

Change-Point Detection in Time Series Data via the Cross-Entropy Method

G. Sofronov ^a, L. Ma ^a

^a*Department of Statistics, Macquarie University, Sydney, Australia*
Email: lijing.ma@students.mq.edu.au

Abstract: In many applications data are collected over time and very often the statistical properties, such as mean or variance, of the data will change along data. In recent years, there has been an increasing interest to the problem, which is known as a change-point problem, where it is necessary to detect the number and locations of change-points of time series processes. The change-point detection problem, which may also be called a segmentation or break-point problem, can be found in a wide range of applications, such as financial time series analysis (e.g. changing volatility), signal processing (e.g. structural analysis of EEG signals), geology data analysis (e.g. analysis of volcanic eruption series) and environmental applications (e.g. detecting changes in ecological systems due to climatic conditions crossing some critical thresholds). For example, many macroeconomic variables such as inflation may be subject to changes in government policy that may cause structural breaks in the data and make highly persistent impacts. An awareness of those changes can assist people to eliminate and manage unnecessary risk and further improve the decision-making.

The main concern of change-point problem is the complexity to build a flexible and effective model to estimate the unknown number and locations of break points in time series. In this paper, we develop an innovative methodology to tackle this problem. We compare the statistical performance of a number of computational methods for estimating unknown parameters of autoregressive data with structural breaks. Specifically, we consider the Cross Entropy method for modelling break points using minimum description length (MDL) information criterion to estimate change-points as well as parameters of the process on each segment. Numerical experiments illustrate the robustness of this approach. We obtain estimates for the locations of change-points in artificially generated sequences and compare the accuracy of these estimates to those obtained with other methods. Finally, we use the proposed method to detect the potential location of change-points in the Australian annual inflation rate data from 1960 to 2016.

Keywords: *Change-point detection, autoregressive time series process, cross-entropy method, minimum description length*

1 INTRODUCTION

A problem of detecting change-points in time series processes, which is also called a segmentation or break-point problem, is an important research question that has been discussed by many authors for more than six decades. A change-point happens when the statistical properties, for example, mean or variance, of observations suddenly change, which may be due to an external stimulus or an inherent mechanism. Change-point detection plays an important role in modelling time series in wide range of scientific endeavours, such as financial time series analysis, econometrics (see Priyadarshana and Sofronov (2012)), signal processing (see Sofronov et al. (2012)), genomics (see Sofronov et al. (2009)), geology data analysis (see Furlan (2010)) and environmental applications (see, for example, Raveendran and Sofronov (2017)). In general, there are two classes of change-point models: retrospective and sequential (or prospective). The difference between these two classes is how data are collected, for the retrospective change-point detection, the process of data acquisition is completed at the moment when the procedure starts while for sequential method the new data continuously come in. In this paper, we focus on retrospective type of change-point analysis, and the objective of this paper is to provide an effective computational method to estimate unknown multiple change-points in the stationary AR(1) process with breaks in mean.

For a given time series or sequence, various methods have been proposed to detect multiple change-points. The general idea is to minimize an objective function by using optimization methods. There are two main streams of algorithms widely applied to objective function, one stream is based on Binary segmentation algorithm originates from the work of Scott and Knott (1974) and Sen and Srivastava (1975), Olshen et al. (2004) further explored it and proposed circular Binary segmentation for analysing the DNA copy number data, and Fryzlewicz (2014) proposed wild binary segmentation approach with 'strengthened Schwarz information criterion'. The another stream is based on dynamic programming algorithm, which involves the literatures by Auger and Lawrence (1989), Bai and Perron (1998), Killick et al. (2012) added pruning technique to it for improving the computation efficiency and proposed pruned exact linear time (PELT) method. In addition, Bansal et al. (2008) applied EM algorithm to change-point problem. Genetic algorithm has been used to identify change-points in time series by Doerr et al. (2016).

