Bayesian Alphabet

* * *

Evangelina López de Maturana & Oscar González-Recio

Topics

Reasoning

Bayesian methods regularize the prediction models

Examples

- Bayes A
- Bayes B
- Bayes Lasso
- Bayes R

Bayesian LASSO

Details on Bayesian LASSO

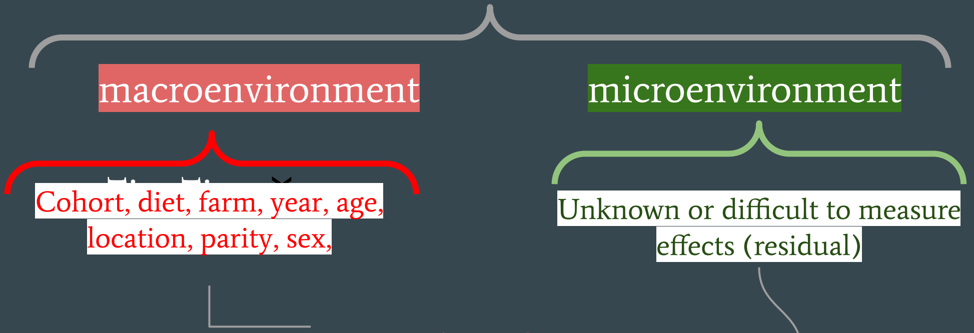
Implementing B Lasso

hands-on

Mathematical representation of biological processes

$$P=G+E$$

$$y_i=g_i+E_i$$



macroenvironment

Cohort, diet, farm, year, age, location, parity, sex,

microenvironment

Unknown or difficult to measure effects (residual)

$$y_i=\text{EnvironmentalEffects} + g_i + e_i$$

$$y_i=X_ib_i + Z_ig_i+e_i$$

Marker regression

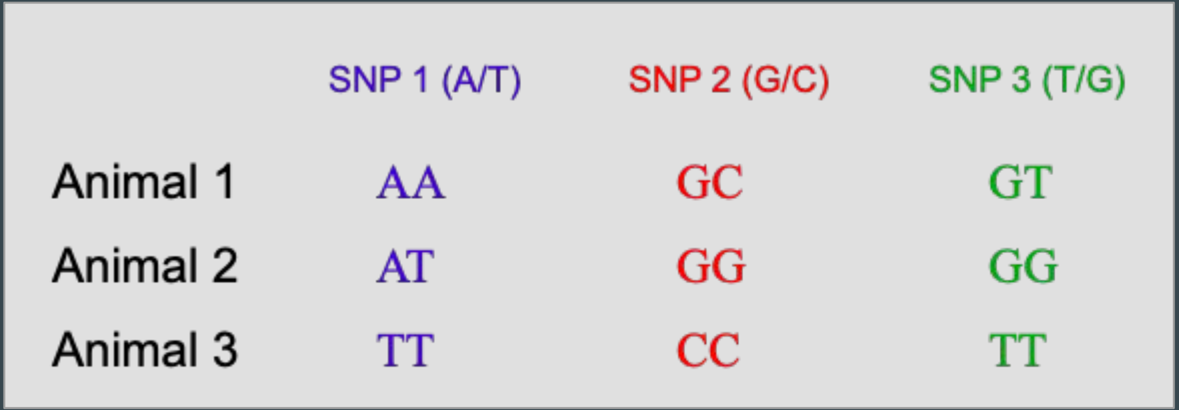
Decomposing the polygenic effect into the sum of SNP (linear) effects

$$y_i = X_i b_i + Z_{ui} u_i + e'_i$$

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{X}\beta_f + snp_1\beta_1 + snp_2\beta_2 + \ldots + snp_p\beta_p + \mathbf{e}$$

$$\mathbf{e} \sim N(0, \sigma_e^2) \qquad \sigma_e^2 \sim \upsilon_e s_e^2 \chi^{-2}_{ss,\upsilon}$$

