

# **SKEW 2026 International Workshop**

Department of Statistical Sciences

University of Padova

**7-9 January 2026**

**ABSTRACT BOOKLET**



---

## Scientific Committee

Antonio Canale  
Daniele Durante  
Marc Genton  
Brunero Liseo  
Bruno Scarpa

## Organizing Committee

Antonio Canale  
Daniele Durante  
Elena Guzzonato  
Bruno Scarpa

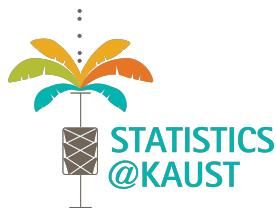
### Organized at



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



### Sponsored by



# Contents

|  |    |
|--|----|
| Session 1 . . . . .                    | 4  |
| Session 2 . . . . .                    | 5  |
| Session 3 . . . . .                    | 6  |
| Session 4 . . . . .                    | 7  |
| Session 5: Capitanio Lecture . . . . . | 9  |
| Session 6 . . . . .                    | 10 |
| Session 7 . . . . .                    | 11 |
| Session 8 . . . . .                    | 12 |
| Posters . . . . .                      | 13 |



---

## Session 1

Chair: Reinaldo Arellano-Valle

### **Multivariate SUN, SUNNY, and SUNSET distributions: Properties and applications**

Marc Genton

*King Abdullah University of Science and Technology (KAUST), Saudi Arabia*

The broad class of multivariate unified skew-normal (SUN) distributions has attracted sustained attention over the past two decades. In this talk, we provide an overview of its multiple origins, highlight its key properties, and describe its geometry under a parametrization with natural interpretability. We emphasize the issue of non-identifiability of the SUN distribution under permutations of its latent variables and propose several strategies to achieve identifiability, including identifiable sub-models such as the SUNNY class. This non-identifiability concern also arises in alternative parametrizations and extends to unified skew-elliptical (SUE) distributions, particularly the unified skew-t (SUT) family. An important feature of SUN distributions is their fundamental conjugacy property, which greatly facilitates Bayesian inference and computation. When used as priors for parameter vectors in general probit, tobit, and multinomial probit models, the resulting posteriors remain within the SUN family. We further discuss extensions of these results to the broader class of multivariate SUE distributions, ensuring conjugacy beyond fully observed, discretized, or censored multivariate normal models. Finally, we illustrate these properties with applications to multivariate SUT distributions and to the multivariate SUN with independent sets of latent variables (SUNSET) family.

### **An overview of SAS distributions and their properties**

Arthur Richard Pewsey

*Universidad de Extremadura, Spain*

As participants at this conference will no doubt be aware, Adelchi's paper "A class of distributions which includes the normal ones" (Scand. J. Statist., 1985) sparked huge renewed interest into models for asymmetrically distributed data. Rather than focusing on a perturbation-based approach, this talk provides an overview of a method of generating skew models from symmetric ones based on sinh-arcsinh (SAS) transformation. The main properties of SAS models obtained by transforming standard normal and standard t random variables are outlined, and their relationships with previously existing models identified. The related issues of: (i) skewness-invariant measures of kurtosis; (ii) the scaling of base random variables prior to SAS transformation so as to obtain appealing asymmetric models; (iii) goodness-of-fit testing, are also considered.



---

## Session 2

Chair: Omilos Papaspiliopoulos

### **Skew-symmetric numerical schemes for stochastic differential equations**

Samuel Livingstone  
*University College London, UK*

I will explain a new simple and explicit numerical scheme for time-homogeneous stochastic differential equations. The scheme is based on sampling increments at each time step from a skew-symmetric probability distribution, with the level of skewness determined by the drift and volatility of the underlying process. I will discuss weak and strong convergence to the underlying diffusion as the step-size is reduced. I will then discuss the problem of simulating from the limiting distribution of an ergodic diffusion process using the numerical scheme with a fixed step-size. The problem is of fundamental interest in Bayesian statistics, where Monte Carlo sampling algorithms based on simulating from the equilibrium distribution of a Markov process play a crucial role. I will show conditions under which the numerical scheme converges to equilibrium at a geometric rate, and quantify the bias between the equilibrium distributions of the scheme and of the true diffusion process (which, in the statistical context, is the posterior distribution of interest). Notably, our results do not require a global Lipschitz assumption on the drift, in contrast to those required for the Euler–Maruyama scheme for long-time simulation at fixed step-sizes (otherwise known as the unadjusted Langevin algorithm, or ULA).