However, common multiple change-point search algorithms base on the assumption of observations are independent and identically distributed, but these methods tend to inaccurately detect the number and locations of change-points due to the endogenous dependency of time series, a stationary time series which present moderate to strong autocorrelation is easily confused with structural breaks, this will cause the overestimate problem by using those common segmentation approaches. This is the motivation that enables us to develop a segmentation method taking into account of the dependent structure of time series data and correctly identify the break-points. Recently, Chakar et al. (2017) proposed a robust approach for estimating change-points in the mean of AR(1) model. In this paper, we use the Cross-Entropy method with different information criteria to estimate the number and the locations of change-points, the autocorrelation parameter and the mean shifts in AR(1) model. The main advantage of CE method is that it combines estimation and optimization together in order to simultaneously estimate all parameters.

The paper is organised as follows. Section 2 introduced the general approach for multiple change-point problem. In Section 3, we explain the proposed methodology. Section 4 shows the results of numerical experiments. Finally, Section 5 provides general discussion and conclusions.

2 THE MULTIPLE CHANGE-POINT PROBLEM

Let $X = (X_1, X_2, \dots, X_T)$ be a data sequence of length T . Then data are segmented by change-points into $N + 1$ segments defined by N change-points $\tau = (\tau_1, \tau_2, \dots, \tau_N)$, $0 = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_N < \tau_{N+1} = T$, where the i -th segment includes observations $(X_{\tau_{i-1}+1}, \dots, X_{\tau_i})$.

In this paper, we consider the segmentation of an AR(1) sequence with homogeneous ρ and σ^2 :

$$X_t = \rho X_{t-1} + \delta_{i+1} \mathcal{I}_{[\tau_i+1 \leq t \leq \tau_{i+1}]} + \varepsilon_t, \quad t = 1, \dots, T \quad (1)$$

with independent and identically distributed $\varepsilon_t \sim N(0, \sigma^2)$, where $\mathcal{I} = 1$ for $\tau_i + 1 \leq t \leq \tau_{i+1}$ and 0 everywhere else, $0 \leq i \leq N$, which allows for shifts in the mean level, $\delta = (\delta_1, \delta_2, \dots, \delta_{N+1})$.

One of the common approaches to the multiple change-point problem is to minimize the following objective

function:

$$\sum_{i=0}^N [\mathcal{C}(X_{(\tau_i+1):\tau_{i+1}})] + \beta f(N) \quad (2)$$

where the \mathcal{C} is a cost function for a segment, e.g. negative log-likelihood, $\beta f(N)$ is a penalty term. The standard penalty functions include Akaike information criterion (AIC), Bayesian information criterion (BIC), modified BIC, etc. In this paper, we also use the minimum description length (MDL) principle (see Davis et al. (2006)) as the penalty term.

The log-likelihood function based on this model is

$$\begin{aligned} \mathcal{L}\mathcal{L}(X \mid N, \tau, \rho, \delta, \sigma^2) &= -\frac{N+1}{2} \ln \left(\frac{2\pi\sigma^2}{1-\rho^2} \right) - \sum_{i=0}^N -\frac{\tau_{i+1} - \tau_i - 1}{2} \ln(2\pi\sigma^2) \\ &\quad - \sum_{i=0}^N \frac{(X_t - \delta_{i+1}/(1-\rho))^2}{2\sigma^2/(1-\rho^2)} - \sum_{i=0}^N \sum_{t=\tau_i+2}^{\tau_{i+1}} \frac{(X_t - \delta_{i+1} - \rho X_{t-1})^2}{2\sigma^2}. \end{aligned}$$

Then we can derive the objective functions (AIC, BIC and MDL) based on (1):

$$\mathbf{F}_{AIC} = -2\mathcal{L}\mathcal{L}(X \mid N, \tau, \rho, \delta, \sigma^2) + 2(N+3), \quad (3)$$

$$\mathbf{F}_{BIC} = -2\mathcal{L}\mathcal{L}(X \mid N, \tau, \rho, \delta, \sigma^2) + \log(T)(N+3), \quad (4)$$

$$\begin{aligned} \mathbf{F}_{MDL} &= -2\mathcal{L}\mathcal{L}(X \mid N, \tau, \rho, \delta, \sigma^2) + 2\log(T) + 2\log(N) \\ &\quad + \sum_{i=0}^N \log(\tau_{i+1} - \tau_i - 1) + 2 \sum_{i=0}^N \log(\tau_{i+1}). \end{aligned} \quad (5)$$