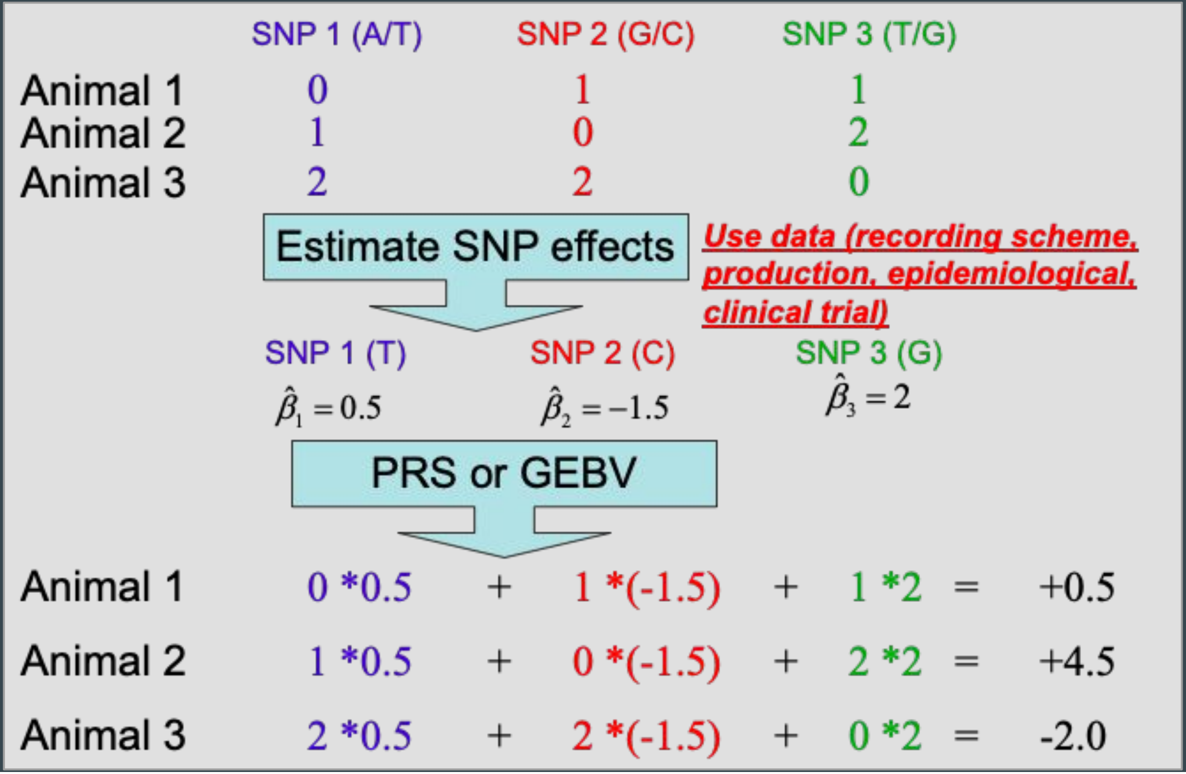
$$\left.\begin{array}{l} \mathbf{b}_i \sim N(0, \sigma_i^2) \\ \sigma_i^2 \sim \chi^{-2}_{(\upsilon,S)} \end{array}\right\} \text{Or alternatives}$$



	SNP 1 (A/T)	SNP 2 (G/C)	SNP 3 (T/G)
Animal 1	AA	GC	GT
Animal 2	AT	GG	GG
Animal 3	TT	CC	TT

	SNP 1 (A/T)	SNP 2 (G/C)	SNP 3 (T/G)
Animal 1	AA	GC	GT
Animal 2	AT	GG	GG
Animal 3	TT	CC	TT

Animal 1	0	1	1
Animal 2	1	0	2
Animal 3	2	2	0



	SNP 1 (A/T)	SNP 2 (G/C)	SNP 3 (T/G)
Animal 1	0	1	1
Animal 2	1	0	2
Animal 3	2	2	0

Estimate SNP effects

Use data (recording scheme, production, epidemiological, clinical trial)

SNP 1 (T)	SNP 2 (C)	SNP 3 (G)
$\hat{\beta}_1 = 0.5$	$\hat{\beta}_2 = -1.5$	$\hat{\beta}_3 = 2$

PRS or GEBV

Animal 1	0 *0.5	+	1 *(-1.5)	+	1 *2	=	+0.5
Animal 2	1 *0.5	+	0 *(-1.5)	+	2 *2	=	+4.5
Animal 3	2 *0.5	+	2 *(-1.5)	+	0 *2	=	-2.0

Marker regression.

Estimation of allele substitution effect.

- Ridge Regression (Whittaker et al., 2000)
- Bayes A
- Bayes B
- Bayes C
- Bayes R
- Bayesian Lasso (Park & Casella, 2007)
    - Statistically more robust.

Marker regression.

Estimation of allele substitution effect.

- Ridge Regression (Whittaker et al., 2000)
- Bayes A
- Bayes B
- Bayes C
- Bayes R
- Bayesian Lasso (Park & Casella, 2007)
    - Statistically more robust.

Very strict priors (4 d.f. !!)

Do not disappear asymptotically (Gianola et al., *p.c.*).

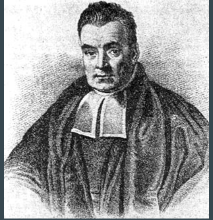
Do not allow Bayesian Learning ⚠

Bayesian brief recap

Assume a model for the data

$$p(\mathbf{y}|\boldsymbol{\theta}) = N(\mathbf{X}\boldsymbol{\beta} + \cdots + \mathbf{Z}\mathbf{u}, \sigma_e^2)$$

Bayes theorem



$$p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y})p(\mathbf{y})$$

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

y=data

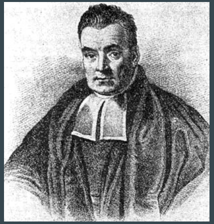
$\boldsymbol{\theta}$= unknown parameters, coefficients, variances, …

Bayesian brief recap

Assume a model for the data

$$p(\mathbf{y}|\boldsymbol{\theta}) = N(\mathbf{X}\boldsymbol{\beta} + \cdots + \mathbf{Z}\mathbf{u}, \sigma_e^2)$$

Bayes theorem



$$p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y})p(\mathbf{y})$$

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

Choose priors

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}, \sigma_e^2)p(\boldsymbol{\beta}|\sigma_b^2)p(\mathbf{u}|\sigma_u^2)p(\sigma_b^2)p(\sigma_u^2)p(\sigma_e^2)$$

Make inferences using McMC algorithms (Gibbs sampling, acceptance rejection, Metropolis-Hasting)

Ridge Regression

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{Wq} + \sum_{j=1}^{q} \mathbf{x}_j \mathbf{b}_j + \mathbf{e}$$

$$GEBV_i = \sum_{j=1}^{q} x_{ij} \hat{b}_j$$

- A priori distribution for SNP effects $b_j \sim N(0, \sigma_b^2)$
  - SNP effects normally distributed

- A priori distribution for SNPs variance $V(GEBV) = \mathbf{XX}'\sigma_b^2$
  - Same variance for all SNP
  - Distributed as inverse chi-squared

$$\sigma_b^2 \sim \chi_{(v,S)}^{-2}$$

Bayes A

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{Wq} + \sum_{j=1}^{q} \mathbf{x}_j \mathbf{b}_j + \mathbf{e}$$

$$GEBV_i = \sum_{j=1}^{q} x_{ij} \hat{b}_j$$

- A priori distribution for SNP effects
  - SNP effects normally distributed

$$b_j \sim N\left(0, \sigma^2_{b_j}\right)$$

- A priori distribution for SNPs variance
  - Different variance for each SNP
  - Same prior variance for all SNP
  - Distributed as inverse chi-squared

$$\sigma^2_{b_j} \sim \chi^{-2}_{(\nu,S)}$$

Bayes B

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{Wq} + \sum_{j=1}^{q} \mathbf{x}_j \mathbf{b}_j + \mathbf{e}$$

$$GEBV_i = \sum_{j=1}^{q} x_{ij} b_j$$

- A priori distribution for SNP effects
  - SNP effects normally distributed

$$b_j \sim N\left(0, \sigma^2_{b_j}\right)$$

- A priori distribution for SNPs variance
  - Different variance for each SNP
  - Same prior variance for all SNP
  - Distributed as a mixture distribution, with inverse chi-squared or zero inflated with p and (1-p) probabilities

$$\sigma_i^2 \sim \begin{cases} = 0 & p = \pi \\ \sim \chi^{-2}_{(\upsilon,S)} & p = (1-\pi) \end{cases}$$

Bayes C

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{Wq} + \sum_{j=1}^{q} \mathbf{x}_j \mathbf{b}_j + \mathbf{e}$$

$$GEBV_i = \sum_{j=1}^{q} x_{ij} b_j$$

- A priori distribution for SNP effects
  - SNP effects distributed as a mixture, with Normal or zero inflated with p and (1-p) probabilities

$$b_j \sim \begin{cases} = 0 & p = \pi \\ \sim N(0, \sigma_j^2) & p = (1 - \pi) \end{cases}$$

- A priori distribution for SNPs variance
  - Different variance for each SNP
  - Same prior variance for all SNP
  - Distributed as inverse chi-squared

$$\sigma_{b_j}^2 \sim \chi_{(\nu, S)}^{-2}$$

Bayes R

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{Wq} + \sum_{j=1}^{q} \mathbf{x}_j \mathbf{b}_j + \mathbf{e}$$

$$GEBV_i = \sum_{j=1}^{q} x_{ij} b_j$$

- A mixture priori distribution for SNP effects
  - SNP effects distributed as a mixture of normal distribution with different variances, based on a dirichlet distribution.

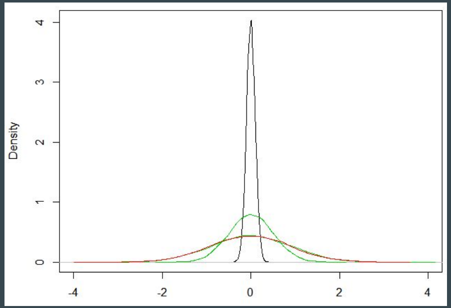
$$N(0, 0\sigma_u^2)$$

$$N(0, 0.0001\sigma_u^2)$$

$$N(0, 0.001\sigma_u^2)$$

$$N(0, 0.01\sigma_u^2)$$

$$\mathbf{P} \sim \text{Dirichlet}(\alpha), \alpha = [1, 1, 1, 1].$$



$$u \sim N(0, \mathbf{G}\sigma_u^2) \quad \sigma_u^2 \sim \chi^{-2}_{(\nu,S)}$$

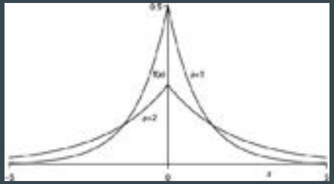
Bayes Lasso

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{Wq} + \sum_{j=1}^{q} \mathbf{x}_j\mathbf{b}_j + \mathbf{e}$$

$$GEBV_i = \sum_{j=1}^{q} x_{ij}b_j$$

- A priori distribution for SNP effects
  - SNP effects distributed as a double exponential, controlled by the lambda parameter (shrinkage).

