SMSTC Statistics Stream

Perth Symposium
5 October 2017

Steve Buckland
School of Mathematics & Statistics
University of St Andrews
steve@st-andrews.ac.uk

Linear model

$$y_i \quad = \quad \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i \quad \varepsilon_i \sim \mathbb{N}(0, \sigma^2)$$

Response variable

Explanatory variables

Parameters

Noise

Linear model

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_p x_{1p} + \varepsilon_1 \quad \varepsilon_1 \sim \mathbb{N}(0, \sigma^2) \\
\vdots &\quad \vdots \\
y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_p x_{np} + \varepsilon_n \quad \varepsilon_n \sim \mathbb{N}(0, \sigma^2)
\end{aligned}
$$

n observations

p explanatory variables

Revision of linear models

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_p x_{1p} + \varepsilon_1 \quad \varepsilon_1 \sim \mathbb{N}(0, \sigma^2) \\ \vdots & \quad \vdots \quad \vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_p x_{np} + \varepsilon_n \quad \varepsilon_n \sim \mathbb{N}(0, \sigma^2) \end{aligned}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim \mathbb{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \qquad \mathrm{RSS}(f) = \sum_{i=1}^{N} (y_i - f(x_i))^2$$

Linear model,
covered in Semester 1
(Regression and
Simulation Methods)

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \qquad \mathrm{RSS}(f) = \sum_{i=1}^{N} (y_i - f(x_i))^2$$

$$\mathrm{RSS}(\beta) = (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)$$

$$\frac{\partial \mathrm{RSS}}{\partial \beta} = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta)$$

$$\frac{\partial^2 \mathrm{RSS}}{\partial \beta \partial \beta^T} = 2\mathbf{X}^T\mathbf{X}.$$

$$\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta) = 0$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

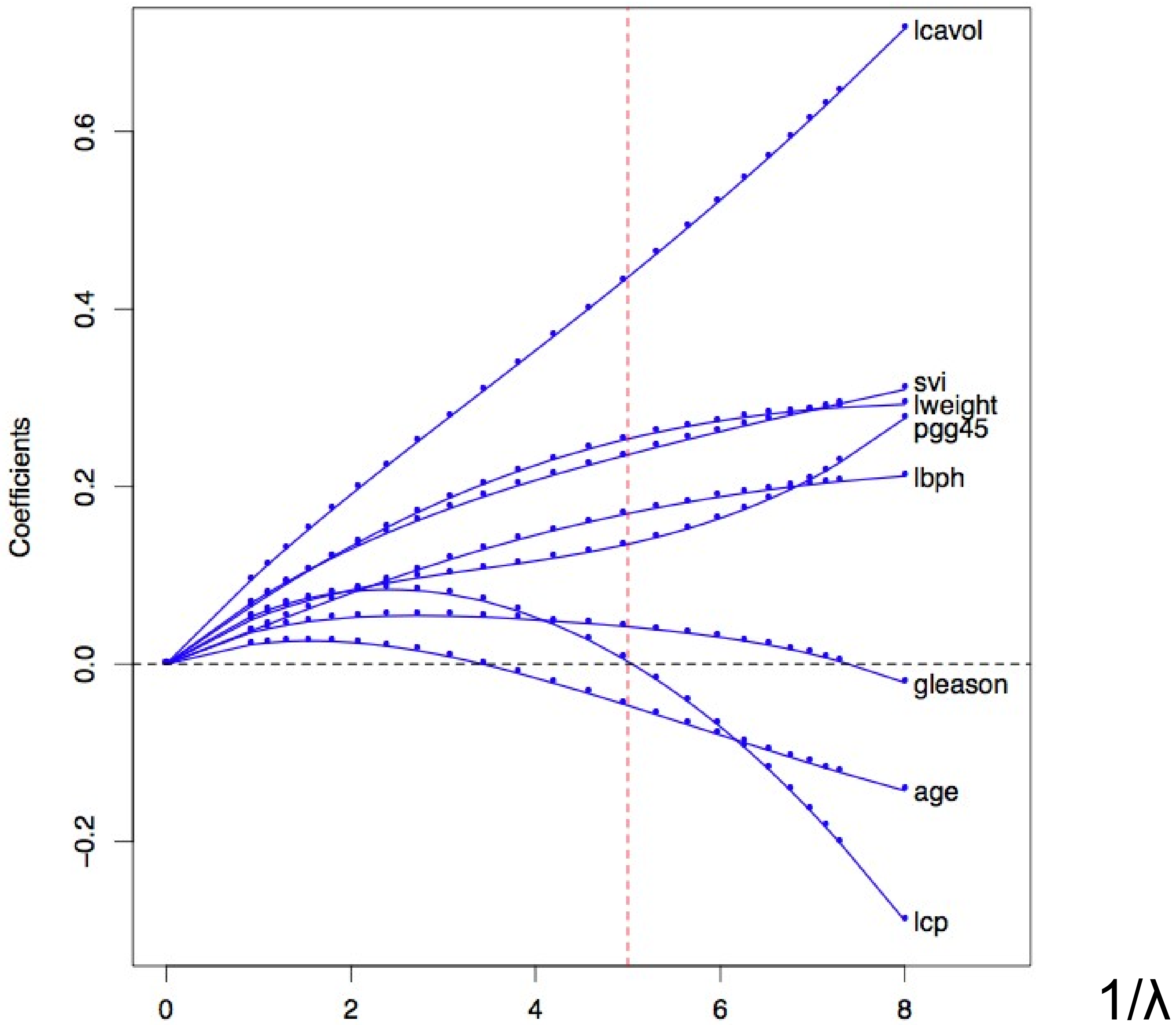
$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

