Gelman-Rubin statistic $\hat{R}$

$$B = \frac{S}{M-1} \sum_{m=1}^{M} \left(\overline{\theta}_m - \overline{\theta}\right)^2, \quad \text{where } \overline{\theta}_m = \frac{1}{S} \sum_{s=1}^{S} \theta_m^{(s)},$$

$$\overline{\theta} = \frac{1}{M} \sum_{m=1}^{M} \overline{\theta}_m$$

$$W = \frac{1}{M} \sum_{m=1}^{M} s_m^2, \quad \text{where } s_m^2 = \frac{1}{S-1} \sum_{s=1}^{S} \left(\theta_m^{(s)} - \overline{\theta}_m\right)^2$$

$$= \frac{1}{M(S-1)} \sum_{m=1}^{M} \sum_{s=1}^{S} \left(\theta_m^{(s)} - \overline{\theta}_m\right)^2$$

The between-sequence variance B contains a factor of S because it is based on the variance of the within-sequence means, $\overline{\theta}_m$, each of which is an average of S values $\theta_m^{(s)}$.

We can estimate $Var(\theta|y)$, the marginal posterior variance of the estimand, by a weighted average of W and B, namely the unbiased estimator:

$$\widehat{Var}(\theta|y) = \frac{S-1}{S} W + \frac{1}{S} B$$

If we divide the equation above by W, we get:

$$\frac{\widehat{Var}(\theta|y)}{W} = \frac{S-1}{S} + \frac{1}{SW} B$$

$$\Rightarrow \hat{R} \equiv \sqrt{\frac{\widehat{Var}(\theta|y)}{W}} = \sqrt{\underbrace{\frac{S-1}{S}}_{1} + \underbrace{\frac{1}{SW}}_{0} B} \xrightarrow[S \to \infty]{} 1$$

Autocorrelation function

Given measurements $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(S)}$ at times $1, 2, \ldots, S$, the lag t autocorrelation function is defined as:

$$\rho_t = \frac{\sum_{i=1}^{S-t} (\theta^{(i)} - \bar{\theta})(\theta^{(i+t)} - \bar{\theta})}{\sum_{i=1}^{S} (\theta^{(i)} - \bar{\theta})^2}$$

It is used to:

1) detect non-randomness in the data
2) identify an appropriate time-series model if the data are not random.

Effective sample size

Effective number of independent simulation draws from the chains of any estimand of interest θ.
In general, in the Markov chains, the simulations of θ within each sequence will be autocorrelated.
Consider to check the statistical efficiency of the average of the simulation θ as an estimate of the posterior mean, $E(\theta|y)$.
Then:

$$\lim_{S \to +\infty} SM \, Var(\bar{\theta}) = \left(1 + 2\sum_{t=1}^{\infty} \rho_t\right) Var(\theta|y),$$

where ρ_t is the autocorrelation of the sequence $\theta_m^{(s)}$ at lag t.
If the S simulations draws from each of the M chains were independent, then:

$$Var(\bar{\theta}) = \frac{1}{SM} Var(\theta|y),$$

and the sample size would be SM (see above).
In presence of autocorrelation, we define the effective sample size as:

$$n_{eff} = \frac{SM}{1 + 2\sum_{t=1}^{\infty} \rho_t} \longrightarrow SM \text{ as } \sum_{t=1}^{\infty} \rho_t \to 0$$