Approximating the likelihood ratio in Linear-Gaussian state-space models for change detection

Kostas Tsampourakis [1] Víctor Elvira [1,2]

[1]School of Mathematics, University of Edinburgh, United Kingdom [2]The Alan Turing Institute, United Kingdom

Introduction

- Change-point detection is the problem of detecting changes in the distribution of an observed or latent signal.
- Used for various tasks, e.g. signal segmentation (offline), process monitoring (online).
- Ubiquitous in science and engineering, with applications in biological science, climate science, speech processing, and finance, to name a few.

Background

Linear-Gaussian State Space Models

A LG-SSM is defined by the equations:

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{q}_t, \tag{1}$$
$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{r}_t, \tag{2}$$

where $\mathbf{A}_t \in \mathbb{R}^{d_x \times d_x}, \mathbf{H}_t \in \mathbb{R}^{d_y \times d_x}, \mathbf{Q}_t \in \mathbb{R}^{d_x \times d_x}, \mathbf{R}_t \in \mathbb{R}^{d_y \times d_y}$ and $\mathbf{q}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t), \mathbf{r}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$

- Parameter $\boldsymbol{\theta}_t = \{\mathbf{A}_t, \mathbf{H}_t, \mathbf{Q}_t, \mathbf{R}_t\}$.
- Marginal likelihood computed recursively by the **Kalman filter**.

Change detection

Null hypothesis:

$$\mathcal{H}^0 : \boldsymbol{\theta}_t = \boldsymbol{\theta}^0, \ \ t = 1, \dots, T \tag{3}$$

Alternative hypotheses:

$$\mathcal{H}_k(\boldsymbol{\theta}^0, \boldsymbol{\theta}^1) : \begin{cases} \boldsymbol{\theta}_t = \boldsymbol{\theta}^0, \ \ t = 1, \dots, k-1 \\ \boldsymbol{\theta}_t = \boldsymbol{\theta}^1, \ \ t = k, \dots, T \end{cases} \tag{4}$$

- Parameters $\boldsymbol{\theta}^0, \boldsymbol{\theta}^1$ known.
- Change location k unknown.
- Notation: $\mathcal{H}^1$ is the case $\boldsymbol{\theta}_t = \boldsymbol{\theta}^1, t = 1, \dots, T$

Likelihood Ratio

The log-likelihood ratio between $\mathcal{H}_k$ and $\mathcal{H}^0$ is,

$$\lambda_{k,T} = \log \frac{p^{(k)}(\mathbf{y}_{1:T})}{p_0(\mathbf{y}_{1:T})} = \sum_{t=k}^{T} \log \frac{p^{(k)}(\mathbf{y}_t|\mathbf{y}_{1:t-1})}{p_0(\mathbf{y}_t|\mathbf{y}_{1:t-1})}, \tag{5}$$

The factors $p^{(k)}(\mathbf{y}_t|\mathbf{y}_{1:t-1})$ are given by a Kalman filter matched to $\boldsymbol{\theta}^0$ for $t < k$ and to $\boldsymbol{\theta}^1$ for $t \geq k$.

Motivation

To decide between $\mathcal{H}^0$ and the composite hypothesis $\cup_k \mathcal{H}_k$, one has to evaluate the log-likelihood ratio $\lambda_{k,T}$ for each value of k, typically $k = 1, \dots, T-1$.

- This involves running a Kalman filter for each candidate change-point location k, resulting in an algorithm that takes $O(T^2)$ Kalman steps.
- In online change-point detection, this has been addressed by limiting the candidate change point in a sliding window of M observations, resulting in a $O(MT)$ algorithm.
- We propose an approximation of the log-likelihood ratio which is computed in $O(T)$ Kalman steps.
- The validity of the proposed approach relies on the stability of the Kalman filter.

Approximate Likelihood Ratio

The following approximation of the LR is the basis of the proposed procedure:

$$\tilde{\lambda}_{k,T} = \sum_{t=k}^{T} \log \frac{p_1(\mathbf{y}_t|\mathbf{y}_{1:t-1})}{p_0(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \tag{6}$$

This approximation takes $O(T)$ Kalman steps to compute.

Main Result

Under Assumption ($\mathfrak{A}$) of the paper, there exist $r > 0$ and $0 < \beta < 1$ such that,

