Empirical Bayes Matrix Factorization

Matthew Stephens, Wei Wang

March 9, 2018

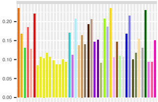
Buzz words

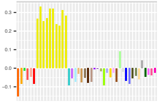
- Principal Components Analysis
- Clustering
- Factor Analysis
- Dimension Reduction
- Latent Dirichlet Allocation / Admixture / Mixed Membership
- Non-negative Matrix Factorization

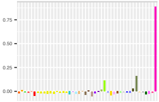
Matrix Factorization

$$Y = \sum_{k=1}^{K} l_k f_k^T + E$$

Example: Genotype effects across tissues

$Y_{i\cdot} = l_{i1}$ 

$+ \, l_{i2}$ 

$+ \, l_{i3}$ 

$+ \, E_{i\cdot}$

One model, many methods (many applications)

$$Y = \sum_{k=1}^{K} l_k f_k^T + E$$

PCA: $l_1, \dots, l_K$ are orthogonal; as are $f_1, \dots, f_K$.
NMF: elements of l_k and f_k are non-negative.
LDA, topic models, admixture: $\sum_k l_{ki} = 1$
Sparse FA: l_k and/or f_k are sparse ...but how sparse?

One model, many methods (many applications)

$$Y = \sum_{k=1}^{K} l_k f_k^T + E$$

PCA: $l_1, \ldots, l_K$ are orthogonal; as are $f_1, \ldots, f_K$.
NMF: elements of l_k and f_k are non-negative.
LDA, topic models, admixture: $\sum_k l_{ki} = 1$
Sparse FA: l_k and/or f_k are sparse ...but how sparse?

One model, many methods (many applications)

$$Y = \sum_{k=1}^{K} l_k f_k^T + E$$

PCA: $l_1, \ldots, l_K$ are orthogonal; as are $f_1, \ldots, f_K$.
NMF: elements of l_k and f_k are non-negative.
LDA, topic models, admixture: $\sum_k l_{ki} = 1$
Sparse FA: l_k and/or f_k are sparse ...but how sparse?

One model, many methods (many applications)

$$Y = \sum_{k=1}^{K} l_k f_k^T + E$$

PCA: $l_1, \dots, l_K$ are orthogonal; as are $f_1, \dots, f_K$.
NMF: elements of l_k and f_k are non-negative.
LDA, topic models, admixture: $\sum_k l_{ki} = 1$
Sparse FA: l_k and/or f_k are sparse ...but how sparse?

One model, many methods (many applications)

$$Y = \sum_{k=1}^{K} l_k f_k^T + E$$

PCA: $l_1, \dots, l_K$ are orthogonal; as are $f_1, \dots, f_K$.
NMF: elements of l_k and f_k are non-negative.
LDA, topic models, admixture: $\sum_k l_{ki} = 1$
Sparse FA: l_k and/or f_k are sparse ...but how sparse?

A simpler problem: the normal means problem

Observe $X_1, \ldots, X_n$ with

$$X_j \sim N(\theta_j, s_j^2).$$

Challenge: Estimate $\theta_1, \ldots, \theta_n$.

Key idea: Shrinkage.

A simpler problem: the normal means problem

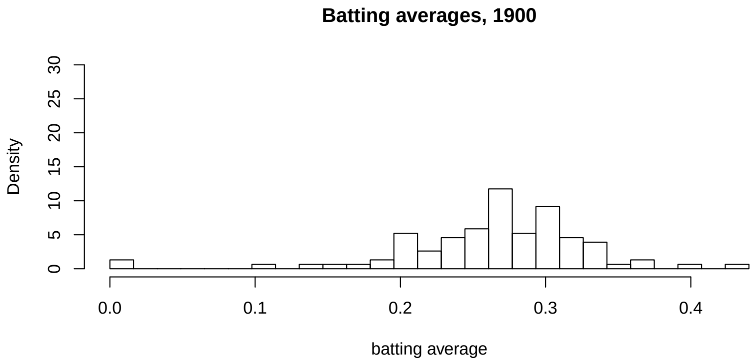
Observe $X_1, \ldots, X_n$ with

$$X_j \sim N(\theta_j, s_j^2).$$

Challenge: Estimate $\theta_1, \ldots, \theta_n$.

