Nonparametric changepoint detection

May 8, 2023

Abstract

Keywords: nonparametric maximum likelihood, PELT, multiple changepoint, penalty selection.

1 Introduction

PELT: (Killick et al., 2012)

Nonparametric: (Ross, 2021; Ross and Adams, 2012; Matteson and James, 2014; Haynes et al., 2017b,a; Pettitt, 1979; Zou et al., 2014)

Empirical CDF: (Zou et al., 2014; Haynes et al., 2017a) Nonparametric-online: (Austin et al., 2023; Romano et al., 2023)

Contributions:

- Compared to Haynes et al. (2017b), we propose a new cost function + penalty based on a proper theory.
- The proposed cost function targets the most general alternative and we show under which alternative our conditional idea works could have higher power than the original one.
- Any improvement of computational complexity? - NOT REALLY, linear in the number of quantiles $O(Kn \log n)$

[1]Department of Mathematical Sciences, Durham University. Email: `hyeyoung.maeng@durham.ac.uk`.
[2]Department of Mathematics and Statistics, Lancaster University. Email: `i.eckley@lancaster.ac.uk`.
[3]Department of Mathematics and Statistics, Lancaster University. Email: `p.fearnhead@lancaster.ac.uk`.

2 Methodology

2.1 Background and motivation

We assume that data points $X_{i,\ i=1,\dots,n}$ are independent and identically distributed with F, the cumulative distribution function (CDF). Under the non-parametric framework, Zou et al. (2014) consider the empirical CDF of the sample

$$\hat{F}_{1:n}(q) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \leq q\}, \tag{1}$$

which leads to

$$n\hat{F}_{1:n}(q) \sim \text{Binomial}(n, F(q)),$$

where q is a fixed quantile. Considering the sample as binary data allows us to write the maximum log-likelihood as

$$\mathcal{L}(X_{1:n}; q) = n\hat{F}_{1:n}(q)\log\{\hat{F}_{1:n}(q)\} + \{n - n\hat{F}_{1:n}(q)\}\log\{1 - \hat{F}_{1:n}(q)\}. \tag{2}$$

We now consider the single change point problem where there is one change point τ such that

$$X_i \sim \begin{cases} F_1(x), & 1 \leq i \leq \tau_1, \\ F_2(x), & \tau_1 + 1 \leq i \leq n. \end{cases} \tag{3}$$

Then the likelihood ratio test (LRT) statistic for detecting a change at a fixed quantile q can be defined using (2) as follows,

$$\max_{1\leq\tau\leq n} 2[\mathcal{L}(X_{1:\tau}; q) + \mathcal{L}(X_{(\tau+1):n}; q) - \mathcal{L}(X_{1:n}; q)]. \tag{4}$$

As q is a fixed constant, for a fixed τ, the test statistic in (4) asymptotically follows a chi-squared distribution with degrees of freedom equals to one under the null hypothesis that there is no change as $n \to \infty$ (Wilks, 1938).
However, the choice of q in (4) hugely affects the detection problem in (3). To overcome this, Zou et al. (2014) and uses the integration over different values of quantile, $q_1, \dots, q_Q$, which can be approximated as follows (Haynes et al., 2017a):

$$T(Q) = \max_{1\leq\tau\leq n} \frac{1}{Q}\sum_{j=1}^{Q} 2\left[\mathcal{L}(X_{1:\tau}; q_j) + \mathcal{L}(X_{(\tau+1):n}; q_j) - \mathcal{L}(X_{1:n}; q_j)\right]. \tag{5}$$

To see how two test statistics in (4) and (5) work differently, we first examine the limiting distributions under the null hypothesis where there is no change. The LRT statistic in (4) can be approximated by the second order Taylor expansion as follows:

$$\begin{aligned} 2[\mathcal{L}(X_{1:\tau};q)+\mathcal{L}(X_{(\tau+1):n};q)-\mathcal{L}(X_{1:n};q)] &\approx \frac{\big[m(\tau,q)-\mathrm{E}\{m(\tau,q)\}\big]^2}{\mathrm{Var}\{m(\tau,q)\}} \\ &\approx \frac{\varepsilon^2}{\mathrm{Var}(\varepsilon)} \end{aligned}$$

