Sparse Parameter Estimation Approach to Change Point Detection

Neubauer, Jiří
University of Defence, Department of Econometrics
Kounicova 65
Brno 662 10, Czech Republic
E-mail: jiri.neubauer@email.cz

Veselý, Vítězslav
Masaryk University, Department of Applied Mathematics and Computer Science
Lipová 41a
Brno 602 00, Czech Republic
E-mail: vesely@econ.muni.cz

Introduction

Detection of changes is a problem of discovering time points at which properties of a stochastic process change. This covers a broad range of real-world problems and has been actively discussed. The standard statistical approach to this problem is described for example in Csörgó and Horváth (1997). The contribution is focused on multiple change point detection in a one-dimensional stochastic process using sparse parameter estimation from an overparametrized model. Detection of changes has originally arisen in the context of quality control. Nowadays, we can find wide range of fields where change point problem is applied, such as epidemiology, medicine (rhythm analysis), ecology, signal processing etc.

Authors' approach to change point detection is quite different from the standard statistical techniques. A stochastic process residing in a bounded interval with changes in the mean is estimated using dictionary (a family of functions, the so-called atoms, which are overcomplete in the sense of being nearly linearly dependent) and consisting of Heaviside functions. Among all possible representations of the process we want to find a sparse one utilizing a significantly reduced number of atoms. This problem can be solved by ℓ_1-minimization (see Bruckstein et al. (2009), Chen et. al (1998)). The basis pursuit algorithm is used to get sparse parameter estimates. Basic properties of this approach were studied in Neubauer and Veselý (2011) and in Neubauer and Veselý (2010) by simulations. According to these results, the basis pursuit approach proposes an alternative technique of the change point detection. In this contribution the authors calculate empirical probability of successful change point detection as a function depending on the number of change points and the level of standard deviation of an additive white noise of the stochastic process. The empirical probability was computed by simulations where locations of change points were chosen randomly from uniform distribution.

Heaviside Dictionary for Change Point Detection

In this paragraph we briefly describe the method based on basis pursuit algorithm(BPA) for the detection of the change point in the sample path $\{y_t\}$ in one dimensional stochastic process $\{Y_t\}$. We assume a deterministic functional model on a bounded interval $\mathcal{I}$ described by the dictionary $G = \{G_j\}_{j\in J}$ with atoms $G_j \in L^2(\mathcal{I})$ and with additive white noise e on a suitable finite discrete mesh $\mathcal{T} \subset \mathcal{I}$:

$$Y_t = x_t + e_t,\ t \in \mathcal{T},$$

where $x \in \mathrm{sp}(\{G_j\}_{j\in J})$, $\{e_t\}_{t\in\mathcal{T}} \sim WN(0, \sigma^2)$, $\sigma > 0$, and J is a big finite indexing set. Smoothed function $\hat{x} = \sum_{j\in J} \hat{\xi}_j G_j =: \mathbf{G}\hat{\xi}$ minimizes on $\mathcal{T}$ ℓ^1-penalized optimality measure $\frac{1}{2}\|\mathbf{y} - \mathbf{G}\xi\|^2$ as

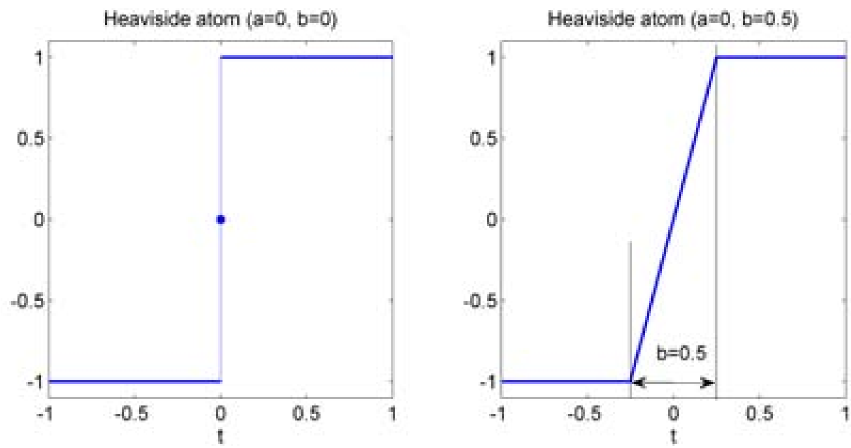

Figure 1: Heaviside atoms with parameters $a = 0, b = 0$ and $a = 0, b = 0.5$

follows:

