

# gkwdist: An R Package for the Generalized Kumaraswamy Distribution Family

José Evandeilton Lopes<sup>1</sup>

<sup>1</sup> Federal University of Paraná (UFPR), Brazil

DOI:

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted:

Published:

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

## Summary

Bounded continuous data on the unit interval  $(0, 1)$  arise frequently in scientific applications. Proportions, rates, indices, and probabilities for instance naturally fall within this domain. While the Beta distribution (Ferrari & Cribari-Neto, 2004) and the two-parameter Kumaraswamy distribution (Jones, 2009; Kumaraswamy, 1980) serve as standard choices for such data, they may lack flexibility for datasets exhibiting heavy tails, bimodality, or complex boundary behaviors. The **gkwdist** package implements the five-parameter Generalized Kumaraswamy (GKw) distribution (Carrasco, Ferrari, & Cordeiro, 2010) and its complete hierarchy of seven nested subfamilies, providing a unified, high-performance framework for modeling bounded data with varying degrees of complexity. All computational routines are implemented in **C++** via **RcppArmadillo** (Eddelbuettel & Sanderson, 2014), ensuring numerical stability and computational efficiency for both interactive analysis and large-scale applications.

## Statement of Need

Practitioners modeling bounded responses require distributions that balance flexibility with parsimony. The GKw family addresses this through a principled hierarchical structure: analysts can begin with simple two-parameter models (Beta, Kumaraswamy) and systematically evaluate more complex specifications (three to five parameters) only when justified by the data. This approach (starting simple and adding complexity as needed) reduces overfitting risk while accommodating challenging data features such as bimodality, asymmetric tails, and boundary concentration.

The target audience includes statisticians and data scientists working with proportions, rates, or indices in fields such as hydrology (the original domain of Kumaraswamy's distribution), econometrics, ecology, health sciences, and finance. By providing exact analytical score vectors and Hessian matrices for all seven subfamilies, **gkwdist** enables efficient maximum likelihood estimation and facilitates principled model selection via likelihood ratio tests or information criteria such as AIC and BIC.

## State of the Field

Existing R packages provide partial coverage of bounded distributions. The **extraDistr** package (Wolodko, 2020) implements basic Kumaraswamy density, distribution, and quantile functions. The **VGAM** package (Yee, 2015) offers Kumaraswamy regression via the **kumar()** family function. The **betareg** package (Cribari-Neto & Zeileis, 2010) provides comprehensive Beta regression modeling with diagnostic tools. The **gamlss.dist** package (Rigby, Stasinopoulos, Heller, & De Bastiani, 2019) includes Beta variants within the GAMLSS framework, supporting distributional regression for location, scale, and shape

parameters. However, none of these packages provides the complete GKw hierarchy covering the seven nested distributions ranging from two to five parameters with analytical derivatives essential for efficient inference. The **gkwdist** package fills this gap by offering: (i) all subfamilies in a unified API following standard R conventions; (ii) exact analytical gradients and Hessians for each distribution; and (iii) numerically stable evaluation across challenging parameter regions via robust logarithmic transformations.

## Mathematical Framework

Let  $X \sim \text{GKw}(\alpha, \beta, \gamma, \delta, \lambda)$  with all parameters positive. The probability density function is:

$$f(x; \theta) = \frac{\lambda \alpha \beta}{B(\gamma, \delta + 1)} x^{\alpha-1} (1 - x^\alpha)^{\beta-1} w^{\gamma\lambda-1} (1 - w^\lambda)^\delta$$

where  $w = 1 - (1 - x^\alpha)^\beta$ ,  $\theta = (\alpha, \beta, \gamma, \delta, \lambda)^\top$ , and  $B(\cdot, \cdot)$  denotes the Beta function. The cumulative distribution function is  $F(x) = I_{w^\lambda}(\gamma, \delta + 1)$ , where  $I_y(a, b)$  represents the regularized incomplete beta function.

Seven subfamilies arise through parameter constraints (Table 1). The Kumaraswamy (Kw) and Exponentiated Kumaraswamy (EKw) subfamilies admit closed-form quantile functions, enabling efficient random variate generation via inversion. The Beta distribution emerges as a special case when  $\alpha = \beta = \lambda = 1$ , establishing a direct connection to the foundational work of Ferrari & Cribari-Neto (2004) with a simple variation in its second parameter  $\alpha = \gamma$  and  $\beta = \delta + 1$ .