Joint work with Yuga Iguchi, Nikolas Nuesken, Giorgos Vasdekis & Ruiyang Zhang

### **Variational Inference based on a Subclass of Closed Skew Normals**

Linda Tan  
*National University of Singapore, Singapore*

Gaussian distributions are widely used in Bayesian variational inference to approximate intractable posterior densities, but the ability to accommodate skewness can improve approximation accuracy significantly, when data or prior information is scarce. We study the properties of a subclass of closed skew normals constructed using affine transformation of independent standardized univariate skew normals as the variational density, and illustrate how it provides increased flexibility and accuracy in approximating the joint posterior in various applications, by overcoming limitations in existing skew normal variational approximations. The evidence lower bound is optimized using stochastic gradient ascent, where analytic natural gradient updates are derived. We also demonstrate how problems in maximum likelihood estimation of skew normal parameters occur similarly in stochastic variational inference, and can be resolved using the centered parameterization.



---

## Session 3

Chair: Nicola Sartori

### **Advances in Clustering of Incomplete Multivariate Data with Skewed and Heavy-Tailed Clusters**

Andriëtte Bekker

*University of Pretoria, South Africa*

This presentation focuses on the significant challenge of clustering incomplete multivariate data, particularly in the presence of skewed, heavy-tailed clusters and mild outliers. Our approach is the first to accommodate missing data under a missing at random (MAR) mechanism for the full finite mixture of scale mixtures of multivariate skew-normal (FMSMSN) family, extending previous work that focused on more restrictive special cases. To mitigate the effect of possible atypical observations within an incomplete dataset, a heavy-tailed extension, the contaminated multivariate skew-normal distribution, will receive attention. We derive the distributional properties of the missing components and develop an augmented EM-type algorithm with closed-form conditional expectations for the E-step. Simulation studies demonstrate strong clustering performance and accurate parameter recovery across varying levels of missingness, sample sizes, and cluster overlap. We also present a real-data application to global  $CO_2$  emissions, illustrating the practical advantages of our method.

### **A quasi-SUN process for nonparametric binary regression**

Brunero Liseo

*Università la Sapienza, Italy*

Conjugate families have been recently introduced in a Bayesian analysis of probit and logit models. Even more recently, some of these results have been extended - in the probit case - to the case of a nonparametric predictor. In this paper, we extend the conjugacy property to the case of a nonparametric predictor of a logit model. We introduce a new class of distributions, named *Quasi SUN* family (qSUN, hereafter), which generalises the SUN class.



---

## Session 4

Chair: Antonio Canale

### **Bayesian repulsive mixture model for functional data: contaminating the normal distribution to repel similarly shaped curves**

Rosangela Loschi

*Universidade Federal de Minas Gerais, Brazil*

We introduce a repulsive mixture model for multivariate functional data having as one of the main goal to clustering curves with similar shapes. We assume that more than one functional curve is observed for each individual. The clustering process takes into account both individual-specific information, as we assume covariate-dependent mixture weights, and the similarity of the curve shape. To facilitate the identification of well-differentiated clusters while avoiding redundant clusters in relation to their curve shapes, we propose a repulsive prior distribution for the component-specific location vector of the B-spline coefficients. The proposed prior distribution is a distortion of the multivariate normal distribution, which is multiplied by a functional repulsive factor which depends on a B-spline curve-tailored distance extending existing repulsive priors to the context of functional data. To sample from the posterior distribution, we propose a MCMC algorithm with a split-merge step. The split-merge step significantly improves the mixing of the chain. Different features of the proposed model, including the effects of repulsion and covariates in the clustering, are investigated through simulation. The model is used to identify different movement characteristics in individuals with Chronic Ankle Instability (CAI data). We also evaluate the effect of repulsion on clustering of functional NBA data which has a high noise level.

