Statistica Sinica **33** (2023), 1-20
doi:https://doi.org/10.5705/ss.202021.0048

MOVING SUM DATA SEGMENTATION FOR STOCHASTIC PROCESSES BASED ON INVARIANCE

Claudia Kirch and Philipp Klein

Otto-von-Guericke-University Magdeburg

Abstract: The segmentation of data into stationary stretches, also known as the multiple change point problem, is important for many applications in time series analysis and signal processing. Based on strong invariance principles, we analyze a data segmentation methodology using moving sum statistics for a class of regime-switching multivariate processes, where each switch results in a change in the drift. In particular, this framework includes the data segmentation of multivariate partial sum, integrated diffusion, and renewal processes, even if the distance between the change points is sublinear. We study the asymptotic behavior of the corresponding change point estimators, show their consistency, and derive the corresponding localization rates, which are minimax optimal in a variety of situations, including an unbounded number of changes in Wiener processes with drift. Furthermore, we derive the limit distribution of the change point estimators for local changes. This result can, in principle, be used to derive confidence intervals for the change points.

Key words and phrases: Change point analysis, data segmentation, invariance principle, moving sum statistics, multivariate processes, regime-switching processes

1. Introduction

Change point analysis aims at detecting and localizing structural breaks in time series data, with applications in a variety of fields, such as neurophysiology (see Messer et al. (2014)), genomics (compare Olshen et al. (2004), Niu and Zhang (2012), Li, Munk and Sieling (2016), Chan and Chen (2017)), finance (Aggarwal, Inclan and Leal (1999), Cho and Fryzlewicz (2012)), astrophysics (see Fisch, Eckley and Fearnhead (2022)), and oceanographics (Killick et al. (2010)).

Early literature focused on testing for a single change point in the mean. Later studies examined changes in more complex data structures, and currently focus on detecting changes in high-dimensional data; see for example Csörgö and Horvàth (1997), Horváth and Rice (2014), and Cho and Kirch (2021).

During the last two decades, interest has shifted from testing to the multiple change problem, aiming at segmenting data into stationary stretches, often fo-

Corresponding author: Claudia Kirch, Institute for Mathematical Statistics, Department of Mathematics, Otto-von-Guericke University, 39106 Magdeburg, Germany. E-mail: claudia.kirch@ovgu.de.