



# Discussion of the paper "Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC"

Théo Moins, Julyan Arbel, Anne Dutfoy, Stéphane Girard

## ► To cite this version:

Théo Moins, Julyan Arbel, Anne Dutfoy, Stéphane Girard. Discussion of the paper "Rank-Normalization, Folding, and Localization: An Improved  $\hat{R}$  for Assessing Convergence of MCMC". Bayesian Analysis, 2021, 16 (2), pp.711–712. 10.1214/20-ba1221 . hal-03222934

**HAL Id: hal-03222934**

**<https://inria.hal.science/hal-03222934>**

Submitted on 10 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Discussion of the paper “Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC” by Vehtari et al.

Théo Moins<sup>1</sup>, Julyan Arbel<sup>1</sup>, Anne Dutfoy<sup>2</sup>, and Stéphane Girard<sup>1</sup>

<sup>1</sup>Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK 38000 Grenoble, France

<sup>2</sup>EDF R&D dept. Périclès 91120 Palaiseau, France

**Abstract:** In this note, we discuss an adaptation of the quantile transformation introduced in [Vehtari et al. \(2021\)](#) to the computation of some new  $\hat{R}$  and analyse associated theoretical properties.

We have highly appreciated this contribution to improve MCMC convergence diagnostics and would like to thank the authors for their insightful paper. In this note we discuss an adaptation of the quantile transformation introduced in [Vehtari et al. \(2021\)](#) to the computation of some new  $\hat{R}$  and analyse associated theoretical properties. More specifically, we propose to compute a *continuous version*  $\hat{R}(x)$  for any level  $x$  based on indicator variables  $I(\theta^{(n,m)} \leq x)$ , rather than on the parameter values  $\theta^{(n,m)}$  themselves or their Gaussian version, see Equation (4.1) in [Vehtari et al. \(2021\)](#). This provides us with a function  $\hat{R}(x)$  defining a local convergence diagnostic for any  $x$ . The rank-normalization step is circumvented since working on Bernoulli random variables  $I(\theta^{(n,m)} \leq x)$  ensures the existence of all moments whatever the  $\theta^{(n,m)}$  distribution is. Assume that all elements  $\theta^{(m,1)}, \dots, \theta^{(m,N)}$  of a given chain  $m$  follow the same distribution  $F_m$  (without independence assumption) which may vary with  $m$ . Under this stationarity assumption, mean and variance of Bernoulli random variables  $I(\theta^{(n,m)} \leq x)$  can be easily written as functions of  $F_m$ , leading to an explicit expression for  $R^2(x)$ , the population version of  $\hat{R}^2(x)$ :

$$R^2(x) = 1 + \frac{1}{M} \frac{\sum_{i=1}^M \sum_{j=i+1}^M (F_i(x) - F_j(x))^2}{\sum_{i=1}^M F_i(x)(1 - F_i(x))}. \quad (1)$$

Clearly,  $R^2(x) \geq 1$  for all  $x \in \mathbb{R}$ , with equality iff all  $F_m$  coincide. Moreover,  $R^2(x) \rightarrow 1$  as  $|x| \rightarrow \infty$ . In order to condense the continuous index (1) into a scalar one, we may also consider its supremum over  $\mathbb{R}$ , denoted by  $R_\infty$ , which is computed in practice at empirical quantiles. Hereafter, we illustrate how the problems of traditional  $\hat{R}$  referred to as items 1. and 2. of Section 1.2 in [Vehtari et al. \(2021\)](#) are bypassed by our proposal on two toy examples. See also Figure 1 for finite sample results.

**Example 1: Same mean and different variances.** Here  $F_m(x) = F(x/\sigma_m)$  where  $\sigma_m$  is a scale parameter. As an example, we consider  $M$  chains uniformly distributed on  $[-\sigma_m, \sigma_m]$  with  $\sigma_m = \sigma \leq \sigma_M$  for all  $m \in \{1, \dots, M-1\}$  to model a lack of convergence. From (1), function  $R^2$  has a maximum reached at  $-\sigma/2$  and  $\sigma/2$ :

$$R_\infty^2 := \sup_{x \in \mathbb{R}} R^2(x) = R^2(\pm\sigma/2) = 1 + \frac{M-1}{M} \left( 1 - \frac{2}{1 + \frac{\sigma_M}{\sigma}} \right).$$

It appears that  $R_\infty^2$  is an increasing function of  $\sigma_M/\sigma$  with  $R_\infty^2 = 1$  iff  $\sigma_M = \sigma$ , and upper bounded by  $2 - 1/M$ .

