The Horseshoe

An (Un)expected Journey

Matthieu Authier

Observatoire Pelagis, UAR3462, CNRS-LRUniv

MINISTÈRE DE LA TRANSITION ÉCOLOGIQUE
Liberté
Égalité
Fraternité

PELAGIS UAR 3462 - LRUniv - CNRS

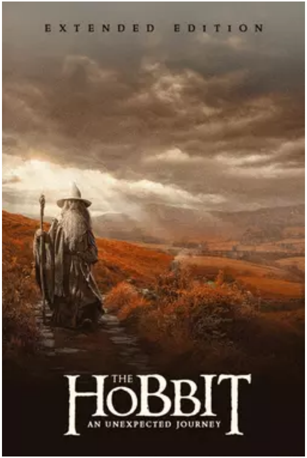
La Rochelle Université

Adera Soutenir les acteurs de l'innovation

Sommaire

2012

2012



The journey begins as a quantitative ecologist

Medium to Small data

1. intensive field work
2. N ranging from < 5 to > 1000
3. many covariates to consider: age, sex, mass, length, environmental covariates...

Statistical ecology was coming of age...

Gimenez et al. (2014)

Statistical ecology was coming of age...

Gimenez et al. (2014)
The community talked a lot about hierarchical models.

Statistical ecology was coming of age...

Gimenez et al. (2014)
The community talked a lot about hierarchical models.

DIC was used extensively (although we also knew about the debate around it; Spiegelhalter et al., 2002; Celeux et al., 2006)

Statistical ecology was coming of age...

Gimenez et al. (2014)
The community talked a lot about hierarchical models.

DIC was used extensively (although we also knew about the debate around it; Spiegelhalter et al., 2002; Celeux et al., 2006)

There was not much about priors in the community.

Bayesian methods in ecology

"[...] in cases with many predictor variables and few observations ecologists should consider regression shrinkage methods (Dahlgren, 2010)."

Sommaire

Linear model setting

$\forall\ i, y_i|\mu_i, \sigma \sim \mathcal{N}(\mu_i, \sigma)$
with $\mu_i = \beta_0 + \sum_{k=1}^{p} \beta_k \times x_i^k$

Linear model setting

$\forall\ i, y_i|\mu_i, \sigma \sim \mathcal{N}(\mu_i, \sigma)$
with $\mu_i = \beta_0 + \sum_{k=1}^{p} \beta_k \times x_i^k$
Hierarchical priors (scale-mixture of normals):

$$\beta_k|\sigma, \lambda_k \sim \mathcal{N}(0, \sigma\lambda_k). \tag{1}$$

λ_k are local scale parameters; sampled from a common distribution.

Linear model setting

$\forall\ i, y_i|\mu_i, \sigma \sim \mathcal{N}(\mu_i, \sigma)$
with $\mu_i = \beta_0 + \sum_{k=1}^{p} \beta_k \times x_i^k$
Hierarchical priors (scale-mixture of normals):

$$\beta_k|\sigma, \lambda_k \sim \mathcal{N}(0, \sigma\lambda_k). \quad (1)$$

λ_k are local scale parameters; sampled from a common distribution.
Conditioning on the residual scale σ guarantees unimodal posteriors (Park & Casella, 2008).

The Bayesian Lasso

Assuming
$\lambda_k^2|\tau \sim \mathcal{E}(\frac{\tau^2}{2})$
leads to the Bayesian Lasso (Park & Casella, 2008).

The Bayesian Lasso

Assuming
$$\lambda_k^2|\tau \sim \mathcal{E}(\frac{\tau^2}{2})$$
leads to the Bayesian Lasso (Park & Casella, 2008).

The choice of an exponential distribution for the local scales leads to a Laplace distribution. The global parameter τ pulls all the weights globally towards zero, while the the local scales λ_k allow some of the weights to escape the shrinkage.

The Bayesian Lasso

Assuming

$$\lambda_k^2|\tau \sim \mathcal{E}(\frac{\tau^2}{2})$$

leads to the Bayesian Lasso (Park & Casella, 2008).

The Laplace distribution has light tails, which can lead to excessive shrinkage for large values of τ^2.

The Horseshoe

Assuming

$$\begin{cases} \beta_k | \sigma, \lambda_k \sim \mathcal{N}(0, \sigma\lambda_k) \\ \lambda_k | \tau \sim \mathrm{C}^+(0, \tau) \\ \tau \sim \mathrm{C}^+(0, 1) \end{cases}$$

leads to the Horseshoe prior (Carvalho et al., 2009, 2010; Polson & Scott, 2012).

