Packages mggd et mcauchyd
Distribution gaussienne généralisée, Distribution de Cauchy multivariée

Pierre Santagostini et **Nizar Bouhlel**

Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, F-49000 Angers

Distribution normale généralisée multivariée

Densité de probabilité d'un vecteur aléatoire de dimension p, de loi normale généralisée multivariée de paramètres $(\boldsymbol{\mu}, \Sigma, \beta)$:

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

Divergence de Kullback-Leibler entre deux vecteurs aléatoires de dimension p, de lois normales multivariées généralisées centrées, de paramètres $(\boldsymbol{0}, \Sigma_1, \beta_1)\ et(\boldsymbol{0}, \Sigma_2, \beta_2)$ [1] :

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

où $\lambda_1 < \dots < \lambda_{p-1} < \lambda_p$ sont les valeurs propres de $\Sigma_1\Sigma_2^{-1}$
et F_D^{p-1} est la fonction D-hypergéométrique de Lauricella.

Distribution de Cauchy multivariée

Densité de probabilité d'un vecteur aléatoire de dimension p, de loi de Cauchy multivariée de paramètres $(\boldsymbol{\mu}, \Sigma)$:

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

Divergence de Kullback-Leibler entre deux vecteurs aléatoires de dimension p, de lois de Cauchy centrées, de paramètres $(\boldsymbol{0}, \Sigma_1)\ et(\boldsymbol{0}, \Sigma_2)$ [2] :

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

où $\lambda_1 < \dots < \lambda_{p-1} < \lambda_p$ sont les valeurs propres de $\Sigma_1\Sigma_2^{-1}$

- Si $\lambda_1 < 1 < \lambda_p$,
  $D = \log\lambda_p - \frac{\partial}{\partial a}\left(F_D^{(p)}\left(a, \frac{1}{2}, \dots, \frac{1}{2}, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p}\right)\right)\Big|_{a=0}$
- Si $\lambda_p < 1$, $D = -\frac{\partial}{\partial a}\left(F_D^{(p)}\left(a, \frac{1}{2}, \dots, \frac{1}{2}; a + \frac{1+p}{2}; 1 - \lambda_1, \dots, 1 - \lambda_p\right)\right)\Big|_{a=0}$
- Si $1 < \lambda_1$, $D = -\prod_{i=1}^{p}\frac{1}{\sqrt{\lambda_i}}\frac{\partial}{\partial a}\left(F_D^{(p)}\left(\frac{1+p}{2}, \frac{1}{2}, \dots, \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{1}{\lambda_1}, \dots, 1 - \frac{1}{\lambda_p}\right)\right)\Big|_{a=0}$

mggd

Distribution normale généralisée de paramètres:

```
> mu <- c(0, 1, 4)
> Sigma <- matrix(c(0.8,0.3,0.2,0.3,0.2,0.1,0.2,0.1,0.2), nrow=3)
> beta <- 0.74
```

Densité de probabilité

```
> mvdggd(x=c(0, 1, 4), mu=mu, Sigma=Sigma, beta=beta)
[1] 0.285622
```

Simulation

```
> set.seed(1234)
> x <- mvrggd(n=1000, mu=mu, Sigma=Sigma, beta=beta)
> colMeans(x)
[1] -0.03557157  0.98157600  3.97975136
```

Estimation des paramètres

```
> estparmvggd(x, eps=1e-0.6)
$mu
[1] -0.03557157  0.98157600  3.97975136

$Sigma
          [,1]      [,2]      [,3]
[1,] 0.9239214 0.3502010 0.2439851
[2,] 0.3502010 0.2425337 0.1221494
[3,] 0.2439851 0.1221494 0.2486933

$beta
[1] 0.7784991

attr(,"epsilon")
[1] 1e-06
attr(,"k")
[1] 8
```

Arguments:

- x: data frame ou matrice numérique
- eps (optionnel): précision souhaitée pour l'estimation du paramètre `$beta`

Résultat:

- Liste de trois éléments (estimation des paramètres) avec deux attributs `"k"` et `"epsilon"` (nombre d'itérations et précision de l'estimation du paramètre `$beta`)

Divergence de Kullback-Leibler

```
> beta1 <- 0.74
> beta2 <- 0.55
> Sigma1 <- matrix(c(0.8,0.3,0.2, 0.3,0.2,0.1,0.2,0.1,0.2), nrow=3)
> Sigma2 <- matrix(c(1,0.3,0.2,0.3,0.5,0.1,0.2,0.1,0.7), nrow=3)
> kldggd(Sigma1, beta1, Sigma2, beta2, eps=1e-06)
[1] 1.662957
attr(,"k")
[1] 30
attr(,"epsilon")
[1] 1e-06
```

Arguments:

- `Sigma1`, `beta1`, `Sigma2`, `beta2`: paramètres des deux distributions
- eps (optionnel): précision du calcul de la fonction D-hypergéométrique de Lauricella

Résultat:

- Valeur numérique avec deux attributs `"k"` et `"epsilon"` (nombre d'itérations pour le calcul de la fonction D-hypergéométrique de Lauricella et précision de ce calcul)

mcauchyd

Distribution de Cauchy de paramètres:

```
> mu <- c(0, 1, 4)
> Sigma <- matrix(c(1,0.6,0.2,0.6,1,0.3,0.2,0.3,1), nrow=3)
```

Densité de probabilité

```
> mvdcd(x=c(1, 2, 3), mu=mu, Sigma=Sigma)
[1] 0.007732682
```

Simulation

```
> set.seed(1234)
> x <- mvrcd(n=1000, mu=mu, Sigma=Sigma)
> apply(x, 2, median)
[1] -0.001186104  0.967051925  3.967185009
```

Divergence de Kullback-Leibler

```
> Sigma1 <- matrix(c(1,0.6,0.2,0.6,1,0.3,0.2,0.3,1), nrow=3)
> Sigma2 <- matrix(c(1,0.3,0.1,0.3,1,0.4,0.1,0.4,1), nrow=3)
> kldcauchy(Sigma1, Sigma2)
[1] 0.05891708
attr(,"epsilon")
[1] 1e-06
attr(,"k")
[1] 15
```

Perspectives

- Un futur package R fournira des outils analogues pour la distribution t multivariée.
- D'autres mesures de distance/divergence seront ajoutées: Rényi, Bhattacharyya…

Bibliographie

[1] Nizar Bouhlel and Ali Dziri. 2019. Kullback-Leibler Divergence Between Multivariate Generalized Gaussian Distributions. IEEE Signal Processing Letters. IEEE Signal Processing Letters 26 (7): 1021–25. https://doi.org/10.1109/LSP.2019.2915000

[2] Nizar Bouhlel 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

[3] Pierre Santagostini and Nizar Bouhlel. 2023a. mcauchyd: Multivariate Cauchy Distribution; Kullback-Leibler Divergence. https://CRAN.R-project.org/package=mcauchyd

[4] Pierre Santagostini and Nizar Bouhlel. 2023b. mggd: Multivariate Generalised Gaussian Distribution; Kullback-Leibler Divergence. https://CRAN.R-project.org/package=mggd