Missing the Point: Non-Convergence in Iterative Imputation Algorithms

Hanne I. Oberman [1] Stef van Buuren [1 2] Gerko Vink [1]

Abstract

Iterative imputation is a popular tool to accommodate missing data. While it is widely accepted that valid inferences can be obtained with this technique, these inferences all rely on algorithmic convergence. There is no consensus on how to evaluate the convergence properties of the method. This paper provides insight into identifying non-convergence in iterative imputation algorithms. Our study found that—in the cases considered—inferential validity was achieved after five to ten iterations, much earlier than indicated by diagnostic methods. We conclude that it never hurts to iterate longer, but such calculations hardly bring added value.

1. Iterative Imputation

Most imputation software packages draw inference from iterative imputation procedures. With iterative imputation, the validity of the inference depends on the state-space of the algorithm at the final iteration. This introduces a potential threat to the validity of the imputations: What if the algorithm has not converged? Are the imputations then to be trusted? And can we rely on the inference obtained on the completed data?

These remain open questions since the convergence properties of iterative imputation algorithms have not been systematically studied (Van Buuren, 2018, § 6.5.2). There is no scientific consensus on how to evaluate the convergence of imputation algorithms (Raghunathan & Bondarenko, 2007; Zhu & Raghunathan, 2015; Takahashi, 2017). The current recommended practice is to visually inspect imputations for signs of non-convergence. This approach may be undesirable for several reasons: 1) it may be challenging to the untrained eye, 2) only severely pathological cases of non-convergence may be diagnosed, and 3) there is not an objective measure that quantifies convergence (Van Buuren, 2018, § 6.5.2). Therefore, a quantitative, diagnostic method to identify non-convergence would be preferred.

2. Identifying Non-Convergence

In our study, we consider two identifiers for non-convergence in iterative algorithms: autocorrelation and potential scale reduction factor $\widehat{R}$ (as recommended by e.g. Cowles & Carlin, 1996). We follow (Lynch, 2007, p. 147) to calculate autocorrelation, and use the recently proposed adapted version of $\widehat{R}$ by (Vehtari et al., 2019, p. 5). Aside from the usual parameters to monitor through visual inspection (i.e., chain means and chain variances), we also look at multivariate parameters to summarize the state-space of the algorithm. We investigate convergence of the parameter of scientific interest, and propose a novel parameter: the first eigenvalue of the variance-covariance matrix after imputation. This novel parameter has the appealing quality that it is not dependent on the model of scientific interest. With that, it suits one of the main advantages of imputation techniques—solving the missing data problem and the substantive scientific problem separately.

We evaluate the performance and plausibility of these methods to identify non-convergence through model-based simulation in `R` (R Core Team, 2020). For reasons of brevity, we only focus on the iterative imputation algorithm implemented in the popular `mice` package in `R` (Van Buuren & Groothuis-Oudshoorn, 2011).

3. Simulation Study

We induce non-convergence in the imputation algorithm using two sets of simulation conditions: early stopping and missingness severity. Early stopping means that we terminate the imputation algorithm

*Equal contribution [1]Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands [2]Netherlands Organisation for Applied Scientific Research TNO, Leiden, The Netherlands. Correspondence to: Hanne Oberman <h.i.oberman@uu.nl>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 108, 2020.

after a different number of iterations in each condition ($T = 1, 2, ..., 100$). The severity of the missingness is determined by the proportion of missing cases ($p_{\text{inc}} = .05, .25, .50, .75, .95$)[1]. For each simulation condition, we evaluate the validity of several statistical inferences, where inferential validity is defined as unbiased estimates and nominal coverage rates across simulation repetitions ($n_{\text{sim}} = 1000$).

The simulation set-up is summarized in pseudo-code (see Algorithm 1). The complete script and technical details are available from github.com/hanneoberman/MissingThePoint.

Algorithm 1 Simulation set-up

```
Simulate data
repeat
  for all missingness conditions do
    Create missingness
    for all early stopping conditions do
      Impute missingness
      Perform analysis of scientific interest
      Compute non-convergence diagnostics
      Pool results across imputations
      Compute performance measures
    end for
  end for
  Combine outcomes of all conditions
until all simulation repetitions are completed
Aggregate outcomes across simulation runs
```

4. Results

Our results demonstrate that inferential validity was achieved after five to ten iterations—much earlier than indicated by the diagnostic methods. For example,

5. Discussion

We have shown that iterative imputation algorithms can yield correct outcomes, even when a converged state has not yet formally been reached. Any further iterations would then burn computational resources without improving the statistical inferences. Our study found that—in the cases considered—inferential validity was achieved after five to ten iterations, much earlier than indicated by the $\widehat{R}$ and AC diagnostics. Of course, it never hurts to iterate longer, but such calculations hardly bring added value.

[1]We only consider a 'missing completely at random' missingness mechanism (Rubin, 1976).