where $m(\tau,q)=\tau\hat{F}_{1:\tau}(q)=\sum_{i=1}^{\tau}\mathbb{1}\{X_i\le q\}$ and $\varepsilon=\{m(\tau,q)-\tau\hat{F}_{1:n}(q)\}/\sqrt{n}$. The details of approximation are addressed in the Appendix. Thus, by the central limit theorem, the test statistic in (4) which is asymptotically follows χ_1^2 under the null hypothesis, while the limiting distribution of the one in (5) is not clear, as the test statistic in (5) is approximated as

$$\frac{1}{Q}\sum_{j=1}^{Q}\frac{\big[m(\tau,q_j)-\mathrm{E}\{m(\tau,q_j)\}\big]^2}{\mathrm{Var}\{m(\tau,q_j)\}}\approx\frac{1}{Q}\sum_{j=1}^{Q}\frac{\varepsilon_j^2}{\mathrm{Var}(\varepsilon_j)},\tag{6}$$

which has the form of average of Q dependent χ_1^2 distributions, where

$$\varepsilon_j=\frac{1}{\sqrt{n}}\big[m(\tau,q_j)-\tau\hat{F}_{1:n}(q_j)\big]=\frac{1}{\sqrt{n}}\left[\sum_{i=1}^{\tau}\mathbb{1}\{X_i\le q_j\}-\frac{\tau}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i\le q_j\}\right].\tag{7}$$

Using the test statistic in (5), Haynes et al. (2017a) detects a changepoint if

$$T(Q)\ge\beta,\tag{8}$$

where β is the threshold for the test and using $\beta=\log(n)$ is recommended by following Yao et al. (1988) which studies the classical least-squares multiple change point detection problem under the parametric framework. Considering the tail behaviour of the chi-squared distribution (Laurent and Massart, 2000), using the thresthold $\beta=\log(n)$ is appropriate when the limiting distribution is well obtained as in (4) where only one quantile is used in the test, while applying the same threshold for the test in (5) could possibly leads to failure in controlling the false positive rate as empirically shown in Ross (2021). This is due to the lack of theoretical justification of the limiting distribution of the test statistic for the single change point problem under the null hypothesis.

Our contribution is filling this gap by proposing a new cost function which has an exact limiting distribution built on a set of quantiles.

2.2 Cost function based on conditional distribution

We propose a new cost function built on a set of quantiles as follows:

$$\mathcal{L}^c(X_{1:n}; q_1, \ldots, q_Q) = \sum_{j=1}^{Q+1} d_j \log \frac{d_j}{n}, \tag{9}$$

where

$$d_j = \sum_{i=1}^{n} \mathbb{1}\{q_{j-1} < X_i \le q_j\}, \tag{10}$$

with $q_0 = 0$ and $q_{Q+1} = \infty$. This conditional log-likelihood is based on the conditional CDF in the sense that d_j in (10) can be rewritten as a function of empirical CDF as follows:

$$d_j = \hat{F}^c_{1:n}(q_j | q_{j-1}) = \hat{F}_{1:n}(q_j) - \hat{F}_{1:n}(q_{j-1}). \tag{11}$$

Then the new test statistic based on the conditional cost function is proposed as follows:

$$T^c(Q) = \max_{1 \le \tau \le n} 2\Big[\mathcal{L}^c(X_{1:\tau}; q_1, \ldots, q_Q) + \mathcal{L}^c(X_{(\tau+1):n}; q_1, \ldots, q_Q) - \mathcal{L}^c(X_{1:n}; q_1, \ldots, q_Q)\Big]. \tag{12}$$

Unlike the existing test statistic in (5) whose cost function is obtained by averaging over the log-likelihood built on a set of quantiles, the one in (12) is based on the conditional log-likelihood which still uses a set of quantiles but in a different way i.e. considering $K+1$ subsegments divided by K quantiles. One of the biggest advantage of using this conditional idea is that the new test statistic in (12) asymptotically follows the chi-squared distribution with degrees of freedom equal to Q under the null hypothesis, which enables us to control the false positive rate. This limiting distribution is achieved by the following approximation,

$$T^c(Q) \approx \boldsymbol{\varepsilon}^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}, \tag{13}$$

where ε_j is defined in (7), $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_Q)^\top$ and $\boldsymbol{\Sigma}_\varepsilon$ is the covariance of $\boldsymbol{\varepsilon}$. As shown in (13), the novelty of our proposal comes from the test statistic built by considering the dependent structure of $\boldsymbol{\varepsilon}$. The details of the approximation in (13) can be found in the Appendix.