What happens when this matrix is singular?

Ridge regression

$$\hat{\beta}^{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$$

$$\text{RSS}(\lambda) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) + \lambda\beta^T\beta$$


Coefficients
0.6
0.4
0.2
0.0
−0.2
0
2
4
6
8
lcavol
svi
lweight
pgg45
lbph
gleason
age
lcp
1/λ

Lasso (Semester 2 – Modern Regression and Bayesian Methods)

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \Big)^2$$

$$\text{subject to } \sum_{j=1}^{p} |\beta_j| \le t.$$

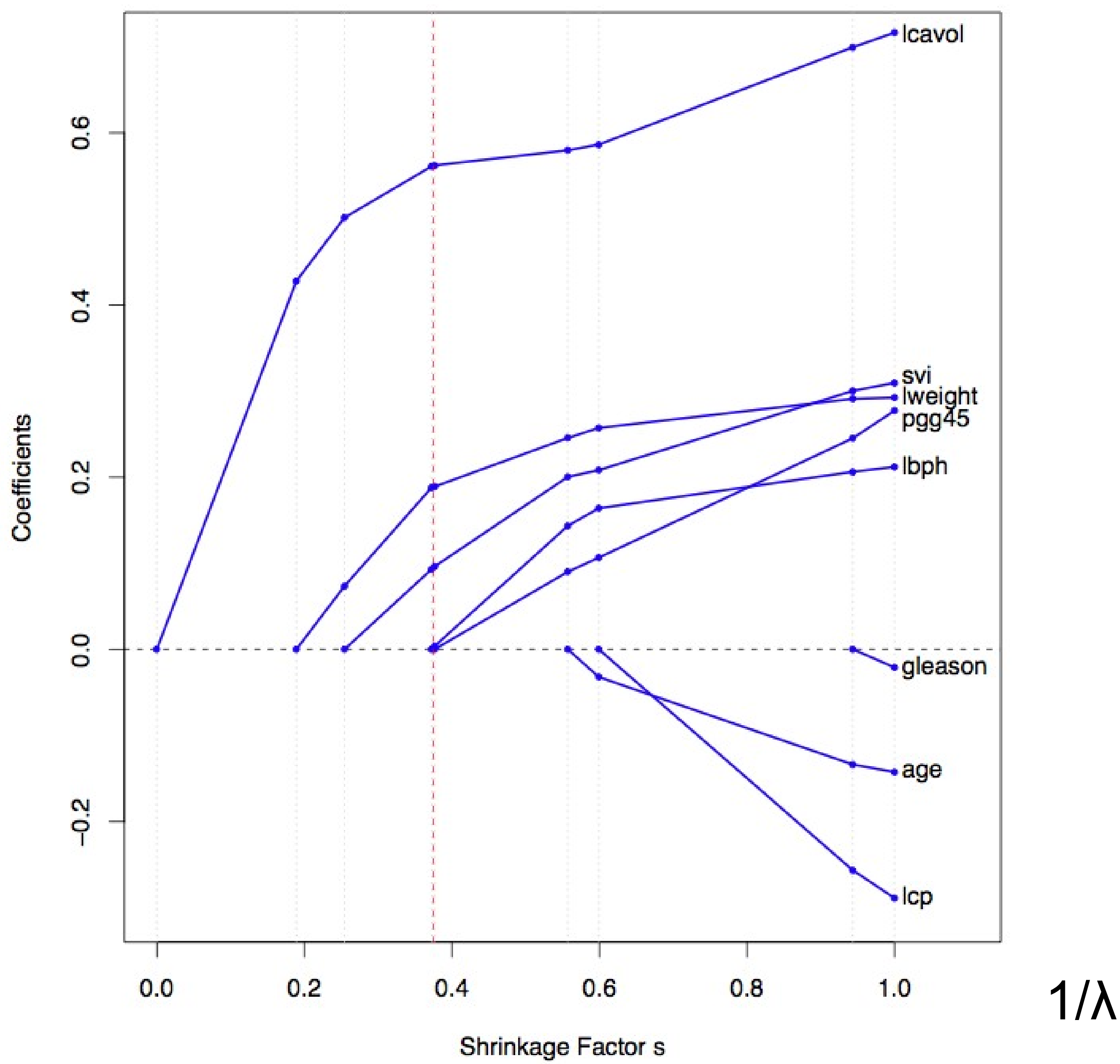
Lasso (Semester 2)