Joint work with Ricardo Cunha Pedroso (INSPER and Universidade Federal de Minas Gerais-Brazil) and Fernando A. Quintana (Pontificia Universidad Católica de Chile)

### **Harnessing Demographic Skewness: Insights and Applications**

Stefano Mazzuco

*Università di Padova, Italy*

Skew-symmetric distributions, such as the skew-normal family, can offer a powerful and parsimonious framework for demographic modeling, moving beyond the constraints of classical distributions. Their utility extends beyond mere flexibility in curve-fitting providing a formal statistical lens through which to interpret the dynamic, often asymmetric, shapes of key demographic profiles. In fertility analysis, the skewness of the age schedule of fertility is highly sensitive to period shocks. During crises (e.g., pandemics, economic recessions), the displacement or postponement of births often manifests as a rapid, transient increase in negative skewness, reflecting a compression of fertility at older reproductive ages. Conversely, a recuperation phase or a surge in "catch-up" births can produce a pronounced positive skew. Analyzing cohort fertility through this lens allows demographers to quantify and separate the tempo and quantum effects of such disruptions more effectively than with symmetric models.



---

In mortality analysis, the skewness of the age-at-death distribution can encapsulate information about the population's health heterogeneity and the process of mortality selection. This study demonstrates the application of skew-symmetric models to both fertility and mortality, using data from multiple national populations.



---

## Session 5: Capitanio Lecture

Chair: Adelchi Azzalini

### **Skewed Bernstein-von Mises theorem and online skew-modal approximations**

Francesco Pozza  
*Università Bocconi, Italy*

Gaussian deterministic approximations are routinely employed in Bayesian statistics to ease inference when the target posterior of direct interest is intractable. Although these approximations are justified, in asymptotic regimes, by Bernstein-von Mises type results, in practice the expected Gaussian behavior may poorly represent the actual shape of the target posterior, thereby affecting approximation accuracy. Motivated by these considerations, we derive an improved class of closed-form and valid deterministic approximations of posterior distributions which arise from a novel treatment of a third-order version of the Laplace method yielding approximations within a tractable family of skew-symmetric distributions. Under general assumptions which also account for misspecified models and non-i.i.d. settings, this novel family of approximations is shown to have a total variation distance from the target posterior whose rate of convergence improves by at least one order of magnitude the one achieved by the Gaussian from the classical Bernstein-von Mises theorem. Specializing such a general result to the case of regular parametric models shows that the same improvement in approximation accuracy can be also established for polynomially bounded posterior functionals, including moments. Unlike other higher-order approximations based on, e.g., Edgeworth expansions, our results prove that it is possible to derive closed-form and valid densities which are expected to provide, in practice, a more accurate, yet similarly-tractable, alternative to Gaussian approximations of the target posterior of direct interest. We strengthen these arguments by developing a practical skew-modal approximation for both joint and marginal posteriors which achieves the same theoretical guarantees of its theoretical counterpart by replacing the unknown model parameters with the corresponding maximum a posteriori estimate. Finally, we demonstrate that our theoretical findings can be also extended to online settings, that is, when data are collected sequentially and the posterior must be updated at each new data point. Simulation studies and real-data applications confirm that our theoretical results closely match the remarkable empirical performance observed in practice, even in finite, possibly small, sample regimes.



---

## Session 6

Chair: Alessandra Rosalba Brazzale

### **An EM algorithm for fitting matrix-variate skew-normal distributions on interval-censored and missing data**

Victor Lachos

*University of Connecticut, USA*

Matrix-variate normal distributions are widely used for modeling three-way data structures that frequently arise in longitudinal studies and multidimensional spatiotemporal analyses. In practice, however, such datasets often contain incomplete information, including censored values—reported only as being above or below detection limits—and missing observations. Moreover, deviations from normality, such as skewness, introduce additional challenges. To address these issues, we propose an efficient EM-type algorithm for maximum likelihood estimation under interval-censored and/or missing data within the matrix-variate skew-normal framework. The algorithm leverages closed-form expressions based on the truncated moments of multivariate skew-normal distributions, which can be efficiently evaluated using available software. Through simulation studies, we demonstrate the limitations of matrix-variate normal models when applied to non-normal data. The practical value of the proposed method is further illustrated with a real-world case study on water quality monitoring.