2.3 Cost function based on ℓ_∞ aggregation

Instead of averaging the cost function over a set of quantiles as in (5), we could detect a change by choosing a maximum as follows

$$T^\infty(Q) = \max_{1 \le \tau \le n} \max_{1 \le j \le Q} 2\Big[\mathcal{L}(X_{1:\tau}; q_j) + \mathcal{L}(X_{(\tau+1):n}; q_j) - \mathcal{L}(X_{1:n}; q_j)\Big], \tag{14}$$

where $\mathcal{L}(X_{i:j}|q_j)$ is defined as in (2).
UP TO HERE

2.4 Multiple change point scenario

We consider the multiple change point problem such that

$$X_i \sim F_k(x), \qquad \text{for} \ \ \tau_k + 1 \leq i \leq \tau_{k+1}, \ i = 1, \dots, n, \ k = 0, \dots, K_n,$$

where F_k is the cummulative distribution function (CDF) of segment $[\tau_k + 1, \tau_{k+1}]$, K_n is the true number of change-points and $\tau_{k, \ k=1,\dots,K_n}$ are the locations of true change points such that $F_k \neq F_{k+1}$. By convention, we denote $\tau_0 = 1$ and $\tau_{K_n+1} = n$.

3 Theoretical results

4 Numerical studies

Competitive methods:

1. NP-PELT: changepoint.np
2. Ross (2021): npwbs
3. James and Matteson (2015): ecp
4. Padilla et al (2021): `https://github.com/hernanmp/`

The empirical results obtained under Dirichlet distribution show that out conditional idea beats the original LRT when the sign of change in probability is somewhat random (rather than locally grouped).
So, simulations could be two parts? one based on LRT statistics for two different types of conditional probabilities generated via Dirichlet distribution and the other using PELT-based algorithm with generated data under change in mean or change in variance or change in distribution.
Quick check shows that underperformance of the original method does not necessarily mean that the sign of change in probability is random. How do we connect these two results?
To do in simulations:

- Find the best working threshold constant for the conditional idea.
- Comparing with the competing methods.

Questions:

- Why max and avg aggregation work well when K is pretty large?

5 Data applications

6 Conclusions

References

Austin, E., Romano, G., Eckley, I. A., and Fearnhead, P. (2023). Online non-parametric changepoint detection with application to monitoring operational performance of network devices. *Computational Statistics & Data Analysis*, 177:107551.

Haynes, K., Eckley, I. A., and Fearnhead, P. (2017a). Computationally efficient changepoint detection for a range of penalties. *Journal of Computational and Graphical Statistics*, 26:134–143.

Haynes, K., Fearnhead, P., and Eckley, I. A. (2017b). A computationally efficient nonparametric approach for changepoint detection. *Statistics and computing*, 27:1293–1305.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107:1590–1598.

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, pages 1302–1338.

Matteson, D. S. and James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109:334–345.

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28:126–135.

Romano, G., Eckley, I., and Fearnhead, P. (2023). A log-linear non-parametric online changepoint detection algorithm based on functional pruning. *arXiv preprint arXiv:2302.02718*.

Ross, G. J. (2021). Nonparametric detection of multiple location-scale change points via wild binary segmentation. *arXiv preprint arXiv:2107.01742*.

Ross, G. J. and Adams, N. M. (2012). Two nonparametric control charts for detecting arbitrary distribution changes. *Journal of Quality Technology*, 44:102–116.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The annals of mathematical statistics*, 9:60–62.

Yao, Y.-C. et al. (1988). Estimating the number of change-points via schwarz'criterion. *Statistics & Probability Letters*, 6:181–189.

Zou, C., Yin, G., Feng, L., and Wang, Z. (2014). Nonparametric maximum likelihood approach to multiple change-point problems.

A Proofs