On the use of a local $\hat{R}$ to improve MCMC convergence diagnostic

Théo Moins[1]

Joint with Julyan Arbel[1], Anne Dutfoy[2] Stéphane Girard[1]

[1]Statify, Inria Grenoble Rhône-Alpes

[2]EDF R&D dept. Périclès

May 17, 2022

Context

Limits of extrapolation associated with Bayesian extreme value models.

Aim: Understand the risks of hazardous meteorological events.

Inondations : le Lot-et-Garonne touché par la "crue la plus importante depuis quarante ans"
(Source: lemonde.fr, Février 2021)

MCMC

Bayesian inference on $\theta \sim \pi \implies$ computation of $\mathbb{E}_\pi[f(\theta)] = \int f(\theta)\pi(\theta)d\theta$.

MCMC (Markov Chain Monte Carlo):

Monte Carlo	Markov Chain
$\mathbb{E}[f(\theta)] \approx \frac{1}{n}\sum_{i=1}^n f(\theta_i)$	$\theta_{i+1} \mid \theta_i \sim P(\theta_i, \cdot)$

MCMC

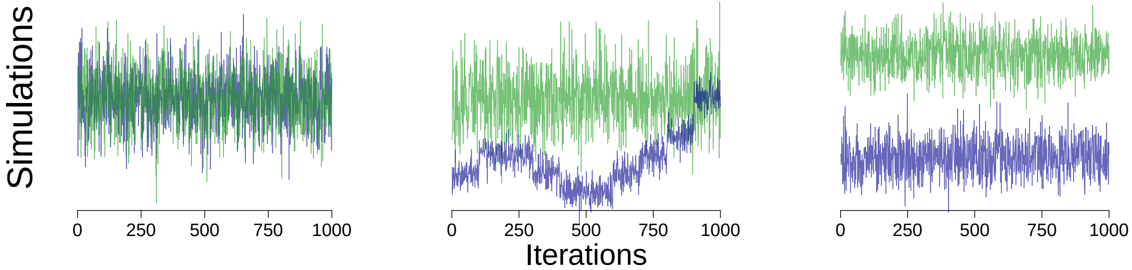
Bayesian inference on $\theta \sim \pi \implies$ computation of $\mathbb{E}_\pi[f(\theta)] = \int f(\theta)\pi(\theta)d\theta$.

MCMC (Markov Chain Monte Carlo):

Monte Carlo	Markov Chain
$\mathbb{E}[f(\theta)] \approx \frac{1}{n}\sum_{i=1}^{n} f(\theta_i)$	$\theta_{i+1} \mid \theta_i \sim P(\theta_i, \cdot)$

- **Algorithms**: Metropolis–Hastings, Gibbs sampling, Hamiltonian Monte Carlo (HMC) (Neal, 2011), No U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014), etc.
- **Librairies**: JAGS (Plummer et al., 2003), Stan (Carpenter et al., 2017), PyMC3 (Salvatier et al., 2016)...

Has the chain(s) converged? Need for multiple chains



$\hat{R}$ (aka potential scale reduction factor)

Introduced by Gelman and Rubin (1992).

Consider m chains of size n, with $\theta^{(i,j)}$ denoting the ith draw from chain j.

Comparison of the **between-variance** B and the **within-variance** W of the chains:

$$\boxed{\hat{R} = \sqrt{\frac{\hat{W}+\hat{B}}{\hat{W}}}}$$

$\hat{R}$ (aka potential scale reduction factor)

Introduced by Gelman and Rubin (1992).
Consider m chains of size n, with $\theta^{(i,j)}$ denoting the ith draw from chain j.
Comparison of the **between-variance** B and the **within-variance** W of the chains:

$$\boxed{\hat{R} = \sqrt{\tfrac{\hat{W}+\hat{B}}{\hat{W}}}}$$

$$\text{Between var}: \hat{B} = \frac{1}{m-1}\sum_{j=1}^{m}(\bar{\theta}^{(.,j)} - \bar{\theta}^{(.,.)})^2, \quad \text{where } \bar{\theta}^{(.,j)} = \frac{1}{n}\sum_{i=1}^{n}\theta^{(i,j)}, \quad \bar{\theta}^{(.,.)} = \frac{1}{m}\sum_{j=1}^{m}\bar{\theta}^{(.,j)},$$

$$\text{Within var}: \hat{W} = \frac{1}{m}\sum_{j=1}^{m}s_j^2, \quad \text{where } s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\theta^{(i,j)} - \bar{\theta}^{(.,j)})^2.$$

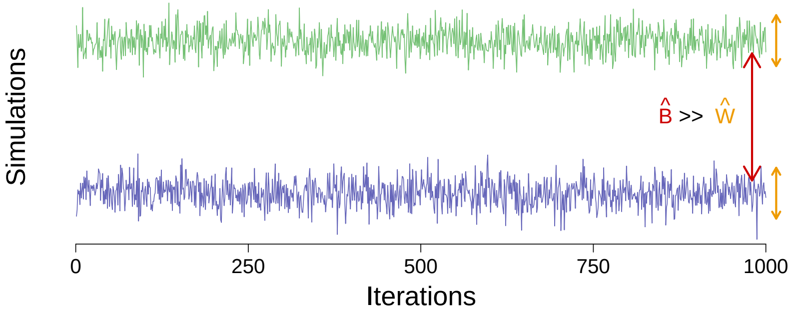
$\hat{R}$ (aka potential scale reduction factor)

Introduced by Gelman and Rubin (1992).

