



HAL
open science

Improving MCMC convergence diagnostic: a local version of R-hat

Théo Moins, Julyan Arbel, Stéphane Girard, Anne Dutfoy

► **To cite this version:**

Théo Moins, Julyan Arbel, Stéphane Girard, Anne Dutfoy. Improving MCMC convergence diagnostic: a local version of R-hat. BayesComp-ISBA workshop: Measuring the quality of MCMC output, Oct 2021, online, France. hal-03372736

HAL Id: hal-03372736

<https://inria.hal.science/hal-03372736>

Submitted on 20 Oct 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.

Improving MCMC convergence diagnostic: a local version of \hat{R}

Inria

Théo Moins¹, Julyan Arbel¹, Stéphane Girard¹, Anne Dutfoy²

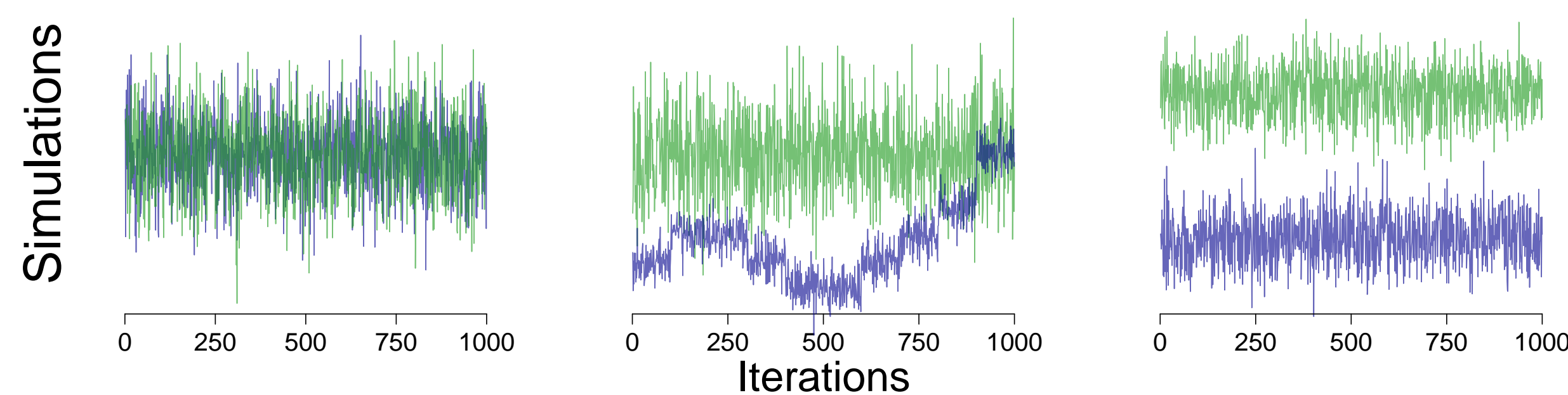
¹Statify, Inria Grenoble Rhône-Alpes,

²EDF R&D dept. Périclès



Overview: has the chain converged?

