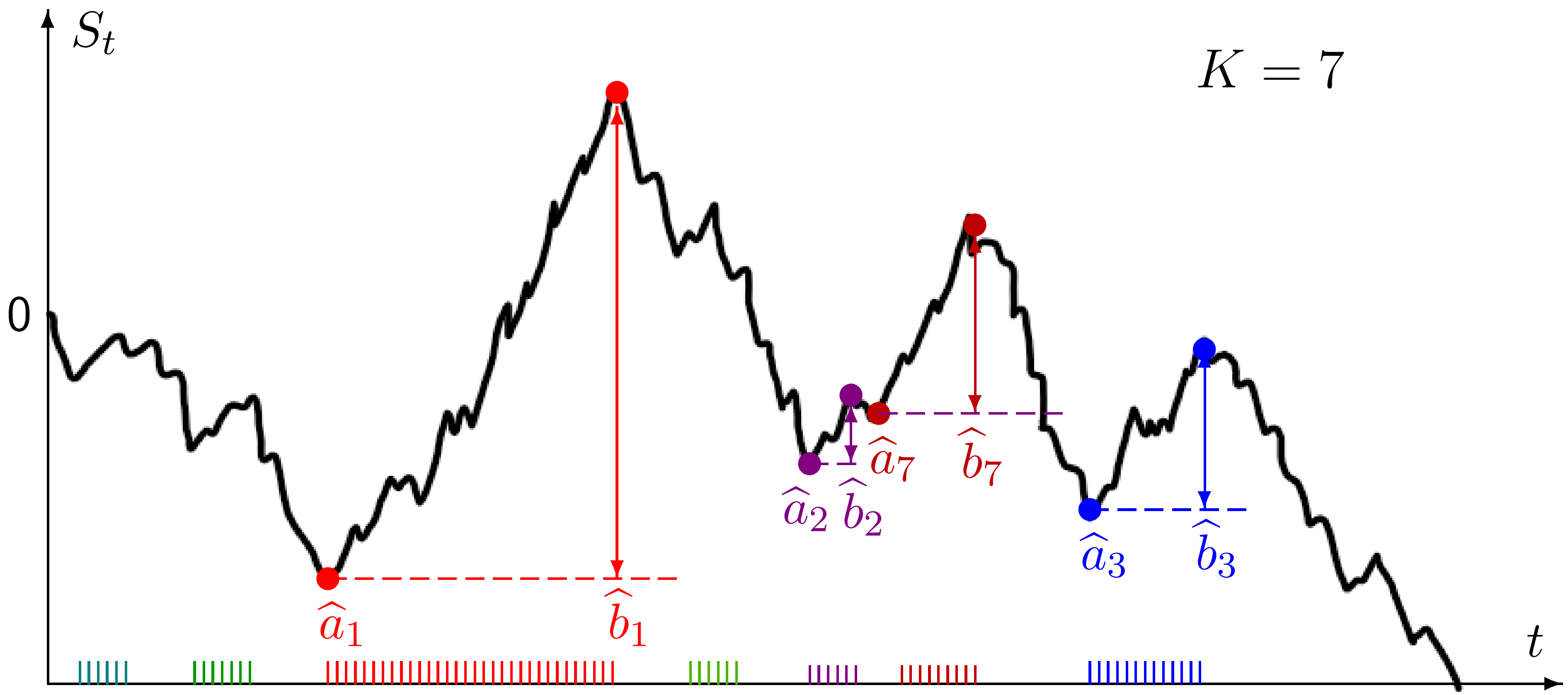
which are K intervals of the biggest growth of S_t.