Consider m chains of size n, with $\theta^{(i,j)}$ denoting the ith draw from chain j.

Comparison of the **between-variance** B and the **within-variance** W of the chains:

$$\boxed{\hat{R} = \sqrt{\frac{\hat{W}+\hat{B}}{\hat{W}}}}$$



$\hat{R}$ (aka potential scale reduction factor)

Introduced by Gelman and Rubin (1992).
Consider m chains of size n, with $\theta^{(i,j)}$ denoting the ith draw from chain j.
Comparison of the **between-variance** B and the **within-variance** W of the chains:

$$\boxed{\hat{R} = \sqrt{\frac{\hat{W}+\hat{B}}{\hat{W}}}}$$

Inference from iterative simulation using multiple sequences
A Gelman, DB Rubin - Statistical science, 1992 - projecteuclid.org
The Gibbs sampler, the algorithm of Metropolis and similar iterative simulation methods are potentially very helpful for summarizing multivariate distributions. Used naively, however, iterative simulation can give misleading answers. Our methods are simple and generally …
Cited by 13999 Related articles All 20 versions Import into BibTeX

Recent improvement: rank-$\hat{R}$ Vehtari et al. (2021)

Limitations of the different $\hat{R}$

To summarize, the main limitations are:

- It does not target a specific quantity of interest.
  Converging according to which quantity?

Limitations of the different $\hat{R}$

To summarize, the main limitations are:

- It does not target a specific quantity of interest.
  Converging according to which quantity?
- It is not robust to certain types of non-convergence.
  $\hat{R}$ and potentially also rank-$\hat{R}$

Limitations of the different $\hat{R}$

To summarize, the main limitations are:

- It does not target a specific quantity of interest.
  Converging according to which quantity?
- It is not robust to certain types of non-convergence.
  $\hat{R}$ and potentially also rank-$\hat{R}$
- It suffers from a lack of interpretability.
  What is R associated to $\hat{R}$?

Limitations of the different $\hat{R}$

To summarize, the main limitations are:

- It does not target a specific quantity of interest.
  Converging according to which quantity?
- It is not robust to certain types of non-convergence.
  $\hat{R}$ and potentially also rank-$\hat{R}$
- It suffers from a lack of interpretability.
  What is R associated to $\hat{R}$?
- It must be compared to an arbitrary chosen threshold.
  $\hat{R} \geq 1.1$? 1.01?

Limitations of the different $\hat{R}$

To summarize, the main limitations are:

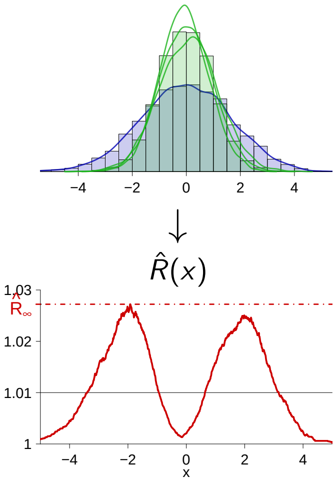
- It does not target a specific quantity of interest.
  Converging according to which quantity?
- It is not robust to certain types of non-convergence.
  $\hat{R}$ and potentially also rank-$\hat{R}$
- It suffers from a lack of interpretability.
  What is R associated to $\hat{R}$?
- It must be compared to an arbitrary chosen threshold.
  $\hat{R} \geq 1.1$? 1.01?
- It is associated with a univariate parameter.
  How to manage multiple parameters?

Local version of $\hat{R}$, or $\hat{R}(x)$

Idea: compute $\hat{R}$ on indicator variables $\mathbb{I}(\theta^{(i,j)} \leq x) \in \{0, 1\}$ for a given quantile x

Local version of $\hat{R}$, or $\hat{R}(x)$

Idea: compute $\hat{R}$ on indicator variables $\mathbb{I}(\theta^{(i,j)} \leq x) \in \{0,1\}$ for a given quantile x



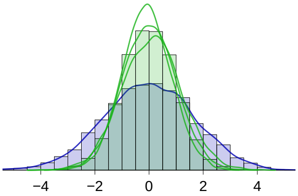
Local version of $\hat{R}$, or $\hat{R}(x)$

Idea: compute $\hat{R}$ on indicator variables $\mathbb{I}(\theta^{(i,j)} \leq x) \in \{0,1\}$ for a given quantile x

Benefits:

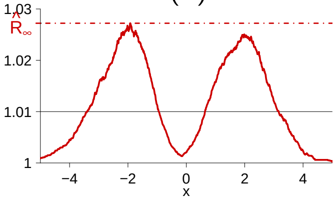
- It is local
  $\implies$ detects (non-)convergence locally
- Bernoulli variables
  $\implies$ all moments exist (no need for ranks)
- Detects many false negatives
- Scalar summary:

$$\hat{R}_\infty = \sup_x \hat{R}(x)$$



$\downarrow$

$\hat{R}(x)$



Limitations of the different $\hat{R}$

$\hat{R}_\infty$

To summarize, the main limitations are:

- **It does not target a specific quantity of interest.**
  Converging according to which quantity?
- It is not robust to certain types of non-convergence.
  $\hat{R}$ and potentially also rank-$\hat{R}$
- It suffers from a lack of interpretability.
  What is R associated to $\hat{R}$?
- It must be compared to an arbitrary chosen threshold.
  $\hat{R} \geq 1.1$? 1.01?
- It is associated with a univariate parameter.
  How to manage multiple parameters?

