Mash vs. Flash

Jason Willwerscheid

4/20/2018

In each case below, I follow the vignettes to produce a MASH fit (I use both canonical and data-driven covariance matrices). I fit a FLASH object (fixing the standard errors) by adding up to 10 factors greedily, then adding p fixed one-hot vectors, and finally backfitting.

The two fits perform similarly. The MASH fit does better on data generated from the MASH model; more surprisingly, it performs comparably to FLASH on data generated from both the standard two-factor FLASH model and the "augmented FLASH model" (described below).

Flash Model

First I simulate from the basic FLASH model $Y = LF + E$ with $E_{ij} \sim N(0,1)$. Here, $Y \in \mathbb{R}^{1000 \times 10}$, $L \in \mathbb{R}^{1000 \times 2}$ has i.i.d. $N(0,1)$ entries, and F is as follows:

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]   10   10   10   10   10   10   10   10   10    10
## [2,]   10    8    6    4    2    0    0    0    0     0
```

The MSE of the FLASH fit is 0.2, vs. 0.21 for the MASH fit. The proportion of 95% confidence intervals that contain the true value LF_{ij} is 0.94 for FLASH and 0.96 for MASH. The true false sign rate vs lfsr appears as follows:
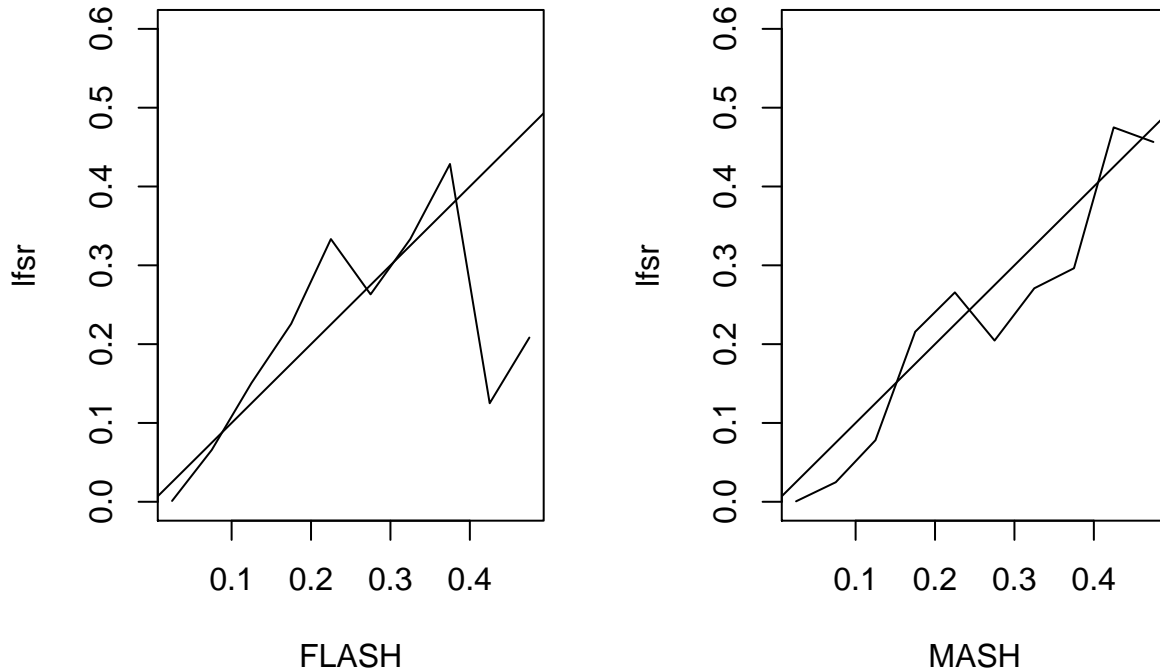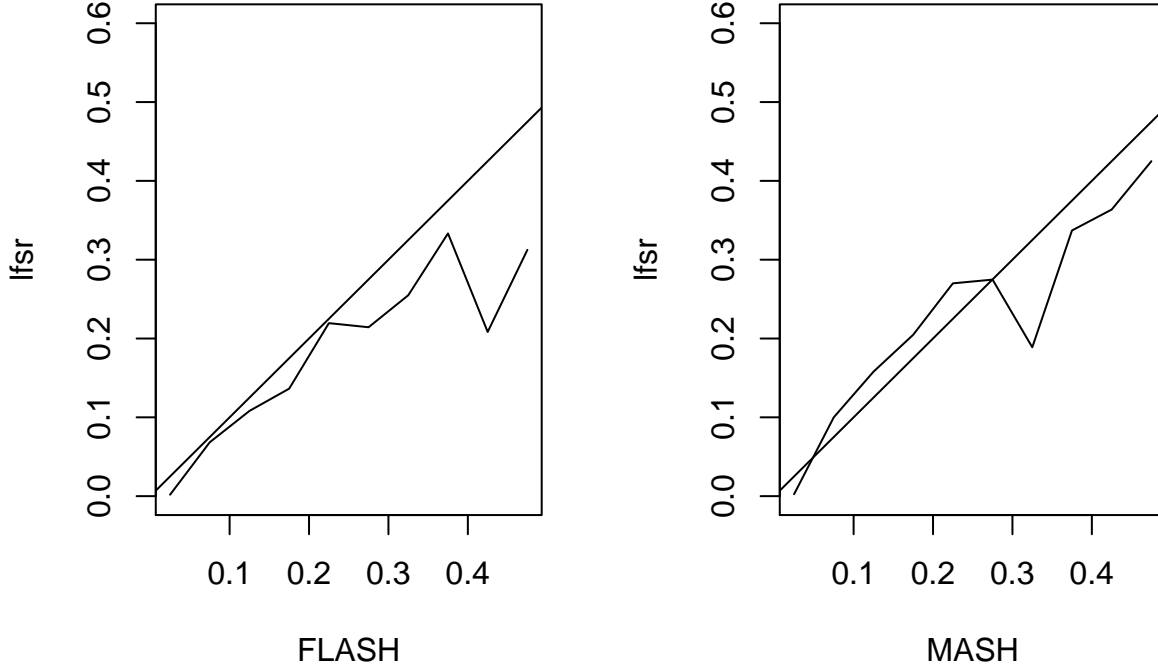
Augmented Flash Model

Next I simulate from the "augmented" FLASH model

$$Y = L \begin{pmatrix} F \\ 3I_{10} \end{pmatrix} + E$$

with F as above.

The MSE of the FLASH fit is 0.93, vs. 1.05 for the MASH fit. The proportion of 95% confidence intervals that contain the true value is 0.94 for FLASH and 0.93 for MASH. The true false sign rate vs lfsr appears as follows:

MASH Model

Finally I simulate from the MASH model

$$X \sim \sum \pi_i N(0, \Sigma_i), \; Y = X + E$$

with $E_{ij} \sim N(0, 1)$. I set Σ_1 to be the all ones matrix, Σ_2 to be a banded covariance matrix with non-zero entries on the first three off-diagonals, and Σ_3 through Σ_{12} to have a single non-zero entry (corresponding to tissue-specific effects). π is set to $(0.3, 0.3, 0.04, 0.04, \dots, 0.04)$.

The MSE of the FLASH fit is 0.56, vs. 0.43 for the MASH fit. The proportion of 95% confidence intervals that contain the true value is 0.9 for FLASH and 0.94 for MASH. The true false sign rate vs lfsr appears as follows: