Lasso and capture-recapture

Olivier Gimenez and Ian Renner

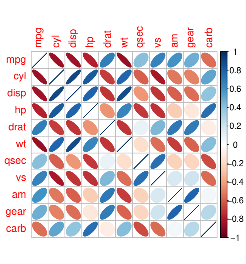
Introduction

What is this about?

- Explain variation in abundance and survival
- Capture-recapture models are often used
- Survival models with imperfect detection
- With technology, come many variables
- Often do not know which ones (not) to include

What are the issues?

Many, possibly correlated, covariates



Correlation $\implies$ numerical instability

Many covariates $\implies$ $\searrow$ precision and predictability

What we usually do

Think hard about which covariates to consider

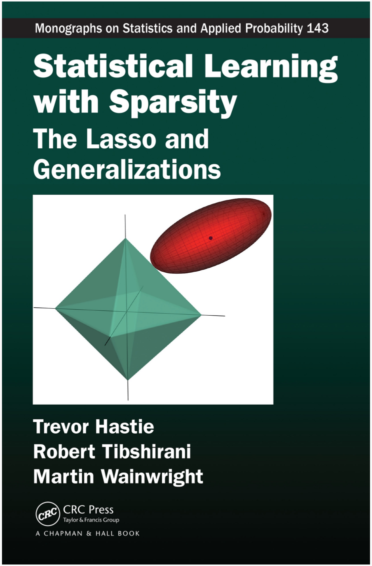
Select covariates using:

- AIC or stepwise procedure
- DIC, SSVS, RJMCMC

This talk: shrink and select model coefficients

Theory

The reference - free book!



It all starts with the ridge regression

Maximize likelihood, penalize magnitude of coeff.

$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) \text{ subject to } \sum_{j=1}^{p} \beta_j^2 < c$$

It all starts with the ridge regression

Maximize likelihood, penalize magnitude of coeff.

$$\widehat{\boldsymbol{\beta}} = \operatorname{argmax} L(\boldsymbol{\beta}) \text{ subject to } \sum_{j=1}^{p} \beta_j^2 < c$$

It all starts with the ridge regression

Maximize likelihood, penalize magnitude of coeff.

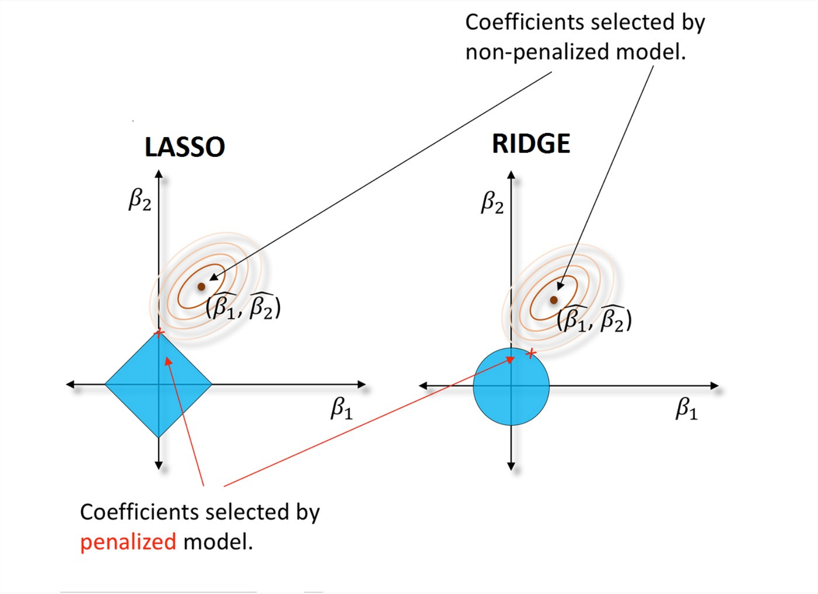
$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) \text{ subject to } \sum_{j=1}^{p} \beta_j^2 < c$$

Lasso = Least Absolute Shrinkage and Selection Operator

Change the constraint: ℓ^2 vs. ℓ^1 norm

$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) \text{ subject to } \sum_{j=1}^{p} |\beta_j| < c$$

Lasso vs. ridge regression, graphically



Lasso: maximizing penalized likelihood

$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) \text{ subject to } \sum_{j=1}^{p} |\beta_j| < c$$