3 THE CROSS-ENTROPY METHOD

In this paper, we consider the multiple change-point problem as a combinatorial optimisation problem of one of the objective functions defined in (3), (4), or (5). To solve this problem we apply the Cross-Entropy (CE) method, which is a stochastic optimisation method developed by Rubinstein and Kroese (2004). Generally speaking, if F denotes an objective (performance) function defined on a finite set \mathcal{Y} , then the idea of the CE method can be described as the following way.

Let us define a set of indicator functions $\{I_{\{F(y) \geq \gamma\}}\}$ on \mathcal{Y} for different levels $\gamma \in R$. Let $\{f(\cdot, u)\}$ be a family of probability density functions (pdfs) on \mathcal{Y} with a real-valued parameter u . Following Rubinstein and Kroese (2004), we associate the optimisation problem with the problem of estimating

$$l(\gamma) = \mathbf{P}_u(F(Y) \geq \gamma) = \sum_x I_{\{F(y) \geq \gamma\}} f(x, u) = \mathbf{E}_u I_{\{F(Y) \geq \gamma\}},$$

where γ is a known or unknown parameter and \mathbf{P}_u is the probability measure under which the random state Y has the pdf $f(\cdot, u)$.

The problem of estimating l is not trivial. Adaptive changes to the pdf are based on the Kullback-Leibler (or the CE) distance. Thus it allows to create a sequence $f(\cdot, u_0), f(\cdot, u_1), \dots, f(\cdot, u^*)$. The final pdf $f(\cdot, u^*)$ corresponds to the degenerate density at an optimal point. This means that the CE method creates a sequence of pairs $\{(\gamma_k, u_k)\}$, which converges quickly to a close neighbourhood of the optimal tuple (γ^*, u^*) . More specifically, we should set up u_0 and simulation parameters, and then we carry out the following procedure Rubinstein and Kroese (2004):

1. **Adaptive updating of γ_k .** For a fixed u_{k-1} , let γ_k be a $(1-\beta)$ -quantile of $\widehat{F}(Y)$ under u_{k-1} . A simple estimator $\widehat{\gamma}_k$ of γ_k is

$$\widehat{\gamma}_t = \widehat{F}_{(\lceil (1-\beta)M \rceil)},$$

where, for a random sample Y_1, \dots, Y_M from $f(\cdot, u_{k-1})$, $\widehat{F}_{(i)}$ is the i -th order statistic of the performances $\widehat{F}(Y_1), \dots, \widehat{F}(Y_M)$.

2. **Adaptive updating of u_k .** For fixed γ_k and u_{k-1} , derive u_k from the solution of the CE program

$$\max_u D(u) = \max_u \mathbf{E}_{u_{k-1}} I_{\{\widehat{F}(Y) \geq \gamma_k\}} \ln f(Y, u).$$

The CE method in Priyadarshana and Sofronov (2015) and Priyadarshana and Sofronov (2012) consider four-parameter beta distributions, whereas in this paper, we use normal distributions. The main steps of our algorithm are as follows:

