Linked shrinkage to improve estimation of interaction effects in regression models

Mark van de Wiel, Joint work with Matteo Amestoy & Jeroen Hoogland

Epidemiology & Data Science, Amsterdam University Medical Centers, The Netherlands

ISCB, Milan, 2023

Amsterdam UMC
University Medical Centers

EDS EPIDEMIOLOGY AND DATA SCIENCE

Setting

- Low-dimensional epidemiological cohort study
  - Response: y, say cholesterol
  - Covariates: age, bmi, smoking, ethnicity, etc.
  - $n \gg p$, say $n = 1,000, p = 14$

Setting

- Low-dimensional epidemiological cohort study
  - Response: y, say cholesterol
  - Covariates: age, bmi, smoking, ethnicity, etc.
  - $n \gg p$, say $n = 1,000, p = 14$

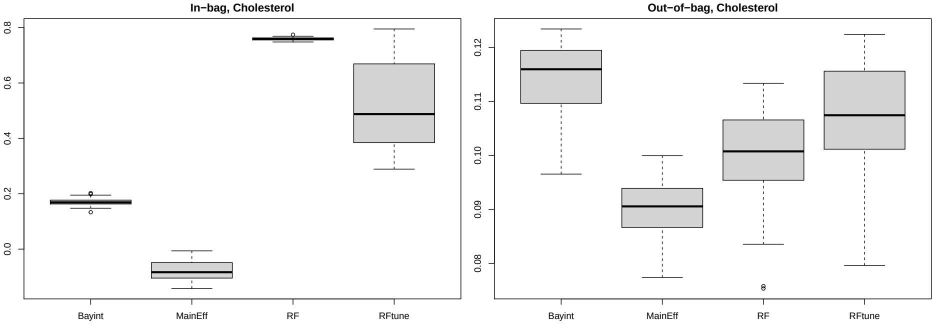
- Aims
  1. Interpretable model that explains y
  2. Variable importance
  3. Competitive prediction

Teaser: prediction

- Prediction evaluated by R^2; Outcome: cholesterol
- Training sets of $n = 1,000$, $p = 14$ (subsets); Complementary test sets
- Regression models: Bayint (interactions), main effects
- Random Forest: default, tuned

Teaser: prediction

- Prediction evaluated by R^2; Outcome: cholesterol
- Training sets of $n = 1,000$, $p = 14$ (subsets); Complementary test sets
- Regression models: Bayint (interactions), main effects
- Random Forest: default, tuned



Model

- Regression model with two-way interactions

$$Y_i = \alpha + \sum_{j=1}^{p} \beta_j x_{ij} + \sum_{j,k:j\neq k} \beta_{jk} x_{ij} x_{ik} + \epsilon_i$$

- Main problem: #parameters increases quadratically with $p \Rightarrow$
- Selection/Shrinkage
  - 2-step: only add interactions of significant main effects [and/or]
  - lassoint: lasso penalty on interactions
  - hierarchical lasso (Bien, Taylor, and Tibshirani 2013)
  - ridge2: differential ridge penalty main effect & interactions
  - Bayloc: Bayesian local shrinkage (Gelman et al. 2008)

Our solution: linked shrinkage (Bayint)

$$Y_i = \alpha + \sum_{j=1}^{p} \beta_j x_{ij} + \sum_{j,k:j \neq k} \beta_{jk} x_{ij} x_{ik} + \epsilon_i, \;\; \epsilon_i \sim N(0, \sigma^2)$$

$$\beta_j \sim N(0, \sigma^2 \tau_j^2), \;\; \beta_{jk} \sim N(0, \sigma^2 \tau_j \tau_k \tau_{int})$$

$$\tau_j \sim C^+(0, 1), \;\; \tau_{int} \sim U(0.01, 1)$$

$$\alpha \sim N(0, 10^2), \sigma^2 \sim IG(1, 0.001)$$

Our solution: linked shrinkage (Bayint)

$$Y_i = \alpha + \sum_{j=1}^{p} \beta_j x_{ij} + \sum_{j,k:j\neq k} \beta_{jk} x_{ij} x_{ik} + \epsilon_i,\ \ \epsilon_i \sim N(0, \sigma^2)$$

$$\beta_j \sim N(0, \sigma^2 \tau_j^2),\ \ \beta_{jk} \sim N(0, \sigma^2 \tau_j \tau_k \tau_{int})$$

$$\tau_j \sim C^+(0, 1),\ \ \tau_{int} \sim U(0.01, 1)$$

$$\alpha \sim N(0, 10^2), \sigma^2 \sim IG(1, 0.001)$$

Coded in R-stan for flexibility

Comparison

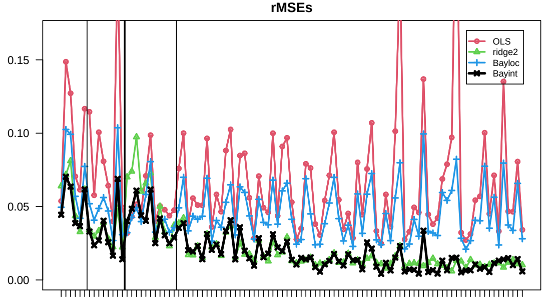
Now: Focus on estimation accuracy

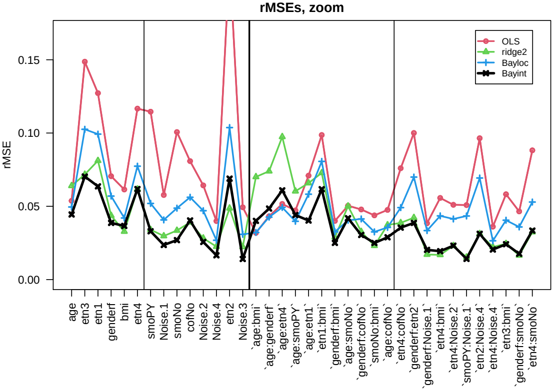
- Benchmark: OLS estimates of β on large data set $(N = 21,570)$
- Data: 25 (near non-overlapping) subsets; $n = 1,000, p = 14, q = 85$.
- Outcome: cholesterol (log)
- Metric: RMSE of $\hat{\beta}_j$ and $\hat{\beta}_{jk}$:

$$\sqrt{\frac{1}{25}\sum_{\ell=1}^{25}(\hat{\beta}_j^{(\ell)} - \beta_j^{\text{bench}})^2}$$