The Horseshoe

Assuming

$$\begin{cases} \beta_k | \sigma, \lambda_k \sim \mathcal{N}(0, \sigma\lambda_k) \\ \lambda_k | \tau \sim \mathrm{C}^+(0, \tau) \\ \tau \sim \mathrm{C}^+(0, 1) \end{cases}$$

leads to the Horseshoe prior (Carvalho et al., 2009, 2010; Polson & Scott, 2012).

λ_k and τ are local and global scale parameters, both sampled from a Cauchy distribution (a.k.a. the Witch of Agnesi; Stigler, 1974), which is heavy-tailed.

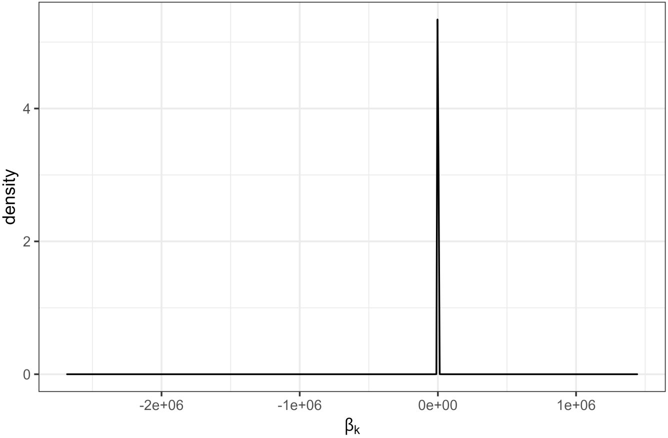
The Horseshoe

Assuming

$$
\begin{cases}
\beta_k|\sigma, \lambda_k \sim \mathcal{N}(0, \sigma\lambda_k) \\
\lambda_k|\tau \sim \mathrm{C}^+(0, \tau) \\
\tau \sim \mathrm{C}^+(0, 1)
\end{cases}
$$

The global parameter τ pulls all the weights globally towards zero, while the thick half Cauchy tails for the local scales λ_k allow some of the weights to escape the shrinkage (Piironen & Vehtari, 2017).

The Horseshoe



Shrinkage coefficient

The shrinkage factor κ_k describes how much coefficient β_k is shrunk towards zero from the maximum likelihood β_k^{ML}:

Shrinkage coefficient

The shrinkage factor κ_k describes how much coefficient β_k is shrunk towards zero from the maximum likelihood β_k^{ML}:

$$\begin{cases} \bar{\beta}_k = (1 - \kappa_k) \times \beta_k^{\mathrm{ML}} \\ \kappa_k = 0, \ \text{no shrinkage} \\ \kappa_k = 1, \ \text{complete shrinkage} \end{cases}$$

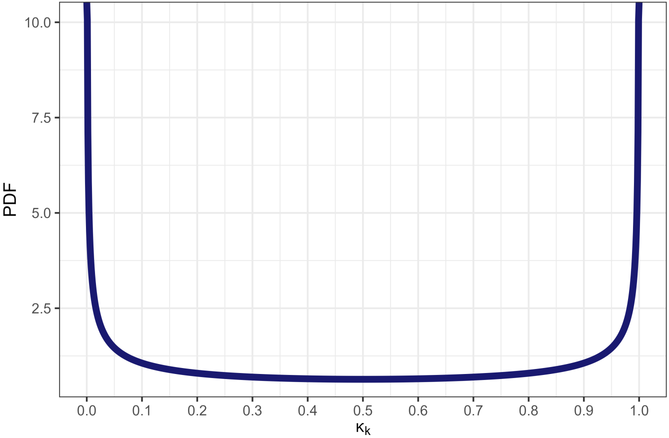
Shrinkage coefficient

The shrinkage factor κ_k describes how much coefficient β_k is shrunk towards zero from the maximum likelihood β_k^{ML}:

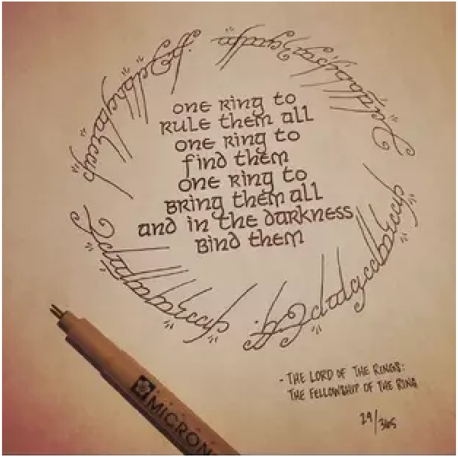
$$\begin{cases} \bar{\beta}_k = (1 - \kappa_k) \times \beta_k^{\mathrm{ML}} \\ \kappa_k = 0, \text{ no shrinkage} \\ \kappa_k = 1, \text{ complete shrinkage} \end{cases}$$