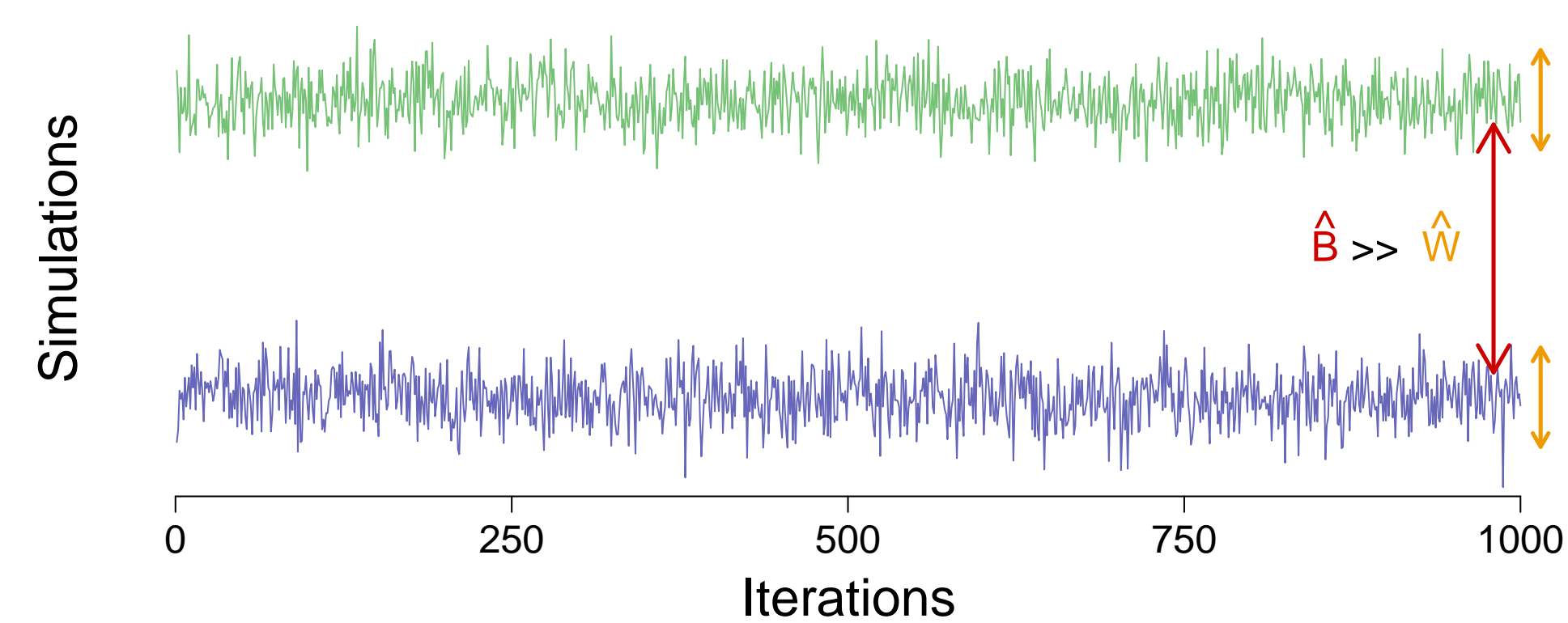
- Diagnosing convergence of Markov chain Monte Carlo (MCMC) is crucial in Bayesian analysis.
- It is frequent to run multiple chains in parallel with different initial values.
- At convergence, all chains should have a similar distribution.



\hat{R} : Potential Scale Reduction Factor [1]

M chains of size N , and $\theta^{(m,n)}$ the n th draw from chain m . Comparison of the **between-variance** B and the **within-variance** W of the chains: if $(\bar{\theta}^{(\cdot,m)}, s_m^2)$ is the mean and variance of $\theta^{(n,m)}$, and $\bar{\theta}^{(\cdot)}$ the mean of $\bar{\theta}^{(\cdot,m)}$,

$$\hat{R} = \sqrt{\frac{\hat{W} + \hat{B}}{\hat{W}}}, \quad \text{with} \quad \begin{cases} \hat{B} = \frac{1}{M-1} \sum_{m=1}^M (\bar{\theta}^{(\cdot,m)} - \bar{\theta}^{(\cdot)})^2, \\ \hat{W} = \frac{1}{M} \sum_{m=1}^M s_m^2. \end{cases}$$

