Changepoint Detection or: How I Learned to Love the Normal Distribution

David Veitch

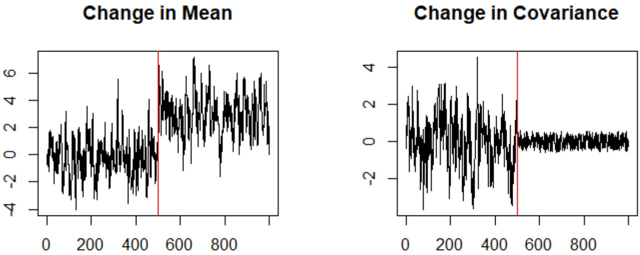
February 5 2021

Agenda

What is Changepoint Detection

$x_1, \ldots, x_n \in \mathbb{R}^p$ are observations which are ordered sequentially.

Changepoint detection seeks to determine at what point did the distribution change from X_i to X_{i+1}.



Changepoint Detection - Areas of Interest

- Univariate vs. multivariate vs. high dimensional
- Detecting one vs. multiple changepoints
- Changepoints in mean, covariance, distribution
- Changepoints in oscillation [Zhou et al., 2020]
- Changepoints for non-stationary data [Zhou, 2013a]
- Multiscale changepoint detection [Wu and Zhou, 2020]
- Changepoints in 'non-traditional data' such as graphs [Chen and Zhang, 2015] or images [Shi et al., 2018]

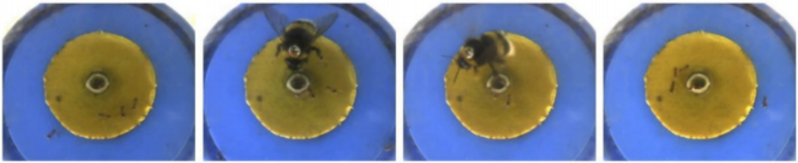
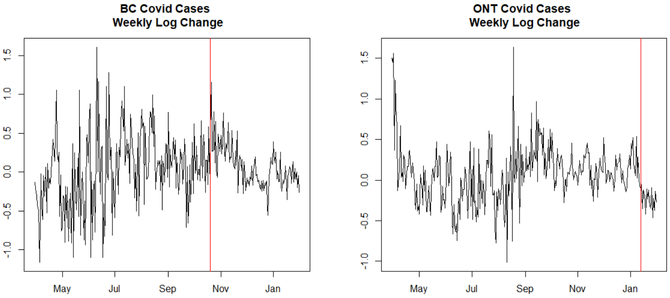


Fig. 1. Extracted frames with dimension 288 $\times$ 352 located at 1, 5, 40, and 49 from the original video 09-10-2010_15h_49_27.17.mpg (faculty.tru.ca/xshi/09-10-2010_15h_49_27.17.mpg). The landing and departure times of the bee are 5 and 41, respectively.

Current Research Project

Multiscale changepoint detection for the means of high-dimensional non-stationary time series.



Do these red lines represent real changepoints or are they just noise?

Data violates many assumptions current methods make.

Gaussian Approximation - Motivation

Why is this related to changepoint detection?

Let $x_i = (x_{i,1}, \ldots, x_{i,10}) \in \mathbb{R}^{10}$ be a vector of the log change of each province's case numbers at date i. We want to detect if one province has a big spike/decrease in this (maybe a policy change radically affected behaviour).

Take a window of size $2m$, if a change occurred at some date in one province's numbers, then the following statistic should be big

$$T = \max_{1 \le j \le n} \max_{1 \le r \le 10} \left| \left(\frac{1}{\sqrt{m}} \sum_{i=j-m-1}^{j-1} x_{i,r} \right) - \left(\frac{1}{\sqrt{m}} \sum_{i=j}^{j+m} x_{i,r} \right) \right| \tag{1}$$

$$T \approx \max_{1 \le j \le n} \max_{1 \le r \le 10} \text{Change in mean of province } r \text{ at date } i \tag{2}$$

Gaussian Approximation - Motivation

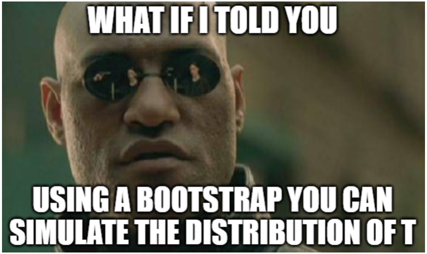
$$T = \max_{1 \le j \le n} \max_{1 \le r \le 10} \left| \left(\frac{1}{\sqrt{m}} \sum_{i=j-m-1}^{j-1} x_{i,r} \right) - \left(\frac{1}{\sqrt{m}} \sum_{i=j}^{j+m} x_{i,r} \right) \right| \tag{3}$$

Problem!
How do we know what a sufficiently large value of T is to reject the null hypothesis that there is no changepoint?

$$T = \max_{1 \leq j \leq n} \max_{1 \leq r \leq 10} \left| \left(\frac{1}{\sqrt{m}} \sum_{i=j-m-1}^{j-1} x_{i,r} \right) - \left(\frac{1}{\sqrt{m}} \sum_{i=j}^{j+m} x_{i,r} \right) \right| \tag{4}$$

Problem!
How do we know what a sufficiently large value of T is to reject the null hypothesis that there is no changepoint?