1. Choose initial sets for $\mu^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_N^{(0)}, \mu_1^{(N+1)}, \mu_2^{(N+2)}, \dots, \mu_N^{(2N+1)})$ and $(\sigma^2)^{(0)} = ((\sigma_1^2)^{(0)}, (\sigma_2^2)^{(0)}, \dots, (\sigma_N^2)^{(0)}, (\sigma_{N+1}^2)^{(0)}, (\sigma_{N+2}^2)^{(0)}, \dots, (\sigma_{2N+1}^2)^{(0)})$. The length of both vectors is $2N + 1$ with the first N components corresponding to N change-points and the last $N + 1$ components corresponding to the vector δ . Set $k = 1$.
2. Generate a random sample $c^{(1)}, c^{(2)}, \dots, c^{(M)}$ from the normal distributions with parameters $(\mu^{(k-1)}, (\sigma^2)^{(k-1)})$, where $c^{(i)} = (c_1^{(i)}, c_2^{(i)}, \dots, c_N^{(i)}, c_{N+1}^{(i)}, c_{N+2}^{(i)}, \dots, c_{2N+1}^{(i)})$, $i = 1, 2, \dots, M$, where the first N components $(c_1^{(i)}, c_2^{(i)}, \dots, c_N^{(i)})$ is a change-point vector.
3. For each $i = 1, 2, \dots, M$, order $(c_1^{(i)}, c_2^{(i)}, \dots, c_N^{(i)})$ from smallest to biggest. After adjusting the mean levels in each segment, estimate the autocorrelation coefficient ρ using the entire data sequence.
4. Evaluate the objective function (either (3), (4) or (5)) for each $c^{(1)}, c^{(2)}, \dots, c^{(M)}$. Define the elite sample, which is the best performing combinations of the change-points. Let $M_{elite} = \beta M$ be the size of the elite sample.
5. For all $j = 1, 2, \dots, 2N + 1$, estimate the parameters $\mu_j^{(k)}$ and $(\sigma_j^2)^{(k)}$ using the elite sample and update the current parameter sets as follows:

$$\mu_j^{(k)} = \frac{\sum_{i \in I} c_j^{(i)}}{N_{elite}}, \quad (\sigma_j^2)^{(k)} = \frac{\sum_{i \in I} (c_j^{(i)} - \mu_j^{(k)})^2}{N_{elite}},$$

where I is the set of indices of the best performing samples.

6. Stopping criterion is $\max_j (\sigma_j^2)^{(k)} < \varepsilon$.
7. If the stopping criterion is met, then stop the process and identify the combination of the positions of change points and the values of mean levels for all segments that minimizes the objective function. Otherwise set $k = k + 1$ and iterate from step 2.

4 NUMERICAL EXPERIMENTS

In this section, we discuss two numerical examples: in the first example, we consider artificially generated sequences with a known distribution, which allows us to provide direct comparison with existing techniques. The second example uses a real data — Australian annual inflation rate from 1960 to 2016.

Example 1: Simulated AR(1) data. Following model (1), we generated 9 sets of 100 AR(1) sequences of length 201 with autocorrelation coefficient ranging from 0.1 to 0.9 and $\sigma = 1$. Then we introduced 3 abrupt change-points with positions $\tau_1 = 40$, $\tau_2 = 70$, and $\tau_3 = 150$ and the mean levels $\delta_1 = 0$, $\delta_2 = 1$, $\delta_3 = -1$, and $\delta_4 = 3$.

We run the CE algorithm with the following simulation parameters: the elite proportion value $\beta = 0.05$ and the sample size $M = 500$. Let CE-AIC, CE-BIC and CE-MDL denote the CE algorithms with the objective functions AIC, BIC and MDL, respectively.

Table 1 compares the number of change-points obtained by the CE-AIC, CE-BIC, CE-MDL algorithms and the Binary Segmentation (BS) method and the PELT method, both of which are implemented in `changePoint` package (see Killick and Eckley (2014)). Both the BS and the PELT method use the BIC as an objective function. It is important to note that we only consider the changes in the mean level, and limit the maximum number of change-points to five in all methods.

Table 1. Estimated number of change-points

Algorithm		≤ 1	2	3	4	≥ 5		≤ 1	2	3	4	≥ 5
BS	$\rho = 0.3$	0	9	56	27	8	$\rho = 0.5$	0	2	18	28	52
PELT		0	6	45	27	22		0	0	8	15	77
CE-AIC		0	2	10	17	71		0	2	9	20	69
CE-BIC		1	20	19	14	46		0	9	17	24	50
CE-MDL		22	33	18	8	19		24	41	14	7	14
BS	$\rho = 0.7$	0	2	3	5	90	$\rho = 0.9$	0	0	0	0	100
PELT		0	0	1	2	97		0	0	0	0	100
CE-AIC		0	2	7	20	71		0	1	9	29	61
CE-BIC		0	17	20	25	38		6	22	28	28	16
CE-MDL		39	41	8	5	7		67	23	5	3	2

Table 1 shows that both the BS and the PELT methods tend to overestimate the number of change-points comparing with the CE method when autocorrelation gets stronger, while they perform relatively well when the autocorrelation ρ is small.