With the Horseshoe, $\kappa_k \sim \mathcal{B}\left(\frac{1}{2}, \frac{1}{2}\right)$

Shrinkage coefficient



One prior to find them all and in the model shrink them



Sommaire

Southern Elephant Seals

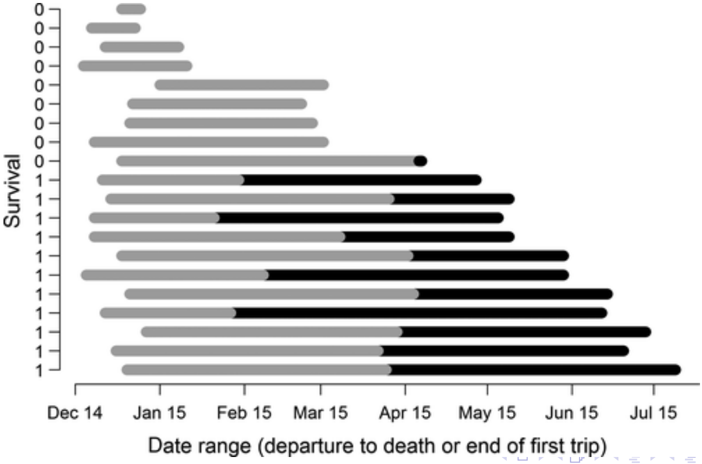
Southern Elephant Seals

Cox et al. (2020)

- Nov-Dec 2014: 10 male and 10 female weanlings ($\approx$ 80 kg and 3 weeks-old)
- Equipped with 2 independent tags (Argos & SPOT)
- Response variable: time to death T_i (simulatenous tag failure), censored
- $p = 48$ covariates, incl. mass, sex, movement, environment, etc.

Southern Elephant Seals

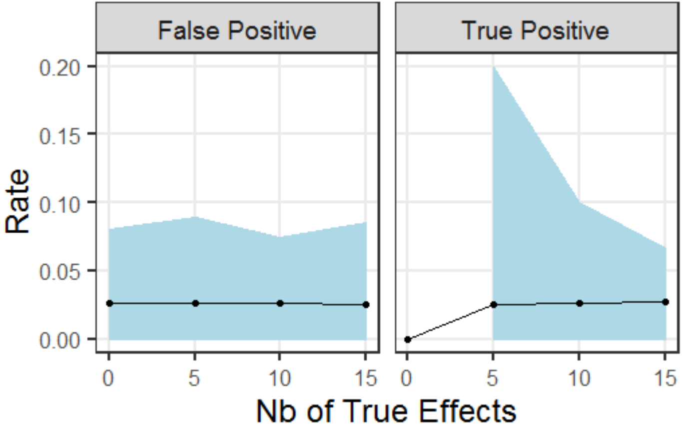
$log(T_i) \sim \mathcal{N}(\mu_i, \sigma)$, if death is observed (censored otherwise; unimodal hazard)



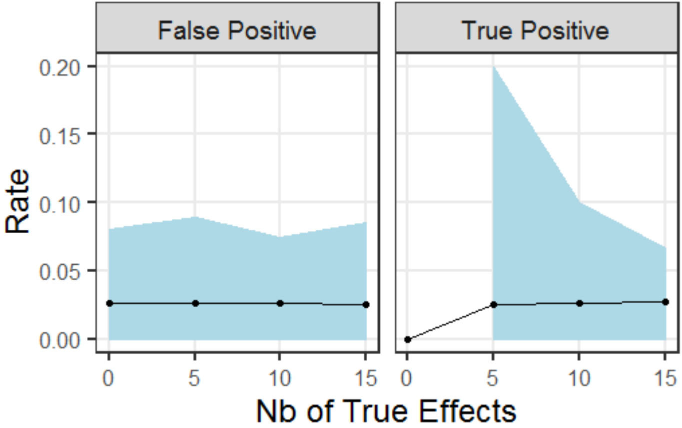
Simulation study

- Sample size: 20 individuals
- $p = 50$ covariates
- true effect size: $\approx 10\%$ of increased/decreased survival time
- true number of active features: 0, 5, 10, 20