References

Cowles, M. K. and Carlin, B. P. Markov chain Monte Carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association, 91(434):883–904, 1996.

Lynch, S. M. Introduction to applied Bayesian statistics and estimation for social scientists. Springer Science & Business Media, 2007.

R Core Team. R: A language and environment for statistical computing. Vienna, Austria, 2020.

Raghunathan, T. and Bondarenko, I. Diagnostics for Multiple Imputations. SSRN Scholarly Paper ID 1031750, Social Science Research Network, Rochester, NY, November 2007. URL https://papers.ssrn.com/abstract=1031750.

Rubin, D. B. Inference and Missing Data. Biometrika, 63(3):581–592, 1976. doi: 10.2307/2335739.

Takahashi, M. Statistical Inference in Missing Data by MCMC and Non-MCMC Multiple Imputation Algorithms: Assessing the Effects of Between-Imputation Iterations. Data Science Journal, 16:37, July 2017. doi: 10.5334/dsj-2017-037.

Van Buuren, S. Flexible imputation of missing data. Chapman and Hall/CRC, 2018.

Van Buuren, S. and Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(1):1–67, December 2011. doi: 10.18637/jss.v045.i03.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. March 2019. URL http://arxiv.org/abs/1903.08008.

Zhu, J. and Raghunathan, T. E. Convergence Properties of a Sequential Regression Multiple Imputation Algorithm. Journal of the American Statistical Association, 110(511):1112–1124, July 2015. doi: 10.1080/01621459.2014.948117.

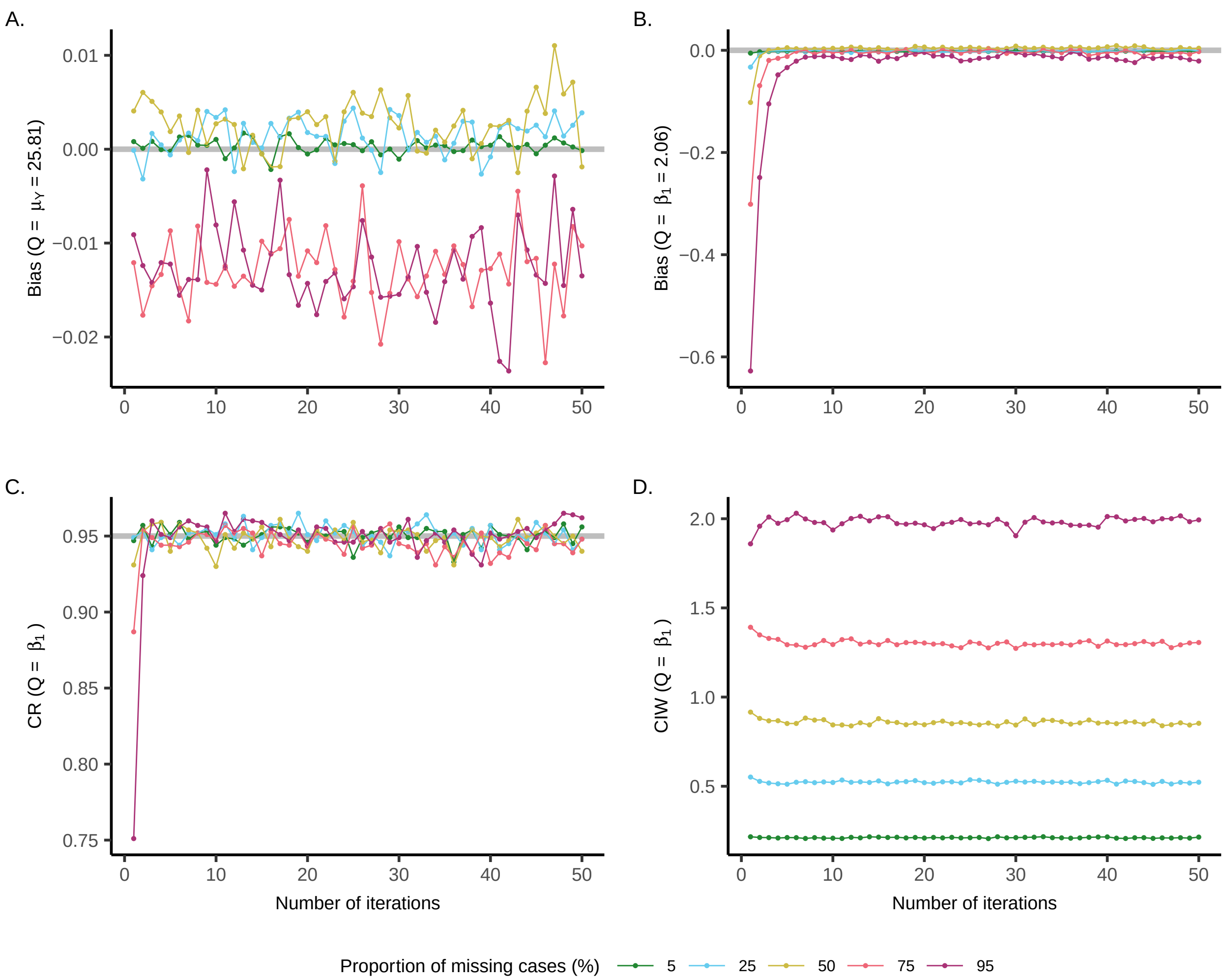

Figure 1. Impact of non-convergence on statistical inferences. Depicted are the bias, coverage rate (CR) and confidence interval width (CIW) of the worst-performing quantities of scientific interest Q in terms of bias. The gray lines represent the objectives: unbiased estimates with nominal coverage.

