AMBARTI models for agricultural experiments

Prado, E. B., Lemos, A. A.
Hamilton Institute & Dept. of Mathematics and Statistics

Maynooth University
National University of Ireland Maynooth

January 11th, 2021 - Group meeting

Agenda

- Additive Main Effect interaction (AMMI) models
- AMMI + BART = AMBARTI
- Some preliminary results

Additive Main effects and Multiplicative Interactions (AMMI)

- Linear–bilinear models are frequently used to analyse two-way data such as genotype-by-environment data;
- A example of this class of models is the AMMI model.

$$y_{ij} = \mu + g_i + e_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}\left(0, \tau^{-1}\right),$$

where

- y - response
- g - genotype
- e - environment
- λ - singular value of the multiplicative component
- γ - genotype singular value
- δ - environment singular value

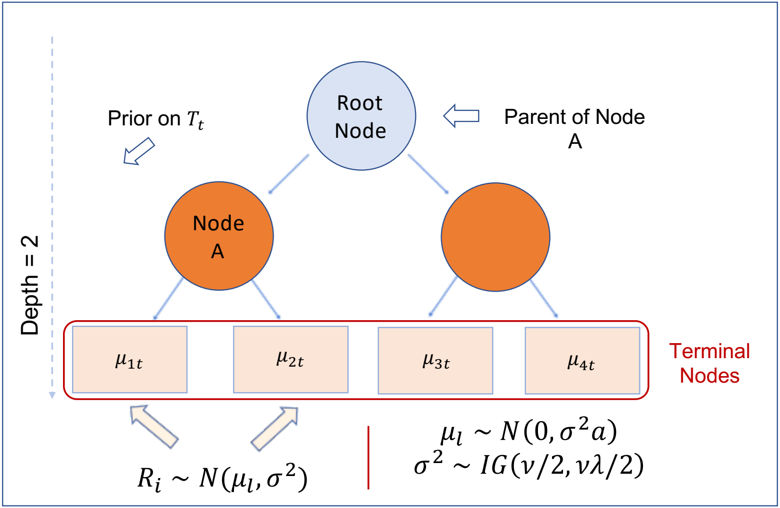
Additive Main effects and Multiplicative Interactions (AMMI)

The following priors are assumed in its Bayesian version (Josse et al, JABES, 2014):

$$
\begin{aligned}
\mu &\sim \mathcal{N}\left(m, s_\mu^2\right) \\
g_i &\sim \mathcal{N}\left(0, s_g^2\right) \\
e_j &\sim \mathcal{N}\left(0, s_e^2\right) \\
\left(\lambda_q\right)_{q=1,\ldots,Q} &\sim \text{ ordered sample of } Q \text{ independent } \mathcal{N}^+\left(0, s_\lambda^2\right) \\
\gamma_{1q} &\sim \mathcal{N}^+(0,1) \quad \text{ for } q = 1, \ldots, Q \\
\gamma_{iq} &\sim \mathcal{N}(0,1) \quad \text{ for } i > 1 \text{ and } q = 1, \ldots, Q \\
\delta_{jq} &\sim \mathcal{N}(0,1) \quad \text{ for } j \geq 1 \text{ and } q = 1, \ldots, Q \\
\sigma_E &\sim \mathcal{U}\left(0, S_{\mathrm{ME}}\right)
\end{aligned}
$$

Additive Main Effect Bayesian Additive Regression Tree interaction models (AMBARTI)

BART



Depth = 2
Prior on T_t
Root Node
Parent of Node A
Node A
μ_{1t}
μ_{2t}
μ_{3t}
μ_{4t}
Terminal Nodes
$R_i \sim N(\mu_l, \sigma^2)$
$\mu_l \sim N(0, \sigma^2 a)$
$\sigma^2 \sim IG(\nu/2, \nu\lambda/2)$

AMMI + BART

BART is a flexible tree-based method that can be used for predicting when there are interactions and non-linear relationships.

$$y_{ij}|\mathbf{x}_{ij}, \mathcal{T}, \mathcal{M}, \Theta, \sigma^2 \sim \mathrm{N}\left(g_i + e_j + \sum_{t=1}^{T} h(\mathbf{x}_{ij}, \mathcal{M}_t, \mathcal{T}_t), \sigma^2\right),$$

where y_{ij} is the yield for genotype i and environment j, and g_i and e_j are the genotype and environment effects, respectively.