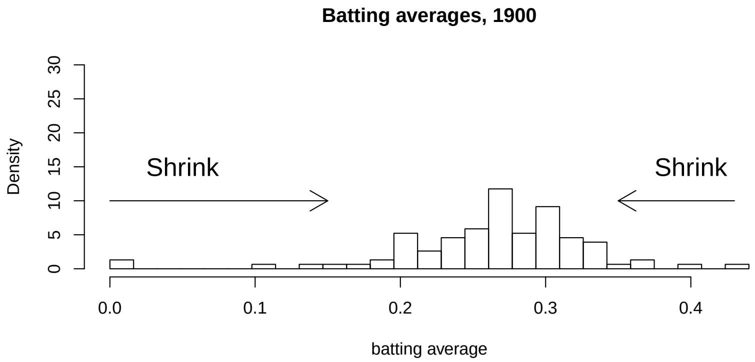
Key idea: Shrinkage.

Shrinkage Estimation[1]



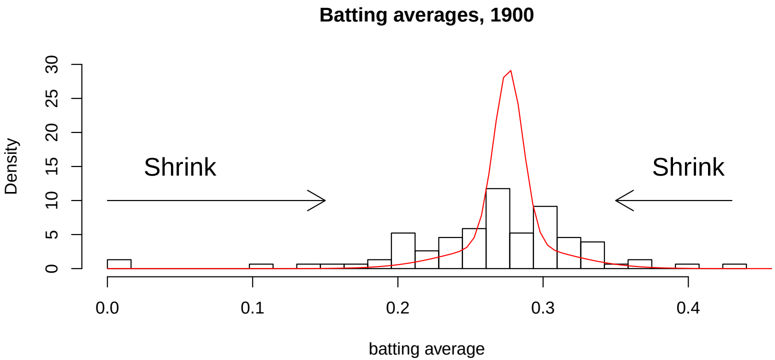
[1] http://varianceexplained.org/r/empirical_bayes_baseball/

Shrinkage Estimation[1]



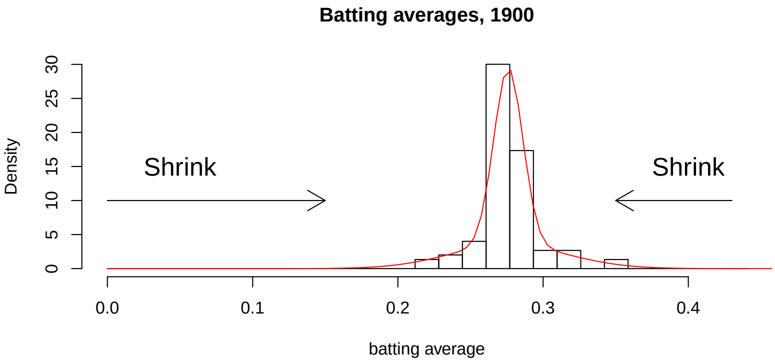
[1] http://varianceexplained.org/r/empirical_bayes_baseball/

Shrinkage Estimation[1]



[1]http://varianceexplained.org/r/empirical_bayes_baseball/

Shrinkage Estimation[1]



[1] http://varianceexplained.org/r/empirical_bayes_baseball/

Empirical Bayes: learn how much to shrink from the data

$$X_j \sim N(\theta_j, s_j^2)$$
$$\theta_j \sim g(\cdot) \in \mathcal{G}$$

Fit in two steps:

1. Estimate g by maximimum likelihood: $\hat{g} = \arg\max_g L(g)$
2. Estimate θ_j using posterior distribution $\theta_j | X_j, \hat{g}$.

...tractable for very flexible $\mathcal{G}$ (e.g. $\mathcal{G}$ = distributions that are unimodal; S., 2016).

Empirical Bayes: learn how much to shrink from the data

$$X_j \sim N(\theta_j, s_j^2)$$
$$\theta_j \sim g(\cdot) \in \mathcal{G}$$

Fit in two steps:

1. Estimate g by maximimum likelihood: $\hat{g} = \arg\max_g L(g)$
2. Estimate θ_j using posterior distribution $\theta_j | X_j, \hat{g}$.

...tractable for very flexible $\mathcal{G}$ (e.g. $\mathcal{G}$ = distributions that are unimodal; S., 2016).

Empirical Bayes Factor Analysis

$$Y = \sum_{k=1}^{K} l_k f_k^T + E$$

$$l_{k1}, \ldots, l_{kn} \sim g_k^l(\cdot) \in \mathcal{G}$$

$$f_{k1}, \ldots, f_{kp} \sim g_k^f(\cdot) \in \mathcal{G}$$

Harder than normal means problem...