Results 1: comparison Bayint with OLS, ridge2, Bayloc

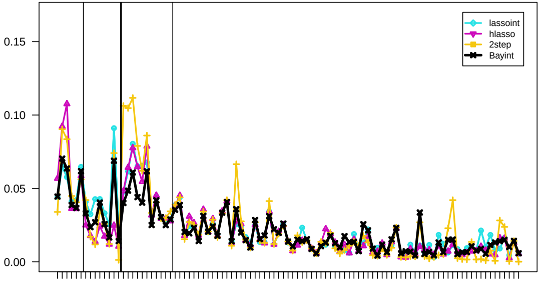
rMSEs for main effects and interactions (after bold vertical line)
Parameters ordered by significance in master set (thin lines demarcates $p < 0.01$)

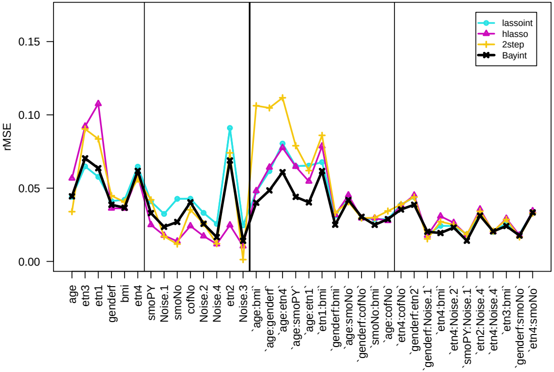




Results 2: comparison Bayint with lassoint, hlasso, 2step

rMSEs for main effects and interactions (after bold vertical line)
Parameters ordered by significance in master set (thin lines demarcates $p < 0.01$)



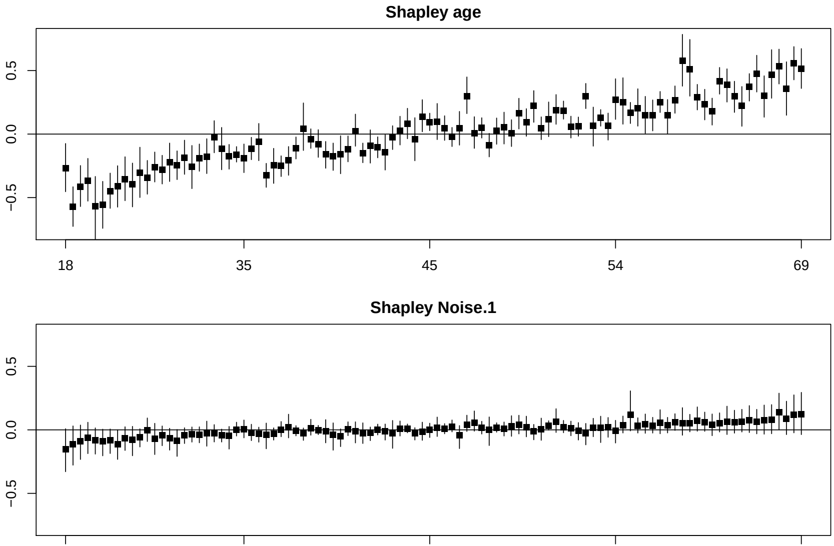


Interpretation: variable importance

- Quantifying variable importance not trivial in regression model with interactions (Afshartous and Preston 2011)
- Shapley values: metric quantifying variable importance per sample (Aas, Jullum, and Løland 2021)
  - Originates from game theory
  - Many desirable properties
  - Applies to complex machine learners
  - Expensive to compute, usually
- Explicit formula for our model: [e.g. $x^*_{ij} = 49$, age (j) for ind i]

$$\phi(x^*_{ij}) = \beta_j x^*_{ij} + \frac{1}{2}\left(\sum_{k:k\neq j} \beta_{jk} x^*_{ij} x^*_{ik} - \sum_{k:k\neq j} \beta_{jk} E[x_{ij} x_{ik}]\right),$$

Shapleys + intervals



That's it!

Conclusion

Bayint + Shapleys: flexible, accurate, good prediction + interpretation

That's it!

Conclusion

Bayint + Shapleys: flexible, accurate, good prediction + interpretation

References

Aas, Kjersti, Martin Jullum, and Anders Løland (2021). "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values". In: *Artificial Intelligence* 298, p. 103502.

Afshartous, D. and R. A Preston (2011). "Key results of interaction models with centering". In: *Journal of Statistics Education* 19.3.

Bien, Jacob, Jonathan Taylor, and Robert Tibshirani (2013). "A lasso for hierarchical interactions". In: *Annals of statistics* 41.3, p. 1111.

Gelman, Andrew et al. (2008). "A weakly informative default prior distribution for logistic and other regression models". In: *The Annals of applied statistics* 2.4, pp. 1360–1383.