### **On robustness of the skew-t model**

Marcia Branco

*Universidade de São Paulo, Brazil*

For a long time the Student-t distribution has been used as an alternative to the normal distribution for robust modelling purpose. This robustness property is associated with the weight of the tails of the Student-t distribution. Distributions with heavier tails deal better with the problem of the influence of outliers on estimates. In the asymmetric world, it seems natural to expect that the skew-t distribution has the same robustness properties. However, this analogy cannot be made without additional caution. What robustness properties can be extended from the Student-t distribution to the skew-t distribution? In what situations does the skew-t distribution fail to handle the presence of outliers? These are the topics I intend to discuss in my presentation. This is a joint work with Simone Harnik and Marc Genton.



---

## Session 7

Chair: Clécio da Silva Ferreira

### **Skewness-based projection pursuit: past, present and future**

Nicola Loperfido  
*Università di Urbino, Italy*

Skewness-based projection pursuit aims at finding interesting data projections by maximizing their third standardized cumulants. The method relies on tensor eigenvectors, which naturally generalize matrix eigenvectors to multi-way arrays: eigenvectors of symmetric tensors of order  $k$  and dimension  $p$  are stationary points of polynomials of degree  $k$  in  $p$  variables on the unit sphere. Dominant eigenvectors of symmetric tensors maximize polynomials in several variables on the unit sphere, while base eigenvectors are roots of polynomials in several variables. Skewness-based projection pursuit finds interesting data projections using the dominant eigenvector of the sample third standardized cumulant to maximize skewness and uses its base eigenvectors to remove skewness, thus facilitating the search for interesting data features other than skewness. The talk outlines the theory and the applications of skewness-based projection pursuit, as well as its limitations.

### **Quantifying homophily through skewed link functions in Bayesian network models: estimating peer influence**

Dipak Dey  
*University of Connecticut, USA*

Research has increasingly focused on examining the role of social networks in shaping human behavior, recognizing their significant influence on contemporary interactions. Several studies have highlighted homophily as a major challenge when attempting to accurately estimate peer effects. Traditional methods typically address homophily in social networks using binary link structures. In this work, we aim to enhance existing methodologies by employing a skewed link function, which allows for more effective management of homophily in the measurement of social influence. This study, situated within the framework of latent space modeling, seeks to integrate two distinct processes—social influence and social selection—into a unified Bayesian model. The model incorporates network structure to assess the relationship between latent homophily among network participants and their behaviors, while quantifying social influence effects. These effects are differentiated based on the strength of node connections. We conducted extensive simulations to evaluate the properties of this approach, followed by its application to real-world data to explore how individual networks influence participation in various life behaviors.



---

## Session 8

Chair: Bruno Scarpa

### Aspects of factor modelling of skew multivariate data

Sylvia Frühwirth-Schnatter  
*Wirtschaftsuniversität, Vienna*

It is well known that Adelchi Azzalini's skew-normal distributions, both for univariate and multivariate variables, has a stochastic representation as a factor model with a single factor following a truncated standard normal distribution and the skewness parameters acting as factor loadings. This representation is particularly useful to perform Bayesian inference using Gibbs sampling. A similar stochastic representation exists for the extended SUN family, which exhibits multiple latent factors of dimension  $m > 1$ . When fitting a SUN distribution to non-Gaussian data, several inference problems arise, such as selecting the dimension  $m$  of the latent factors and estimating the matrix of skewness parameters, together with all other model parameters. The present presentation discusses the idea to exploit recent advances in sparse Bayesian factor analysis for Gaussian data to resolve these challenging inference problems in factor modelling of non-Gaussian data, including the use of exchangeable shrinkage process (ESP) priors to select  $m$ .