$$\pi(b \mid \sigma_e^2) = \prod_{j=1}^{q} \frac{\lambda}{2\sqrt{\sigma_e^2}} e^{-\lambda|b_j|/\sigma_e}$$



- A priori distribution for lambda
  - Gamma on λ^2 with hyperparameters, with shape and rate depending on hyperparameters p, r, τ and δ.

$$\pi(\lambda^2) = \frac{\delta^r}{\Gamma(r)}(\lambda^2)^{r-1}e^{-\delta\lambda^2}, \qquad \lambda^2 > 0\ (r > 0, \delta > 0),$$

- A priori distribution for residual variance
  - Uninformative marginal prior, scale invariant (inverted chi-squared or inverted gamma)

$$\sigma_{b_j}^2 \sim \chi_{(\nu,S)}^{-2}$$

Marker regression with residual polygenic effect

Decomposing the polygenic effect into the sum of SNP (linear) effects

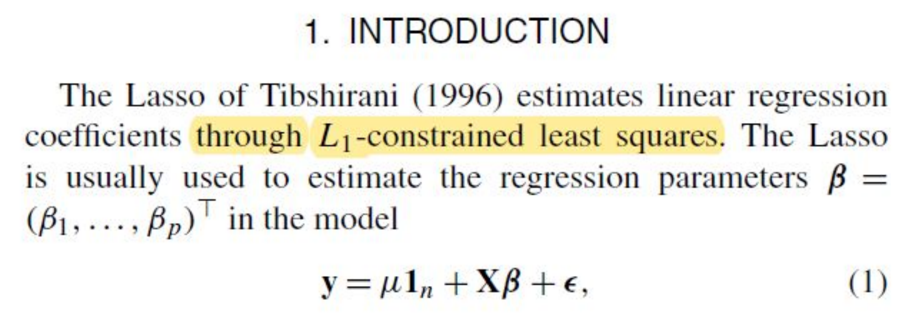
$$y_i = X_i b_i + snp_{i1}\beta_{i1} + snp_{i2}\beta_{i2} + \ldots + snp_{ip}\beta_{ip} + Z_{ui}u_{ir} + e'_i$$

- Choose model (distribution) for SNP effects
- A priori distribution for additive and residual variances
  - Distributed as inverse chi-squared

$u \sim N(0, \mathbf{G}\sigma_u^2)$ $\sigma_u^2 \sim \chi^{-2}_{(\nu,S)}$

Bayes Lasso

- The original LASSO (Tibshirani, 1996)

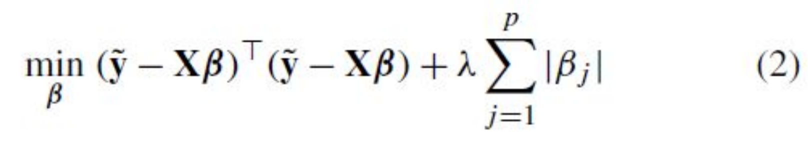


1. INTRODUCTION

The Lasso of Tibshirani (1996) estimates linear regression coefficients through L_1-constrained least squares. The Lasso is usually used to estimate the regression parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ in the model

$$\mathbf{y} = \mu \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad (1)$$

L_1 penalized function

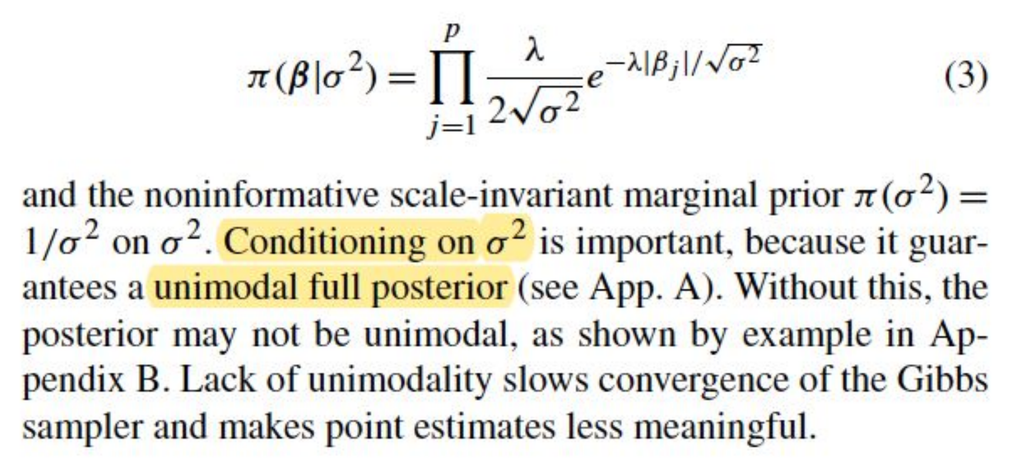


$$\min_{\boldsymbol{\beta}} (\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})^\top (\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j| \qquad (2)$$

The whole model can be efficiently computed through a modification of LARS algorithm for all lambda values.

