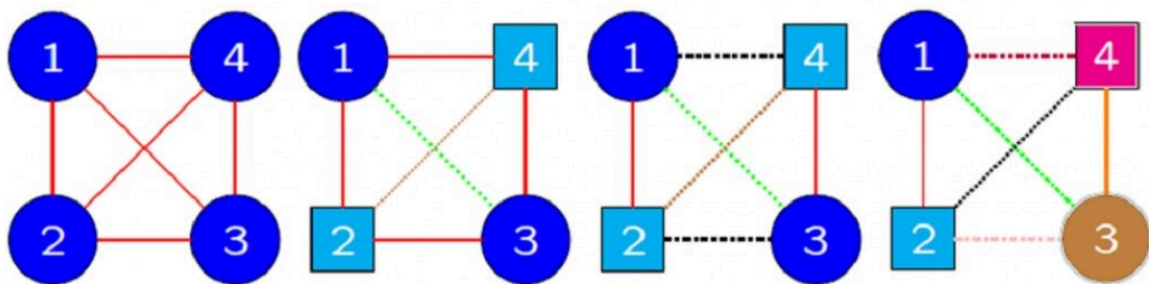


# New method to identify symmetries in data using Bayesian statistics

September 13 2022



Examples of colored graphs designating symmetries of four-dimensional data: Vertices and edges of the same color and shape in a graph are mapped to each other by a symmetry permutation preserving the structure of data. Credit: Hideyuki Ishi, Osaka Metropolitan University

An international research team led by scientists from Osaka Metropolitan University has developed a method to identify symmetries in multi-dimensional data using Bayesian statistical techniques.

This statistical approach requires complex calculations of integrals, which are often considered approximations only. In their new study, the research team successfully derived new exact integral formulas. Their findings contribute to improving the accuracy of methods to identify data symmetries, possibly extending their applications to wider areas of interest, such as genetic analysis.

Symmetries in nature make things beautiful; symmetries in data make data handling efficient. However, the complexity of identifying such patterns in data has always bedeviled researchers. Scientists from Osaka Metropolitan University and their colleagues have taken a major step towards detecting symmetries in multi-dimensional data by utilizing Bayesian statistics. Their findings were published in *The Annals of Statistics*.

Bayesian statistics has been in the spotlight in recent years due to improvements in computer performance and its potential applications in artificial intelligence. Bayesian statistics is a statistical approach that, even when data are insufficient, derives the probability of an event occurring by first setting a prior probability and then, whenever new information is obtained, calculating a posterior probability—an update to the prior probability—that the event will occur. The calculation of posterior probabilities requires complex calculations of integrals and therefore is often considered an approximation only.

The international team including Professor Hideyuki Ishi from Osaka Metropolitan University, Professor Piotr Graczyk from the University of Angers, Professor Bartosz Kołodziejek from Warsaw University of Technology, and the late Professor H el ene Massam from York

University (Toronto) has succeeded in deriving new exact integral formulas, and in developing a method to search for symmetries in multi-dimensional data using Bayesian statistical techniques.

When the amount of data to be handled increases, the optimal pattern must be selected from a vast number of patterns, making it difficult to solve the problem precisely. Addressing this challenge, the team has also developed an [efficient algorithm](#) for obtaining an approximate solution even in such cases.

In the words of Professor Ishi, "Symmetries in data are ubiquitous in a wide variety of models. Once symmetries are identified, the number of parameters required to display the structure of the data, and the number of samples required to determine the parameters, can be significantly reduced. In the future, the results of this research are expected to contribute to [genetic analysis](#), discovering chromosomes that have the same function in different locations."

**More information:** Piotr Graczyk et al, Model selection in the space of Gaussian models invariant by symmetry, *The Annals of Statistics* (2022). [DOI: 10.1214/22-AOS2174](https://doi.org/10.1214/22-AOS2174)

Provided by Osaka Metropolitan University

Citation: New method to identify symmetries in data using Bayesian statistics (2022, September 13) retrieved 25 April 2024 from <https://phys.org/news/2022-09-method-symmetries-bayesian-statistics.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.