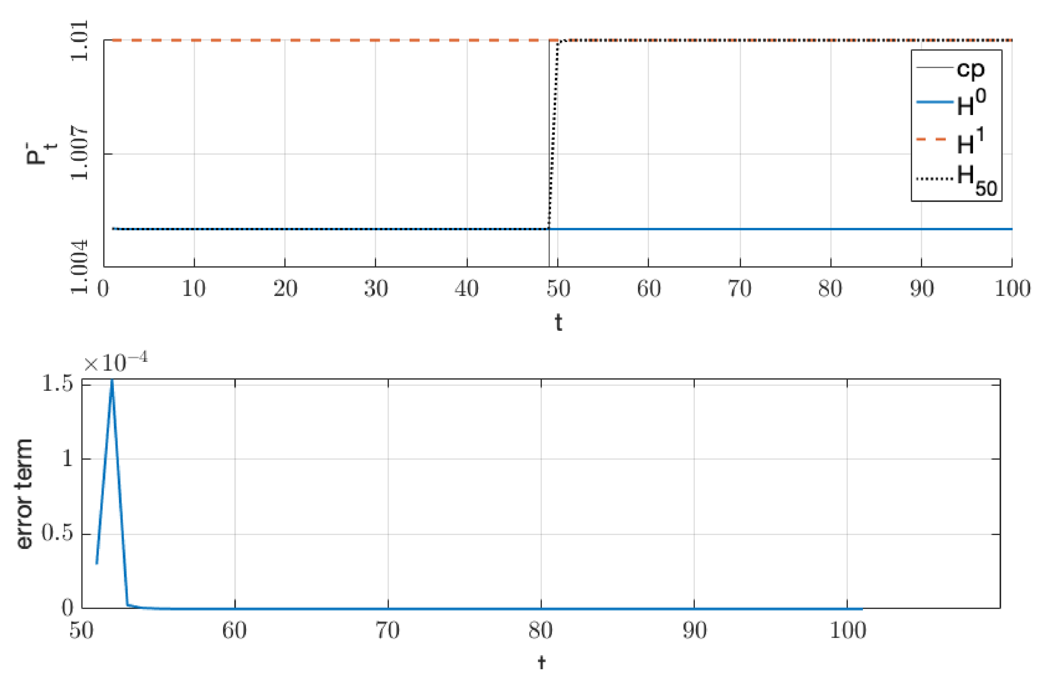
$$|\lambda_{k,T} - \tilde{\lambda}_{k,T}| \leq r \sum_{j=1}^{T-k} \beta^{j-1}. \tag{7}$$

Numerical Examples

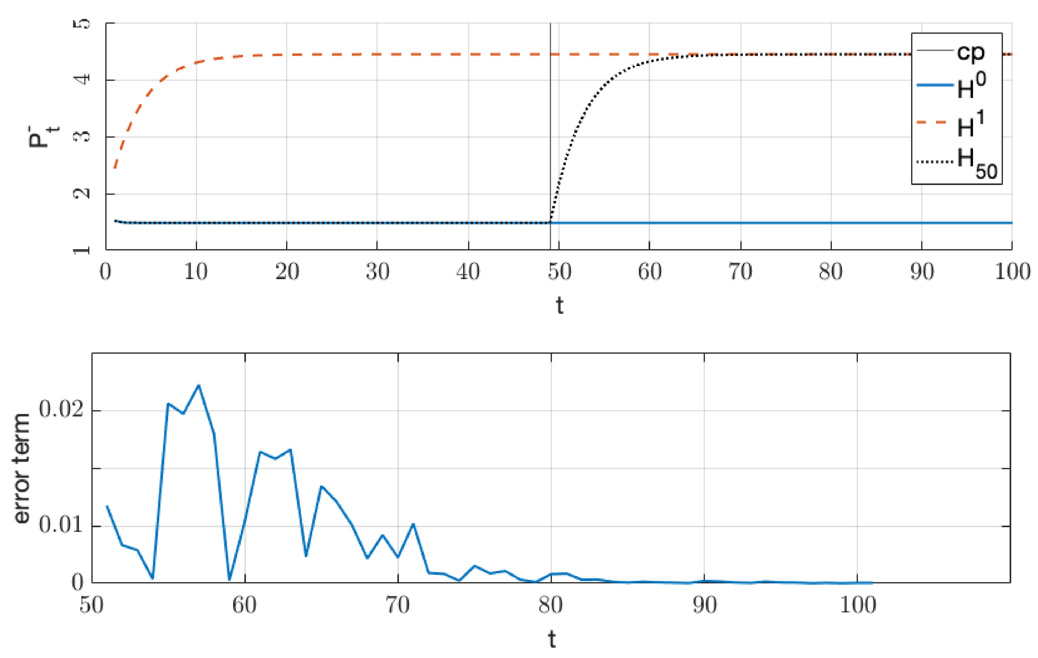
To illustrate the results above we generate time series with $T = 100$ time steps, inducing a the changepoint at $k = 50$ i.e., the set of model parameters is $\boldsymbol{\theta}^0$ for $1 \leq t < 50$, and $\boldsymbol{\theta}^1$ for $50 \leq t \leq 100$. We consider two examples, one with fast convergence and one with slow convergence:

- Fast: $a = 0.1, h = 1, \sigma_q = 1$, and $\sigma_r^0 = 1, \sigma_r^1 = 100$
- Slow: $a = 0.9, h = 1, \sigma_q = 1$, and $\sigma_r^0 = 1, \sigma_r^1 = 100$

Fast Convergence



Slow Convergence



Conclusions

- We propose an approximation to the LR that is $O(T)$ to the number of observations.
- We have provided an upper bound for the error of the approximation.
- The approximation error depends on the rate of convergence of the Kalman filters.
- This approximation can be used in the context of the GLR with more than one change-point.

References

[1] Graham C Goodwin and Kwai Sang Sin. *Adaptive filtering prediction and control.* Courier Corporation, 2014.

[2] Tze Leung Lai. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(4):613–644, 1995.

[3] Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. *Sequential analysis: Hypothesis testing and changepoint detection.* CRC Press, 2014.

kostas.tsampourakis@ed.ac.uk