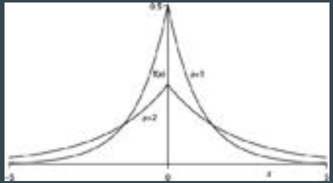
Bayes Lasso

- Park and Casella proposed a fully Bayesian analysis using a conditional (to the residual variance) Laplace prior on lambda (λ)



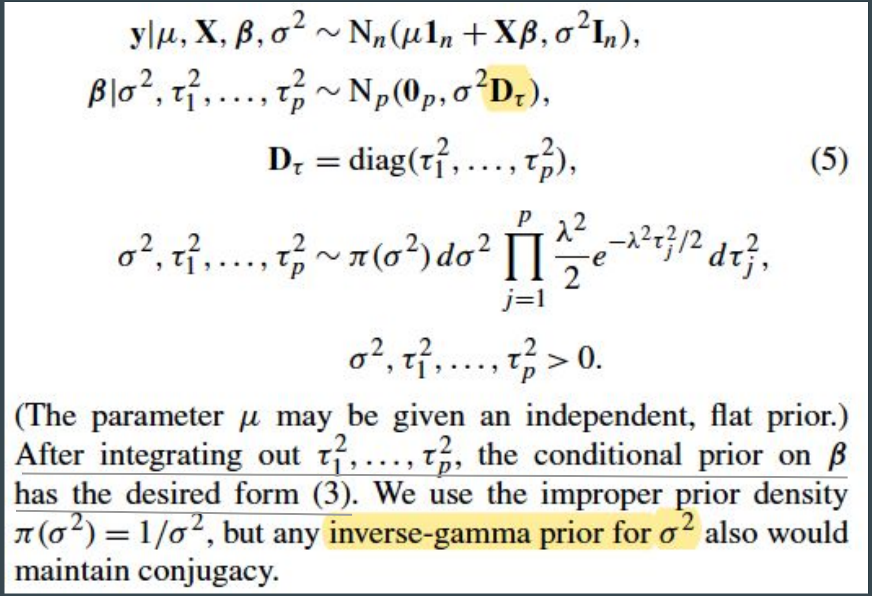
$$\pi(\beta|\sigma^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}} \qquad (3)$$

and the noninformative scale-invariant marginal prior $\pi(\sigma^2) = 1/\sigma^2$ on σ^2. Conditioning on σ^2 is important, because it guarantees a unimodal full posterior (see App. A). Without this, the posterior may not be unimodal, as shown by example in Appendix B. Lack of unimodality slows convergence of the Gibbs sampler and makes point estimates less meaningful.



Bayes Lasso

- Gibbs sampler. The hierarchical representation of the full model is as follows:

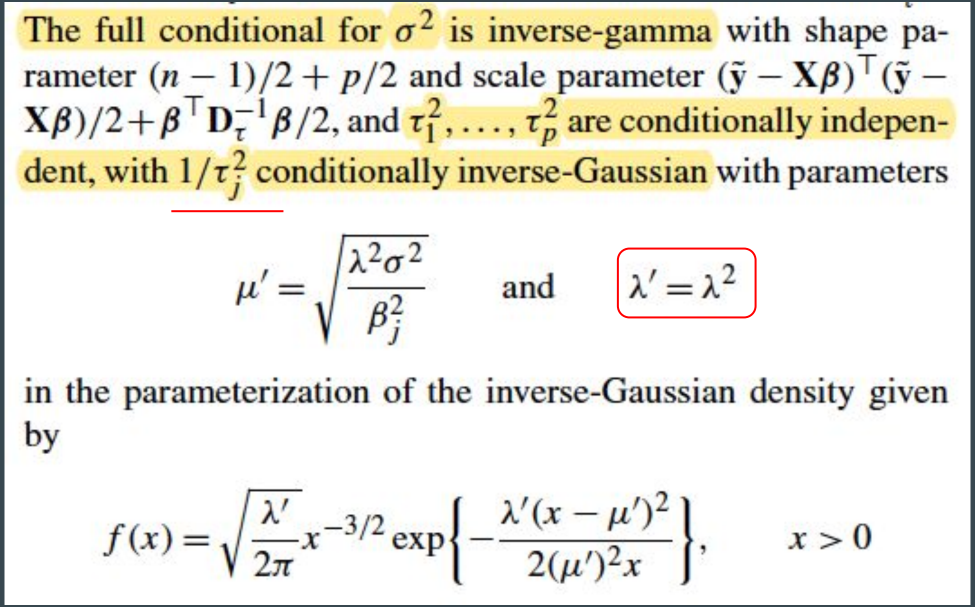


$$
\begin{aligned}
\mathbf{y}|\mu, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 &\sim \mathrm{N}_n(\mu \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n),\\
\boldsymbol{\beta}|\sigma^2, \tau_1^2, \ldots, \tau_p^2 &\sim \mathrm{N}_p(\mathbf{0}_p, \sigma^2 \mathbf{D}_\tau),\\
\mathbf{D}_\tau &= \mathrm{diag}(\tau_1^2, \ldots, \tau_p^2), \qquad (5)\\
\sigma^2, \tau_1^2, \ldots, \tau_p^2 &\sim \pi(\sigma^2)\, d\sigma^2 \prod_{j=1}^{p} \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2/2}\, d\tau_j^2,\\
\sigma^2, \tau_1^2, \ldots, \tau_p^2 &> 0.
\end{aligned}
$$