**Example 2: Heavy tails and different locations.** Here  $F_m(x) = F(x - \mu_m)$ , with  $F$  a heavy-tailed distribution and  $\mu_m$  a location parameter. We consider Pareto distributed chains with shape parameter  $\alpha > 0$ , and starting point  $\eta > 0$  for  $(M-1)$  chains and  $\eta_M \geq \eta$  for the remaining one. In that case,  $R_\infty^2$  exists for any tail-index  $\alpha > 0$ :

$$R_\infty^2 = R^2(\eta_M) = 1 + \frac{1}{M} \left( \left( \frac{\eta_M}{\eta} \right)^\alpha - 1 \right).$$

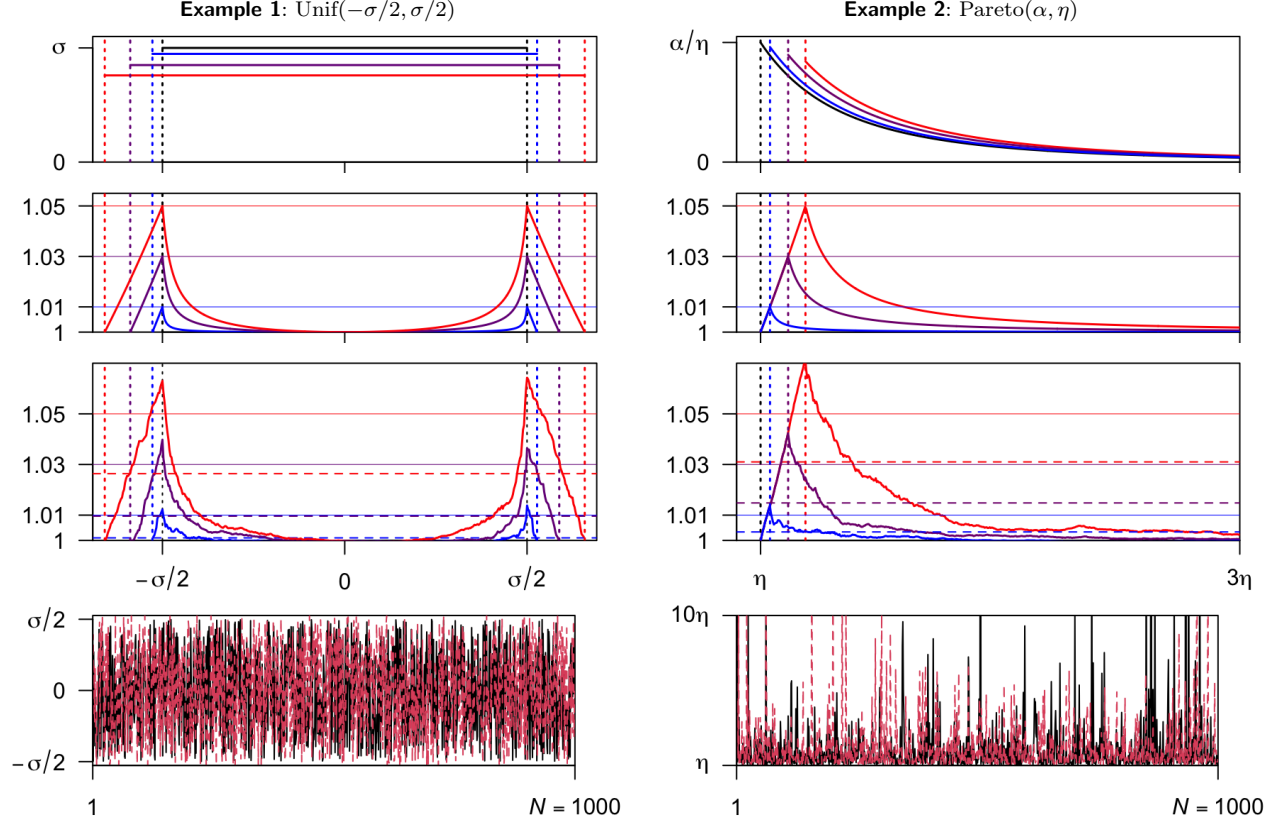


FIG 1. Illustrations with  $M = 4$  chains,  $N = 1000$  iterations each. Colors blue, violet and red resp. correspond to choices of 'diverging' distributions  $F_M$  such that our scalar diagnostic  $R_\infty$  matches with 1.01, 1.03 and 1.05. Top row: density of distributions  $F_m$  used, uniform (left) and Pareto (right). Colors for 'diverging'  $F_M$  and black for others. Second row: population (theoretical) version  $R(x)$  of our index. Third row: empirical version  $\hat{R}(x)$  of our index on simulated data; discussed rank- $\hat{R}$  shown as colored dashed lines. Bottom row: traceplots of one 'converging' chain ( $m = 1$ , black) and the 'diverging' chain ( $m = M$ , red case,  $R_\infty = 1.05$ ).

Thus,  $R_\infty$  is an increasing function of  $\eta_M/\eta$ , with  $R_\infty = 1$  iff  $\eta_M = \eta$ .

To conclude, it appears on these two examples that the proposed local version  $\hat{R}(x)$  allows both to localize the convergence of the MCMC in different quantiles of the distribution, and at the same time to handle the problems not detected by classical  $\hat{R}$  pointed out in the article. Compared to the rank- $\hat{R}$  proposed in Vehtari et al. (2021), our scalar version  $\hat{R}_\infty$  seems to be more conservative (see third row of Figure 1). Further investigation has to be done on a range of real world problems.

## References

- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). "Rank-Normalization, Folding, and Localization: An Improved  $\hat{R}$  for Assessing Convergence of MCMC." *Bayesian Analysis*, 1 – 28.  
 URL <https://doi.org/10.1214/20-BA1221>