Limitations of the different $\hat{R}$

To summarize, the main limitations are:

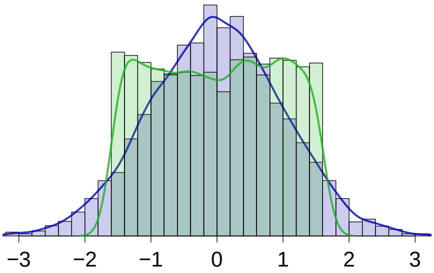
$\hat{\boldsymbol{R}}_\infty$

✓

- **It does not target a specific quantity of interest.**
  Converging according to which quantity?
- It is not robust to certain types of non-convergence.
  $\hat{R}$ and potentially also rank-$\hat{R}$
- It suffers from a lack of interpretability.
  What is R associated to $\hat{R}$?
- It must be compared to an arbitrary chosen threshold.
  $\hat{R} \geq 1.1$? 1.01?
- It is associated with a univariate parameter.
  How to manage multiple parameters?

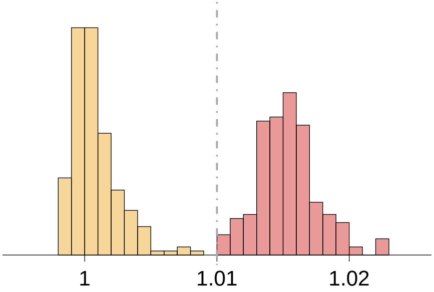
$\hat{R}_\infty$ where Rank-$\hat{R}$ is fooled

Uniform and Normal densities



$\rightarrow$

200 replications of Rank-$\hat{R}$ and $\hat{R}_\infty$



https://theomoins.github.io/localrhat/Simulations.html

Limitations of the different $\hat{R}$

To summarize, the main limitations are:

$\hat{R}_\infty$

✔

- It does not target a specific quantity of interest.
  Converging according to which quantity?
- **It is not robust to certain types of non-convergence.**
  $\hat{R}$ **and potentially also rank-**$\hat{R}$
- It suffers from a lack of interpretability.
  What is R associated to $\hat{R}$?
- It must be compared to an arbitrary chosen threshold.
  $\hat{R} \geq 1.1$? 1.01?
- It is associated with a univariate parameter.
  How to manage multiple parameters?

Limitations of the different $\hat{R}$

To summarize, the main limitations are:

$\hat{\boldsymbol{R}}_{\boldsymbol{\infty}}$

- It does not target a specific quantity of interest. ✓
  Converging according to which quantity?
- **It is not robust to certain types of non-convergence.** ✓
  $\hat{R}$ and potentially also rank-$\hat{R}$
- It suffers from a lack of interpretability.
  What is R associated to $\hat{R}$?
- It must be compared to an arbitrary chosen threshold.
  $\hat{R} \geq 1.1$? 1.01?
- It is associated with a univariate parameter.
  How to manage multiple parameters?

Theoretical properties

Assume chain $Z = j$ has distribution F_j (stationarity assumption, to focus on mixing). Then,

$$\mathbb{E}[I(\theta \leq x) \mid Z = j] = F_j(x), \quad \text{and} \quad \mathrm{Var}\left[I(\theta \leq x) \mid Z = j\right] = F_j(x) - F_j^2(x)$$

Theoretical properties

Assume chain $Z = j$ has distribution F_j (stationarity assumption, to focus on mixing). Then,

$$\mathbb{E}[I(\theta \leq x) \mid Z = j] = F_j(x), \quad \text{and} \quad \mathrm{Var}\,[I(\theta \leq x) \mid Z = j] = F_j(x) - F_j^2(x)$$

Theoretical $B(x)$ and $W(x)$:

$$B(x) = \frac{1}{m}\sum_{j=1}^{m} F_j^2(x) - \left(\frac{1}{m}\sum_{j=1}^{m} F_j(x)\right)^2, \qquad W(x) = \frac{1}{m}\sum_{j=1}^{m} \left(F_j(x) - F_j^2(x)\right).$$

Theoretical properties

Assume chain $Z = j$ has distribution F_j (stationarity assumption, to focus on mixing). Then,

$$\mathbb{E}[I(\theta \leq x) \mid Z = j] = F_j(x), \quad \text{and} \quad \mathrm{Var}\left[I(\theta \leq x) \mid Z = j\right] = F_j(x) - F_j^2(x)$$

Theoretical $B(x)$ and $W(x)$:

$$B(x) = \frac{1}{m}\sum_{j=1}^{m} F_j^2(x) - \left(\frac{1}{m}\sum_{j=1}^{m} F_j(x)\right)^2, \qquad W(x) = \frac{1}{m}\sum_{j=1}^{m}\left(F_j(x) - F_j^2(x)\right).$$

Proposition (Moins et al., 2022)