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \Big)^2$$

$$\text{subject to } \sum_{j=1}^{p} |\beta_j| \leq t.$$

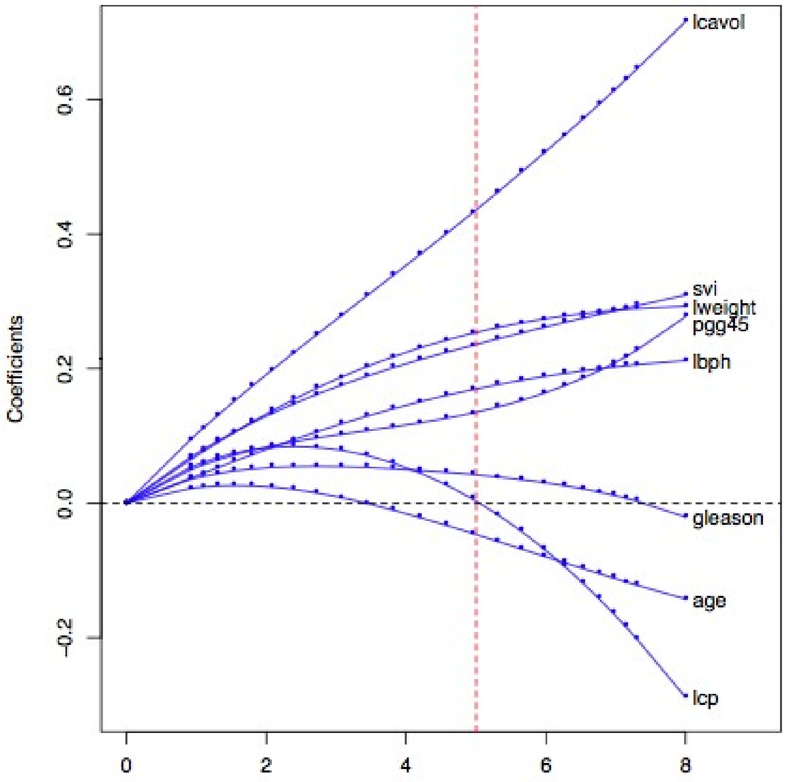
$$\hat{\beta}^{\text{lasso}} =$$

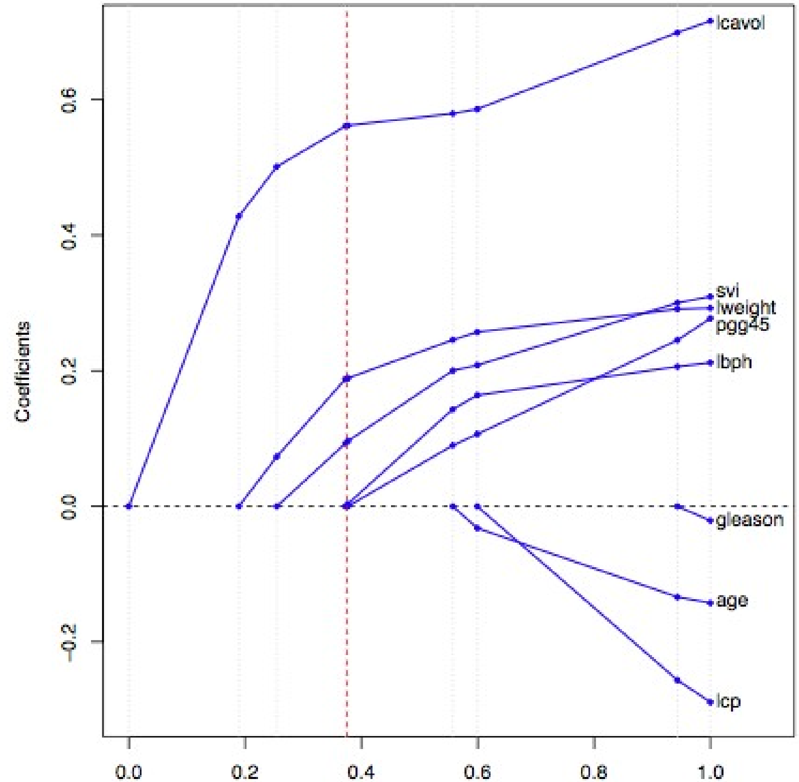
$$\underset{\beta}{\operatorname{argmin}} \Big\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \Big\}$$


lcavol
svi
lweight
pgg45
lbph
gleason
age
lcp
Coefficients
-0.2
0.0
0.2
0.4
0.6
0.0
0.2
0.4
0.6
0.8
1.0
Shrinkage Factor s
1/λ

