

SKEWED PIVOT-BLEND MODELING WITH APPLICATIONS TO SEMICONTINUOUS OUTCOMES

Yiyuan She^{*1}, Xiaoqiang Wu¹, Lizhu Tao² and Debajyoti Sinha¹

¹*Florida State University and* ²*Sichuang University*

Abstract: Skewness is a common occurrence in statistical applications. In recent years, various distribution families have been proposed to model skewed data by introducing unequal scales based on the median or mode. However, we argue that the point at which unbalanced scales occur may be at any quantile and cannot be reparametrized as an ordinary shift parameter in the presence of skewness. In this paper, we introduce a novel skewed pivot-blend technique to create a skewed density family based on any continuous density, even those that are asymmetric and nonunimodal. Our framework enables the simultaneous estimation of scales, the pivotal point, and other location parameters, along with various extensions. We also introduce a skewed two-part model tailored for semicontinuous outcomes, which identifies relevant variables across the entire population and mitigates the additional skewness induced by commonly used transformations. Our theoretical analysis reveals the influence of skewness without assuming asymptotic conditions. Experiments on synthetic and real-life data demonstrate the excellent performance of the proposed method.

Key words and phrases: Composite models, semicontinuous outcomes, skewed data, two-part models, two-piece densities, variable selection.

1. Introduction

Statisticians frequently encounter skewed data in biomedical, econometric, environmental, and social research. Commonly used models, such as linear regression, least absolute deviations, and robust regression, presume symmetric errors and are prone to significant distortions when confronted with skewness. To mitigate the issue, many researchers prefer transforming the data beforehand, with logarithmic-type transformations being among the most popular choices. Alternatively, some researchers use modal regression (Lee, 1989) or median-based methods, which are less sensitive to the assumption of symmetric errors. However, these approaches do not explicitly account for and describe skewness.

To comprehensively address this issue, adopting a “joint” modeling approach becomes essential and beneficial. This paper simultaneously estimates location, scale, and skewness parameters, thus avoiding the risk of either concealing true skewness (*masking*) or erroneously detecting spurious skewness (*swamping*). This

^{*}Corresponding author. E-mail: sheyiyuan@westlake.edu.cn