$R(x)$, the population version of $\hat{R}(x)$, can be written

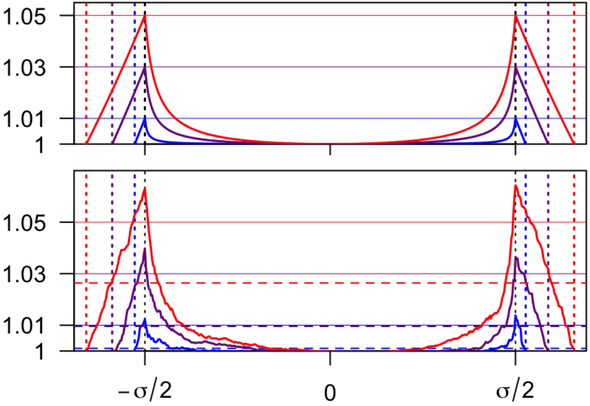
$$R(x) := \sqrt{\frac{W(x) + B(x)}{W(x)}} = \sqrt{1 + \frac{\sum_{j=1}^{m}\sum_{k=j+1}^{m}\left(F_k(x) - F_j(x)\right)^2}{m\sum_{j=1}^{m} F_j(x)(1 - F_j(x))}}.$$

Population $R(x)$

$$\boxed{R(x) = \sqrt{1 + \frac{\sum_{j=1}^{m}\sum_{k=j+1}^{m}\left(F_k(x)-F_j(x)\right)^2}{m\sum_{j=1}^{m}F_j(x)(1-F_j(x))}}}$$

Population $R(x)$

$$R(x) = \sqrt{1 + \frac{\sum_{j=1}^{m} \sum_{k=j+1}^{m} \left(F_k(x) - F_j(x)\right)^2}{m \sum_{j=1}^{m} F_j(x)(1 - F_j(x))}}$$

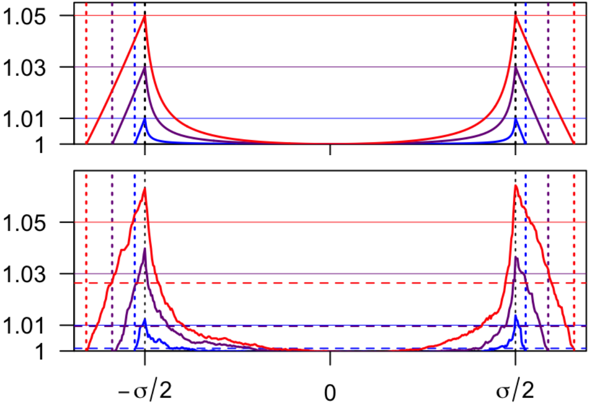


Population $R(x)$

$$R(x) = \sqrt{1 + \frac{\sum_{j=1}^{m}\sum_{k=j+1}^{m}\left(F_k(x) - F_j(x)\right)^2}{m\sum_{j=1}^{m} F_j(x)(1 - F_j(x))}}$$

Properties:

- $R \equiv 1 \iff$ all F_j are equal
- $R \geq 1$
- $\lim_{\pm\infty} R = 1$
- R_∞ invariant to monotone transformation



Limitations of the different $\hat{R}$

To summarize, the main limitations are: $\hat{R}_\infty$

- It does not target a specific quantity of interest. ✔
  Converging according to which quantity?
- It is not robust to certain types of non-convergence. ✔
  $\hat{R}$ and potentially also rank-$\hat{R}$
- **It suffers from a lack of interpretability.**
  What is R associated to $\hat{R}$?
- It must be compared to an arbitrary chosen threshold.
  $\hat{R} \geq 1.1?$ $1.01?$
- It is associated with a univariate parameter.
  How to manage multiple parameters?

Limitations of the different $\hat{R}$

To summarize, the main limitations are: $\hat{R}_\infty$

- It does not target a specific quantity of interest. ✔
  Converging according to which quantity?
- It is not robust to certain types of non-convergence. ✔
  $\hat{R}$ and potentially also rank-$\hat{R}$
- **It suffers from a lack of interpretability.** ✔
  What is R associated to $\hat{R}$?
- It must be compared to an arbitrary chosen threshold.
  $\hat{R} \geq 1.1?\ 1.01?$
- It is associated with a univariate parameter.
  How to manage multiple parameters?

Convergence properties of $\hat{R}(x)$

Assumption of a Markov chain central limit theorem:

$$\sqrt{nm}(\hat{F}(x) - F(x)) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(x)\right), \quad \text{with} \quad \hat{F}(x) = \frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}\mathbb{I}\{\theta^{(i,j)} \leq x\}$$

Convergence properties of $\hat{R}(x)$

Assumption of a Markov chain central limit theorem:

$$\sqrt{nm}(\hat{F}(x) - F(x)) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(x)\right), \quad \text{with} \quad \hat{F}(x) = \frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}\mathbb{I}\{\theta^{(i,j)} \leq x\}$$

Define a local effective sample size $\quad \mathrm{ESS}(x) := nm\dfrac{F(x)(1-F(x))}{\sigma^2(x)}$

$\hookrightarrow$ Number of samples to obtain the same variance in the i.i.d case.

Convergence properties of $\hat{R}(x)$

Assumption of a Markov chain central limit theorem:

$$\sqrt{nm}(\hat{F}(x) - F(x)) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(x)\right), \quad \text{with} \quad \hat{F}(x) = \frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}\mathbb{I}\{\theta^{(i,j)} \leq x\}$$

Define a local effective sample size $\quad \mathrm{ESS}(x) := nm\dfrac{F(x)(1 - F(x))}{\sigma^2(x)}$

$\hookrightarrow$ Number of samples to obtain the same variance in the i.i.d case.