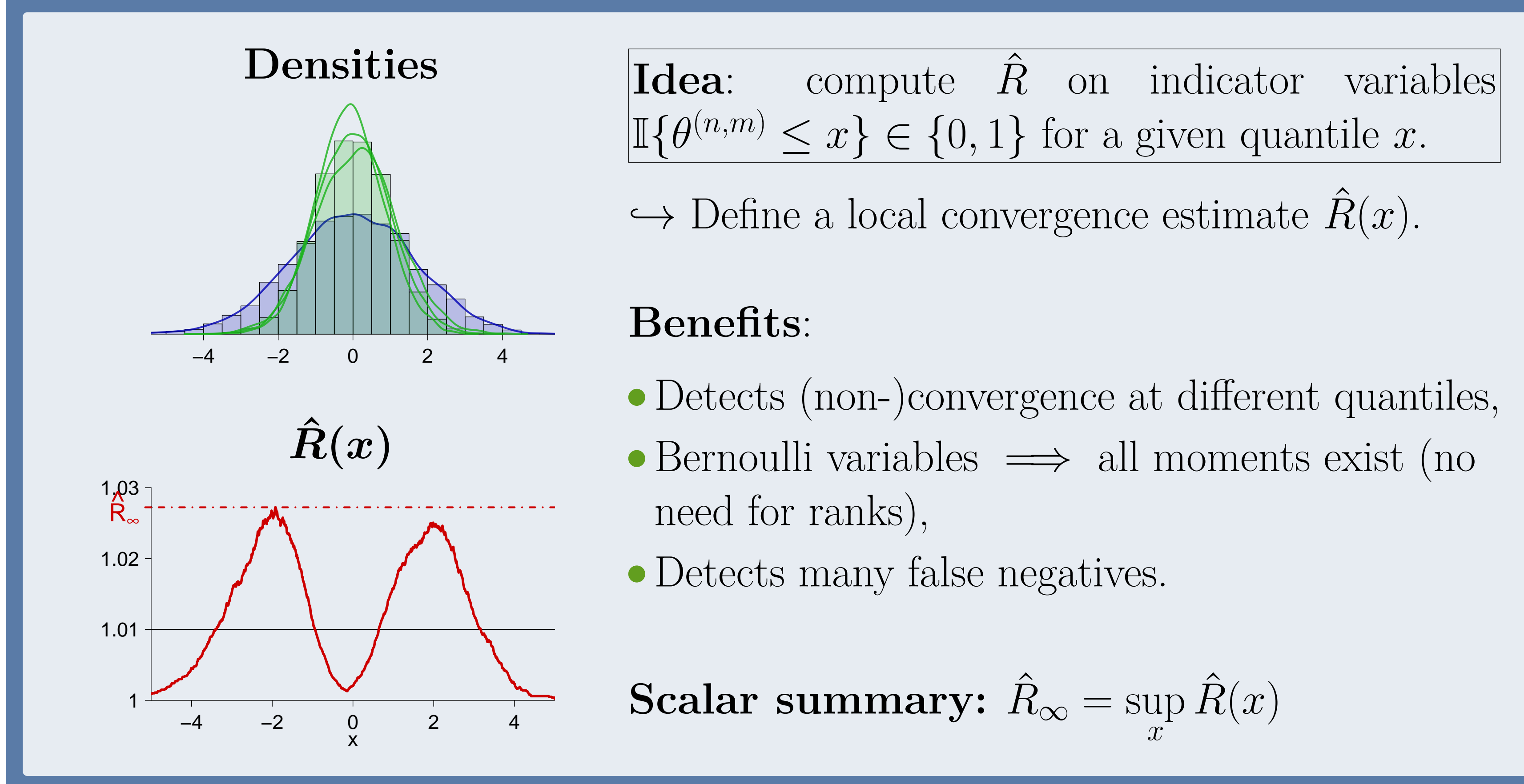


If $\hat{R} > 1.01$: convergence issue detected.

Ways of fooling \hat{R} and improvements [2]

- Chains with same mean and different variances: $\hat{R} \approx 1$
 \hookrightarrow **Bulk- \hat{R}** : \hat{R} on $z^{(n,m)}$, the normally transformed ranks of $\theta^{(m,n)}$.
- Chains with infinite mean and different locations: $\hat{R} \approx 1$
 \hookrightarrow **Tail- \hat{R}** : \hat{R} on $\zeta^{(n,m)}$, the deviations from the median.
 \implies $\text{Rank-}\hat{R} = \max(\text{Bulk-}\hat{R}, \text{Tail-}\hat{R})$

