Annals of the Institute of Statistical Mathematics (2025) 77:597–625
https://doi.org/10.1007/s10463-025-00925-0

# Selection-bias-adjusted inference for the bivariate normal distribution under soft-threshold sampling

**Joseph B. Lang**[1]

Received: 22 February 2024 / Revised: 28 August 2024 / Accepted: 9 January 2025 /
Published online: 2 May 2025


## Abstract

The problem of estimating parameters and predicting outcomes of a bivariate Normal distribution is more challenging when, owing to data-dependent selection (or missingness or dropout), the available data are not a representative sample of bivariate realizations. This problem is addressed using an observation model that is induced by a combination of a multivariate Normal "science" model and a realistic "soft-threshold selection" model with unknown truncation point. This observation model, which is expressed using an intuitive selection subset notation, is a generalization of existing "hard-threshold" models. It affords simple-to-compute selection-bias-adjusted estimates of both the regression (conditional mean) parameters and the bivariate correlation. In addition, a simple bootstrap approach for computing both confidence and prediction intervals in the soft-threshold selection setting is described. Simulation results are promising. To motivate this research, two illustrative examples describe a setting where selection bias is an issue of concern.

**Keywords** Inverse Mill's ratio · Missing data · Selection bias · Truncated normal distribution

The online version of this article contains supplementary material.

✉ Joseph B. Lang
joseph-lang@uiowa.edu

[1] Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242, USA