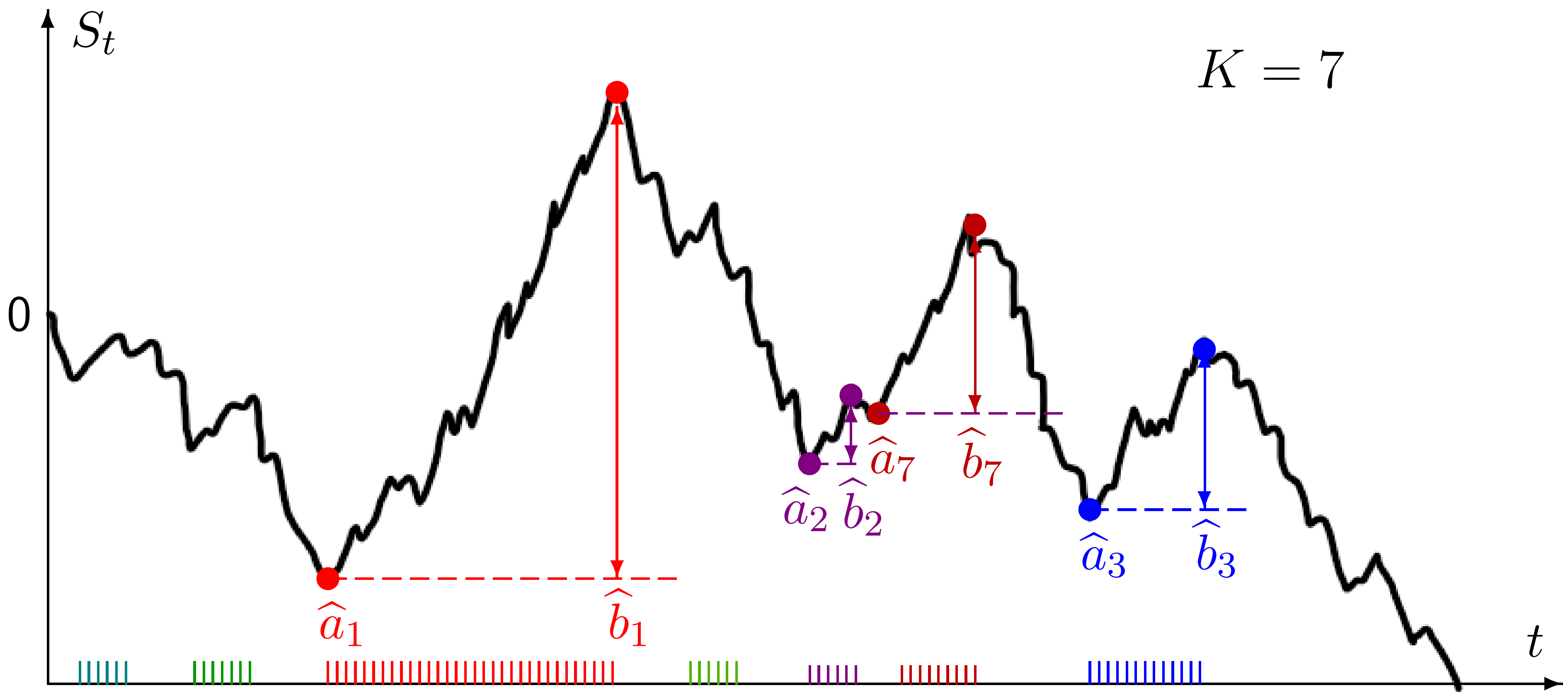

S_t
$K = 1$
0
$\widehat{a}$
$\widehat{b}$
t


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 the rate of false alarms

- Some detected change-points may be *false alarms*.
- $[\widehat{a}_k, \widehat{b}_k]$ is a *false alarm* if $[\widehat{a}_k, \widehat{b}_k] \cap (\cup [a_j, b_j]) = \varnothing$
- Goal: control the *familywise* rate of false alarms,

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

- A *false adjustment* occurs when $[\widehat{b}_k, \widehat{a}_{k+1}] \cap (\cup [b_j, a_{j+1}]) = \varnothing$
- Control $\mathbb{P}\left\{ \cup_k \left([\widehat{b}_k, \widehat{a}_{k+1}] \ \cap \ (\cup_j [b_j, a_{j+1}]) \ = \ \varnothing \right) \right\} \ \leq \ \beta$

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 $k = 1, \ldots, 2K$.

Detecting disorders with familywise FAR $\le \alpha$ and detecting adjustments with familywise FAR $\le \beta$.

No Bonferroni or Holm type correction is needed!

Estimation precision. How accurate are $\widehat{\theta}_k = (\widehat{a}_k, \widehat{b}_k)$?

Local estimators... For any $t \in [a_k, b_k]$, let

$$\tilde{a}_t = t - \arg\min_{0 \le i \le t} S_{t-i} \quad \text{and} \quad \tilde{b}_t = t + \arg\max_{0 \le i \le n-t} \widetilde{S}_{t+i}$$

These $\tilde{a}_t$ and $\tilde{b}_t$ are independent, with distributions

$$\mathbb{P}(\widetilde{b}_t = b + r) = \begin{cases} \widetilde{R}_{G,b-t}(0) R_{F,n-b}(0) & \text{for } r = 0 \\ \int_0^\infty \widetilde{R}_{G,b-t}(x) B_{F,r,n-b-r}(x) dx & \text{for } r > 0 \\ \int_0^\infty R_{F,n-b}(x) \widetilde{B}_{G,-r,b-t+r}(x) dx & \text{for } r < 0 \end{cases}$$

$$\mathbb{P}(\widetilde{a}_t = a + l) = \begin{cases} R_{F,a}(0) \widetilde{R}_{G,t-a}(0) & \text{for } l = 0 \\ \int_0^\infty \widetilde{R}_{G,\gamma-a}(x) B_{F,-l,a+l}(x) dx & \text{for } l < 0 \\ \int_0^\infty R_{F,a}(x) \widetilde{B}_{G,l,\gamma-a-l}(x) dx & \text{for } l > 0 \end{cases}$$