Simulation study

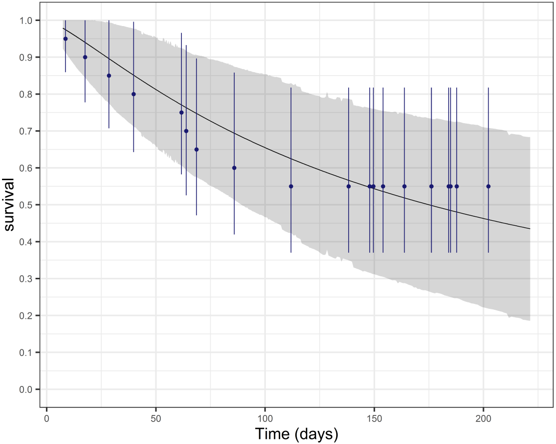


Simulation study



Any selected feature has a 'probability' of 0.5 to be right or wrong.

Survival results



Survival results

- lower male survival
- decreased survival with increased daily maximum speeds and distances traveled.
- Individuals with swim efforts that increased through time were more likely to die than those whose swim efforts did not.

Survival results

- lower male survival
- decreased survival with increased daily maximum speeds and distances traveled.
- Individuals with swim efforts that increased through time were more likely to die than those whose swim efforts did not.

Kind of expected...

Sommaire

The Regularized Horseshoe

a.k.a. the Finnish Ponyshoe

Piironen & Vehtari (2017) discuss a regularized version of the Horseshoe to guarantee that the prior always shrinks the coefficients at least by a small amount.

The Regularized Horseshoe

a.k.a. the Finnish Ponyshoe
Piironen & Vehtari (2017) discuss a regularized version of the Horseshoe to guarantee that the prior always shrinks the coefficients at least by a small amount.

The Horseshoe favors either no or complete shrinkage. While this guarantees that the strong signals will not be overshrunk, this property can also be harmful, especially when the parameters are weakly identified.

An example of such case is the flat likelihood arising in logistic regression with separable data.

The Regularized Horseshoe

$$\begin{cases} \beta_k | \tau, \tilde{\lambda}_k \sim \mathcal{N}(0, \tau \tilde{\lambda}_k) \\ \tilde{\lambda}_k = \dfrac{c \times \lambda_k}{\sqrt{c^2 + \lambda_k^2}} \\ c | \nu, s \sim \mathrm{T}_{\nu}{}^{+}(0, s) \\ \lambda_k | \tau \sim \mathrm{C}^{+}(0, 1) \\ \tau | \tau_0 \sim \mathrm{C}^{+}(0, \tau_0) \\ \tau_0 = \dfrac{p_0}{p - p_0} \times \dfrac{\sigma}{\sqrt{n}} \end{cases}$$

c controls regularization for large effects, p_0 is the prior guess for relevant features (Piironen & Vehtari, 2017).

Sommaire

High bycatch of short-beaked common dolphin in the Bay of Biscay

High bycatch of short-beaked common dolphin in the Bay of Biscay

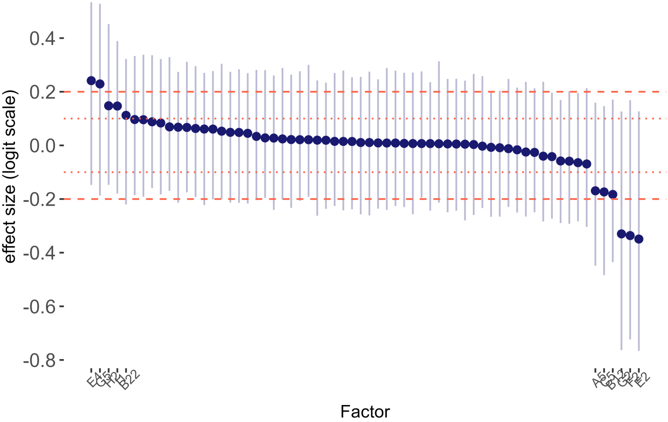
Interviews with fishermen in 2019, $N = 96$

- "How many dolphins do you bycatch on average every year?"
- $p = 64$ features: harbour, gear, métiers, vessel length