Solution!



Gaussian Approximation

Many of the ideas used to accomplish this originate in the follow paper Chernozhukov et al. [2013]:

The Annals of Statistics
2013, Vol. 41, No. 6, 2786–2819
DOI: 10.1214/13-AOS1161

GAUSSIAN APPROXIMATIONS AND MULTIPLIER BOOTSTRAP FOR MAXIMA OF SUMS OF HIGH-DIMENSIONAL RANDOM VECTORS

BY VICTOR CHERNOZHUKOV[1], DENIS CHETVERIKOV[1] AND KENGO KATO[2]

Massachusetts Institute of Technology, University of California, Los Angeles, and University of Tokyo

Gaussian Approximation - Setup

$x_1, \ldots, x_n \in \mathbb{R}^p$ be **independent** random variables that are centred $\mathbb{E}[x_i] = 0$. Let $x_{i,j}$ be the j-th coordinate of x_i.

$$X = (X_1, \ldots, X_p)^T = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i. \tag{5}$$

$y_1, \ldots, y_n \in \mathbb{R}^p$ are **Gaussian** $\mathcal{N}(0, \mathbb{E}[x_i x_i^T])$

$$Y = (Y_1, \ldots, Y_p)^T = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_i. \tag{6}$$

By CLT one would think that $X \overset{d}{\approx} Y$.

Gaussian Approximation - Test Statistics

Define the following test statistics

$$T_0 = \max_{1 \leq j \leq p} X_j \tag{7}$$

$$Z_0 = \max_{1 \leq j \leq p} Y_j. \tag{8}$$

The **Kolmogorov distance** between the distributions T_0 and Z_0 as follows

$$\rho = \sup_{t \in \mathbb{R}} |P(T_0 \leq t) - P(Z_0 \leq t)|. \tag{9}$$

Gaussian Approximation - Quality of Gaussian Approximation as Function of n, p

Lemma 2.3 (A Simple GAR) of the paper states that for sufficiently 'nice' random variables we have

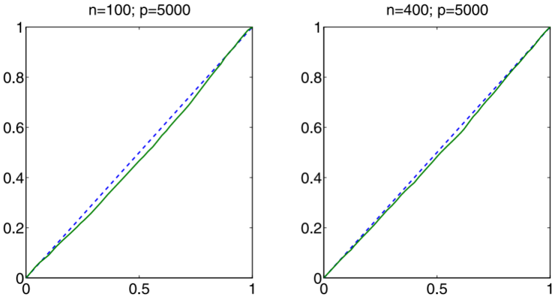
$$\rho = \sup_{t \in \mathbb{R}} |P(T_0 \leq t) - P(Z_0 \leq t)| \leq C \left(\frac{(\log(pn))^7}{n} \right)^{1/8}. \tag{10}$$

This means if $p >> n$ a Gaussian approximation should still work.

Gaussian Approximation - Quality of Gaussian Approximation as Function of n, p

n	30	30	100	100	1,000	1,000	1,000
p	10	30	50	100	250	1,000	100,000
Bound	1.92C	2.08C	2.41C	2.33C	2.34C	2.49C	3.59C

So if you have quite a good approximation at $n = 30, p = 10$ then you should still have a decent approximation for $n = 1,000, p = 100,000$!



Gaussian Approximation - The Catch

So the above says for all types of independent random variables we can approximate the distribution of

$$T_0 = \max_{1 \le j \le p} X_j. \tag{11}$$

Gaussian Approximation - The Catch

So the above says for all types of independent random variables we can approximate the distribution of

$$T_0 = \max_{1 \le j \le p} X_j. \tag{12}$$

But only if we have a sequence of Gaussian random variables $y_1, \ldots, y_n \in \mathbb{R}^p$ where each y_i's covariance matrix is the same as x_i.

What do we do?

We can approximate Z_0 with another statistic!