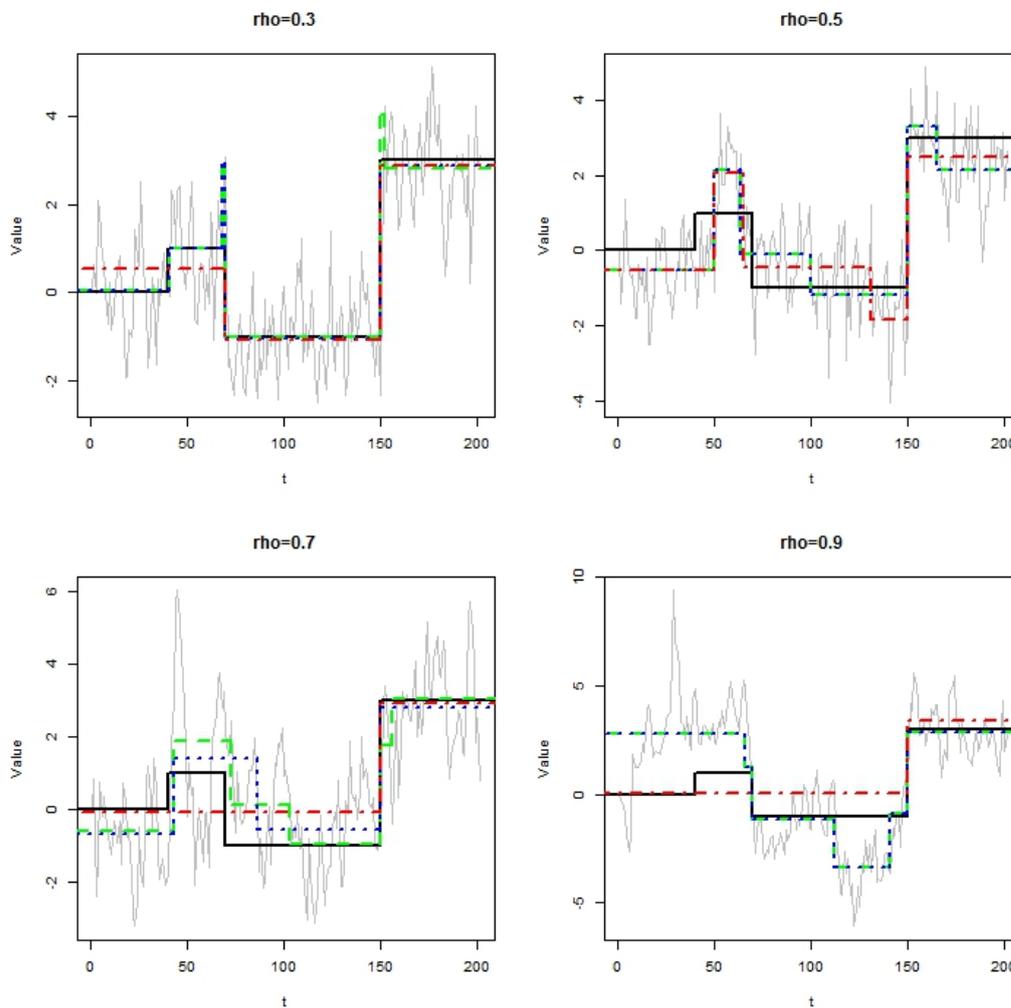


Figure 1. The true profile (black solid line) and the profile plots estimated by the CE-AIC algorithm (green dashed line), the CE-BIC algorithm (blue dotted line), and the CE-MDL (red dot-dashed line)

Figure 1 displays 4 profile plots for the first sequence out of 100 generated sequences for $\rho = 0.3, 0.5, 0.7$ and 0.9 . It is clear that the CE-AIC and the CE-BIC overestimate while the CE-MDL underestimates the correct number of change-points. Other than that the methods are in very good agreement with each other.