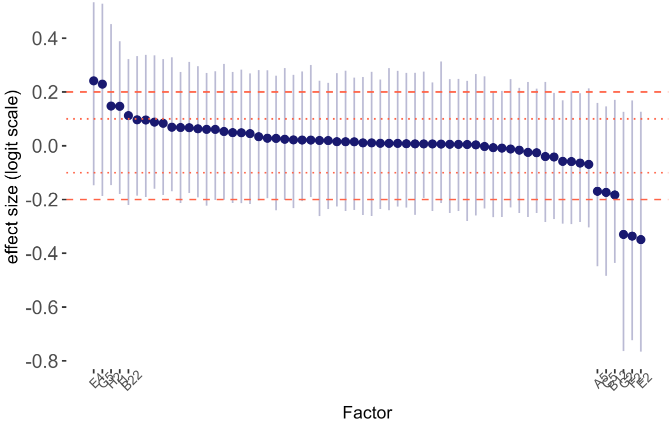
Two logistic regressions ($p_0 = 10$, T_3^+ priors for scales Piironen & Vehtari, 2017):

1. willingness to answer (even to answer "none")
2. at least one dolphin bycaught

Willingness to answer (N = 96)

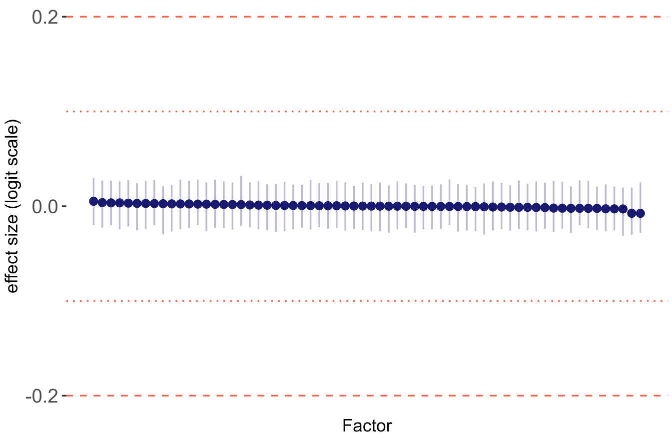


Willingness to answer (N = 96)

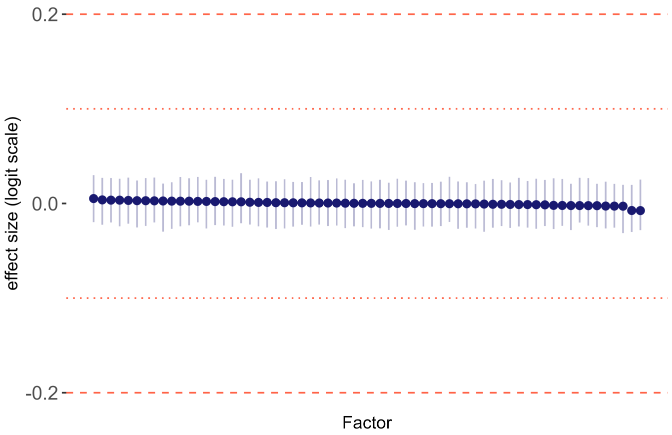


Gillnetters and pair-trawlers are respectively more and less likely to answer

Bycatch incidence ($N = 79$)



Bycatch incidence ($N = 79$)



Empirical average in the sample is a whopping 0.86

Interview results

- pair-trawlers less likely to answer (already identified)
- gill-netters were under the radar
- bycatch occurrence is widespread

Interview results

- pair-trawlers less likely to answer (already identified)
- gill-netters were under the radar
- bycatch occurrence is widespread
- these results did not make it to the final report...

Interview results

- pair-trawlers less likely to answer (already identified)
- gill-netters were under the radar
- bycatch occurrence is widespread
- these results did not make it to the final report...

Kind of expected...

Sommaire

2022

2022

The Horseshoe in Ecology?

- easy to use (Thanks Stan!)
- very useful imho to get reliable results

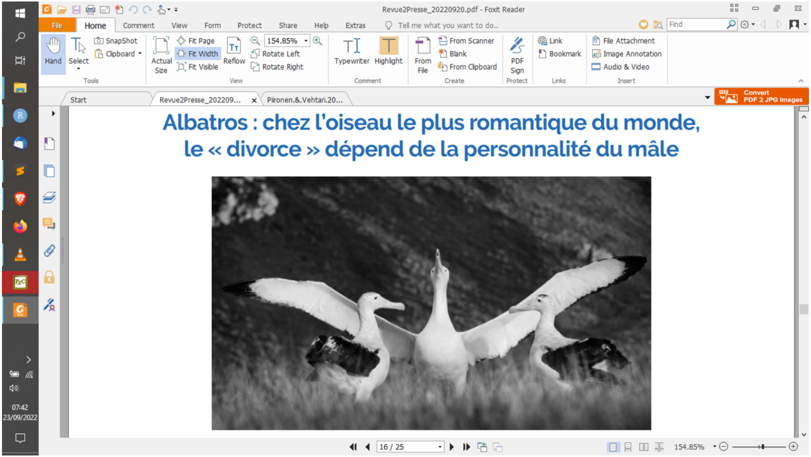
The Horseshoe in Ecology?

