Discussion of Multiscale Change Point Inference by Frick, Munk and Sieling

Yining Chen, Rajen D. Shah and Richard J. Samworth†
University of Cambridge, UK

We congratulate the authors for their stimulating paper; our comments focus on possible avenues for further development.

A Gaussian quasi-SMUCE (GQSMUCE).

In the paper it is assumed that Y comes from a known exponential family, which may not be realistic. But consider the following setting:

$$Y_i = \vartheta(i/n) + \epsilon_i, \qquad \text{for } i = 1, \ldots, n,$$

where ϵ_i are i.i.d. with $\mathbb{E}\epsilon_1 = 0$ and $\mathbb{E}\epsilon_1^2 = \sigma^2$. The number of change points can still be estimated by solving

$$\inf_{\vartheta \in \mathcal{S}} \#J(\vartheta) \quad \text{s.t.} \quad T_n(Y, \vartheta, c_n) \leq q$$

with

$$T_n(Y, \vartheta, c_n) = \max_{\substack{1 \leq i < j \leq n \\ \vartheta(t) = \theta \text{ for } t \in [i/n, j/n] \\ (j-i+1)/n \geq c_n}} \left(\frac{\left| \sum_{l=i}^{j} (Y_l - \theta) \right|}{\sigma \sqrt{j - i + 1}} - \sqrt{2 \log \frac{en}{j - i + 1}} \right).$$

Scrutiny of the proof of Theorem 2.1 shows that the result continues to hold for the *Gaussian quasi-likelihood* (GQSMUCE) estimator above provided there exists $s_0 > 0$ such that $\mathbb{E}e^{s\epsilon_1} < \infty$ for $|s| < s_0$. An analogue of Theorem 2.2 can also be proved, establishing model selection consistency.

To examine the performance of GQSMUCE in a non-Gaussian setting, and similar to Section 5.1, we let

$$Y_i = \vartheta(i/n) + \sigma\epsilon_i, \qquad \text{for } i = 1, \ldots, 497, \tag{1}$$

where $\epsilon_1, \ldots, \epsilon_{497}$ have a shifted and scaled Beta$(2, 2)$ distribution with zero mean and unit variance. Results are summarised in Table 1. We see that GQSMUCE outperforms CBS (Olshen *et.al.*, 2004) at lower noise levels $\sigma = 0.1$ and 0.2, but tends to underestimate the number of change points when $\sigma = 0.3$. These findings are qualitatively similar to results in Section 5.1.

†*Address for correspondence:* Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, UK. CB3 0WB.
E-mail: {Y.Chen, R.Shah, R.Samworth}@statslab.cam.ac.uk

Table 1. Relative frequencies of estimated numbers of change points by model selection for GQSMUCE and CBS (Olshen *et.al.*, 2004) in 500 Monte Carlo simulations. The true signals have six change points.

	σ	≤ 4	5	6	7	≥ 8
GQSMUCE ($1-\alpha=0.55$)	0.1	0.000	0.000	**0.988**	0.012	0.000
CBS (Olshen *et. al.*, 2004)	0.1	0.000	0.000	0.924	0.036	0.040
GQSMUCE ($1-\alpha=0.55$)	0.2	0.000	0.000	**0.994**	0.006	0.000
CBS (Olshen *et. al.*, 2004)	0.2	0.000	0.000	0.872	0.100	0.028
GQSMUCE ($1-\alpha=0.55$)	0.3	0.012	0.248	0.772	0.018	0.000
CBS (Olshen *et. al.*, 2004)	0.3	0.000	0.010	**0.806**	0.148	0.036

Table 2. Empirical coverage and probability of correctly estimating the number of change points K obtained from 500 simulations.

α	SMUCE		SMUCE_2	
	Coverage of $\mathcal{C}(q_{1-\alpha})$	$P(\hat{K}(q_{1-\alpha})=K)$	Coverage of $\mathcal{C}(q^*, q_{1-\alpha})$	$P(\hat{K}(q^*)=K)$
0.90	0.874	0.978	**0.882**	0.986
0.95	0.874	0.914	**0.944**	0.986
0.99	0.728	0.738	**0.974**	0.986

More generally, we believe multiscale methods for change point inference (or appropriately defined 'regions of interest' in multivariate settings) offer great potential even with more complex data-generating mechanisms, and we await future methodological, theoretical and computational developments with interest.

Coverage of confidence sets.

One attractive feature of SMUCE is the fact that confidence sets can be produced for ϑ. However, Table 5 in the paper shows in the Gaussian example with unknown mean that even when the sample size is as large as 1500, a nominal 95% confidence set only has only 55% coverage; even more strikingly, a nominal 80% coverage set has 84% coverage!

This phenomenon, where larger nominal coverage may reduce actual coverage, is caused by the choice of $1-\alpha$ determining not only the nominal coverage but also $\hat{K}$, the estimated number of change points.

As an alternative, consider the confidence set

$$\mathcal{C}(q^*, q_{1-\alpha}) := \Big\{\vartheta \in \mathcal{S} : \#J(\vartheta) = \hat{K}(q^*) \text{ and } T_n(Y, \vartheta) \leq q_{1-\alpha}\Big\},$$

where q^* can be chosen as suggested in Section 4, for example. We compare this approach (SMUCE_2) with that proposed in the paper for the simulation setting (1), where here we take $\epsilon_i \sim N(0, 0.05)$; results are presented in Table 2. As well as giving better coverage here, the new confidence sets have the reassuring property that $\mathcal{C}(q^*, q_{1-\alpha'}) \supseteq \mathcal{C}(q^*, q_{1-\alpha})$ for $\alpha' \leq \alpha$.