Example 2: Australian inflation rate series. In this example we analyse Australian annual inflation rate data from 1960 to 2016 (see, for example, Reserve Bank of Australia (2017)). Figure 2 shows the estimated segments obtained by the CE methods with AIC, BIC and MDL information criteria. Since the actual number and the locations of change-points are unknown, we look for the agreement between the methods. The CE-AIC method estimates the 5 possible change-points in 1974, 1984, 1986, 1991 and 2009, while the CE-BIC and the CE-MDL methods present the same results with 3 change-points occurred in 1974, 1984 and 1991. All methods have correctly identified the period from 1974 to 1991 with high annual inflation rate. For the model with 5 change-points, the estimated ρ is equal to 0.65, which is quite similar to 0.61, the estimated value of ρ for the model with 3 change-points.

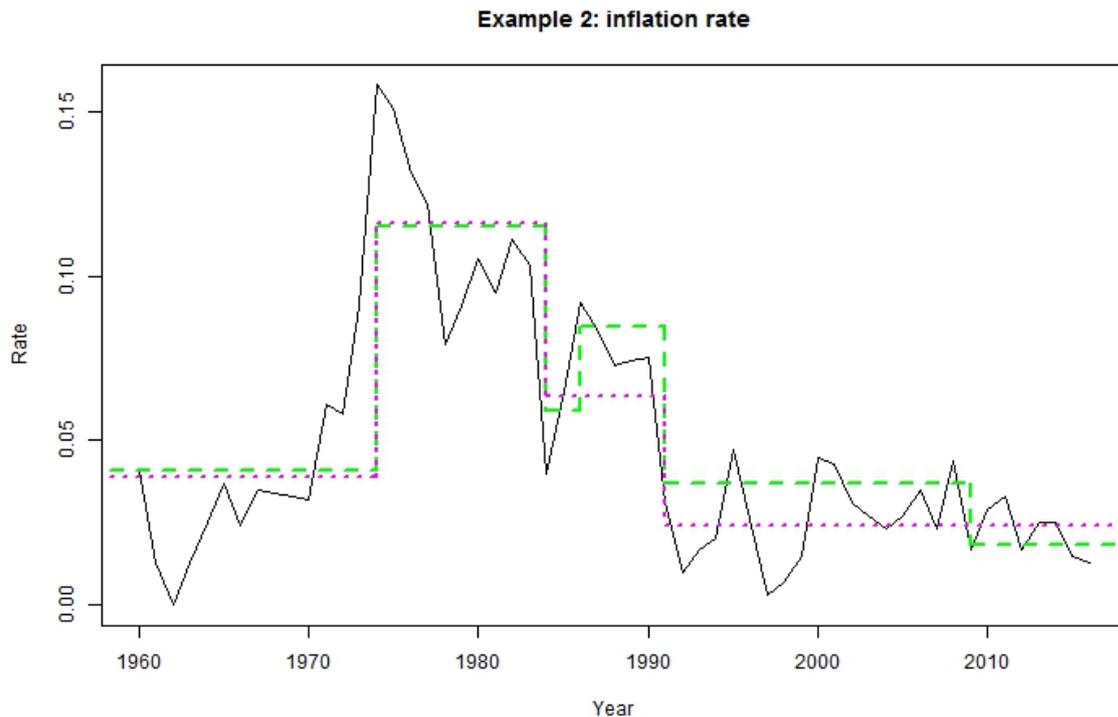


Figure 2. Annual inflation rate and profile plots of the CE-AIC algorithm (green dashed line) and the CE-BIC/MDL algorithms (magenta dotted line)

