A Local Version of R to Improve MCMC Convergence Diagnostic

Théo Moins, Julyan Arbel, Stéphane Girard, Anne Dutfoy

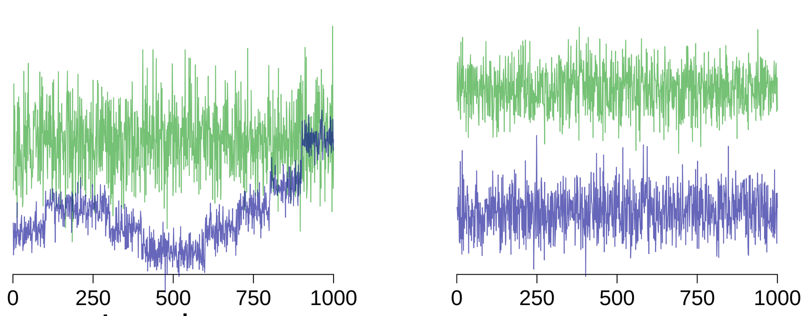
A Local Version of $\hat{R}$ to Improve MCMC Convergence Diagnostic

Théo Moins[1], Julyan Arbel[1], Stéphane Girard[1], Anne Dutfoy[2]

[1]Statify, Inria Grenoble Rhône-Alpes, [2]EDF R&D dept. Périclès

Overview: has the chain converged?

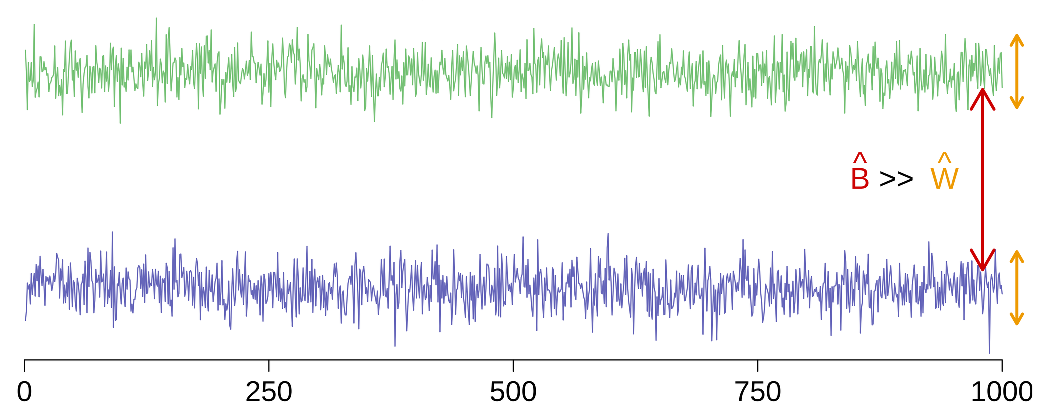
- Diagnosing convergence of Markov chain Monte Carlo (MCMC) is crucial in Bayesian analysis.
- It is frequent to run multiple chains in parallel with different initial values.
- At convergence, all chains should have a similar distribution.



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

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

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



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

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

Two cases where $\hat{R}$ fails:

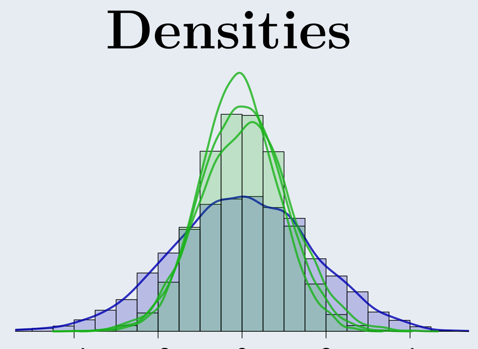
- Chains with infinite mean and different locations: $\hat{R} \approx 1$
  $\hookrightarrow$ **Bulk-$\hat{R}$**: $\hat{R}$ on $z^{(n,m)}$, the normally transformed ranks of $\theta^{(m,n)}$.
- Chains with same mean and different variances: $\hat{R} \approx 1$
  $\hookrightarrow$ **Tail-$\hat{R}$**: $\hat{R}$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