Proposition (Moins et al., 2022)

Assume that all m chains are mutually independent and have converged to a common distribution F. Then for any $x \in \mathbb{R}$,

$$\mathrm{ESS}(x)(\hat{R}^2(x) - 1) \xrightarrow{d} \chi^2_{m-1} \quad \text{as} \quad n \to \infty.$$

Threshold elicitation: $\hat{R}(x)$

Let $z_{m-1,1-\alpha}$ be the quantile of level $1-\alpha$ of the χ^2_{m-1} distribution, and introduce the associated threshold (type I error)

$$R_{\lim,\alpha}(x) := \sqrt{1 + \frac{z_{m-1,1-\alpha}}{\mathrm{ESS}(x)}} \quad \Longrightarrow \quad \mathbb{P}(\hat{R}(x) \geq R_{\lim,\alpha}(x)) \simeq \alpha.$$

Threshold elicitation: $\hat{R}(x)$

Let $z_{m-1,1-\alpha}$ be the quantile of level $1-\alpha$ of the χ^2_{m-1} distribution, and introduce the associated threshold (type I error)

$$R_{\mathrm{lim},\alpha}(x) := \sqrt{1 + \frac{z_{m-1,1-\alpha}}{\mathrm{ESS}(x)}} \quad \Longrightarrow \quad \mathbb{P}(\hat{R}(x) \geq R_{\mathrm{lim},\alpha}(x)) \simeq \alpha.$$

$\mathrm{ESS}(x)$	α	m	$R_{\mathrm{lim},\alpha}(x)$
400	0.05	2	1.005
		4	1.010
		8	1.017
		15	1.029
		50	1.080
		100	1.144

Threshold elicitation: $\hat{R}(x)$

Let $z_{m-1,1-\alpha}$ be the quantile of level $1-\alpha$ of the χ^2_{m-1} distribution, and introduce the associated threshold (type I error)

$$R_{\mathrm{lim},\alpha}(x) := \sqrt{1 + \frac{z_{m-1,1-\alpha}}{\mathrm{ESS}(x)}} \quad \Longrightarrow \quad \mathbb{P}(\hat{R}(x) \geq R_{\mathrm{lim},\alpha}(x)) \simeq \alpha.$$

$\mathrm{ESS}(x)$	α	m	$R_{\mathrm{lim},\alpha}(x)$
400	0.05	2	1.005
		4	1.010
		8	1.017
		15	1.029
		50	1.080
		100	1.144

$\hookrightarrow$ 1.01 seems reasonable in the most common configurations.

Threshold elicitation: $\hat{R}_\infty$?

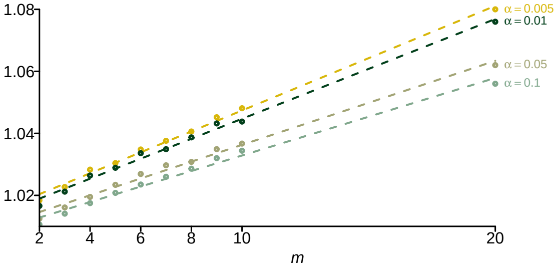
A threshold for $\hat{R}_\infty = \sup_x \hat{R}(x)$ require a result of convergence of the empirical process $\hat{R}(\cdot)$.

Threshold elicitation: $\hat{R}_\infty$?

A threshold for $\hat{R}_\infty = \sup_x \hat{R}(x)$ require a result of convergence of the empirical process $\hat{R}(\cdot)$.

Estimation using replications:

m	0.005	0.01	0.05	0.1
2	1.018	1.016	**1.012**	1.010
3	1.023	1.022	1.016	1.014
4	1.027	1.025	**1.020**	1.018
8	1.038	1.037	**1.031**	1.028
10	1.043	1.041	1.036	1.033
20	1.080	1.076	1.062	1.056



Limitations of the different $\hat{R}$

To summarize, the main limitations are:

$\hat{\boldsymbol{R}}_\infty$

- It does not target a specific quantity of interest. ✔
  Converging according to which quantity?
- It is not robust to certain types of non-convergence. ✔
  $\hat{R}$ and potentially also rank-$\hat{R}$
- It suffers from a lack of interpretability. ✔
  What is R associated to $\hat{R}$?
- **It must be compared to an arbitrary chosen threshold.**
  $\hat{R} \geq 1.1$? 1.01?
- It is associated with a univariate parameter.
  How to manage multiple parameters?

Limitations of the different $\hat{R}$

To summarize, the main limitations are:	$\boldsymbol{\hat{R}_\infty}$
• It does not target a specific quantity of interest. Converging according to which quantity?	✓
• It is not robust to certain types of non-convergence. $\hat{R}$ and potentially also rank-$\hat{R}$	✓
• It suffers from a lack of interpretability. What is R associated to $\hat{R}$?	✓
• **It must be compared to an arbitrary chosen threshold.** $\hat{R} \geq 1.1$? 1.01?	✓
• It is associated with a univariate parameter. How to manage multiple parameters?	

Preprint! 🎉

On the use of a local $\hat{R}$
to improve MCMC convergence diagnostic

Théo Moins * Julyan Arbel * Anne Dutfoy † Stéphane Girard *

March 7, 2022

Abstract

Diagnosing convergence of Markov chain Monte Carlo is crucial and remains an essentially unsolved problem. Among the most popular methods, the potential scale reduction factor, commonly named $\hat{R}$, is an indicator that monitors the convergence of output chains to a target distribution, based on a comparison of the between- and within-variances. Several improvements have been suggested since its introduction in the 90s. Here, we aim at better understanding the $\hat{R}$ behavior by proposing a localized version that focuses on quantiles of the target distribution. This new version relies on key theoretical properties of the associated population value. It naturally leads to proposing a new indicator $\hat{R}_\infty$, which is shown to allow both for localizing the Markov chain Monte Carlo convergence in different quantiles of the target distribution, and at the same time for handling some convergence issues not detected by other $\hat{R}$ versions.

