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.
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 \( (\hat{\theta}^{(m,n)}, s^m_n) \) is the mean and variance of \( \theta^{(m,n)} \), and \( \hat{\theta} \) the mean of \( \theta^{(m)} \),

\[
\hat{R} = \left( \frac{W + B}{W} \right) \quad \text{with} \quad B = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\theta}^{(m)} - \hat{\theta})^2, \quad W = \frac{1}{M} \sum_{n=1}^{N} s^m_n.
\]

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

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

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

\[ \text{Rank-} \hat{R} = \text{max(Bulk-} \hat{R}, \text{Tail-} \hat{R}) \]

A local version of \( \hat{R} \)

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

\[ \Rightarrow \text{Define a local convergence estimate } \hat{R}(x). \]

Benefits:
- Detects (non-)convergence at different quantiles.
- Bernoulli variables \( \Rightarrow \) all moments exist (no need for ranks).
- Detects many false negatives.

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

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

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

Uniform and Normal densities

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

\[ \text{Population version} \]

Assume chain \( m \) has stationary distribution \( F_{\theta^m} \).

As \( \mathbb{E}[\mathbb{1}\{\theta^{(m,n)} \leq x\}] = F_{\theta^m}(x) \), and \( \text{Var}[\mathbb{1}\{\theta^{(m,n)} \leq x\}] = F_{\theta^m}(x) - F_{\theta^m}^2(x) \), theoretical within-variance \( W(x) \) and between-variance \( B(x) \) are

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

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

\[
\hat{R}(x) = \left( \frac{W(x) + B(x)}{W(x)} \right) \quad \text{with} \quad B(x) = \left( \frac{1}{M} \sum_{m=1}^{M} F_{\theta^m}(x) - \left( \frac{1}{M} \sum_{m=1}^{M} F_{\theta^m}(x) \right)^2 \right)^2.
\]

Properties:
- \( \hat{R} \geq 1 \), and \( \hat{R} \equiv 1 \) ⇐ all \( F_{\theta^m} \) are equal.
- \( \hat{R}_\infty \) invariant with respect to reparameterisation.

References