- easy to use (Thanks Stan!)
- very useful imho to get reliable results
- not used

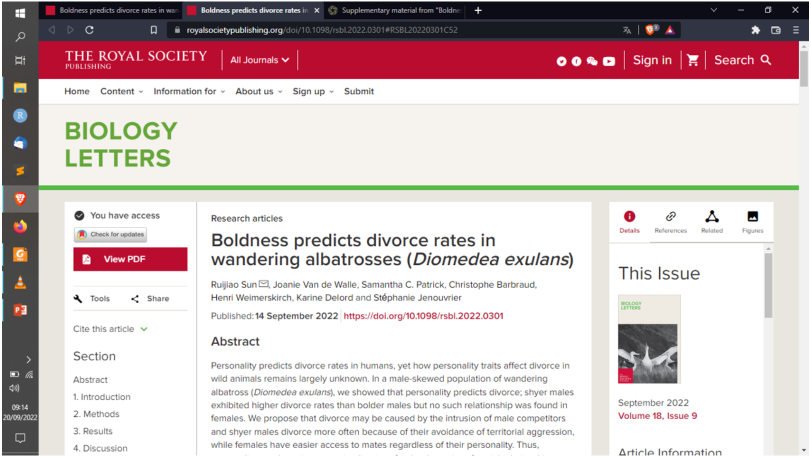
The Horseshoe in Ecology?

- easy to use (Thanks Stan!)
- very useful imho to get reliable results
- not used
- "boring" results?

Wandering albatross (*Diomedea exulans*)



Breaking news this very week



THE ROYAL SOCIETY PUBLISHING

All Journals | Sign in | Search

Home | Content | Information for | About us | Sign up | Submit

BIOLOGY LETTERS

You have access

Check for updates

View PDF

Tools | Share

Cite this article

Section

Abstract

1. Introduction

2. Methods

3. Results

4. Discussion

Research articles

Boldness predicts divorce rates in wandering albatrosses (*Diomedea exulans*)

Ruijiao Sun, Joanie Van de Walle, Samantha C. Patrick, Christophe Barbraud, Henri Weimerskirch, Karine Delord and Stéphanie Jenouvrier

Published: 14 September 2022 | https://doi.org/10.1098/rsbl.2022.0301

Abstract

Personality predicts divorce rates in humans, yet how personality traits affect divorce in wild animals remains largely unknown. In a male-skewed population of wandering albatross (*Diomedea exulans*), we showed that personality predicts divorce; shyer males exhibited higher divorce rates than bolder males but no such relationship was found in females. We propose that divorce may be caused by the intrusion of male competitors and shyer males divorce more often because of their avoidance of territorial aggression, while females have easier access to mates regardless of their personality. Thus,

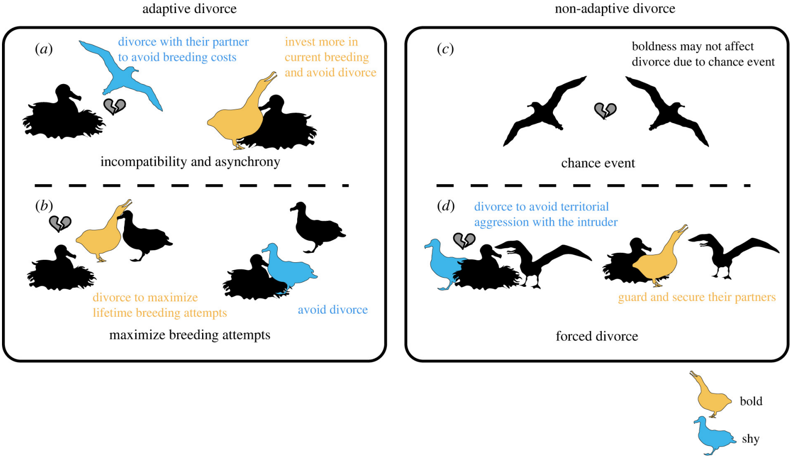
Details | References | Related | Figures

This Issue

September 2022

Volume 18, Issue 9

Article Information


adaptive divorce
non-adaptive divorce
(a)
divorce with their partner to avoid breeding costs
invest more in current breeding and avoid divorce
incompatibility and asynchrony
(b)
divorce to maximize lifetime breeding attempts
avoid divorce
maximize breeding attempts
(c)
boldness may not affect divorce due to chance event
chance event
(d)
divorce to avoid territorial aggression with the intruder
guard and secure their partners
forced divorce
bold
shy

