Detectability of Changepoints Using the Likelihood Ratio Test Statistic

Katharina Ring[1]; Supervisors: Thomas Grundy[2], Mirjam Kirchner[2]

[1]LMU Munich, Department of Statistics, [2]STOR-i, Lancaster University

September 6, 2019

Changepoints

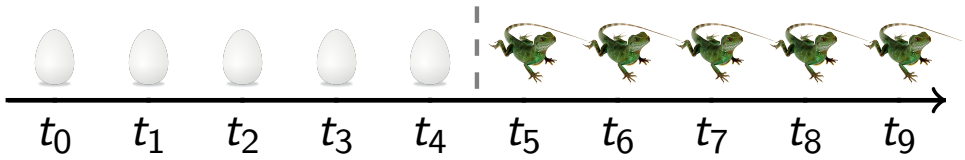


Figure 1: A time series exhibiting a change at t_4

- changepoint detection using the Likelihood Ratio Test
- find detection boundary, separating detectable from undetectable changes
- through a simulation study:
  - before the change: $y \overset{iid}{\sim} N(0,1)$
  - after the change: $y \overset{iid}{\sim} N(\mu, \sigma^2)$ with $\mu \neq 0$ and/or $\sigma^2 \neq 1$

Likelihood Ratio Test

The Likelihood Ratio Test as a binary classifier:

$$H_0 : \text{no changepoint}$$
$$H_1 : \text{one changepoint}$$

$$\lambda = 2 \left(\max_{\tau} \underbrace{\left[\log p(y_{1:\tau}|\hat{\theta}_1) + \log p(y_{\tau+1:n}|\hat{\theta}_2) \right]}_{\text{Log-Likelihood with changepoint } \tau} - \underbrace{\log p(y_{1:n}|\hat{\theta})}_{\substack{\text{Log-Likelihood without} \\ \text{a changepoint}}} \right)$$

$\Rightarrow$ Reject H_0 if λ surpasses a given threshold c.

Detectability

Definition of detectability:

c is chosen s. t. the true positive rate (empirical power) is 80 %
$\Rightarrow$ false positive rate (empirical type I error) is at most 5 %.

Get points for estimation:

- *points of interest* where the ROC passes the boundary region, i. e. the point (0.05, 0.8) with distance < 0.01
- find these points using Surrogate Model Bayesian Optimization

$\rightarrow$ take median of points to estimate boundary

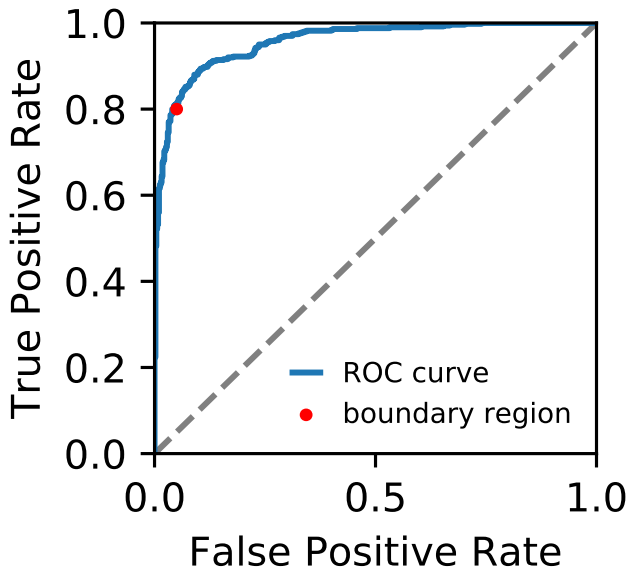
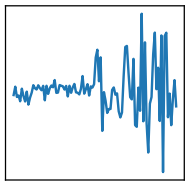


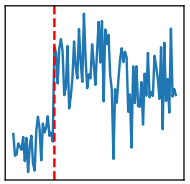
Figure 2: ROC curve example

Univariate Analysis & Results – Estimation of Boundary

Grid Variables:

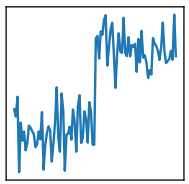


variance change



location

Target:



mean change

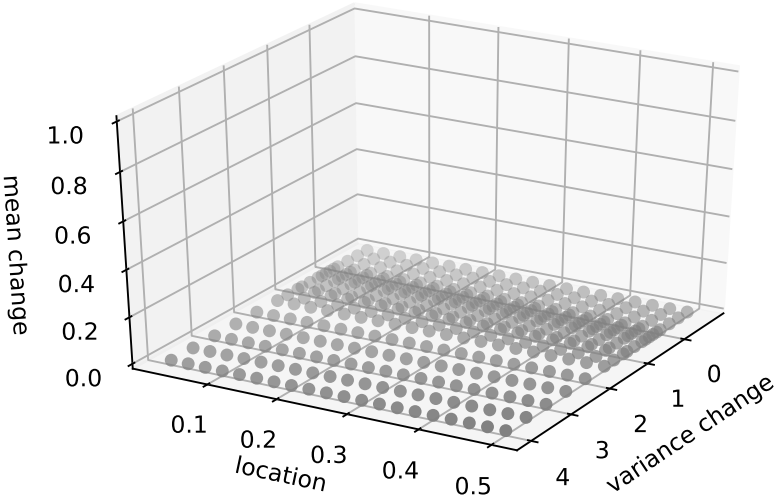


Figure 3: Grid Points

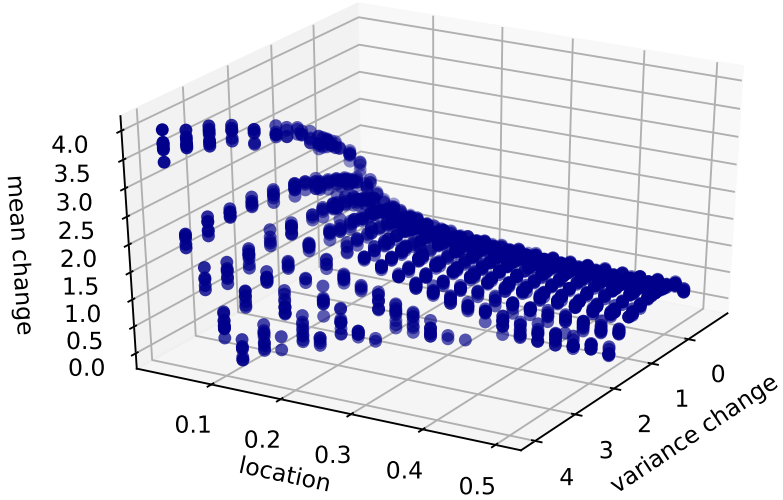
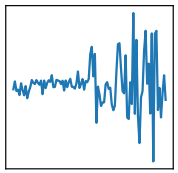


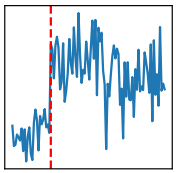
Figure 4: Points of interest

Univariate Analysis & Results

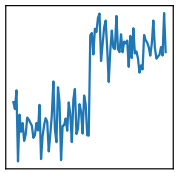
Variables of interest:



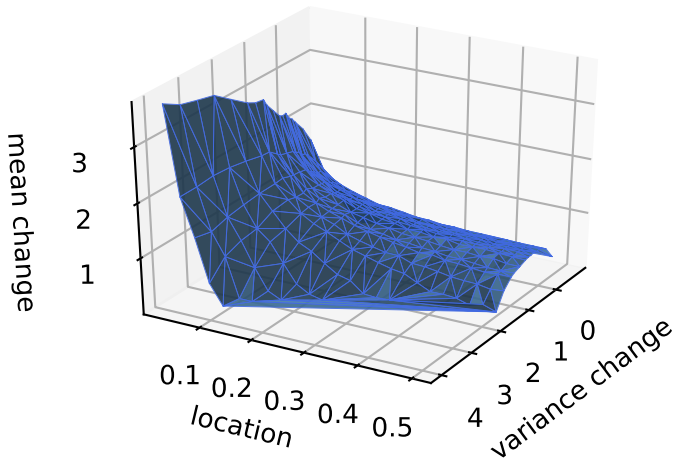
variance change



location



mean change



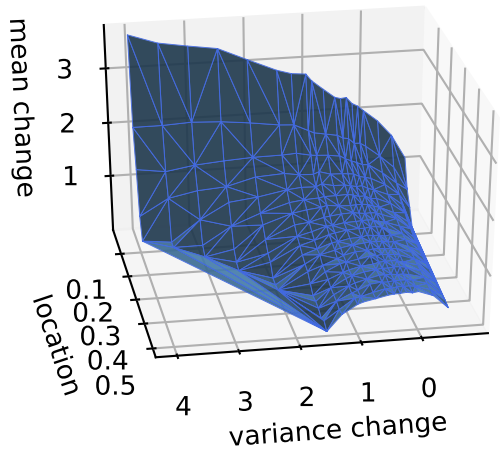
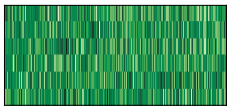


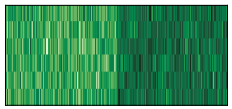
Figure 5: The detection boundary for the univariate case

Multivariate Analysis & Results

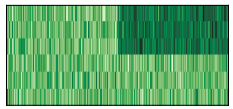
Variables of interest:



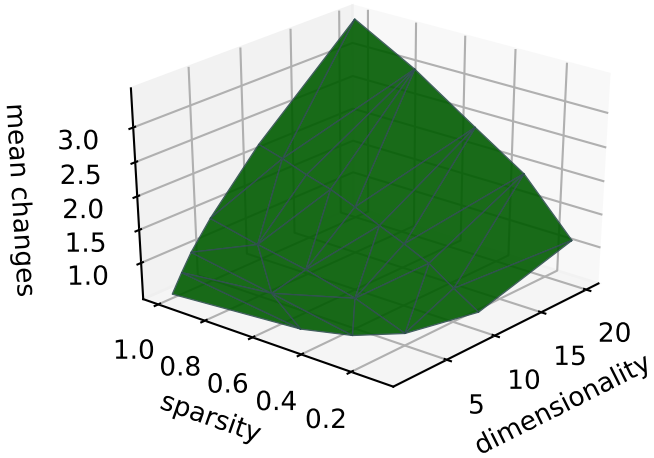
dimensionality



mean changes



sparsity



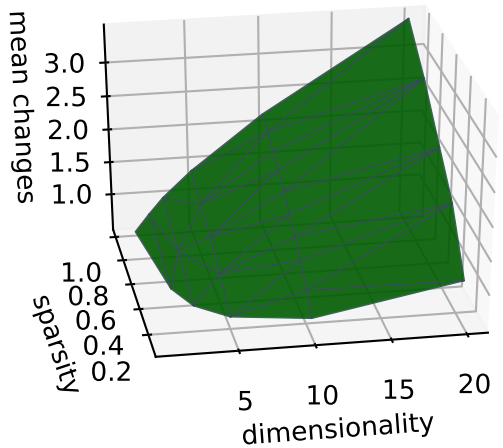


Figure 6: The detection boundary for the multivariate case

Future Work

More variables of interest can be investigated:

- univariate case: length of time series
- multivariate case: length of time series, size of variance change, and location

Additionally: multiple changepoints per time series.

Problem: computationally expensive

Possible solutions:

- better optimization algorithm
- more efficient LRT calculation, e. g. cusum

References

Eckley, I., Fearnhead, P., and Killick, R. (2011). Analysis of Changepoint Models.
In *Bayesian time series models*, Cambridge University Press.

Cai, T., Jeng, J., and Jin, J. (2011). *J. R. Statist. Soc. B*, 73(5): 629–662.
Optimal detection of heterogeneous and heteroscedastic mixtures.