$$\hat{\xi} = \text{argmin}_{\xi\in\ell^2(J)} \frac{1}{2}\|\mathbf{y} - \mathbf{G}\xi\|^2 + \lambda\|\xi\|_1,\ \|\xi\|_1 := \sum_{j\in J} \|G_j\|_2\xi_j,$$

where $\lambda = \sigma\sqrt{2\,\ln(\text{card}\,J)}$ is a smoothing parameter chosen according to the soft-thresholding rule commonly used in wavelet theory. This choice is natural because one can prove that with any orthonormal basis $G = \{G_j\}_{j\in J}$ the shrinkage via soft-thresholding produces the same smoothing result $\hat{x}$. (see Bruckstein et al. (2009)). Such approaches are also known as basis pursuit denoising (BPDN).

Solution of this minimization problem with λ close to zero may not be sparse enough: we are searching small $F \subset J$ such that $\hat{x} \approx \sum_{j\in F} \hat{\xi}_j G_j$ is a good approximation. The procedure of BPDN is described in Neubauer and Veselý (2011).

We build our dictionary from heaviside-shaped atoms on $L^2(\mathbb{R})$ derived from a fixed 'mother function' via shifting and scaling following the analogy with the construction of wavelet bases.

We construct an oversized shift-scale dictionary $G = \{G_{a,b}\}_{a\in\mathcal{A},b\in\mathcal{B}}$ derived from the 'mother function' by varying the shift parameter a and the scale (width) parameter b between values from big finite sets $\mathcal{A} \subset \mathbb{R}$ and $\mathcal{B} \subset \mathbb{R}^+$, respectively ($J = \mathcal{A} \times \mathcal{B}$), on a bounded interval $\mathcal{I} \subset \mathbb{R}$ spanning the space $H = \text{sp}(\{G_{a,b}\})_{a\in\mathcal{A},b\in\mathcal{B}}$, where

$$G_{a,b}(t) = \begin{cases} 1 & \text{for} \quad t - a > b/2, \\ 2(t-a)/b & \quad |t-a| \le b/2, b > 0, \\ 0 & \quad t = a, b = 0, \\ -1 & \quad \text{otherwise.} \end{cases}$$

In the simulations below $\mathcal{I} = [0, 1]$, $\mathcal{T} = \{t/T\}$ (typically with mesh size $T = 100$), $\mathcal{A} = \{t/T\}_{t=t_0}^{T-t_0}$ (t_0 is a boundary trimming, $t_0 = 5$ was used in the simulations) and scale b fixed to zero ($\mathcal{B} = \{0\}$). Clearly the atoms of such Heaviside dictionary are normalized on $\mathcal{I}$, i.e. $\|G_{a,0}\|_2 = 1$. Some examples of Heaviside functions are displayed in the figure 1.

Change Point Detection by Basis Pursuit

Neubauer and Veselý (2011) proposed the method of change point detection if there is just one change point in a one-dimensional stochastic process (or in its sample path). We briefly describe a given method. We would like to find a change point in a stochastic process

$$(1) \qquad Y_t = \begin{cases} \mu + \epsilon_t & t = 1, 2, \ldots, c \\ \mu + \delta + \epsilon_t & t = c+1, \ldots, T, \end{cases}$$

where $\mu, \delta \neq 0, t_0 \leq c < T - t_0$ are unknown parameters and ϵ_t are independent identically distributed random variables with zero mean and variance σ^2. The parameter c indicates the change point in the process. Using the basis pursuit algorithm we obtain some significant atoms, we calculate correlation between significant atoms and analyzed process. The shift parameter of the most significant atom or the atom with the highest correlation is taken as an estimator of the change point c. In case of one change point, the estimator based on highest correlation performed slightly better.