### Skew-symmetric representations of posterior densities

Daniele Durante  
*Università Bocconi, Italy*

The perturbation of symmetric densities via suitably designed skewing functions yields a broad and similarly tractable class of asymmetric extensions, known as skew-symmetric family. Although such a family has been mostly overlooked in Bayesian statistics, its potential impact could be substantial. As shown in this presentation, any generic posterior density can, in fact, be re-expressed in a skew-symmetric form by decomposing it as the product of (i) a symmetrized version of such a density and (ii) a tractable closed-form skewing function that only depends on the un-normalized posterior. This representation opens the avenues for the design of broadly applicable strategies perturbing, at no additional optimization cost, any symmetric posterior approximation (e.g., from the Laplace method, variational Bayes and expectation-propagation) to obtain similarly tractable skew-symmetric alternatives. This improved class of deterministic approximations provably enhances the finite sample accuracy of the original symmetric counterparts and, under suitable assumptions, it improves the convergence rate to the exact posterior by at least a  $\sqrt{n}$  factor, in asymptotic regimes.



---

## Posters

### Scale mixture of skew-normal additive partially linear mixed models with within-subject serial dependence

Clécio da Silva Ferreira

*Federal university of Juiz de Fora*

In longitudinal studies, repeated measures are collected over time, and hence they tend to be serially correlated. Also, some explanatory variables present unknown non-linear relationships with the response variable. In this work, we extend the popular linear mixed models using a family of scale mixtures of skew normal distribution, where the error term has a dependence structure, such as damped exponential correlation or autoregressive correlation of  $p$  order. The nonlinearity is modeled using B-spline functions. The proposed model provides flexibility in capturing the effects of skewness and heavy tails simultaneously when continuous repeated measures are serially correlated. For this robust model, we present an efficient EM-type algorithm for parameters estimation via maximum likelihood, and the observed information matrix is derived analytically to account for standard errors. The methodology is illustrated through an application to real data and some simulation studies.

### Finite Mixture of Regression Models based on Multivariate Scale Mixtures of Skew-Normal Distributions

Camila Zeller

*Statistics, UFJF, Brazil*

The traditional estimation of mixture regression models is based on the assumption of component normality (or symmetry), making it sensitive to outliers, heavy-tailed errors, and asymmetric errors. In this work, we propose addressing these issues simultaneously by considering a finite mixture of regression models with multivariate scale mixtures of skew-normal distributions. This approach provides greater flexibility in modeling data, accommodating both skewness and heavy tails. Additionally, the proposed model allows to use of a specific vector of regressors for each dependent variable. The main advantage of using the mixture of regression models under the class of multivariate scale mixtures of skew-normal distributions is their convenient hierarchical representation, which allows easy implementation of inference. We develop a simple expectation–maximization (EM) type algorithm to perform maximum likelihood inference for the parameters of the proposed model. The observed information matrix is derived analytically to calculate standard errors. Some simulation studies are also presented to examine the robustness of this flexible model against outlying observations. Finally, a real dataset is analyzed, demonstrating the practical value of the proposed method. The R scripts implementing our methods are available on the GitHub repository at <https://bit.ly/3CLcI1W>

Joint work with Luis Benites, Víctor H. Lachos, Héleno Bolfarine and Camila B. Zeller

### Skew Meets Transport: Smart Approximations of the Bayesian Discrepancy Measure

Elena Bortolato

*Universitat Pompeu Fabra*

The Bayesian Discrepancy Measure (BDM) provides a robust method for testing precise statistical hypotheses by evaluating how central a hypothesized value is within the posterior distribution. However, standard first-order approximations of the BDM, which often assume a symmetric Gaussian posterior, can be inaccurate when the posterior distribution exhibits skewness—a common feature in many statistical models. This is particularly challenging in multivariate settings where defining the "center" or "tail areas" is not straightforward. We employ skewed approximations to better capture the actual shape of the posterior distribution, moving beyond the restrictive assumption of symmetry. For multidimensional parameters, we utilize a tractable multivariate skew-normal (SN) distribution, which provides a much more accurate representation of the posterior at a manageable computational cost. We propose a formal and rigorous definition of the



---

multivariate BDM by using center-outward quantile functions, which are derived from an Optimal Transport (OT) map. This approach provides a principled way to generalize the concept of quantiles to multiple dimensions. Specifically, we demonstrate that by approximating the posterior with an SN distribution, we can construct an explicit OT map that transforms the skewed approximation into a standard multivariate normal distribution. This crucial step allows for the direct and efficient computation of the BDM for composite hypotheses.

