0. 論文

DETECTION OF MULTIPLE CHANGE-POINTS IN MULTIVARIATE TIME SERIES

M. Lavielle

Laboratoire de Mathématiques, Université René Descartes et Université Paris–Sud
(e-mail: marc.lavielle@math.u-psud.fr)

G. Teyssière

Statistique Appliquée et MOdélisation Stochastique, CES, Université Paris 1 Panthéon–Sorbonne

タイトル：DETECTION OF MULTIPLE CHANGE-POINTS IN MULTIVARIATE TIME SERIES
著者： Marc Lavielle, Gilles Teyssiere

arXiv投稿日：
学会/ジャーナル： Lithuanian Mathematical Journal 2006

1. どんなもの？

- 変化点数が未知の場合の多次元系列の多重変化点検出を考える
- 強依存過程の系列データ
- 注目するのは変化点で急激に分散構造が変化するものを変化点とする
- 提案する手法は系列間に対して，iid, 弱，強依存な場合のデータに対して変化点を検出できる(これ以外の設定を知らない)
- 実応用としては相関のある市場の動きの変化点検出を考えているみたい

2. 先行研究

- バイナリーセグメンテーション
  - 変化点が存在しなくなるまでセグメンテーションを行う
    - 変化点数を過剰に評価する問題点がある
- 他にもかなり多くの変化点検出論文がある
  - この中で系列の選択について議論しているものはなさそう
- 元論文
  - Detection of multiple changes in a sequence of dependent variables
  - これは変化点数が分かっている場合の検出手法になっている
  - 多次元系列の平均や分散構造の変化を捉える実用的な手法になっている

3. コアアイディア

コスト関数を分散行列に対して定義

- 平均が変化しないモデルは以下の関数を使う

$$J(\bm\tau, \bm Y) = \frac{1}{n}\sum_{k=1}^{K} n_k \log|\hat{\Sigma}_{\tau_k}|$$

- ただし，$\hat{\Sigma}_{\tau_k} = \frac{1}{n_k}\sum_{t=\tau_{k-1}+1}^{\tau_k}(\bm{Y} - \bar{\bm{Y}})(\bm{Y} - \bar{\bm{Y}})^{\top}$で$\bar{\bm{Y}}=\frac{1}{n}\sum_{t=1}^n\bm{Y}_t$として推定する(分散構造のみの変化に注目している)
- 平均も変化している場合は$\bar{\bm{Y}}$の部分が$\bar{\bm{Y}}_{\tau_k}$となる
- $\bar{\bm{Y}}_{\tau_k}=\frac{1}{n_k}\sum_{t=\tau_{k-1}+1}^{\tau_k}\bm{Y}_t$となる

罰則を定義

- イメージ的にはBICとAICと同様に変化点数に関して罰則を付ける
- 罰則項は次の通り

$$\boxed{\beta\left\{K(\bm{\tau})\left(1+c\log \frac{n}{K(\bm{\tau})}\right)\right\}}$$

- $\beta = \frac{2\sigma^2}{n}$
- cはハイパーパラメータでσ^2は系列が持っている分散の値？

最終的な最適化問題

$$\hat{\bm\tau}(\beta) = \argmin_{\bm\tau} J(\bm\tau, \bm Y) + \beta pen(\bm\tau)$$

- 変化点数はどうやらp値を見て決めるみたい

$$\mathcal{P}_{K_i} = P\left(e_{K_i-1} \geqslant J_{K_i-1} - \hat{c}_1(K_i-1) + \hat{c}_2(K_i-1)\log(K_i-1)\right),$$

4. どうやって有効だと検証した？

- 人工的に分散構造を変化させたデータセット
  - Schwarz criteria vs 提案手法で精度の比較を行う
  - Schwarz criteria

Table 1. Average number of detected change-points and their location using the Schwarz criteria, $n = 500$, $\tau_1 = 200$, $\tau_2 = 350$ (based on 5000 replications). Standard errors are given in parentheses.