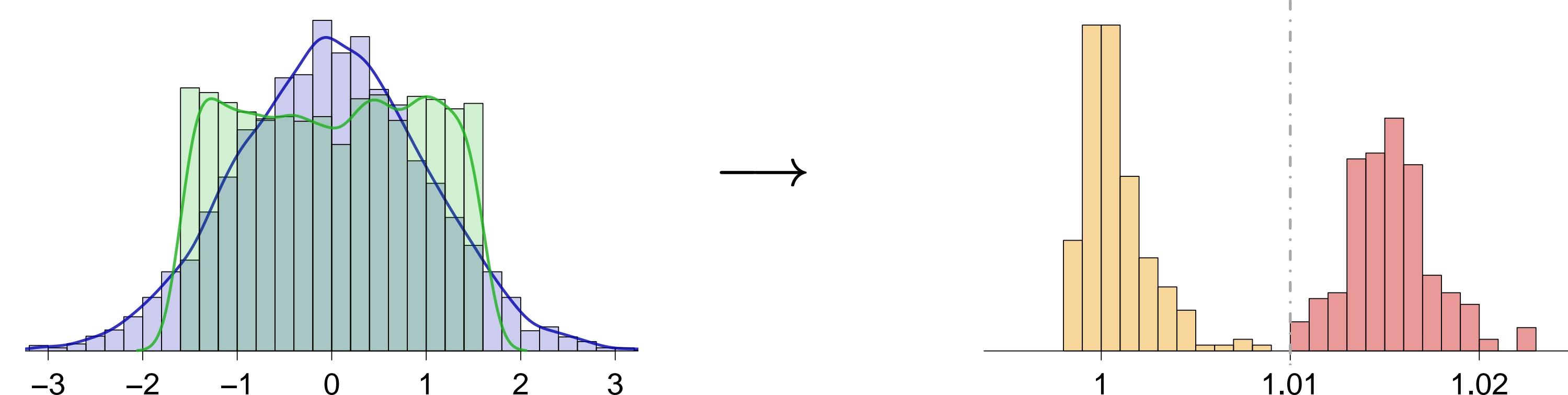
A local version of \hat{R}



Fooling rank- \hat{R} , robustness of \hat{R}_∞

Idea: find two distributions such that the variables $z^{(n,m)}$ and $\zeta^{(n,m)}$ share the same mean, to fool Bulk- \hat{R} and Tail- \hat{R} at the same time \implies Rank- $\hat{R} < 1.01$, and $\hat{R}_\infty > 1.01$.

Uniform and Normal densities 200 replications of Rank- \hat{R} and \hat{R}_∞



Population version

Assume chain m has stationary distribution F_m . As $\mathbb{E}[\mathbb{I}\{\theta^{(n,m)} \leq x\}] = F_m(x)$, and $\text{Var}[\mathbb{I}\{\theta^{(n,m)} \leq x\}] = F_m(x) - F_m^2(x)$, theoretical within-variance $W(x)$ and between-variance $B(x)$ are

$$W(x) = \frac{1}{M} \sum_{m=1}^M (F_m(x) - F_m^2(x)), \quad \text{and} \quad B(x) = \frac{1}{M} \sum_{m=1}^M F_m^2(x) - \left(\frac{1}{M} \sum_{m=1}^M F_m(x) \right)^2.$$

\implies population version of $\hat{R}(x)$ is

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

- Properties:**
- $R \geq 1$, and $R \equiv 1 \iff$ all F_m are equal,
 - R_∞ invariant with respect to reparameterisation.

Multivariate extension

Idea: compute \hat{R} on $\mathbb{I}\{\theta_1^{(n,m)} \leq x_1, \dots, \theta_d^{(n,m)} \leq x_d\}$.

- Eq. (1) remains valid for $R(\mathbf{x})$ with $\mathbf{x} = (x_1, \dots, x_d)$.
- Invariance to reparametrization: if margins are equal, we can assume uniform margins and compute $R(\mathbf{x})$ on M copulas (instead of M CDFs).

\hookrightarrow Two-step algorithm for multivariate diagnostic:

1. Compute the univariate $\hat{R}_{\infty,i}$ separately on each coordinate i .
2. If $\hat{R}_{\infty,i} < 1.01$ for all i , compute the multivariate \hat{R}_∞ to check convergence of the dependence structure.

- With $M = 2$, copulas C_1 and C_2 and $\mathbf{u} \in [0, 1]^d$:

$$R(\mathbf{u}) = \sqrt{1 + \frac{1}{2} \frac{(C_1(\mathbf{u}) - C_2(\mathbf{u}))^2}{C_1(\mathbf{u})(1 - C_1(\mathbf{u})) + C_2(\mathbf{u})(1 - C_2(\mathbf{u}))}}.$$

Upper bound in the multivariate case

Denote by $R_\infty(C_1, C_2)$, the value of R_∞ for C_1 and C_2 ($M = 2$). Let (C_-, C_+) two bounding copulas:

$$\begin{cases} C_-(\mathbf{u}) \leq C_1(\mathbf{u}) \leq C_+(\mathbf{u}) \\ C_-(\mathbf{u}) \leq C_2(\mathbf{u}) \leq C_+(\mathbf{u}) \end{cases} \quad \forall \mathbf{u} \in [0, 1]^d.$$

Then

$$R_\infty(C_1, C_2) \leq R_\infty(C_-, C_+).$$

Corollary: if W_d and M_d are the lower and upper Fréchet-Hoeffding copulas in dimension d ,

$$R_\infty(C_1, C_2) \leq R_\infty(W_d, M_d) = \sqrt{\frac{d+1}{2}}.$$

References

- [1] A. Gelman and D. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.
- [2] A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner. Rank-Normalization, Folding, and Localization: An Improved \hat{R} for Assessing Convergence of MCMC. *Bayesian Analysis*, 2021.
- [3] T. Moins, J. Arbel, A. Dutfoy, and S. Girard. Discussion of [2]. *Bayesian Analysis*, 2021.