5 DISCUSSION AND CONCLUSIONS

In this paper, we have proposed to use the Cross-Entropy method for identifying change-points in AR(1) sequences with shifts in the mean level. The important feature of the proposed method is that it allows us to estimate all parameters of the model such as the mean level at each segment, the autocorrelation for the whole data as well as the number and the locations of change-points. In order to estimate the correct number of change-points we use information criteria such as the AIC, the BIC and the MDL. We should note that no previous study has compared the performance of these information criteria in change-point estimation. The numerical results show that the standard information criteria, the AIC and the BIC, overestimate the number of change-points since both of them are designed for normal independent data. In contrast, the MDL tends to underestimate the number of change-points. These findings may be somewhat limited by the inevitable problem of analysing artificially generated data. The issue is that the sample means could heavily deviate from the theoretical means at each segment, leading to inaccurate estimates of the model parameters, which include the number and the positions of change-points. On the other hand, the underestimation of the number of change-points by the MDL may be due to the fact that we assume the autocorrelation coefficient is the same for all segments, which may contradict with the assumption that adjacent segments are independent (see Davis et al. (2006)). This implies that the formula for the MDL (5) may not be accurate. Therefore, there is a clear need to develop a new criterion or modify an existing one to address this issue, which is a matter for our future research.

REFERENCES

- Auger, I. E. and C. E. Lawrence (1989). Algorithms for the optimal identification of segment neighborhoods. *Bulletin of mathematical biology* 51(1), 39–54.
- Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47–78.
- Bansal, N. K., H. Du, and G. Hamedani (2008). An application of EM algorithm to a change-point problem. *Communications in Statistics-Theory and Methods* 37(13), 2010–2021.
- Chakar, S., E. Lebarbier, C. Lévy-Leduc, and S. Robin (2017). A robust approach for estimating change-points in the mean of an AR(1) process. *Bernoulli* 23(2), 1408–1447.
- Davis, R. A., T. C. M. Lee, and G. A. Rodriguez-Yam (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association* 101(473), 223–239.
- Doerr, B., P. Fischer, A. Hilbert, and C. Witt (2016). Detecting structural breaks in time series via genetic algorithms. *Soft Computing*, 1–14.
- Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics* 42(6), 2243–2281.
- Furlan, C. (2010). Extreme value methods for modelling historical series of large volcanic magnitudes. *Statistical Modelling* 10(2), 113–132.
- Killick, R. and I. Eckley (2014). changepoint: An R package for changepoint analysis. *Journal of Statistical Software, Articles* 58(3), 1–19.
- Killick, R., P. Fearnhead, and I. A. Eckley (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* 107(500), 1590–1598.
- Olshen, A. B., E. Venkatraman, R. Lucito, and M. Wigler (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5(4), 557–572.
- Priyadarshana, M. and G. Sofronov (2012). A modified cross-entropy method for detecting change-points in the Sri-Lankan stock market. In *Proceedings of the IASTED International Conference on Engineering and Applied Science, EAS 2012*, pp. 319–326.
- Priyadarshana, W. and G. Sofronov (2015). Multiple break-points detection in array CGH data via the cross-entropy method. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 12(2), 487–498.
- Raveendran, N. and G. Sofronov (2017). Binary segmentation methods for identifying boundaries of spatial domains. In *2017 Federated Conference on Computer Science and Information Systems, FedCSIS 2017*, pp. 1–10.
- Reserve Bank of Australia (2017). *Measures of Consumer Price Inflation*. <http://www.rba.gov.au/inflation/measures-cpi.html>.
- Rubinstein, R. Y. and D. P. Kroese (2004). *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning*. Springer Science & Business Media.
- Scott, A. J. and M. Knott (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 507–512.
- Sen, A. and M. S. Srivastava (1975). On tests for detecting change in mean. *The Annals of statistics*, 98–108.
- Sofronov, G., T. Polushina, and M. Priyadarshana (2012). Sequential change-point detection via the cross-entropy method. In B. Reljin and S. Stankovic (Eds.), *The 11th Symposium on Neural Network Applications in Electrical Engineering (NEUREL2012)*, pp. 185–188.
- Sofronov, G. Y., G. E. Evans, J. M. Keith, and D. P. Kroese (2009). Identifying change-points in biological sequences via sequential importance sampling. *Environmental Modeling and Assessment* 14(5), 577–584.