(The parameter μ may be given an independent, flat prior.) After integrating out $\tau_1^2, \ldots, \tau_p^2$, the conditional prior on $\boldsymbol{\beta}$ has the desired form (3). We use the improper prior density $\pi(\sigma^2) = 1/\sigma^2$, but any inverse-gamma prior for σ^2 also would maintain conjugacy.

Bayes Lasso

- Gibbs sampler. The hierarchical representation of the full model is as follows:



The full conditional for σ^2 is inverse-gamma with shape parameter $(n-1)/2 + p/2$ and scale parameter $(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})^\top(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})/2 + \boldsymbol{\beta}^\top \mathbf{D}_\tau^{-1}\boldsymbol{\beta}/2$, and $\tau_1^2, \ldots, \tau_p^2$ are conditionally independent, with $\underline{1/\tau_j^2}$ conditionally inverse-Gaussian with parameters

$$\mu' = \sqrt{\frac{\lambda^2\sigma^2}{\beta_j^2}} \quad \text{and} \quad \boxed{\lambda' = \lambda^2}$$

in the parameterization of the inverse-Gaussian density given by

$$f(x) = \sqrt{\frac{\lambda'}{2\pi}}\, x^{-3/2} \exp\left\{-\frac{\lambda'(x-\mu')^2}{2(\mu')^2 x}\right\}, \qquad x > 0$$

Comparison between methods

- De los Campos et al. (2013)

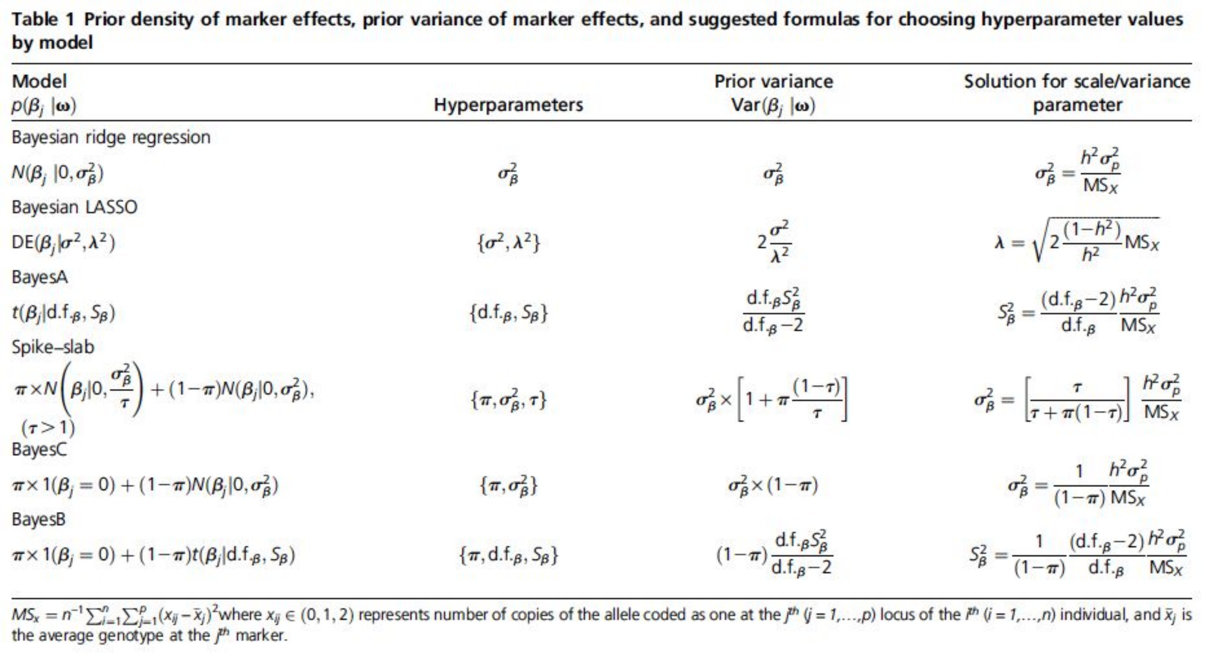


Table 1 Prior density of marker effects, prior variance of marker effects, and suggested formulas for choosing hyperparameter values by model