Constrained optimization not easy

Rewrite the problem with Lagrange multipliers

$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j|$$

Adaptive lasso penalty to achieve oracle properties

Lasso: maximizing penalized likelihood

$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) \text{ subject to } \sum_{j=1}^{p} |\beta_j| < c$$

Constrained optimization not easy

Rewrite the problem with Lagrange multipliers

$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j|; \text{ capture-recapt. lik.}$$

Adaptive lasso penalty to achieve oracle properties

Capture–recapture likelihood

- 1s and 0s for detections and non-detections
- For example, animal i may be $h_i = 101$
- Denote ϕ^t survival prob between t and $t+1$ and p^t recapture prob at t
- Contribution of animal i to likelihood is $\Pr(h_i) = \phi^1(1-p^2)\phi^2 p^3$
- $\text{logit}(\phi^t) = \beta_0 + \beta_1 x_1^t + \ldots + \beta_K x_K^t$
- Likelihood is $\prod_i \Pr(h_i)$ for all animals i

How to choose the penalty term λ?

- Usually, cross-validation techniques
- Build a grid of values for λ
- Repeat optimization for each value of the grid
- Pick λ corresponding to model with lowest BIC

Simulations

Setting: Capture-recapture model

- Sample size: 15 occasions with 15 new ind.
- Detection is 0.9, mean survival is 0.8
- Covariates: $X_1 \sim N(-0.6, \sigma = 1)$, $X_2 \sim N(0, \sigma = 1)$
- Apply Lasso; fit 4 models, compare with AIC
- Repeat 100 times

Simulation results

- Lasso selects correct model (X_1 only) 80%
- Comparable to variable selection using AIC
- Further simulations show similar results

Application

White storks wintering in Sahel

Capture-recapture data over 16 years

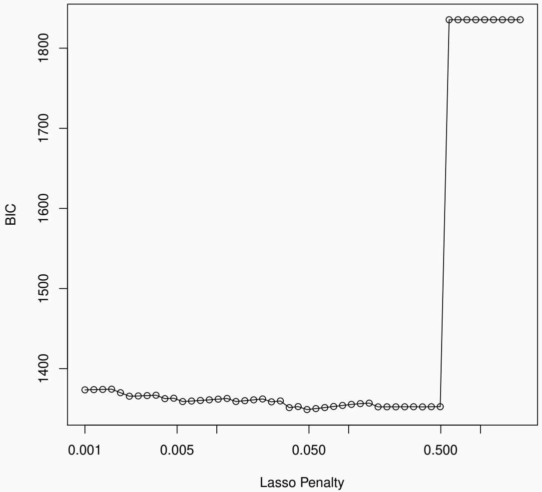
Rainfall was measured at 10 meteo stations in Sahel

Is adult white stork survival affected by rainfall?

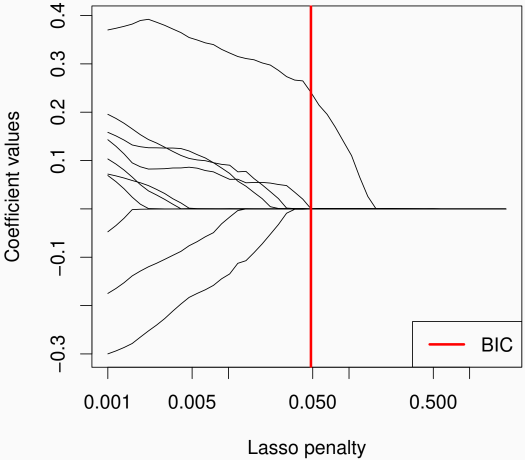
$\text{logit}(\phi^t) = \beta_0 + \beta_1 x_1^t + \ldots + \beta_{10} x_{10}^t$

Do we need to consider 2^{10} candidate models?

Choosing the Lasso penalty using BIC



Exploring regularization path



Rainfall effect at all weather stations

Station	Estimate
Diourbel	7.47×10^{-5}
Gao	-2.99×10^{-5}
Kayes	1.3×10^{-4}
Kita	0.24
Maradi	-1.3×10^{-4}
Mopti	3.5×10^{-4}
Ouahigouya	-5.9×10^{-5}
Segou	1.7×10^{-5}
Tahoua	1.2×10^{-4}
Tombouctou	-2.3×10^{-4}

Conclusions and perspectives

Conclusions and perspectives

- From selecting variables to shrinking estimates
- Penalized likelihood easy to implement
- Ongoing work with Bayesian flavor

Questions