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

\[
\hat{R} = \sqrt{\frac{W + B}{W}}, \quad \text{with} \quad B = \frac{1}{M-1} \sum_{m=1}^{M} (\theta^- - \theta^{(m)}), \quad W = \frac{1}{M-1} \sum_{m=1}^{M} s_m^2.
\]

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

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

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

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

Idea: compute \( \hat{R} \) on indicator variables \( I[\theta^{(m)} \leq x] \) for a given quantile \( x \).

\( \Rightarrow \) 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 \( z^{(n,m)} \) share the same mean, to fool Bulk-\( \hat{R} \) and Tail-\( \hat{R} \) at the same time \( \Rightarrow \) Rank-\( \hat{R} < 1.01 \), and \( \hat{R}_\infty > 1.01 \).

Uniform and Normal densities

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

Population version

Assume chain \( m \) has stationary distribution \( F_m \).

As \( \mathbb{E}[I[\theta^{(m)} \leq x]] = F_m(x) \), and \( \text{Var}[I[\theta^{(m)} \leq x]] = F_m(x) - F_m^2(x) \), theoretical

within-variance \( W(x) \) and between-variance \( B(x) \) are

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

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

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

Properties:

- \( \hat{R} \geq 1 \), and \( \hat{R} \equiv 1 \) \( \iff \) all \( F_m \) are equal.
- \( \hat{R}_\infty \) invariant with respect to reparametrisation.

Multivariate extension

Idea: compute \( \hat{R} \) on \( I[\theta_1^{(m)} \leq x_1, \ldots, \theta_d^{(m)} \leq x_d] \)

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

\( \Rightarrow \) 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.

- With \( M = 2 \), copulas \( C_1 \) and \( C_2 \) and \( u \in [0,1]^d \),

\[
\hat{R}(u) = \sqrt{1 + \frac{1}{2C_1(u)C_2(u) - C_1(u) - C_2(u)}}
\]

Upper bound in the multivariate case

Denote by \( R_{\infty}(C_1, C_2) \), the value of \( R_{\infty} \) for \( C_1 \) and \( C_2 \) \( (M = 2 \) ). Let \( (C_{-i}, C_i) \) two bounding copulas:

- \( C_{-i}(u) \leq C_1(u) \leq C_{-i}(u) \) \( \forall u \in [0,1]^d \).

Then

\[
R_{\infty}(C_1, C_2) \leq R_{\infty}(C_{-i}, C_{-i}).
\]

Corollary: if \( W_{\text{d}} \) and \( M_{\text{d}} \) are the lower and upper Fréchet–Hoëlding copulas in dimension \( d \),

\[
R_{\infty}(C_1, C_2) \leq R_{\infty}(W_{\text{d}}, M_{\text{d}}) = \sqrt{\frac{d}{d-1}}.
\]

References