Recap

1. Want to know distribution of

$$T_0 = \max_{1 \le j \le p} X_j$$

2. Approximate distribution of T_0 using Gaussians by

$$Z_0 = \max_{1 \le j \le p} Y_j$$

3. Approximate distribution of Z_0 by W_0
4. Then

$$W_0 \overset{d}{\approx} T_0$$

The Multiplier Bootstrap

Treat your data $x_1, \ldots, x_n$ as fixed

$$
\begin{aligned}
W &= (W_1, \ldots, W_p) && (13)\\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i e_i && (14)\\
e_i &\sim \mathcal{N}(0, 1) && (15)\\
W_0 &= \max_{1 \leq j \leq p} W_j. && (16)
\end{aligned}
$$

The Multiplier Bootstrap

W should have a covariance structure that is similar to $\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i$ since conditioned on the data, for two coordinates a, b

$$\mathrm{Cov}(W_a, W_b) = \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n x_{i,a} e_i\right)\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n x_{i,b} e_i\right)\right] \tag{17}$$

$$= \frac{1}{n} \sum_{i=1}^n x_{i,a} x_{i,b} \tag{18}$$

$$\mathrm{Cov}(X_a, X_b) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[x_{i,a} x_{i,b}] \tag{19}$$

And by the law of large numbers (Theorem 2.2.4 in Durrett!)

$$\frac{1}{n} \sum_{i=1}^n x_{ia} x_{ib} \xrightarrow{p} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[x_{ia} x_{ib}] \tag{20}$$

The Multiplier Bootstrap - Key Result

Main Result 2 (Validity of Bootstrap for high dimensional means)

$$\sup_{\alpha \in (0,1)} |P(T_0 \leq c_{W_0}(\alpha)) - \alpha| \leq \rho_{\ominus} + \rho \tag{21}$$

Essentially, the quantiles you get by repeatedly resampling W_0 are approximately the quantiles of the test statistic T_0 of the original data.

The Multiplier Bootstrap - Extensions

- [Zhou, 2013b] uses a multiplier bootstrap to estimate the covariance of a nonstationary time series.
- [Zhang et al., 2017] extend Gaussian approximations to high dimensional stationary time series where they use a block method to estimate the long run variance matrix. Here they use it to estimate $|X|_\infty$ (which is closer to what we our seeking to use).

Gaussian Approximation and Changepoint Detection

1. Have shown that Gaussian approximation for sums of high dimensional vectors still works for $p >> n$
2. Extensions would suggest it could still work if there is dependence in the data, and even if dependence is changing over time

Changepoint Detection and Gaussian Approximation

Can do something like the following, for some window m let

$$H_m(j,r) = \left| \left(\frac{1}{\sqrt{m}} \sum_{i=j}^{j+m} x_{i,r} \right) - \left(\frac{1}{\sqrt{m}} \sum_{i=j-m-1}^{j-1} x_{i,r} \right) \right| \tag{22}$$

be the evidence a change in the mean of a high dimensional time series occurred in dimension r at time j. Under null hypothesis that no change points occured $H_m(j,r) \approx 0 \; \forall \; j, r$.

Then create a very big vector $\mathbf{H} \in \mathbb{R}^{(n-2m)\times r}$ and do Gaussian approximation on it to estimate the quantiles under the null hypothesis of

$$H_0 = \max_{m \le j \le n-m, \; 1 \le r \le p} H_m(j,r). \tag{23}$$

Can then reject null hypothesis of no changepoint if H_0 is too big.

Thank you!

References I

Hao Chen and Nancy Zhang. Graph-based change-point detection. *Ann. Statist.*, 43(1):139–176, 02 2015. doi: 10.1214/14-AOS1269. URL `https://doi.org/10.1214/14-AOS1269`.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.*, 41(6):2786–2819, 12 2013. doi: 10.1214/13-AOS1161. URL `https://doi.org/10.1214/13-AOS1161`.

Xiaoping Shi, Yuehua Wu, and Calyampudi Radhakrishna Rao. Consistent and powerful non-euclidean graph-based change-point test with applications to segmenting random interfered video data. *Proceedings of the National Academy of Sciences*, 115(23):5914–5919, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1804649115. URL `https://www.pnas.org/content/115/23/5914`.

Weichi Wu and Zhou Zhou. Multiscale jump testing and estimation under complex temporal dynamics, 2020.

Danna Zhang, Wei Biao Wu, et al. Gaussian approximation for high dimensional time series. *Annals of Statistics*, 45(5):1895–1919, 2017.

References II

Zhou Zhou. Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association*, 108 (502):726–740, 2013a. ISSN 01621459. URL http://www.jstor.org/stable/24246477.

Zhou Zhou. Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association*, 108 (502):726–740, 2013b.

Zhou Zhou, Yang-Guan-Jian Guo, and Hau-Tieng Wu. Frequency detection and change point estimation for time series of complex oscillation, 2020.