Model $p(\beta_j \mid \boldsymbol{\omega})$	Hyperparameters	Prior variance $\mathrm{Var}(\beta_j \mid \boldsymbol{\omega})$	Solution for scale/variance parameter
Bayesian ridge regression			
$N(\beta_j \mid 0, \sigma_\beta^2)$	σ_β^2	σ_β^2	$\sigma_\beta^2 = \frac{h^2 \sigma_p^2}{\mathrm{MS}_X}$
Bayesian LASSO			
$\mathrm{DE}(\beta_j \mid \sigma^2, \lambda^2)$	$\{\sigma^2, \lambda^2\}$	$2\frac{\sigma^2}{\lambda^2}$	$\lambda = \sqrt{2\frac{(1-h^2)}{h^2}\mathrm{MS}_X}$
BayesA			
$t(\beta_j \mid \mathrm{d.f.}_\beta, S_\beta)$	$\{\mathrm{d.f.}_\beta, S_\beta\}$	$\frac{\mathrm{d.f.}_\beta S_\beta^2}{\mathrm{d.f.}_\beta - 2}$	$S_\beta^2 = \frac{(\mathrm{d.f.}_\beta - 2)}{\mathrm{d.f.}_\beta}\frac{h^2 \sigma_p^2}{\mathrm{MS}_X}$
Spike–slab			
$\pi \times N\left(\beta_j \mid 0, \frac{\sigma_\beta^2}{\tau}\right) + (1-\pi)N(\beta_j \mid 0, \sigma_\beta^2)$, $(\tau > 1)$	$\{\pi, \sigma_\beta^2, \tau\}$	$\sigma_\beta^2 \times \left[1 + \pi\frac{(1-\tau)}{\tau}\right]$	$\sigma_\beta^2 = \left[\frac{\tau}{\tau + \pi(1-\tau)}\right]\frac{h^2 \sigma_p^2}{\mathrm{MS}_X}$
BayesC			
$\pi \times 1(\beta_j = 0) + (1-\pi)N(\beta_j \mid 0, \sigma_\beta^2)$	$\{\pi, \sigma_\beta^2\}$	$\sigma_\beta^2 \times (1-\pi)$	$\sigma_\beta^2 = \frac{1}{(1-\pi)}\frac{h^2 \sigma_p^2}{\mathrm{MS}_X}$
BayesB			
$\pi \times 1(\beta_j = 0) + (1-\pi)t(\beta_j \mid \mathrm{d.f.}_\beta, S_\beta)$	$\{\pi, \mathrm{d.f.}_\beta, S_\beta\}$	$(1-\pi)\frac{\mathrm{d.f.}_\beta S_\beta^2}{\mathrm{d.f.}_\beta - 2}$	$S_\beta^2 = \frac{1}{(1-\pi)}\frac{(\mathrm{d.f.}_\beta - 2)}{\mathrm{d.f.}_\beta}\frac{h^2 \sigma_p^2}{\mathrm{MS}_X}$

$MS_x = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij} - \bar{x}_j)^2$ where $x_{ij} \in (0,1,2)$ represents number of copies of the allele coded as one at the j^{th} $(j = 1,\ldots,p)$ locus of the i^{th} $(i = 1,\ldots,n)$ individual, and $\bar{x}_j$ is the average genotype at the j^{th} marker.

Comparison between methods

- De los Campos et al. (2013)

Larger shrinkage usually work better

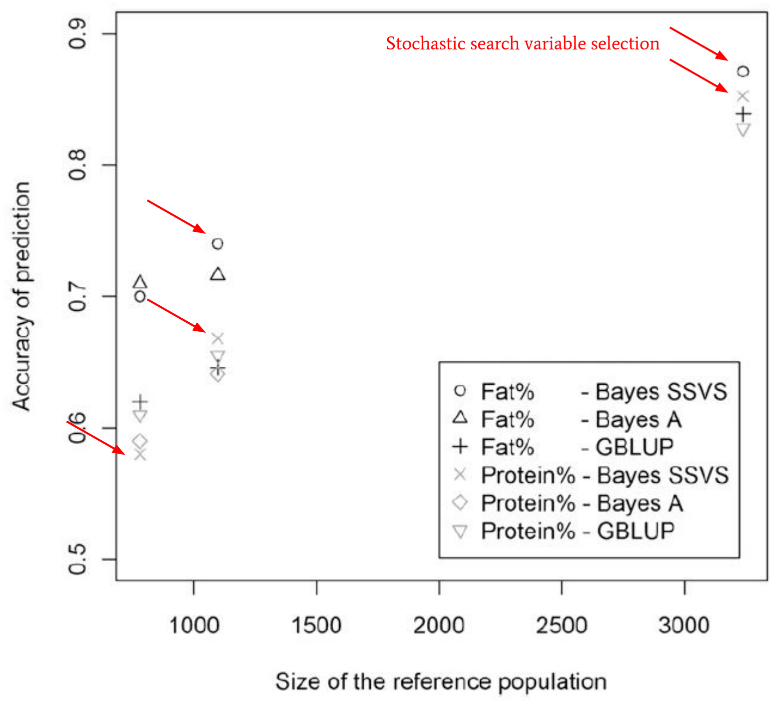


Figure 4 Accuracies of G-BLUP, BayesA, and Bayes SSVS models for fat and protein percentage, estimated using three different Holstein–Friesian reference populations (Hayes *et al*. 2009b; Verbyla *et al*. 2009; de Roos *et al*. 2011). Note that the data used by Hayes *et al*. (2009b) are a subset of the data used by Verbyla *et al*. (2009).