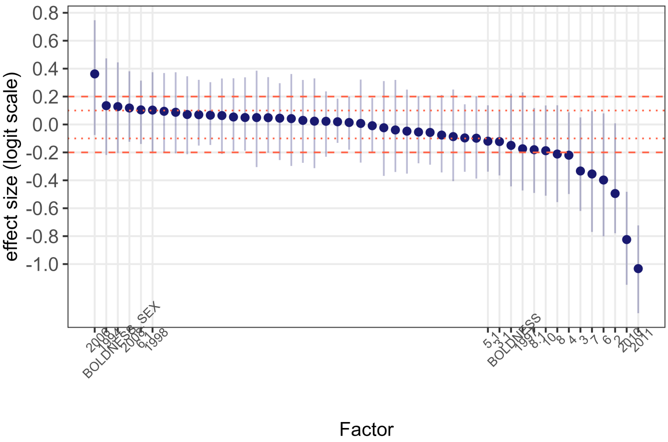
Long-term study intiated in 1959, $N = 1112$, 44 years

- $p = 48$: year, breeding experience, breeding success, pair bond duration, boldness, sex
- "pseudo-replication": same individual contribute several data
- focus on personality (shy-bold), as assessed from the behavioural response of individuals to a human approach (from 5 m)
- response variable: did the pair break-up that year? (note both male and female are included)

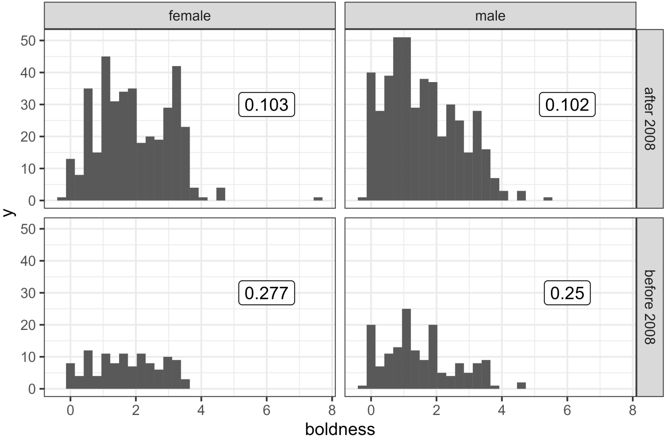
Long-term study intiated in 1959, $N = 1112$, 44 years

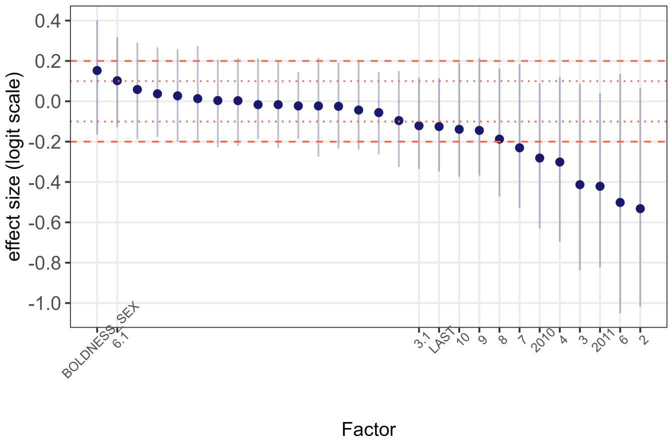
- $p = 48$: year, breeding experience, breeding success, pair bond duration, boldness, sex
- "pseudo-replication": same individual contribute several data
- focus on personality (shy-bold), as assessed from the behavioural response of individuals to a human approach (from 5 m)
- response variable: did the pair break-up that year? (note both male and female are included)

Assume $p_0 = 10$

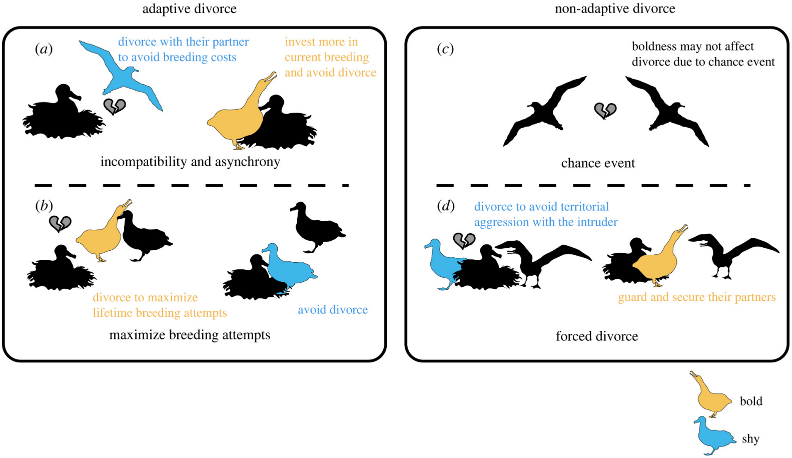

effect size (logit scale)
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
SEX
BOLDNESS
2010
2011
Factor