**Table 1:** GKw subfamily hierarchy with parameter constraints and quantile tractability.

Subfamily	Parameters	Constraint	Closed Quantile
GKw ( <b>gkw</b> )	5	—	No
BKw ( <b>bkw</b> )	4	$\lambda = 1$	No
KKw ( <b>kkw</b> )	4	$\gamma = 1$	Yes
EKw ( <b>ekw</b> )	3	$\gamma = 1, \delta = 0$	Yes
Mc ( <b>mc</b> )	3	$\alpha = \beta = 1$	No
Kw ( <b>kw</b> )	2	$\gamma = \delta = 0, \lambda = 1$	Yes
Beta ( <b>beta_</b> )	2	$\alpha = \beta = \lambda = 1$	No

## Software Design

The package prioritizes computational efficiency and numerical stability through complete C++ implementation via **RcppArmadillo** (Eddelbuettel & Sanderson, 2014). Key design decisions include:

- Unified API:** All distribution functions (**dgkw**, **pgkw**, **qgkw**, **rgkw** and analogous functions for each subfamily) and inference functions (negative log-likelihood **llgkw**, gradient **grgkw**, Hessian **hsgkw**) follow standard R naming conventions.
- Analytical derivatives:** Exact score vectors and Hessian matrices enable gradient-based optimization via **optim()** with method "BFGS", "L-BFGS-B" among others, substantially improving convergence speed compared to numerical differentiation.
- Numerical stability:** Logarithmic transformations handle extreme parameter values and boundary conditions, preventing underflow/overflow in density calculations.

```
library(gkwdist)
x <- rkw(1000, alpha = 2, beta = 3)
fit <- optim(c(1, 1), fn = llkw, gr = grkw, data = x, method = "BFGS")
```

Mathematical derivations of all analytical gradients and Hessians are documented in the package vignettes available at the package website.

## Research Impact Statement

The `gkwdist` package serves as the computational foundation for `gkwreg` (Lopes, 2025), which implements regression models for bounded responses using the GKw family. The package has been available on CRAN since November 2025, with source code hosted on GitHub under the MIT license. The complete test suite achieves over 95% code coverage, and all analytical derivatives have been validated against numerical differentiation via the `numDeriv` package with near zero precision.

Applications span hydrology (Kumaraswamy, 1980), econometrics, ecology, and health sciences, domains where bounded data modeling is essential. The hierarchical structure facilitates systematic model comparison, allowing researchers to identify the most parsimonious adequate model for their specific application.

## AI Usage Disclosure

Documentation and manuscript text were refined with assistance from Claude (Anthropic, version Opus 4.5). Mathematical derivations of score vectors and Hessian matrices were independently verified by systematic comparison with numerical differentiation via the `numDeriv` package across multiple parameter configurations. The core C++ implementation represents substantial original development effort by the author.

## Acknowledgements

The author thanks Prof. Wagner Hugo Bonat (UFPR) for guidance on statistical methodology and the R community for valuable feedback during package development.

## References

- Carrasco, J. M. F., Ferrari, S. L. P., & Cordeiro, G. M. (2010). A new generalized Kumaraswamy distribution. Retrieved from <https://arxiv.org/abs/1004.0911>
- Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34(2), 1–24. doi:10.18637/jss.v034.i02
- Eddelbuettel, D., & Sanderson, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics & Data Analysis*, 71, 1054–1063. doi:10.1016/j.csda.2013.02.005
- Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799–815. doi:10.1080/0266476042000214501
- Jones, M. C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, 6(1), 70–81. doi:10.1016/j.stamet.2008.04.001
- Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1–2), 79–88. doi:10.1016/0022-1694(80)90036-0
- Lopes, J. E. (2025). *gkwreg: Generalized kumaraswamy regression models for bounded data*. Retrieved from <https://CRAN.R-project.org/package=gkwreg>

- Rigby, R. A., Stasinopoulos, M. D., Heller, G. Z., & De Bastiani, F. (2019). *Distributions for modeling location, scale, and shape: Using GAMLSS in R*. Boca Raton: Chapman; Hall/CRC. doi:[10.1201/9780429298547](https://doi.org/10.1201/9780429298547)
- Wolodzko, T. (2020). *extraDistr: Additional univariate and multivariate distributions*. Retrieved from <https://CRAN.R-project.org/package=extraDistr>
- Yee, T. W. (2015). *Vector generalized linear and additive models: With an implementation in R*. New York: Springer. doi:[10.1007/978-1-4939-2818-7](https://doi.org/10.1007/978-1-4939-2818-7)