



A novel size-biased distribution: regression model, INAR(1) process and applications in environmental and medical sciences

Na Elah¹ · Peer Bilal Ahmad¹

Received: 28 June 2025 / Revised: 28 June 2025 / Accepted: 20 September 2025 /

Published online: 26 November 2025

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2025

Abstract

A new statistical distribution, termed as Size-Biased Poisson Xgamma (SBPXG) distribution, is introduced in this study to enhance modeling and interpretation of data arising from size-biased sampling schemes. In such settings, the probability of selecting a unit is proportional to its size, resulting in the overrepresentation of larger observations. In this work, we consider the specific case where the size of a unit is defined to be the observed value of the response variable, thereby focusing on size-biased distributions where selection is directly linked to outcome magnitude. We derive some structural features of the model and a thorough reliability analysis is also carried out. For parameter estimation, Maximum likelihood estimation (MLE) and method of moments (MoM) are used. As the MLE is not available in closed form, the likelihood function is maximized numerically and a simulation analysis based on MLE further validates the robustness of the model. Additionally, we extend the usefulness of the SBPXG model in predictive analytics by developing an associated regression model for applications with relevant covariate data. Furthermore, we present an INAR(1) process under the SBPXG model to handle count data scenarios. We investigate its special characteristics including conditional maximum likelihood (CML) method. By examining three different datasets, we show that the SBPXG model continuously performs better in terms of goodness-of-fit than other models. Additionally, the Vuong's Likelihood Ratio Test is applied to formally compare non-nested models, and the results confirm that the SBPXG regression model offers a significantly better fit than conventional alternatives. The findings demonstrate the model's flexibility and strong potential for analyzing size-biased and overdispersed count data, particularly in healthcare and environmental domains.

Handling Editor: Omer Ozturk.

Extended author information available on the last page of the article