$$\begin{aligned}
\mu_{t\ell}|\mathcal{T}_t &\sim \mathrm{N}(\mu_\mu = 0, \sigma^2_\mu), \\
g_i|\mathcal{T}_t &\sim \mathrm{N}(\mu_g, \sigma^2_g), \\
e_j|\mathcal{T}_t &\sim \mathrm{N}(\mu_e, \sigma^2_e), \\
\sigma^2_g &\sim \mathrm{IG}(a_g, b_g), \\
\sigma^2_e &\sim \mathrm{IG}(a_e, b_e), \\
\sigma^2 &\sim \mathrm{IG}(a, b).
\end{aligned}$$

Some preliminary results

Simulated example

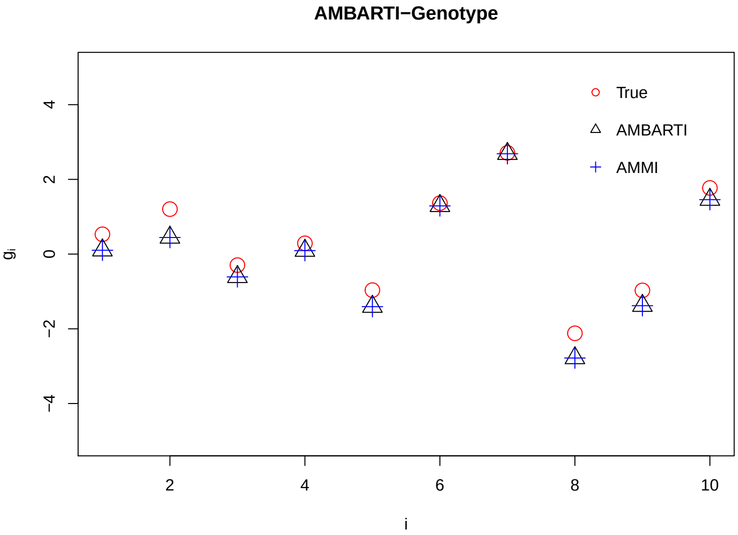
We consider the following setting to generate a simulated data set.

$$y_{ij} = \mu + g_i + e_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}\left(0, \tau^{-1}\right),$$

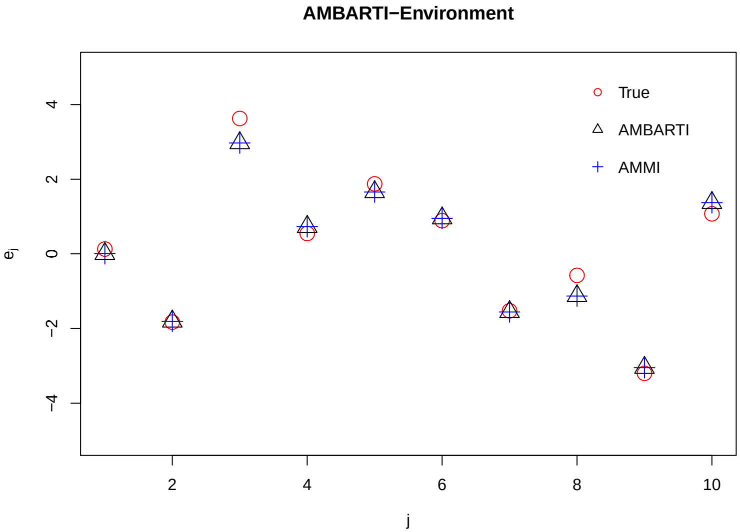
where

- $Q = 1$
- $I = J = 10$
- $\mu = 10$
- $g = \mathrm{N}(0, s_g)$
- $e = \mathrm{N}(0, s_e)$
- $s_g = 2,\ s_e = 2$
- $\lambda = 12$
- $\tau = 1$

Preliminary results



Preliminary results



Some remarks and next steps

- Interestingly, BART is doing a good job...

```
cor(y, y_hat_AMBARTI) = 0.98
cor(y, y_hat_AMMI) = 0.93
```

- More simulations
- Calculate the interactions
- Real datasets

That's all, folks! Thank you!

This work was supported by a Science Foundation Ireland Career Development Award grant number: 17/CDA/4695

References