...but approximate iterative solution based on normal means.

Empirical Bayes Factor Analysis

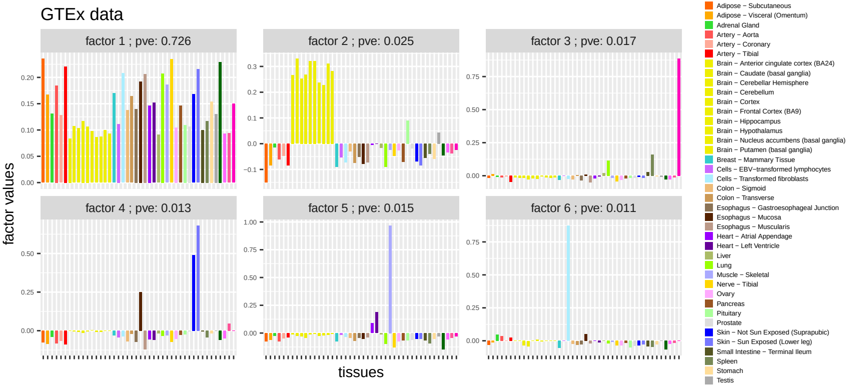
$$Y = \sum_{k=1}^{K} l_k f_k^T + E$$

$$l_{k1}, \ldots, l_{kn} \sim g_k^l(\cdot) \in \mathcal{G}$$

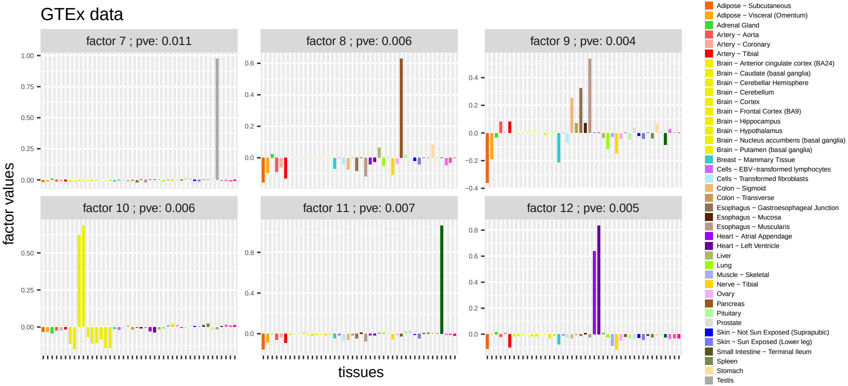
$$f_{k1}, \ldots, f_{kp} \sim g_k^f(\cdot) \in \mathcal{G}$$

Harder than normal means problem...
...but approximate iterative solution based on normal means.

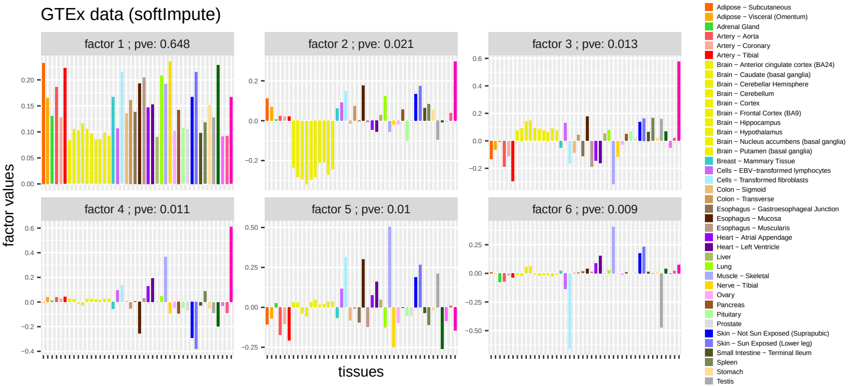
GTEx data: first 6 factors



GTEx data: next 6 factors



Comparison: softImpute (nuclear norm penalty)



What next?

- Faster
- Bigger
- Visualizing Results
- Non-negative Constraint

Acknowledgements

- Wei Wang
- Funding: GTEX, Moore Foundation

`flashr: https://www.github.com/stephenslab/flashr`

`workflowr: https://github.com/jdblischak/workflowr`

Rank 3 simulations