T. Moins, J. Arbel, A. Dutfoy & S. Girard. (2022+) "*On the use of a local R-hat to improve MCMC convergence diagnostic*" https://hal.inria.fr/hal-03600407/document

References I

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2017). Stan: a probabilistic programming language. *Grantee Submission*, 76(1):1–32.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

Hoffman, M. D. and Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Moins, T., Arbel, J., Dutfoy, A., and Girard, S. (2022). On the use of a local $\hat{R}$ to improve MCMC convergence diagnostic.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11):2.

References II

Plummer, M. et al. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, page 10. Vienna, Austria.

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2:e55.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2):667 – 718.

Multivariate case

If parameter θ is d-dimensional: simple multivariate extension by computing $\hat{R}$ on indicator variables $I(\theta_1^{(i,j)} \leq x_1, \ldots, \theta_d^{(i,j)} \leq x_d)$

Multivariate case

If parameter θ is d-dimensional: simple multivariate extension by computing $\hat{R}$ on indicator variables $I(\theta_1^{(i,j)} \leq x_1, \ldots, \theta_d^{(i,j)} \leq x_d)$

As before, population version $R(\boldsymbol{x})$, with $\boldsymbol{x} = (x_1, \ldots, x_d)$:

$$R(\boldsymbol{x}) = \sqrt{1 + \frac{\sum_{j=1}^{m} \sum_{k=j+1}^{m} \left(F_j(\boldsymbol{x}) - F_k(\boldsymbol{x})\right)^2}{m \sum_{j=1}^{m} F_j(\boldsymbol{x})(1 - F_j(\boldsymbol{x}))}}.$$

Multivariate case

If parameter θ is d-dimensional: simple multivariate extension by computing $\hat{R}$ on indicator variables $I(\theta_1^{(i,j)} \leq x_1, \ldots, \theta_d^{(i,j)} \leq x_d)$

As before, population version $R(\boldsymbol{x})$, with $\boldsymbol{x} = (x_1, \ldots, x_d)$:

$$R(\boldsymbol{x}) = \sqrt{1 + \frac{\sum_{j=1}^{m} \sum_{k=j+1}^{m} \left(F_j(\boldsymbol{x}) - F_k(\boldsymbol{x})\right)^2}{m \sum_{j=1}^{m} F_j(\boldsymbol{x})(1 - F_j(\boldsymbol{x}))}}.$$

- $R \equiv 1 \iff$ all F_j are equal
- $R \geq 1$
- R_∞ invariant to monotone transformation $\implies$ if convergence of margins, we can compute R on M copulas (instead of M CDFs)

Multivariate case: upper bound

Assume $m = 2$ chains, with copulas C_1 and C_2 (in dim d), index denoted by $R_\infty(C_1, C_2)$.

Multivariate case: upper bound

Assume $m = 2$ chains, with copulas C_1 and C_2 (in dim d), index denoted by $R_\infty(C_1, C_2)$.

Lemma

Let (C_-, C_+) two bounding copulas in the sense that

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

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

Multivariate case: upper bound

Assume $m = 2$ chains, with copulas C_1 and C_2 (in dim d), index denoted by $R_\infty(C_1, C_2)$.

Lemma

Let (C_-, C_+) two bounding copulas in the sense that

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

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

Proposition (Moins et al., 2022)

Let W_d and M_d the lower and upper Fréchet–Hoeffding copulas in dimension d. Then

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

Multivariate case: bound refinement

Fréchet–Hoeffding copula bounds (comonotone random variables):

$$W_d(\boldsymbol{u}) := \max\left\{1 - d + \sum_{i=1}^{d} u_i, 0\right\} \quad \text{and} \quad M_d(\boldsymbol{u}) := \min\{u_1, \ldots, u_d\}.$$

Multivariate case: bound refinement

Fréchet–Hoeffding copula bounds (comonotone random variables):

$$W_d(\boldsymbol{u}) := \max\left\{1 - d + \sum_{i=1}^{d} u_i, 0\right\} \quad \text{and} \quad M_d(\boldsymbol{u}) := \min\{u_1, \dots, u_d\}.$$

Let us refine the upper bound by comparing with the independent copula $\Pi_d(\boldsymbol{u}) := \prod_{i=1}^{d} u_i$:

- Positive Lower Orthant Dependence (PLOD) copula:
  $\Pi_d(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq M_d(\boldsymbol{u})$ for all $\boldsymbol{u} \in [0,1]^d$
- Negative Lower Orthant Dependence (NLOD) copula:
  $W_d(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq \Pi_d(\boldsymbol{u})$ for all $\boldsymbol{u} \in [0,1]^d$

Multivariate case: bound refinement

Fréchet–Hoeffding copula bounds (comonotone random variables):

$$W_d(\boldsymbol{u}) := \max\left\{1 - d + \sum_{i=1}^{d} u_i, 0\right\} \quad \text{and} \quad M_d(\boldsymbol{u}) := \min\{u_1, \ldots, u_d\}.$$

Let us refine the upper bound by comparing with the independent copula $\Pi_d(\boldsymbol{u}) := \prod_{i=1}^{d} u_i$:

- Positive Lower Orthant Dependence (PLOD) copula:
  $\Pi_d(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq M_d(\boldsymbol{u})$ for all $\boldsymbol{u} \in [0,1]^d$
- Negative Lower Orthant Dependence (NLOD) copula:
  $W_d(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq \Pi_d(\boldsymbol{u})$ for all $\boldsymbol{u} \in [0,1]^d$

⚠ This does not define a total order on copulas!

Multivariate case: bound refinement

Let's stay in the case $m = 2$ chains.

Corollary (Moins et al., 2022)

For any two PLOD d-variate copulas C_1 and C_2, $R_\infty(C_1, C_2) \leq R_\infty(\Pi_d, M_d)$ with

$$\begin{cases} R_\infty(\Pi_2, M_2) = \sqrt{\frac{1}{2} + \frac{1}{\sqrt{3}}} \approx 1.038 & \text{if } d = 2, \\ \sqrt{\frac{d}{2\log d}}(1 + o(1)) \leq R_\infty(\Pi_d, M_d) \leq \sqrt{\frac{d+1}{2}} & \text{as } d \to \infty. \end{cases}$$

Multivariate case: bound refinement

Let's stay in the case $m = 2$ chains.

Corollary (Moins et al., 2022)

For any two PLOD d-variate copulas C_1 and C_2, $R_\infty(C_1, C_2) \leq R_\infty(\Pi_d, M_d)$ with

$$\begin{cases} R_\infty(\Pi_2, M_2) = \sqrt{\frac{1}{2} + \frac{1}{\sqrt{3}}} \approx 1.038 & \text{if } d = 2, \\ \sqrt{\frac{d}{2\log d}}(1 + o(1)) \leq R_\infty(\Pi_d, M_d) \leq \sqrt{\frac{d+1}{2}} & \text{as } d \to \infty. \end{cases}$$

Corollary (Moins et al., 2022)

For any two NLOD d-variate copulas C_1 and C_2, $R_\infty(C_1, C_2) \leq R_\infty(\Pi_d, W_d)$ with

$$R_\infty(\Pi_d, W_d) = \sqrt{1 + \frac{1}{2}\frac{1}{\left(1 - \frac{1}{d}\right)^{-d} - 1}}.$$

Multivariate case: bound refinement

Asymmetric behaviour:

- $R_\infty(\Pi_d, M_d)$ diverges with d at the (almost) same rate as $R_\infty(M_d, W_d)$,
- $R_\infty(\Pi_d, W_d) \xrightarrow[d\to\infty]{} 1.136$.

Illustration with $m = 2$ chains with bivariate normal distributions:

$$\boldsymbol{\theta}^{(i,1)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right), \quad \boldsymbol{\theta}^{(i,2)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\\rho & 1\end{pmatrix}\right), \quad \text{with} \quad \rho \in (-1, 1).$$

Multivariate case: bound refinement

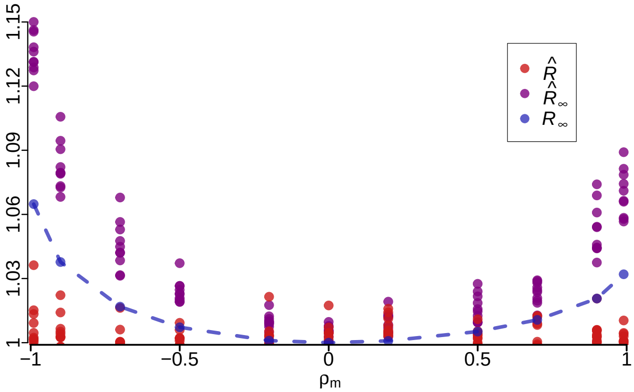
Illustration with bivariate normal distributions:

$$\boldsymbol{\theta}^{(i,1)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right), \quad \boldsymbol{\theta}^{(i,2)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\\rho & 1\end{pmatrix}\right), \quad \text{with} \quad \rho \in (-1,1).$$

Multivariate case: bound refinement

Illustration with bivariate normal distributions:

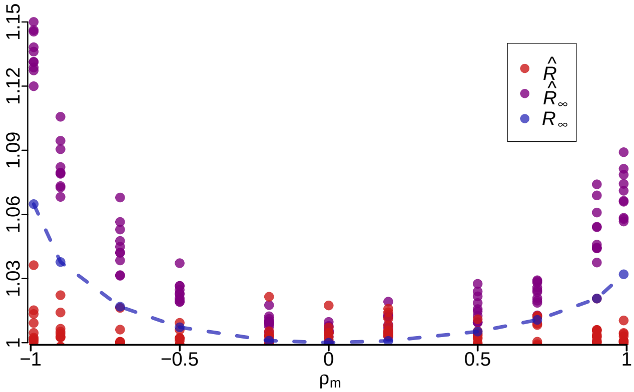
$$\boldsymbol{\theta}^{(i,1)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right), \quad \boldsymbol{\theta}^{(i,2)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\\rho & 1\end{pmatrix}\right), \quad \text{with} \quad \rho \in (-1,1).$$



Multivariate case: bound refinement

Illustration with bivariate normal distributions:

$$\boldsymbol{\theta}^{(i,1)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right), \quad \boldsymbol{\theta}^{(i,2)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\\rho & 1\end{pmatrix}\right), \quad \text{with} \quad \rho \in (-1,1).$$

