Distribution du rapport de deux variables gaussiennes corrélées ou indépendantes

Pierre Santagostini[1] Nizar Bouhlel[2]

Résumé

Le package **gaussratiovegind** a été créé pour modéliser le rapport de deux variables aléatoires normales indépendantes et estimer les paramètres de sa densité de probabilité par deux méthodes : EM [1] et variationnelle [2]. Une nouvelle version de ce package étend cela au rapport de deux variables normales corrélées, avec une nouvelle méthode d'estimation pour estimer en plus le coefficient de corrélation. Une application intéressante est fournie par l'étude d'indices de végétation, par exemple les indices calculés sur des images de fluorescence de chlorophylle.

Mots-clefs : Ratio de gaussiennes - Estimation - Imagerie de fluorescence de chlorophylle - Epidémiologie végétale - Package

Rapport de deux variables aléatoires normales corrélées

Soient X et Y deux variables aléatoires normalement distribuées : $(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

de vecteur moyenne $\boldsymbol{\mu} = \left(\mu_x, \mu_y\right)^T$ et de matrice de variance $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & Cov(X,Y) \\ Cov(X,Y) & \sigma_y^2 \end{pmatrix}$

Nous nous intéressons au rapport $Z = \frac{X}{Y}$.

La densité de probabilité $f_Z(z) = f_Z\left(z; \beta, \rho, \delta_y, r\right)$ est définie par les paramètres $\beta = \frac{\mu_x}{\mu_y}$, $\rho = \frac{\sigma_y}{\sigma_x}$, $\delta_y = \frac{\sigma_y}{\mu_y}$ et $r = Cor(X, Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$ ($r \in [-1, 1]$) et est donnée par :

$$f_Z(z) = \frac{\rho\sqrt{1-r^2}}{\pi(\rho^2 z^2 - 2r\rho z + 1)} e^{\frac{-1}{2(1-r^2)}\left(\frac{\beta^2\rho^2 - 2r\beta\rho + 1}{\delta_y^2}\right)} {}_1F_1\left(1, \frac{1}{2}; \frac{1}{2\delta_y^2(1-r^2)} \frac{\left(\beta\rho^2 z - \rho r(z+\beta) + 1\right)^2}{\rho^2 z^2 - 2\rho r z + 1}\right)$$

où ${}_1F_1$ est la fonction hypergéométrique confluente de Kummer : ${}_1F_1(a, b; x) = \sum_{n=0}^{+\infty} \frac{(a)_n}{(b)_n} \frac{x^n}{n!}$,

$(\cdot)_n$ étant le symbole de Pochhammer : $(x)_n = \frac{\Gamma(x+n)}{\Gamma(x)} = x(x+1)...(x+n-1)$

Si X et Y sont indépendantes ($r = 0$), la densité de probabilité de $Z = \frac{X}{Y}$ devient [1] :

$$f_Z\left(z; \beta, \rho, \delta_y, r = 0\right) = \frac{\rho}{\pi(\rho^2 z^2 + 1)} e^{\frac{-1}{2}\left(\frac{\beta^2\rho^2 + 1}{\delta_y^2}\right)} {}_1F_1\left(1, \frac{1}{2}; \frac{1}{2\delta_y^2} \frac{\left(\beta\rho^2 z + 1\right)^2}{\rho^2 z^2 + 1}\right)$$

[1]Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, F-49000 Angers, France, pierre.santagostini@institut-agro.fr

[2]Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, F-49000 Angers, France, nizar.bouhlel@institut-agro.fr

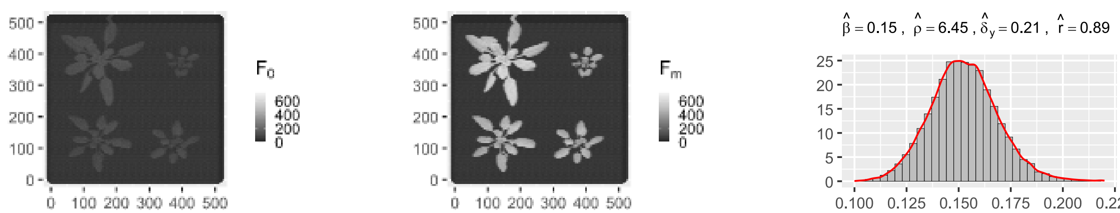
Estimation des paramètres

Le modèle de la loi de probabilité du rapport $Z = \frac{X}{Y}$ est identifiable. Les paramètres β, ρ, δ_y et r sont à estimer par la méthode EM (espérance-maximisation).

Tandis que la précédente version du package **gaussratiovegind** permettait l'estimation des paramètres β, ρ et δ_y dans le cas où X et Y sont independantes, la nouvelle version **gaussratiovegind_3.0.0** [3] permet aussi l'estimation des paramètres de Z quand X et Y sont corrélées.

Application : indice de fluorescence de chlorophylle

La fluorescence de chlorophylle est une technique d'imagerie utilisée pour le phénotypage des plantes. Un flash lumineux est envoyé sur une feuille, et on mesure la fluorescence émise par la chlorophylle. Les biologistes s'intéressent à l'indice de fluorescence de chlorophylle (*maximum yield of photosystem*, PSII) : $\frac{F_v}{F_m} = \frac{F_m - F_0}{F_m} = 1 - \frac{F_0}{F_m}$, où F_0 est le minimum de fluorescence (mesuré avant le flash, Figure 1a) et F_m est le maximum de fluorescence émise après le flash (Figure 1b). Ce rapport donne des informations sur l'état physiologique de la feuille. Les paramètres F_0 et F_m sont normalement distribués. Le rapport $\frac{F_0}{F_m}$ est donné par la Figure 1c.



(a) Paramètre F_0

(b) Paramètre F_m

(c) Histogramme de $\frac{F_0}{F_m}$, densité de probabilité estimée

Figure 1. – Fluorescence de chlorophylle

Bibliographie

[1] A. El Ghaziri, N. Bouhlel, N. Sapoukhina, et D. Rousseau, « On the Importance of Non-Gaussianity in Chlorophyll Fluorescence Imaging », *Remote Sensing*, vol. 15, nº 2, p. 528-529, 2023, doi: https://doi.org/10.3390/rs15020528.

[2] N. Bouhlel, F. Mercier, A. El Ghaziri, et D. Rousseau, « Parameter Estimation of the Normal Ratio Distribution with Variational Inference », in *2023 31st European Signal Processing Conference (EUSIPCO)*, 2023, p. 1823-1827. doi: 10.23919/EUSIPCO58844.2023.10290111.

[3] P. Santagostini, N. Bouhlel, A. El Ghaziri, et D. Rousseau, « gaussratiovegind: Distribution of Gaussian Ratios ». 2026. doi: 10.32614/CRAN.package.gaussratiovegind.