Joint work with: Monica Musio (University of Cagliari) and Laura Ventura (University of Padova)

### **Handling missing data, skewness, and outliers in medical research: a robust factor analysis approach using the canonical fundamental skew-t distribution**

Tsung-I Lin

*National Chung Hsing University*

Addressing incomplete and non-normally distributed multivariate data poses significant challenges in medical research, particularly when the interest is in discovering underlying data structures. This paper introduces a robust factor analysis framework in the presence of missing data employing the Canonical Fundamental Skew-t Factor Analysis (CFUSTFA) model, which incorporates the canonical fundamental skew-t (CFUST) distribution into the latent factors and error terms. This versatile framework considers skewness, heavy tails, and missing data, enabling the model to capture complex structures commonly observed in biomedical datasets. To estimate the model parameters under a missing at random mechanism, we develop a computationally efficient alternating expectation conditional maximization algorithm for maximum likelihood estimation. This approach facilitates the simultaneous imputation of missing values and the extraction of low-dimensional factor representations. Standard errors for parameter estimates were also derived using a general information matrix-based approach. The proposed methodology is evaluated through a simulation study and applied to a hepatitis C virus laboratory dataset exhibiting skewness, excess kurtosis, and missingness. Our findings underscore the capability of the CFUSTFA model to robustly model complex, incomplete, and asymmetric biomedical data, offering enhanced inference and interpretability over the existing factor analysis approaches.

### **Finite Mixture of Multivariate Contaminated Normal Censored Regression Model with Its Application**

Wan-Lun Wang

*National Cheng Kung University*

The complexity of model-based clustering grows as outliers become more prevalent, compounded by restrictions imposed by the detection of quantification. This article introduces a finite mixture of multivariate contaminated normal censored regression (FM-MCNCR) model tailored for handling censored data in a linear regression scenario. For the estimation of model parameters, we devise a computationally analytical alternating expectation conditional maximization (AECM) algorithm. Additionally, we present an information matrix-based formula to approximate the asymptotic standard errors of parameter estimates. Importantly, the AECM algorithm serves a dual role by not only facilitating parameter estimation but also providing methods to recover censored measurements and detect outlier data points as a by-product when it converges. The efficacy and advantages of the proposed methodology are illustrated through a simulation study and an example related to U.S. women's labor force participation data.

### **Robust Semiparametric Graphical Models with Skew-Elliptical Distributions**

Gabriele Di Luzio

*Sapienza, University of Rome*

We propose semiparametric estimators called elliptical skew-(S)KEPTIC for efficiently and robustly estimating non-Gaussian graphical models, relaxing the assumption of semiparametric elliptical distributions to the family of meta skew-elliptical, which accommodates a skewness component. Theoretically, we demonstrate that the elliptical skew-



---

(S)KEPTIC estimators achieve robust convergence rates in both graph recovery and parameters estimation. We conduct numerical simulations that prove the reliable graph recovery performance of the elliptical skew-(S)KEPTIC estimators. Finally, the new method is applied to the daily log-returns of the stocks of the S&P500 index and shows higher sparsity compared to the Gaussian copula graphical models.