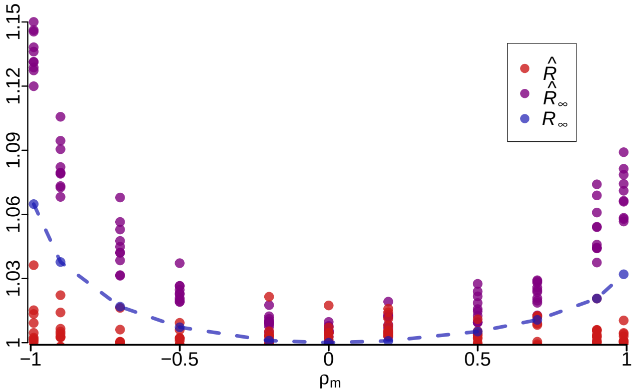


- PLOD and NLOD bounds when $|\rho| \to 1$,

Multivariate case: bound refinement

Illustration with bivariate normal distributions:

$$\boldsymbol{\theta}^{(i,1)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right), \quad \boldsymbol{\theta}^{(i,2)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\\rho & 1\end{pmatrix}\right), \quad \text{with} \quad \rho \in (-1,1).$$

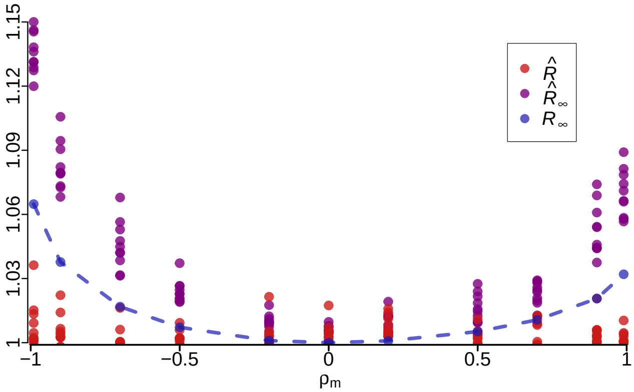


- PLOD and NLOD bounds when $|\rho| \to 1$,
- Asymmetry which favour NLOD when $d = 2$,

Multivariate case: bound refinement

Illustration with bivariate normal distributions:

$$\boldsymbol{\theta}^{(i,1)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right), \quad \boldsymbol{\theta}^{(i,2)} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\\rho & 1\end{pmatrix}\right), \quad \text{with} \quad \rho \in (-1,1).$$



- PLOD and NLOD bounds when $|\rho| \to 1$,
- Asymmetry which favour NLOD when $d = 2$,
- It can be inverted by computing $\hat{R}_\infty^-$ on $\mathbb{I}\{\theta_1^{(\cdot)} \leq x_1, \theta_2^{(\cdot)} \geq x_2\}$.

Multivariate case: bound refinement

$\hat{R}_{\infty}^{(\mathrm{max})} := \max(R_{\infty}^{+}, R_{\infty}^{-})$ consider symmetrically both directions of dependencies...

Multivariate case: bound refinement

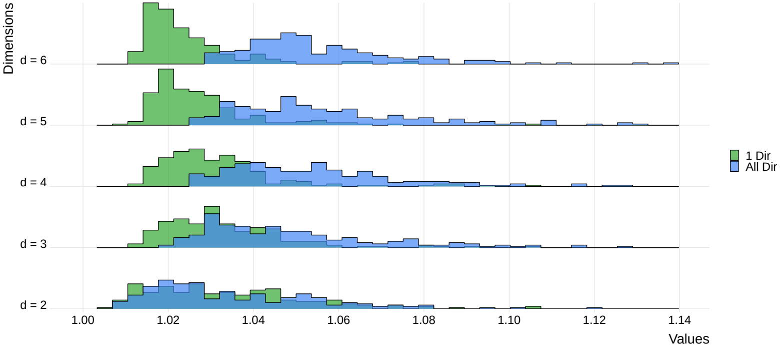
$\hat{R}_\infty^{(\max)} := \max(R_\infty^+, R_\infty^-)$ consider symmetrically both directions of dependencies...

... but in dimension d, 2^{d-1} different R_∞ to compute!

Multivariate case: bound refinement

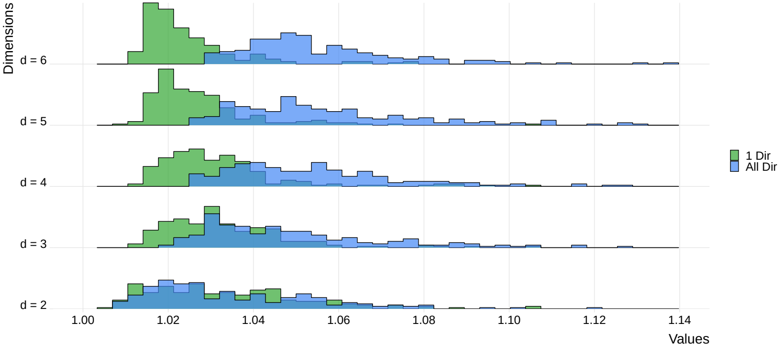
$\hat{R}_\infty^{(\max)} := \max(R_\infty^+, R_\infty^-)$ consider symmetrically both directions of dependencies...

... but in dimension d, 2^{d-1} different R_∞ to compute!



Multivariate case: bound refinement

$\hat{R}_{\infty}^{(\max)} := \max(R_{\infty}^{+}, R_{\infty}^{-})$ consider symmetrically both directions of dependencies...
... but in dimension d, 2^{d-1} different R_{∞} to compute!



Alternative: computation of $\hat{R}_{\infty}$ for a univariate function of the parameters