Considerations

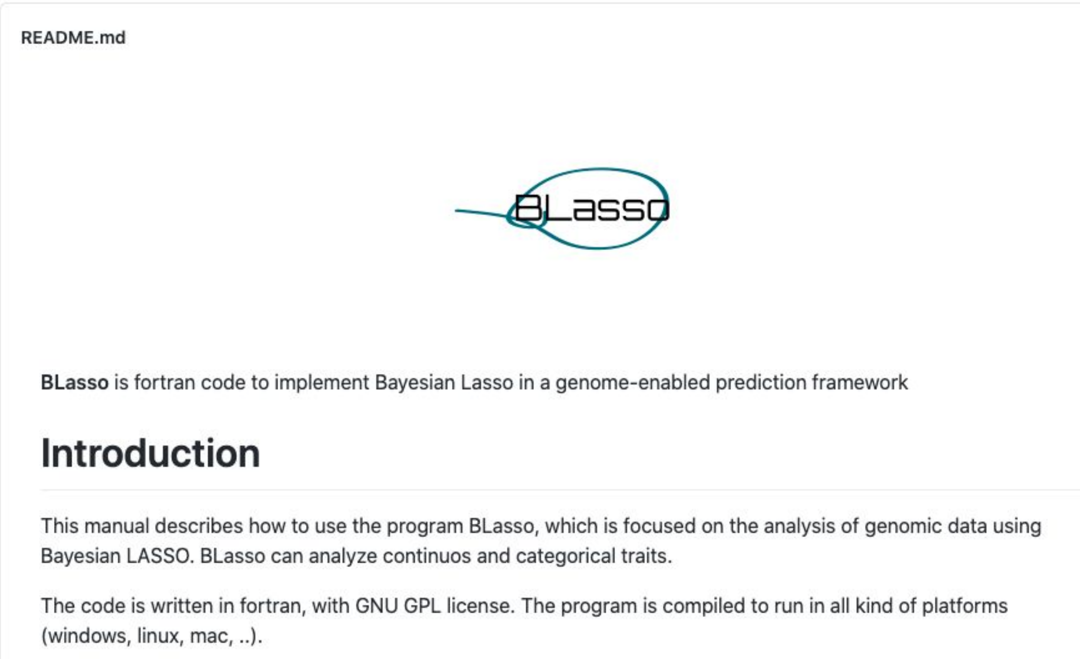
- 🚫 Bayes A and Bayes B use strong priors on the SNP variance, with 4 d.f. that do not allow bayesian learning.
- ⚠ Bayes B and Bayes R show difficult convergence in the McMC implementation.
- Prediction accuracy usually better than GBLUP, and variability (often very minor) depending on the data set, and type of implementation.
- ⚠ Prediction is not inference. Usually not very accurate at detecting QTLs with small effect.

Software

- BlupF90 (Misztal and col. UGA). GBLUP, ssGBLUP, backsolving for SNP effects: http://nce.ads.uga.edu/wiki/doku.php
- GCTA (Yang and col. Westlake Uni). GBLUP, SNP-BLUP: https://cnsgenomics.com/software/gcta/#Overview
- BGLR package in R (de los Campos and Pérez. Michigan). GBLUP, RKHS, Ridge Regression, Bayesian LASSO. https://cran.r-project.org/web/packages/BLR/index.html
- GS3 (Legarra and col. INRA). SNP-BLUP, BayesCPi, Bayesian LASSO. https://github.com/alegarra/gs3
- BayesR (Erbe, Goddard, Hayes and col). BayesR (different versions).

BLASSO

https://github.com/ogrecio/BLasso

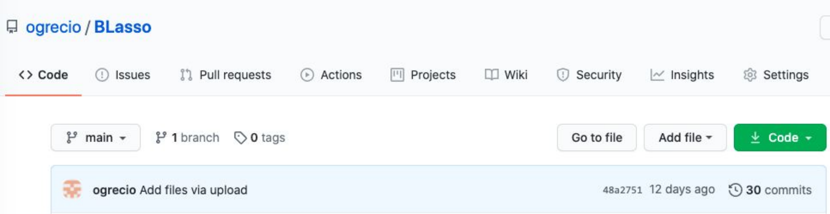

README.md

BLasso is fortran code to implement Bayesian Lasso in a genome-enabled prediction framework

Introduction

This manual describes how to use the program BLasso, which is focused on the analysis of genomic data using Bayesian LASSO. BLasso can analyze continuos and categorical traits.

The code is written in fortran, with GNU GPL license. The program is compiled to run in all kind of platforms (windows, linux, mac, ..).



ogrecio Add files via upload 48a2751 12 days ago 30 commits