

Fábio Veríssimo Jaques da Silveira

New normal-based classes of probability distributions

Recife

January 10, 2022



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

New normal-based classes of probability distributions

Tese julgada adequada para obtenção do título de Doutor em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 10/01/2022 pela comissão examinadora

Área de concentração: Biometria e Estatística Aplicada

Orientador: Dr. Frank Sinatra Gomes da Silva

Coorientador: Dr. Cícero Carlos Ramos de Brito

Recife

January 10, 2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

S587n

Silveira, Fábio Veríssimo Jaques da

New normal-based classes of probability distributions / Fábio Veríssimo Jaques da Silveira. - 2022.
76 f. : il.

Orientador: Frank Sinatra Gomes da Silva.

Coorientador: Cicero Carlos Ramos de Brito.

Inclui referências e apêndice(s).

Tese (Doutorado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, 2022.

1. Classes de distribuições de probabilidade. 2. Distribuição normal. 3. Máxima verossimilhança. 4. Simulação de Monte Carlo. 5. Modelagem de dados reais. I. Silva, Frank Sinatra Gomes da, orient. II. Brito, Cicero Carlos Ramos de, coorient. III. Título

CDD 519.5

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

New normal-based classes of probability distributions

Fábio Veríssimo Jaques da Silveira

Tese julgada adequada para obtenção do título de Doutor em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 10/01/2022 pela comissão examinadora

Orientador:

Dr. Frank Sinatra Gomes da Silva
Orientador

Banca examinadora:

Dr. Abraão David Costa do
Nascimento
DE-UFPE

Dra. Maria do Carmo Soares de Lima
DE-UFPE

Dr. Cícero Carlos Ramos de Brito
IFPE

Dr. Josimar Mendes de Vasconcelos
DEINFO-UFRPE

*I dedicate this doctoral thesis to my
beloved mother, Mirna Barbosa Jaques.*

Acknowledgements

Thanks be given to God Almighty and to my Lord and Savior Jesus Christ for taking care of me during the days of adversity and for granting me balance and strength to stay steady.

I am grateful to my dear family, especially to Mirna (my mother), Breno and Júlia (my siblings), Beatriz and Bianca (my nieces) and Edilene (my aunt). Their love and prayers helped me to carry on.

I thank Flávia for her love and support, especially in the hardest days.

I thank my advisor, Dr. Frank Sinatra Gomes da Silva, and co-advisor, Dr. Cícero Carlos Ramos de Brito. I also thank Dr. Jader Jale, Dr. Sílvio Xavier-Júnior and Felipe Gusmão for the valuable contributions to this work.

I am grateful to my colleagues from the Federal Institute of Education, Science and Technology of Paraíba at João Pessoa, especially to Hélder Alves de Oliveira, for supporting my qualification leave. The last four years would have been much harder without their precious help.

I thank the honorable members of the examining committee for the comments and suggestions.

Finally, I am thankful to all the staff of the Department of Statistics and Informatics for keeping the excellence of the Postgraduate Program of Biometrics and Applied Statistics at this Federal University.

*“Não é possível decidir racionalmente
se estamos na estrada certa, a não ser que
estejamos dispostos a considerar a possibilidade
de estarmos na estrada errada.”
(Phillip E. Johnson)*

Abstract

In this work we present three classes of probability distributions built upon an innovative method that considers the Lebesgue integration of a cumulative distribution function. For the three classes we considered the standard normal cumulative distribution function, differing in the limits of integration. The upper limit of integration of the second class is unprecedented in the literature focused on statistical distributions and the third class admits two different baselines, being a competing alternative to two-component mixture models. The classes have the advantage of demanding no extra parameters, besides those of the baseline(s), providing thereby parsimonious models. Under certain conditions, the distributions generated by the new classes are identifiable. We present some mathematical properties of the classes, like the linear representation of the probability density function, the moments and the moment generating function. Monte Carlo simulation studies are performed to investigate the behavior of the maximum likelihood estimates of the parameters for some submodels of the classes. The same submodels examined in the simulations are also used in applications to real data sets. Information criteria and formal goodness-of-fit statistics (Cramér-von Mises and Anderson-Darling) are used as criteria for model selection in comparisons considering the studied distributions emerged from the classes and some well-known distributions commonly employed in modelling similar data. The results suggest that the distributions from the proposed classes outperform the competing models.

Keywords: classes of probability distributions; normal distribution; maximum likelihood; Monte Carlo simulation; real data modelling.

Resumo

Neste trabalho, são apresentadas três classes de distribuições de probabilidade construídas através de um método inovador que considera a integração de Lebesgue de uma função de distribuição acumulada. Para as três classes, foi considerada a função de distribuição acumulada da distribuição normal padrão, diferindo nos limites de integração. O limite superior de integração da segunda classe é inédito na literatura que aborda distribuições estatísticas e a terceira classe admite duas distribuições-base diferentes, sendo uma alternativa que compete com os modelos de misturas com duas componentes. As classes têm a vantagem de não exigirem parâmetros adicionais, além daqueles que pertencem à distribuição-base, fornecendo assim modelos parcimoniosos. Sob certas condições, as distribuições geradas pelas novas classes são identificáveis. São apresentadas algumas propriedades matemáticas das classes, como a representação linear da função densidade de probabilidade, os momentos e a função geratriz de momentos. São realizadas simulações de Monte Carlo para investigar o comportamento das estimativas de máxima verossimilhança dos parâmetros para alguns submodelos das classes. Os mesmos submodelos examinados nas simulações são também usados em aplicações a conjuntos de dados reais. Critérios de informação e estatísticas de qualidade do ajuste (Cramér-von Mises e Anderson-Darling) são usadas como critérios para seleção do modelo em comparações considerando as distribuições estudadas oriundas das classes e algumas distribuições já conhecidas e comumente empregadas na modelagem de dados similares. Os resultados sugerem que o desempenho de modelagem das distribuições geradas pelas classes propostas é superior ao dos outros modelos usados na comparação.

Palavras-chave: classes de distribuições de probabilidade; distribuição normal; máxima verossimilhança; simulação de Monte Carlo; modelagem de dados reais.

Divulgação não autorizada por se tratar de material temporariamente restrito.

Artigos publicados:

Silveira, F.V.J.; Gomes-Silva, F.; Brito, C.C.R.; Cunha-Filho, M.; Gusmão, F.R.S.; Xavier-Júnior, S.F.A. Normal-G Class of Probability Distributions: Properties and Applications. *Symmetry* 2019, 11, 1407. <https://doi.org/10.3390/sym11111407>.

Silveira, F.; Gomes-Silva, F.; Brito, C.; Cunha-Filho, M.; Jale, J.; Gusmão, F.; Xavier-Júnior, S. The normal-tangent-G class of probabilistic distributions: properties and real data modelling. *Pakistan Journal of Statistics and Operation Research* 2020, 16(4), 827-838. <https://doi.org/10.18187/pjsor.v16i4.3443>.

Preprint postado no repositório arXiv:

Silveira, F.V.J.; Gomes-Silva, F.; Brito, C.C.R.; Jale, J.S.; Gusmão, F.R.S.; Xavier-Júnior, S.F.A.; Rocha, J.S. Modelling wind speed with a univariate probability distribution depending on two baseline functions [2101.03622v1].