DGP	Number of change-points	$\hat{\tau}_1$	$\hat{\tau}_2$
DGP 0	0.2590 (0.59)	—	—
DGP 1	2.3148 (0.67)	191.1580 (52.11)	324.4140 (59.33)
DGP 2	2.3310 (0.66)	187.6320 (41.33)	325.5490 (62.23)
DGP 3	2.1626 (1.47)	—	—
DGP 4	3.8324 (1.55)	145.6920 (73.07)	243.7830 (100.99)

  - 提案手法

Table 4. Average number of detected change-points and their location using the adaptive method, $n = 500$, $\tau_1 = 200$, $\tau_2 = 350$ (based on 5000 replications). Standard errors are given in parentheses.

DGP	Number of change-points	$\hat{\tau}_1$	$\hat{\tau}_2$
DGP 0	0.1248 (0.62)	—	—
DGP 1	1.7974 (0.52)	236.1160 (69.32)	345.6080 (27.48)
DGP 2	1.8290 (0.61)	196.7920 (25.61)	342.9910 (43.26)
DGP 3	0.2962 (0.90)	—	—
DGP 4	1.5650 (0.83)	217.1770 (64.31)	330.1390 (61.25)

  - 変化点の数もより正確に推定できている
  - パラメータを変えても同じ
- real financial markets データセット

- 2変量の系列データ
- 変化点数の推定を行うと$K = 6$と推定された

K_i	β_i	β_{i-1}	l_i	P_i
1	692.97	Inf	Inf	–
4	203.22	598.20	394.98	1.76 e-011
6	85.57	203.22	117.65	8.81 e-010
11	41.46	65.88	24.42	0.0134
19	29.68	37.03	7.35	0.2808
24	22.18	27.60	5.42	0.0068

-
- 結果

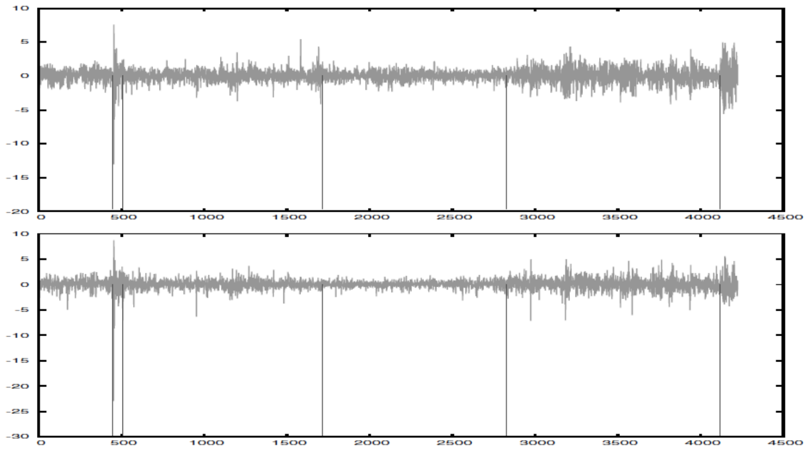

Figure 3. Adaptive detection. Above: The series of returns on FTSE 100 with the estimated change-points represented by vertical lines; Below: The series of returns on S&P 500 with the estimated change-points represented by vertical lines.

5. データセット

- 市場データ FTSE 100
- S&P500と呼ばれるデータ

6. 疑問点

- なぜ分散構造が変化していることをかんがえているのか？
  - 平均が変化していないor変化しているは考えているのに
  - 選考研究がどこかにあるかもしれない
  - 拡張としては簡単にできそうだが
- 見た感じこれも系列の選択はできない
  - 多次元系列の変化点検出は全ての系列に変化がある仮定が一般的なのかもしれない

7. 次に読むべき論文は？

- この系統をよんでも仕方ないのでcusum系の論文を読む

キーワード

- 分散の変化点検出
- 多次元系列データ