Raw data: boldness measurement started in 2008 (scale between 0 and 5)




effect size (logit scale)
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
BOLDNESS
SEX
6.1
3.1
LAST
10
9
8
7
2010
4
3
2011
6
2
Factor

Sommaire


adaptive divorce
non-adaptive divorce
(a)
divorce with their partner to avoid breeding costs
invest more in current breeding and avoid divorce
incompatibility and asynchrony
(b)
divorce to maximize lifetime breeding attempts
avoid divorce
maximize breeding attempts
(c)
boldness may not affect divorce due to chance event
chance event
(d)
divorce to avoid territorial aggression with the intruder
guard and secure their partners
forced divorce
bold
shy

The Horseshoe in Ecology?

- easy to use (Thanks Stan!, Carpenter et al., 2017)
- very useful imho for exploratory studies: just one model to fit

The Horseshoe in Ecology?

- easy to use (Thanks Stan!, Carpenter et al., 2017)
- very useful imho for exploratory studies: just one model to fit
- need to be used more to avoid brittle inferences
- the horseshoe to rein in the wild horses out there (against our tendency to overfit with all-encompassing theories)?

Thanks Eric!

$$\beta_k \sim \mathcal{N}(0, \lambda_k), \ \lambda_k \sim \mathrm{C}^+(0, \tau), \ \tau \sim \mathrm{C}^+(0, 1)$$

References I

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 76. URL https://www.jstatsoft.org/article/view/v076i01.

Carvalho, C., Polson, N. & Scott, J. (2009). Handling Sparsity via the Horseshoe. *Journal of Machine Learning Research (Proceedings of the 12th International Conference on Artificial Intelligence and Statistics)* 5 73–80.

Carvalho, C., Polson, N. & Scott, J. (2010). The Horseshoe Estimator for Sparse Signals. *Biometrika* 97 465–480.

Celeux, G., Forbes, F., Robert, C. & Titterington, D. (2006). Deviance Information Criteria for Missing Data Models. *Bayesian Analysis* 1 651–674.

References II

Cox, S. L., Authier, M., Orgeret, F., Weimerskirch, H. & Guinet, C. (2020). High Mortality Rates in a Juvenile Free-Ranging Marine Predator and Links to Dive and Forage Ability. *Ecology and Evolution* .

Dahlgren, J. (2010). Alternative Regression Methods are not Considered in Murtaugh (2009) or by Ecologists in General. *Ecology Letters* 13 E7–E9.

Gimenez, O., Buckland, S. T., Morgan, B. J., Bez, N., Bertrand, S., Choquet, R., Dray, S., Etienne, M. P., Fewster, R., Gosselin, F., Mérigot, B., Monestiez, P., Morales, J. M., Mortier, F., Munoz, F., Ovaskainen, O., Pavoine, S., Pradel, R., Schurr, F. M., Thomas, L., Thuiller, W., Trenkel, V., de Valpine, P. & E., R. (2014). Statistical Ecology Comes of Age. *Biology Letters* 10 1–4.

Park, T. & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* 103 681–686.

References III

PIIRONEN, J. & VEHTARI, A. (2017). Sparsity Information and Regularization in the Horseshoe and Other Shrinkage Priors. *Electronic Journal of Statistics* 11 5018–5051. URL https://doi.org/10.1214/17-EJS1337SI.

POLSON, N. & SCOTT, J. (2012). Local Shrinkage Rules, Lévy Processes and Regularized Regression. *Journal of the Royal Statistical Society series B* 74 287–311.

SPIEGELHALTER, D., BEST, T., CARLIN, B. & VAN DER LINDE, A. (2002). Bayesian Measures of Model Complexity and Fit (with Discussion). *Journal of the Royal Statistical Society series B* 64 583–640.

STIGLER, S. (1974). Studies in the History of Probability and Statistics. XXXIII Cauchy and the Witch of Agnesi: An Historical Note on the Cauchy Distribution. *Biometrika* 61 375–380.