### **Omitting continuous covariates in binary regression models: Implications for sensitivity and mediation analysis**

Matteo Gasparin

*Università della Svizzera Italiana*

In statistical analysis, Cochran's formula plays a crucial role in disentangling the relationships between marginal and conditional regression coefficients. However, its results and implications are valid only within the linear case. Despite this, due to its simplicity and interpretability, practitioners often continue to use Cochran's formula also outside linear models. With reference to binary outcome models, we derived the approximated expression of the marginal regression coefficient when marginalization is performed over a continuous covariate and show that it mimics Cochran's formula under certain simplifying assumptions. We first postulate a logistic link function and then generalize it, linking the results to skew-symmetric distributions. We then explore the implications of this formula in the context of sensitivity analysis and causal mediation analysis, thereby enlarging the number of circumstances where explicit parametric formulations can be used to evaluate causal direct and indirect effects, otherwise computed via numerical integration. Simulations show that our proposed estimators perform equally well as others based on numerical methods and that the additional interpretability of the explicit formulas does not compromise their precision.

### **Non-linear Transformation Methods to Speed Up Markov Chain Monte Carlo Sampling Algorithms**

Yuxin Liu

*University College London*

We propose a novel nonlinear transformation method to improve the robustness and efficiency of MCMC algorithms. Our approach approximates target distributions using skew normal, Johnson SU, or sinh-arcsinh distributions, then based on this approximation, constructs transformation maps to convert them into new transformed distributions that are more symmetric. We sample in the transformed space and transform samples back to obtain samples from the original distribution. Through simulation studies on various distributions with different skewness levels, tail weights, and dimensions, our methods demonstrate improved efficiency compared to standard samplers in all cases. The methods provides a practical approach to robust sampling in challenging skewed settings

### **Benford's Law and Skewed Generalized Error Distributions (SGED): A Numerical and Permutation-Based Perspective**

Gianfranco Piscopo and Massimiliano Giacalone

*University of Naples Federico II and Università della Campania Luigi Vanvitelli*

The Newcomb–Benford Law (NBL) describes the distribution of the first significant digit in many real-world datasets, following a decreasing logarithmic distribution. Recent studies show that flexible models such as the Generalized Error Distribution (GED) and the Skewed Generalized Error Distribution (SGED), characterized by the shape parameter  $p$  and skewness, are particularly effective in modeling heavy tails and heterogeneity in higher moments in financial and cryptocurrency data. In this work, we build on the contribution of Formann [Formann(2010)] and adopt the parametrization with  $p$  proposed in (Cerqueti, Giacalone & Mattera 2020) extending the analysis to the SGED model. Through numerical simulations and empirical comparisons—particularly involving financial risk data, portfolios, cryptocurrencies, and time series clustering—we examine how the parameters  $p$  and skewness influence the distribution of the first digit and the compliance with NBL. In addition, we introduce a nonparametric permutation-based framework for comparing first-digit



---

distributions across groups and assessing the statistical significance of deviations, using the Nonparametric Combination (NPC) methodology. We also present Monte Carlo evidence that skewness and the presence of heavy tails support adherence to Benford's Law, but that overly extreme asymmetry can lead to systematic deviations. The permutation framework enhances our ability to formally compare empirical distributions under different SGED scenarios.

Joint work with Valerio Ficcadenti (London South Bank University, UK) & Maria Longobardi (University of Naples Federico II, Italy)

## **Python packages for the skew normal-distribution: a comparative analysis with R and a proposal for an integrated approach**

Mattia Casetta  
*University of Padova*

Skew-normal (SN), skew-t (ST), and unified skew-normal (SUN) distributions are widely used tools for modelling asymmetric data and play an important role in bayesian inference. While the `sn` package in R represents the current state of the art in terms of completeness and reliability, the corresponding ecosystem in Python remains fragmented and limited. This contribution provides a systematic comparison of Python implementations for SN, ST, and SUN distributions with the `sn` package in R, which is taken as a reference. The comparison considers consistency with R results, availability and usability of the libraries, completeness of the implemented algorithms, update frequency, and community support. The analysis shows that no existing Python implementation is currently comparable to `sn`. However, SciPy emerges as a solid basis for the univariate SN distribution. Moreover, `skewt-scipy` provides a preliminary implementation of the univariate ST, although further optimization is required. Since no general-purpose Python implementation of the SUN distribution is available, a preliminary feasibility study based on SciPy is presented. Despite limitations related to numerical stability and the lack of a truncated multivariate normal generator, the results indicate that SciPy represents a promising foundation for future developments.