$\Downarrow$

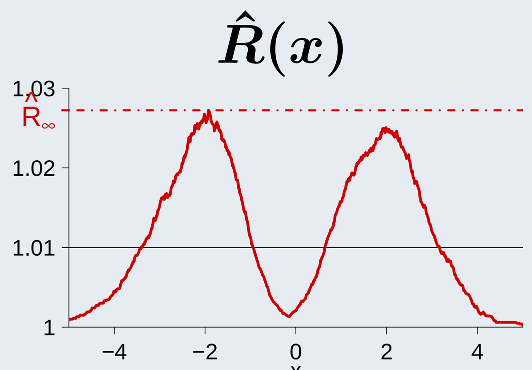
$\boxed{\text{Rank-}\hat{R} = \max(\text{Bulk-}\hat{R}, \text{Tail-}\hat{R})}$

A local version of $\hat{R}$

Densities



$\hat{\boldsymbol{R}}(\boldsymbol{x})$



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

Benefits:

- Detects (non-)convergence at different quantiles,
- Bernoulli variables $\implies$ all moments exist (no need for ranks),
- Detects many false negatives.

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

Population version

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

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

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

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

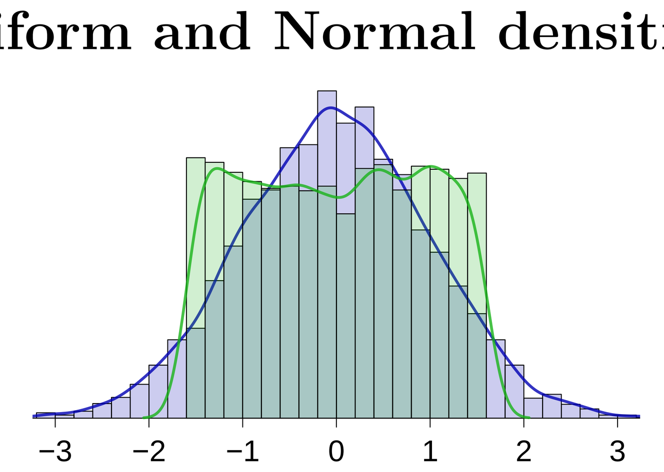
Properties:

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

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

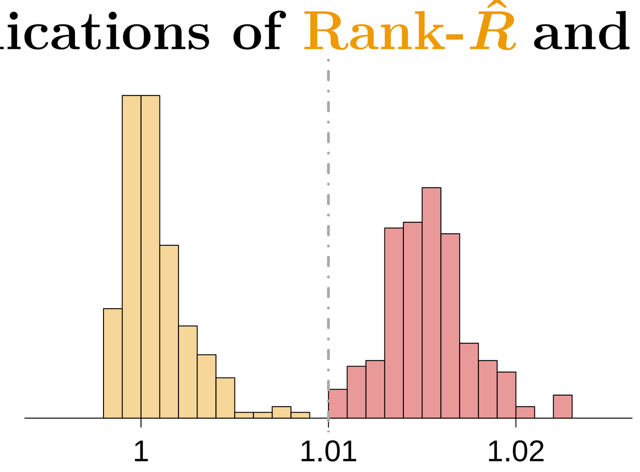
Idea: find two distributions such that the variables $z^{(n,m)}$ and $\zeta^{(n,m)}$ share the same mean, to fool Bulk-$\hat{R}$ and Tail-$\hat{R}$ at the same time $\implies$ Rank-$\hat{R} < 1.01$, and $\hat{R}_\infty > 1.01$.

Uniform and Normal densities



$\rightarrow$

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



Threshold elicitation

Assume all $\hat{F}_j \to F$ with a Markov chain central limit theorem:

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

Define $\text{ESS}(x) := nm\frac{F(x)(1-F(x))}{\sigma^2(x)}$. Then for any $x \in \mathbb{R}$,

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

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

With $\alpha = 0.05$, $m = 4$, $\text{ESS}(x) = 400$: $R_{\text{lim},\alpha}(x) \approx 1.010$

Multivariate extension

$\boxed{\textbf{Idea: } \text{compute } \hat{R} \text{ on } \mathbb{I}\left\{\theta_1^{(n,m)} \leq x_1, \ldots, \theta_d^{(n,m)} \leq x_d\right\}.}$

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

$\hookrightarrow$ 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 $\hat{R}_\infty$ to check convergence of the dependence structure.

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 W_d and M_d Fréchet–Hoeffding copulas in dimension d:

$$W_d(\boldsymbol{u}) \leq C_i(\boldsymbol{u}) \leq M_d(\boldsymbol{u}), \forall \boldsymbol{u} \in [0,1]^d.$$

Then

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

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. On the use of a local $\hat{R}$ to improve MCMC convergence diagnostic. *Preprint, https://arxiv.org/abs/2205.06694*, 2022.