where

$$M_k = \max(0, S_1, \ldots, S_k),$$
$$\widetilde{M}_k = \max(0, \widetilde{S}_1, \ldots, \widetilde{S}_k)$$

$$R_{F,k}(x) = \mathbb{P}_F(M_k \le x),$$
$$\widetilde{R}_{F,k}(x) = \mathbb{P}_F(\widetilde{M}_k \le x)$$

$$\begin{aligned} B_{F,k,s}(y)dy &= \mathbb{P}_F(\arg\max_{0 \le i \le k+s} S_i = k, S_k \in [y, y+dy)) \\ &= \mathbb{P}_F(W_k = 0, S_k \in [y, y+dy))\mathbb{P}_F(M_s = 0) \end{aligned}$$

$$\begin{aligned} \widetilde{B}_{G,k,s}(y)dy &= \mathbb{P}_G(\arg\max_{0 \le i \le k+s} \widetilde{S}_i = k, \widetilde{S}_k \in [y, y+dy)) \\ &= \mathbb{P}_G(\widetilde{W}_k = 0, \widetilde{S}_k \in [y, y+dy))\mathbb{P}_G(\widetilde{M}_s = 0). \end{aligned}$$

Uniform probability bounds

- $\sup_{t\in[a,b]} \mathbb{P}(\tilde{b}_t = b + r)$

$$\geq q_r = \begin{cases} \exp\left(-\sum_{m=1}^{\infty} \frac{1}{m}\left(\mathbb{P}_F\left(\sum_{j=1}^{m} Y_j \geq 0\right) + \mathbb{P}_G\left(\sum_{j=1}^{m} Y_j \leq 0\right)\right)\right) & \text{for } r = 0 \\ \int_0^\infty R_{F,\infty}(x)\widetilde{B}_{G,-r,\infty}(x)dx & \text{for } r < 0 \\ \int_0^\infty \widetilde{R}_{G,\infty}(x)B_{F,r,\infty}(x)dx & \text{for } r > 0 \end{cases}$$

- $\sup_{t\in[a,b]} \mathbb{P}(\tilde{a}_t = a + l)$

$$\geq p_l = \begin{cases} \exp\left(-\sum_{m=1}^{\infty} \frac{1}{m}\left(\mathbb{P}_G\left(\sum_{j=1}^{m} Y_j \geq 0\right) + \mathbb{P}_F\left(\sum_{j=1}^{m} Y_j \leq 0\right)\right)\right) & \text{for } l = 0 \\ \int_0^\infty \widetilde{R}_{G,\infty}(x)B_{F,-l,\infty}(x)dx & \text{for } l < 0 \\ \int_0^\infty R_{F,\infty}(x)\widetilde{B}_{G,l,\infty}(x)dx & \text{for } l > 0 \end{cases}$$

Hinkley (1970)

Estimation precision

Now consider events

$$A_k = \bigcup_{j=1}^{K} [a_j, b_j) \cap [\nu_{2k-1}, \nu_{2k}) \neq \varnothing; \quad \text{with } \mathbb{P}(\cap A_k) \geq 1-\alpha$$

$$B_k = \bigcup_{j=1}^{K-1} [b_j, a_{j+1}) \cap [\nu_{2k}, \nu_{2k+1}) \neq \varnothing \quad \text{with } \mathbb{P}(\cap B_k) \geq 1-\beta$$

On A_k, there exists $t \in [a_j, b_j) \cap [\nu_{2k-1}, \nu_{2k})$ for some j.
On B_k, there exists $t \in [a_j, b_j) \cap [\nu_{2k}, \nu_{2k+1})$ for some j.
Hence, for each $k \in \{1, \ldots, K\}$

$$\mathbb{P}(\inf_j |\hat{a}_k - a_j| > u) \leq 1 - \sum_{-u}^{u} p_i + \alpha \quad \text{for } 0 \leq u \leq \min\{a_1, n-a_K\}$$

$$\mathbb{P}(\inf_j |\hat{b}_k - b_j| > v) \leq 1 - \sum_{-v}^{v} q_i + \beta \quad \text{for } 0 \leq v \leq \min\{b_1, n-b_K\}$$

for all θ. One-sided probabilities are bounded as

$$\mathbb{P}(\inf_j(\hat{a}_k - a_j) > u) \leq 1 - \sum_{-a_j}^{u} p_i + \alpha, \qquad \mathbb{P}(\inf_j(\hat{a}_k - a_j) < -u) \leq 1 - \sum_{-u}^{n-a_j} p_i + \alpha,$$

$$\mathbb{P}(\inf_j(\hat{b}_k - b_j) > v) \leq 1 - \sum_{-b_j}^{v} q_i + \beta, \qquad \mathbb{P}(\inf_j(\hat{b}_k - b_j) < -v) \leq 1 - \sum_{-v}^{n-b_j} q_i + \beta.$$

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.

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

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

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika 57*, 1–17.

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!