

Research on Statistical Methods for Visualizing Data Distributions

Name YAMAMOTO Keiko **E-mail** kei_yama@hakodate-ct.ac.jp

Status Professor

Affiliations

- The Japan Statistical Society
- The Japanese Society of Computational Statistics

Keywords Nonparametric Statistical Analysis, Data Visualization

Technical Support Skills

- Data Distribution Visualization
- Statistical Data Analysis
- Nonparametric Density Estimation



Research Contents Research on Statistical Visualization and Probability Density Estimation

This research focuses on data visualization and probability density estimation, particularly on developing methods that combine the advantages of histograms and Kernel Density Estimation (KDE). Histograms are computationally efficient but sensitive to bin width and bin location, whereas KDE provides smooth density estimates at the cost of higher computational complexity and bandwidth dependence.

To address these limitations, we proposed the Generalized Frequency Polygon (GFP), a nonparametric density estimation method that combines the efficiency of histograms with the accuracy of KDE. We have investigated its theoretical properties and demonstrated that GFP can provide accurate density estimates with relatively low computational cost.

Figure 1 illustrates histograms constructed from the same dataset using different bin settings. The resulting shapes vary considerably depending on the choice of bin width and bin location, highlighting the importance of appropriate parameter selection for accurately visualizing data distributions.

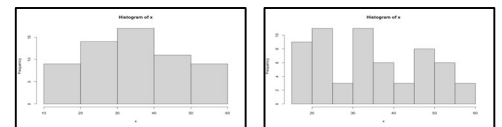


Fig1. Histograms of the same

Figure 2 compares GFP with Frequency Polygon-based methods and KDE for a bimodal distribution. The results show that GFP effectively captures the underlying distributional structure while requiring less computational effort than KDE.

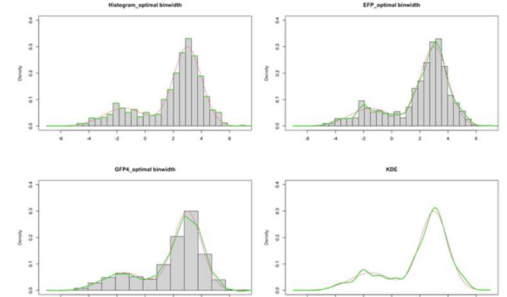


Fig2. Comparison of FP-Based Estimation and KDE

The proposed method is being applied to educational data, questionnaire survey data, and IoT sensor data. By visualizing and quantifying distributional characteristics such as skewness, multimodality, and outliers, this research aims to support more effective and interpretable data analysis.

The results are expected to contribute to a wide range of applications, including data science, educational technology, IoT analytics, quality control, and anomaly detection.

Available Facilities and Equipment