Now let us assume the model with p change points

$$(2) \quad Y_t = \begin{cases} \mu + \epsilon_t & t = 1, 2, \dots, c_1 \\ \mu + \delta_1 + \epsilon_t & t = c_1 + 1, \dots, c_2, \\ \cdots & \cdots \\ \mu + \delta_p + \epsilon_t & t = c_p + 1, \dots, T, \end{cases}$$

where $\mu, \delta_1, \dots, \delta_p \neq 0, t_0 \leq c_1 < \cdots < c_p < T - t_0$ are unknown parameters and ϵ_t are independent identically distributed random variables with zero mean and variance σ^2.

We use the method of change point estimation described above for detection of p change points $c_1, \dots, c_p$ in the model (2). Instead of finding only one significant atom or an atom with the highest correlation with the process Y_t we can identify p significant atoms or atoms with the highest correlation. The shift parameters of these atoms determine estimators for the change points $c_1, \dots, c_p$. Another possibility is to apply the procedure of one change point detection p-times in sequence. In the first step we identify one change point in the process Y_t, then we subtract such significant atom from the process (by linear regression)

$$Y_t = \beta G_{0,\hat{c}_1} + e_t,$$
$$Y'_t = Y_t - \hat{\beta} G_{0,\hat{c}_1},$$

and we apply the method to the new process Y'_t. This sequence is repeated until we get p estimates of change point. The shift parameters of selected atoms are again identifiers of the change points $c_1, \dots, c_p$.

Simulation Study

For the purpose of performance study of the proposed method of multiple change point detection we use simulations of the process (2). We put $T = 100$, $\mu = 0$, the error terms are independent normally distributed with zero mean and the standard deviations $\sigma = 0.2$ and 0.5. Locations of change points were chosen at random uniformly in the interval $[5, 95]$, $\delta_i, (i = 1, \dots, p)$, had values 1 or -1 (again randomly generated). For each number of change points ($p = 1, \dots, 15$) 500 simulations of the process (2) were calculated. We applied 4 methods to detect change points:

- the method based on the most significant atoms (according to ℓ_1-minimization),
- the method based on the significant atoms with highest correlation,
- the iterative method based on the most significant atoms,
- the iterative method based on the significant atom with highest correlation.

We calculate empirical probability of successful change point detection as

$$\text{empirical probability} = \frac{\text{number of successful detections}}{\text{number of simulations}}.$$

The detection was a success provided that all change points could be correctly identified. The results are summarized in graphs (see figure 2). The empirical probability of the iterative methods is generally

