

Packages **mggd** et **mcauchy** – Distribution gaussienne généralisée multivariée, distribution de Cauchy multivariée

Pierre Santagostini 1*

Nizar Bouhlef 2†

Résumé

Nous présentons deux packages qui fournissent des outils pour des lois de probabilité multivariées :

- **mggd** : distributions gaussiennes généralisées multivariées
- **mcauchy** : distributions de Cauchy multivariées

Ces deux packages permettent le calcul de la divergence de Kullback-Leibler entre deux densités de probabilité multivariées, en fonction de leurs paramètres. Pour cela, on utilise une expression analytique des divergences, grâce à des calculs publiés par Bouhlef and Dziri (2019) et Bouhlef and Rousseau (2022).

En outre, ces packages fournissent des outils pour ces distributions de probabilité multivariées : valeur de la densité de probabilité, estimation des paramètres, simulation d'échantillons, graphiques.

Mots-clefs : Package – Loi gaussienne généralisée multivariée – Loi de Cauchy multivariée – Divergence de Kullback-Leibler

Distribution gaussienne généralisée multivariée

Soit $\mathbf{X}$ un vecteur aléatoire réel distribué selon la loi gaussienne généralisée multivariée (multivariate generalised Gaussian distribution, MGGD) à p variables, de vecteur moyenne $\boldsymbol{\mu}$, de matrice de dispersion Σ symétrique définie positive de dimensions $p \times p$ et de paramètre de forme $\beta > 0$. Il a pour densité de probabilité :

$$f(\mathbf{x}|\boldsymbol{\mu}, \Sigma, \beta) = \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}} \Gamma\left(\frac{p}{2\beta}\right)} \frac{\beta}{2^{\frac{p}{2\beta}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}((\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu}))^\beta}$$

$\Gamma(\cdot)$ étant la fonction gamma.

Divergence de Kullback-Leibler

La fonction **kldggd** du package **mggd** calcule la divergence de Kullback-Leibler (KL) entre deux MGGD centrées de paramètres $(\mathbf{0}, \Sigma_1, \beta_1)$ et $(\mathbf{0}, \Sigma_2, \beta_2)$. Son expression est donnée par Bouhlef and Dziri (2019) :

$$KL(\mathbf{X}_1||\mathbf{X}_2) = \ln \left(\frac{\beta_1 |\Sigma_1|^{-1/2} \Gamma\left(\frac{p}{2\beta_2}\right)}{\beta_2 |\Sigma_2|^{-1/2} \Gamma\left(\frac{p}{2\beta_1}\right)} \right) + \frac{p}{2} \left(\frac{1}{\beta_2} - \frac{1}{\beta_1} \right) \ln 2 - \frac{p}{2\beta_1} + 2^{\frac{\beta_2}{\beta_1}-1} \frac{\Gamma\left(\frac{\beta_2}{\beta_1} + \frac{p}{2\beta_1}\right)}{\Gamma\left(\frac{p}{2\beta_1}\right)} \lambda_p^{\beta_2} \\ \times F_D^{(p-1)} \left(-\beta_2, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_{p-1}, \frac{p}{2}; 1 - \frac{\lambda_{p-1}}{\lambda_p}, \dots, 1 - \frac{\lambda_1}{\lambda_p} \right)$$

où $\lambda_1 < \dots < \lambda_{p-1} < \lambda_p$ sont les valeurs propres de la matrice carrée $\Sigma_1 \Sigma_2^{-1}$ et $F_D^{(p-1)}$ est la fonction D -hypergéométrique de Lauricella définie comme suit :

$$F_D^{(n)}(a, b_1, \dots, b_n; g; x_1, \dots, x_n) = \sum_{m_1 \geq 0} \dots \sum_{m_n \geq 0} \frac{(a)_{m_1+\dots+m_n} (b_1)_{m_1} \dots (b_n)_{m_n}}{(g)_{m_1+\dots+m_n}} \frac{x_1^{m_1}}{m_1!} \dots \frac{x_n^{m_n}}{m_n!} \text{ si } |x_i| < 1, i = 1, \dots, n$$

*Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, F-49000 Angers, France, pierre.santagostini@agrocampus-ouest.fr

†Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, F-49000 Angers, France, nizar.bouhlef@agrocampus-ouest.fr

Manipulation de distributions gaussiennes généralisées multivariées

Le package **mggd** fournit aussi des outils pour la manipulation de MGGD :

- **mvddgd** : fonction de densité de probabilité
- **mvrddgd** : simulation d'un échantillon aléatoire
- **estparmvddgd** : estimation des paramètres $\boldsymbol{\mu}$, Σ et β
- **plotmvddgd** et **contourmvddgd** : représentation graphique d'une densité bivariable

Distribution de Cauchy multivariée

Soit $\mathbf{X}$ un vecteur aléatoire réel distribué selon la loi de Cauchy multivariée (multivariate Cauchy distribution, MCD) à p variables, de paramètre de position $\boldsymbol{\mu}$ et de matrice d'échelle Σ symétrique définie positive $p \times p$. Sa densité de probabilité est :

$$f(\mathbf{x}|\boldsymbol{\mu}, \Sigma) = \frac{\Gamma\left(\frac{1+p}{2}\right)}{\pi^{\frac{p}{2}} \Gamma\left(\frac{1}{2}\right) |\Sigma|^{\frac{1}{2}} [1 + (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})]^{\frac{1+p}{2}}}$$

Divergence de Kullback-Leibler

La fonction **kldcauchy** du package **mcauchy** calcule la divergence de Kullback-Leibler entre deux MCD de paramètres $(\mathbf{0}, \Sigma_1)$ et $(\mathbf{0}, \Sigma_2)$. Cette divergence est donnée par Bouhlel and Rousseau (2022).

Soient $\lambda_1 < \dots < \lambda_{p-1} < \lambda_p$ les valeurs propres de la matrice carrée $\Sigma_1 \Sigma_2^{-1}$,

$$KL(\mathbf{X}_1 || \mathbf{X}_2) = -\frac{1}{2} \ln \prod_{i=1}^p \lambda_i + \frac{1+p}{2} \times D$$

$$\text{avec : } D = \begin{cases} \ln \lambda_p - \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left(a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p} \right) \right\} \Big|_{a=0} & \text{si } \lambda_1 < 1, \lambda_p > 1 \\ -\frac{\partial}{\partial a} \left\{ F_D^{(p)} \left(a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1+p}{2}; 1 - \lambda_1, \dots, 1 - \lambda_p \right) \right\} \Big|_{a=0} & \text{si } \lambda_p < 1 \\ -\prod_{i=1}^p \frac{1}{\sqrt{\lambda_i}} \times \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left(\frac{1+p}{2}, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1+p}{2}; 1 - \frac{1}{\lambda_1}, \dots, 1 - \frac{1}{\lambda_p} \right) \right\} \Big|_{a=0} & \text{si } \lambda_1 > 1 \end{cases}$$

Manipulation de distributions de Cauchy multivariées

Le package **mcauchy** fournit aussi les mêmes outils que **mggd**, pour la manipulation de MCD :

- **mvdc** : fonction de densité de probabilité
- **mvrvc** : simulation d'un échantillon aléatoire
- **plotmvc** et **contourmvc** : représentation graphique d'une densité bivariable

Références

- Bouhlel, Nizar, and Ali Dziri. 2019. "Kullback-Leibler Divergence Between Multivariate Generalized Gaussian Distributions. *IEEE Signal Processing Letters* 26 (7): 1021–25. <https://doi.org/10.1109/LSP.2019.2915000>.
- Bouhlel, Nizar, and David Rousseau. 2022. "A Generic Formula and Some Special Cases for the Kullback–Leibler Divergence Between Central Multivariate Cauchy Distributions." *Entropy* 24 (6): 838. <https://doi.org/10.3390/e24060838>.
- Santagostini, Pierre, and Nizar Bouhlel. 2023a. *mcauchy: Multivariate Cauchy Distribution; Kullback-Leibler Divergence*. <https://CRAN.R-project.org/package=mcauchy>.
- . 2023b. *mggd: Multivariate Generalised Gaussian Distribution; Kullback-Leibler Divergence*. <https://CRAN.R-project.org/package=mggd>.