Early Comparisons

David Gerard

2016-03-30

Abstract

I perform early comparisons of cp-rank 1 FLASH against cp-rank 1 least squares and a cp-rank 1 Bayesian approach from Hoff (2011). T-FLASH performs the best.

Data Generation and Simulation Study

$$\mathcal{Y}_{p_1 \times p_2 \times p_3} = \mathcal{X}_{p_1 \times p_2 \times p_3} + \mathcal{E}_{p_1 \times p_2 \times p_3} \quad (1)$$

$$\mathcal{X} = \boldsymbol{u}_1 \circ \boldsymbol{u}_2 \circ \boldsymbol{u}_3 \quad (2)$$

$$p(u_{ij}) = \pi_1 N(u_{ij}|0, \tau_1^2) + \pi_2 N(u_{ij}|0, \tau_2^2) + \pi_3 N(u_{ij}|0, \tau_3^2) \quad (3)$$

$$e_{ijk} \sim N(0, \sigma^2). \quad (4)$$

In this simulation study,

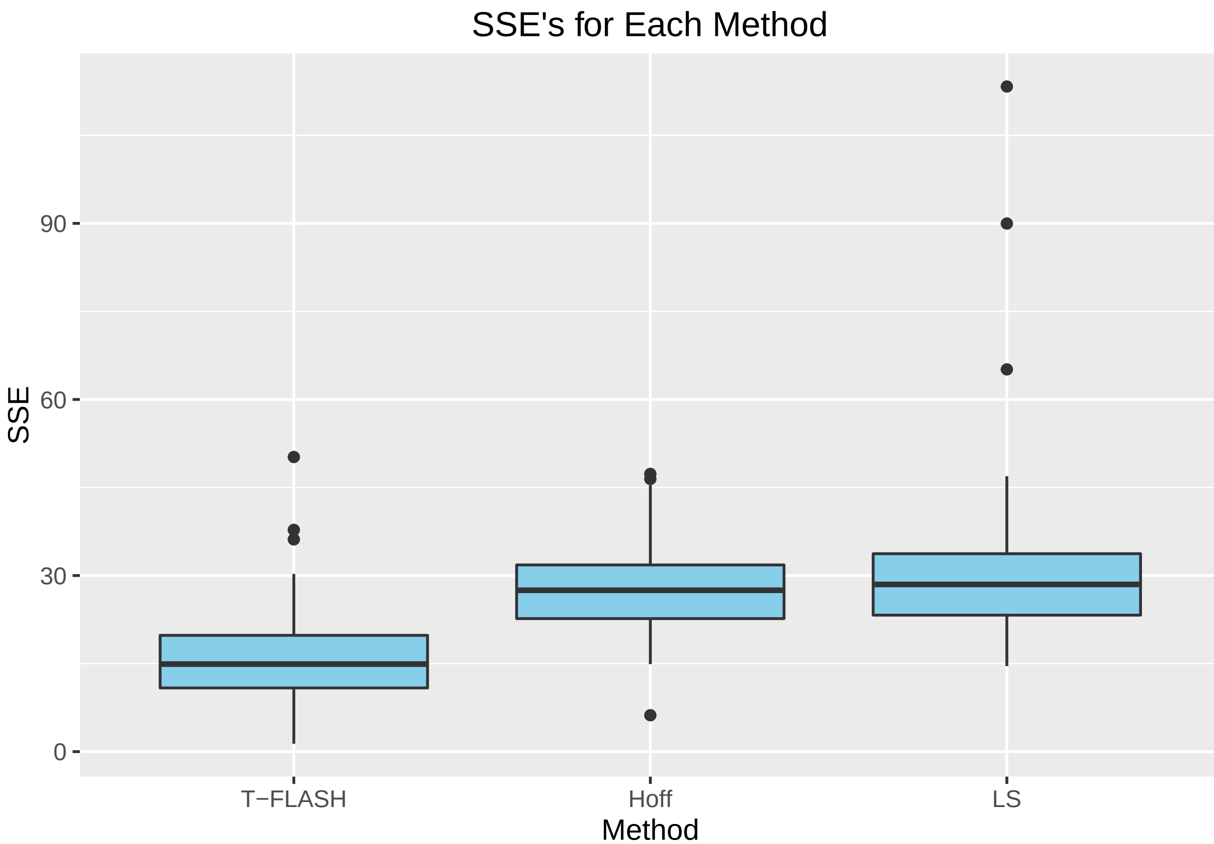
- $p_1 = p_2 = p_3 = 10$,
- $\sigma = 1$,
- $\tau_1 = 0$, $\tau_2 = 1$, $\tau_3 = 5$,
- $\pi_1 = 0.5$, $\pi_2 = 0.3$, $\pi_3 = 0.2$.

I ran through 100 iterations of data generation and fitting three competing methods, generating a new mean $\mathcal{X}$ at each iteration. The three methods were

- T-FLASH,
- The least squares cp-rank 1 tensor, calculated using the `rTensor` package, and
- A hierarchical Bayesian cp-rank 1 tensor mean model as implemented in Hoff (2011).

At each iteration, I calculated the sum of squared errors for each method. A boxplot of these are below. T-FLASH performed the best.

For T-FLASH, I also calculated two-way tables at each iteration for whether or not a component was zero vs the indicator that the posterior probability that the component was zero was greater than 1/2. The mean of these two-way tables over the 100 iterations is in Table 1. Most of the time, T-FLASH works really well. But every once in awhile, it will estimate all of the components to be zero. This is the reason why the bottom left number in Table 1 is non-trivially large.



% latex table generated in R 3.2.4 by xtable 1.8-2 package % Wed Mar 30 16:59:05 2016

	Not Zero	Zero
p < 0.5	0.33	0.01
p > 0.5	0.13	0.53

Table 1: Two way table for posterior-probability of being zero greater than 0.5 vs the truth being zero.

References

Hoff, Peter D. 2011. "Hierarchical Multilinear Models for Multiway Data." *Computational Statistics & Data Analysis* 55 (1). Elsevier: 530–43.