

SpringerBriefs in Statistics
JSS Research Series in Statistics

Shohei Shimizu

Statistical Causal Discovery: LiNGAM Approach



 Springer

SpringerBriefs in Statistics

JSS Research Series in Statistics

Editors-in-Chief

Naoto Kunitomo, The Institute of Mathematical Statistics, Tachikawa, Tokyo, Japan

Akimichi Takemura, The Center for Data Science Education and Research, Shiga University, Hikone, Shiga, Japan

Series Editors

Genshiro Kitagawa, Meiji Institute for Advanced Study of Mathematical Sciences, Nakano-ku, Tokyo, Japan

Shigeyuki Matsui, Graduate School of Medicine, Nagoya University, Nagoya, Aichi, Japan

Manabu Iwasaki, School of Data Science, Yokohama City University, Yokohama, Kanagawa, Japan

Yasuhiro Omori, Graduate School of Economics, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

Masafumi Akahira, Institute of Mathematics, University of Tsukuba, Tsukuba, Ibaraki, Japan

Masanobu Taniguchi, School of Fundamental Science and Engineering, Waseda University, Shinjuku-ku, Tokyo, Japan

Hiroe Tsubaki, The Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan

Satoshi Hattori, Faculty of Medicine, Osaka University, Suita, Osaka, Japan

Kosuke Oya, School of Economics, Osaka University, Toyonaka, Osaka, Japan

Taiji Suzuki, School of Engineering, University of Tokyo, Tokyo, Japan

The current research of statistics in Japan has expanded in several directions in line with recent trends in academic activities in the area of statistics and statistical sciences over the globe. The core of these research activities in statistics in Japan has been the Japan Statistical Society (JSS). This society, the oldest and largest academic organization for statistics in Japan, was founded in 1931 by a handful of pioneer statisticians and economists and now has a history of about 80 years. Many distinguished scholars have been members, including the influential statistician Hirotugu Akaike, who was a past president of JSS, and the notable mathematician Kiyosi Itô, who was an earlier member of the Institute of Statistical Mathematics (ISM), which has been a closely related organization since the establishment of ISM. The society has two academic journals: the Journal of the Japan Statistical Society (English Series) and the Journal of the Japan Statistical Society (Japanese Series). The membership of JSS consists of researchers, teachers, and professional statisticians in many different fields including mathematics, statistics, engineering, medical sciences, government statistics, economics, business, psychology, education, and many other natural, biological, and social sciences. The JSS Series of Statistics aims to publish recent results of current research activities in the areas of statistics and statistical sciences in Japan that otherwise would not be available in English; they are complementary to the two JSS academic journals, both English and Japanese. Because the scope of a research paper in academic journals inevitably has become narrowly focused and condensed in recent years, this series is intended to fill the gap between academic research activities and the form of a single academic paper. The series will be of great interest to a wide audience of researchers, teachers, professional statisticians, and graduate students in many countries who are interested in statistics and statistical sciences, in statistical theory, and in various areas of statistical applications.

Shohei Shimizu

Statistical Causal Discovery: LiNGAM Approach

 Springer

Shohei Shimizu
Faculty of Data Science
Shiga University and RIKEN
Hikone, Shiga, Japan

ISSN 2191-544X ISSN 2191-5458 (electronic)
SpringerBriefs in Statistics
ISSN 2364-0057 ISSN 2364-0065 (electronic)
JSS Research Series in Statistics
ISBN 978-4-431-55783-8 ISBN 978-4-431-55784-5 (eBook)
<https://doi.org/10.1007/978-4-431-55784-5>

© The Author(s), under exclusive licence to Springer Japan KK 2022, corrected publication 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Japan KK, part of Springer Nature. The registered company address is: Shiroshima Trust Tower, 4-3-1 Toranomon, Minato-ku, Tokyo 105-6005, Japan

Preface

This monograph discusses statistical causal discovery methods for inferring causal relationships from data derived primarily from non-randomized experiments. Specifically, I focus on a non-Gaussian approach named LiNGAM. The LiNGAM approach uses a non-Gaussian data structure for model identification and can identify a much broader range of causal relationships than classic methods. This book aims to provide a concise summary of the basic ideas of the LiNGAM approach. More information on recent advances, applications, and available code packages can be found on the following website: <https://www.shimizulab.org/lingam/lingampapers>. I would be delighted if readers could get more familiar with the LiNGAM approach and become interested in working in the field of causal discovery.

Hikone and Osaka, Japan
January 2022

Shohei Shimizu

Acknowledgements I am deeply thankful to all my collaborators. I particularly extend my heartfelt gratitude to Aapo Hyvärinen, Patrik O. Hoyer, Takashi Nicholas Maeda, Yan Zeng, Patrick Blöbaum, Tatsuya Tashiro, Takashi Ikeuchi, Keisuke Kiritoshi, Kento Uemura, Hidetoshi Shimodaira, and Yutaka Kano.

The original version of the book was revised: Missing mathematical symbols have been updated throughout. The correction to the book is available at https://doi.org/10.1007/978-4-431-55784-5_7

Contents

1	Introduction	1
1.1	A Starting Point for Causal Inference	1
1.2	Framework of Causal Inference	2
1.3	Identification and Estimation of the Magnitude of Causation	4
1.4	Identification and Estimation of Causal Structures	6
1.5	Concluding Remarks	9
	References	10
Part I Basics of LiNGAM Approach		
2	Basic LiNGAM Model	15
2.1	Independent Component Analysis	15
2.2	LiNGAM Model	21
2.3	Identifiability of the LiNGAM model	23
2.4	Concluding Remarks	28
	References	28
3	Estimation of the Basic LiNGAM Model	31
3.1	ICA-Based LiNGAM Algorithm	31
3.2	DirectLiNGAM Algorithm	34
3.3	Multigroup Analysis	42
3.3.1	LiNGAM Model for Multiple Groups	42
3.3.2	DirectLiNGAM Algorithm for Multiple LiNGAMs	43
3.4	Concluding Remarks	45
	References	46
4	Evaluation of Statistical Reliability and Model Assumptions	49
4.1	Evaluation of Statistical Reliability	49
4.1.1	A Bootstrap Approach	49
4.1.2	Bootstrap Probability	50
4.1.3	Multiscale Bootstrap for LiNGAM	52

4.2 Evaluation of Model Assumptions 55
References 56

Part II Extended Models

5 LiNGAM with Hidden Common Causes 61
5.1 Identification and Estimation of Causal Structures
of Confounded Variables 61
5.1.1 LiNGAM Model with Hidden Common Causes 64
5.1.2 Identification Based on Independent Component
Analysis 66
5.1.3 Estimation Based on Independent Component Analysis 70
5.2 Identification and Estimation of Causal Structures
of Unconfounded Variables 71
5.3 Other Hidden Variable Models 75
5.3.1 LiNGAM Model for Latent Factors 75
5.3.2 LiNGAM Model in the Presence of Latent Classes 78
References 80

6 Other Extensions 83
6.1 Cyclic Models 83
6.2 Time-Series Models 87
6.3 Nonlinear Models 90
6.4 Discrete Variable Models 91
References 92

Correction to: Statistical Causal Discovery: LiNGAM Approach C1