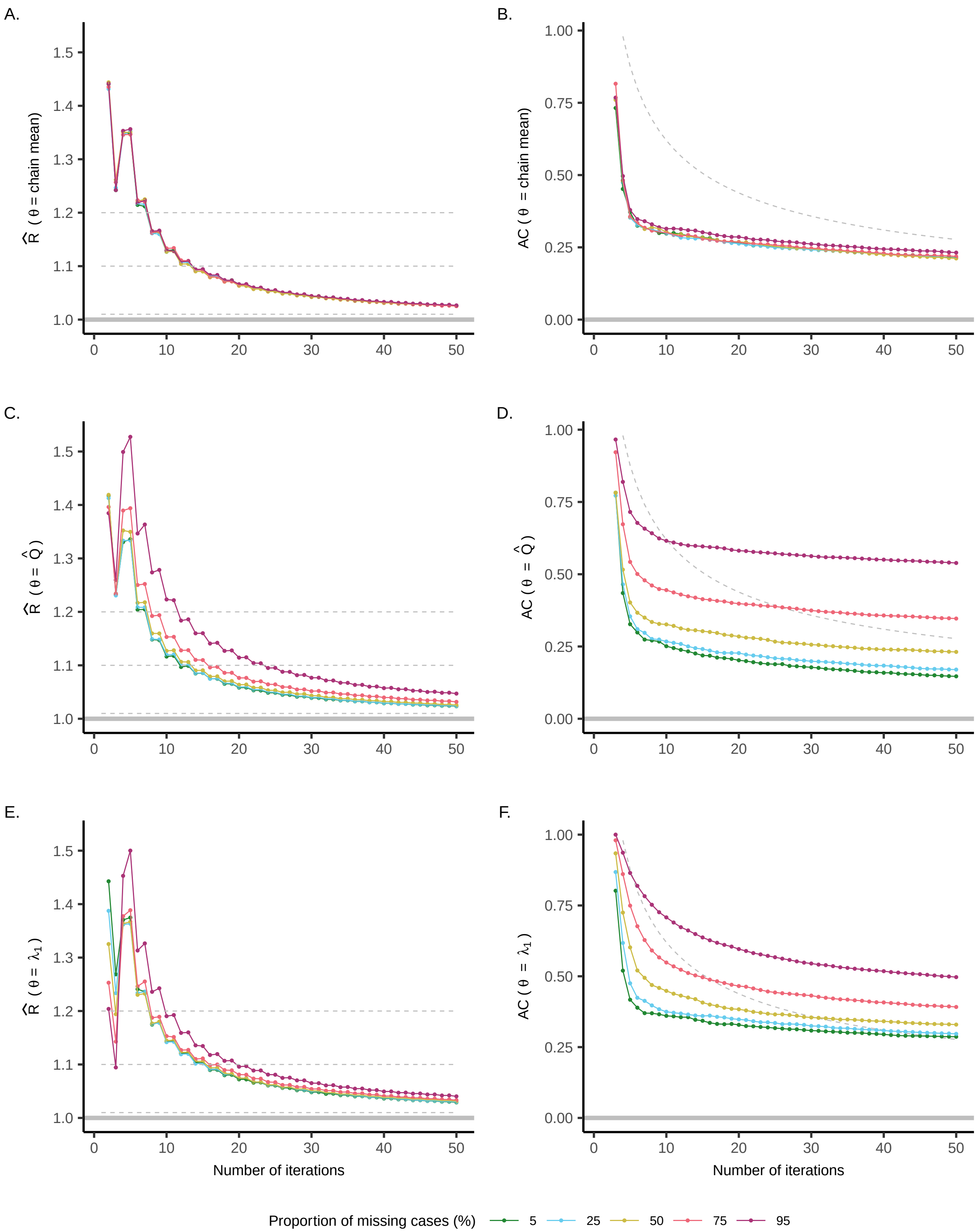

Figure 2. Non-convergence identified by diagnostic methods: $\widehat{R}$ and AC applied on several θs. The left-hand side of the figure contains $\widehat{R}$-values, and the right-hand side contains AC-values. Depicted in the rows are the scalar summaries θ: chain mean, chain variance, the quantity of scientific interest $\hat{Q}$, and the first eigenvalue of the variance-covariance matrix λ_1.