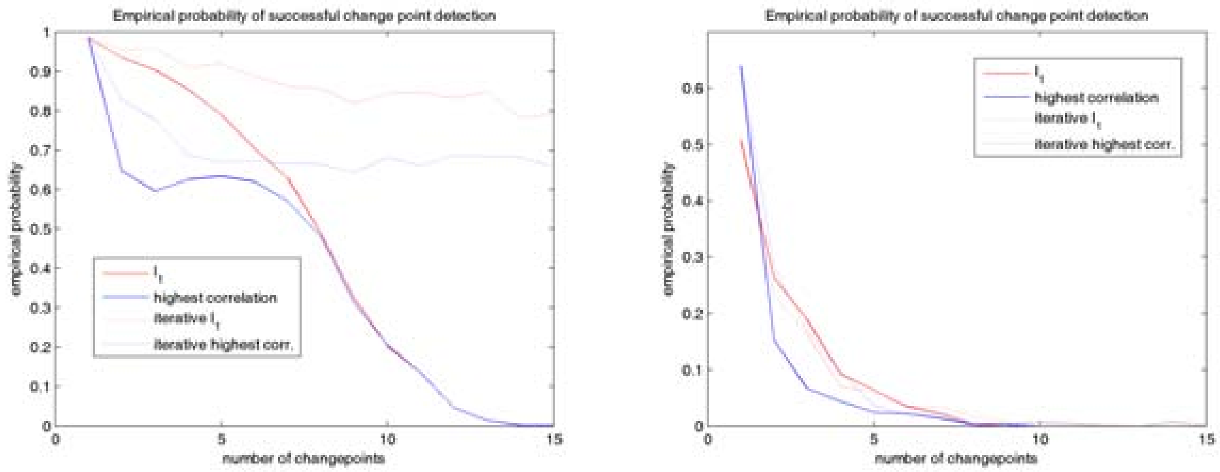

Figure 2: Empirical probability of successful change point detection for $\sigma = 0.2$ and 0.5

higher than the probability of the direct detection methods. We expected that the difference between two neighbouring atoms is at least 5. In the case of iterative methods we exclude all atoms detected in preceding iterations.

Conclusion

According to the simulation results the basis pursuit approach proposes a reasonable detection method of change points in one-dimensional process. In the case of one change point, the method based on significant atoms with highest correlation with the process yields better results than the method based on pure ℓ_1 minimization when the standard deviation of the white noise is large. This is not valid for more than one change point which can be drawn from the graphs of the empirical probability of successful change point detection. The iterative methods performed better than the direct ones. With the increasing number of change points the empirical probability decreases as one can expect. The standard deviation of an additive white noise of the process also affects the probability of successful detection. Higher values of the standard deviation cause lower probability of successful detection.

The change point detection techniques may be useful in a broad range of real-world problems, for instance in modeling of economical or environmental time series where jumps can occur.

Acknowledgment: supported by grants GAČR P402/10/P209 and MSM0021622418

REFERENCES

Bruckstein, A. M. et al. (2009). From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Review*, 51 (1), 34–81.

Chen, S. S. et al. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20 (1), 33-61 (2001 reprinted in *SIAM Review*, 43 (1), 129–159).

Csörgö, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Problem.* Wiley, New York.

Neubauer, J. and Veselý, V. (2011). Change Point Detection by Sparse Parameter Estimation. *Informatica*, 22 (1), 149–164.

Neubauer, J. and Veselý, V. (2010). Multiple Change Point Detection by Sparse Parameter Estimation. In *Proceedings of COMPSTAT'2010, 19th International Conference on Computational Statistics.* p.1453–1459.

RÉSUMÉ (ABSTRACT)

The contribution is focused on multiple change point detection in a one-dimensional stochastic process by sparse parameter estimation from an overparametrized model. Detection of changes has originally arisen in the context of quality control. Nowadays, we can find wide range of fields where change point problem is applied, such as epidemiology, medicine (rhythm analysis), ecology, signal processing etc. Standard statistical approach to change point analysis is described, for instance, in Csörgó and Horváth (1997). Authors' approach to change point detection is quite different. A stochastic process residing in a bounded interval with changes in the mean is estimated using dictionary (a family of functions, the so-called atoms, which are overcomplete in the sense of being nearly linearly dependent) and consisting of Heaviside functions. Among all possible representations of the process we want to find a sparse one utilizing a significantly reduced number of atoms. This problem can be solved by ℓ_1*-minimization (see Bruckstein et al. (2009), Chen et. al (1998)). The basis pursuit algorithm is used to get sparse parameter estimates. Basic properties of this approach were studied in Neubauer and Veselý (2011) and in Neubauer and Veselý (2010) by simulations. According to these results, the basis pursuit approach proposes an alternative technique of the change point detection. In this contribution the authors calculate empirical probability of successful change point detection as a function depending on the number of change points and the level of standard deviation of an additive white noise of the stochastic process. The empirical probability was computed by simulations where locations of change points were chosen randomly from uniform distribution. Such probability decreases with increasing number of change points and/or standard deviation of white noise.*