Ridge regression

Lasso

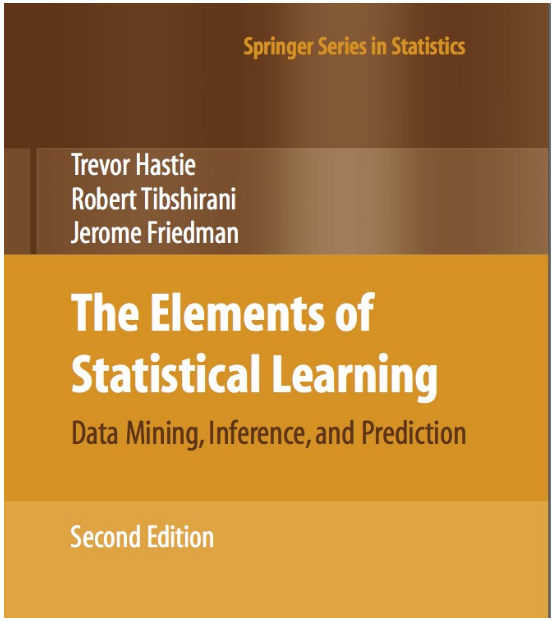




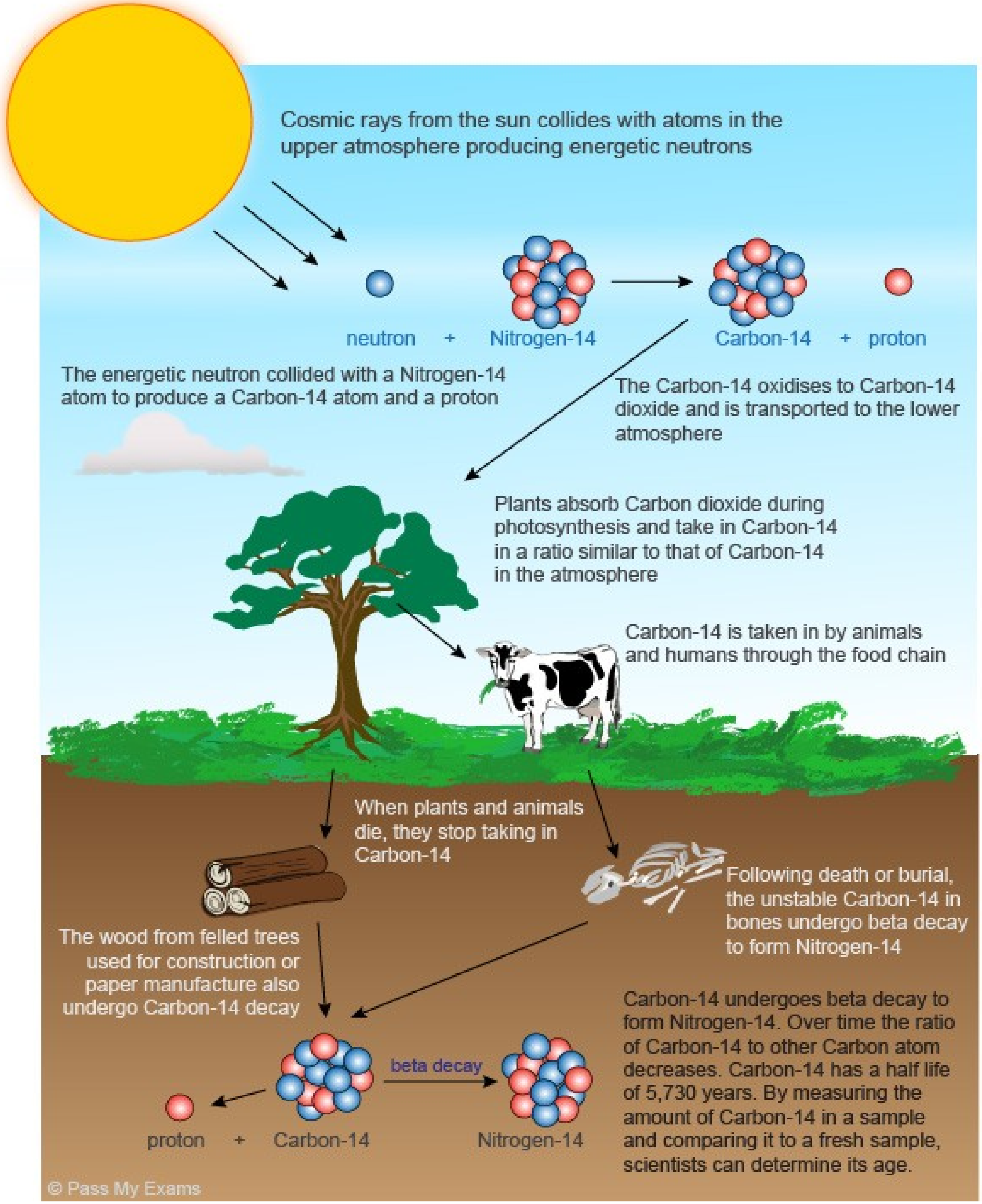
1/λ

Classical Lasso

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \left\{ (\mathbf{y} - \mathbf{X}\mathbf{w})^2 + \lambda \sum_{j=1}^{J} |w_j| \right\}$$



Methodology covered in detail in Semester 2 of the Statistics stream



Cosmic rays from the sun collides with atoms in the upper atmosphere producing energetic neutrons
neutron + Nitrogen-14
Carbon-14 + proton
The energetic neutron collided with a Nitrogen-14 atom to produce a Carbon-14 atom and a proton
The Carbon-14 oxidises to Carbon-14 dioxide and is transported to the lower atmosphere
Plants absorb Carbon dioxide during photosynthesis and take in Carbon-14 in a ratio similar to that of Carbon-14 in the atmosphere
Carbon-14 is taken in by animals and humans through the food chain
When plants and animals die, they stop taking in Carbon-14
Following death or burial, the unstable Carbon-14 in bones undergo beta decay to form Nitrogen-14
The wood from felled trees used for construction or paper manufacture also undergo Carbon-14 decay
Carbon-14 undergoes beta decay to form Nitrogen-14. Over time the ratio of Carbon-14 to other Carbon atom decreases. Carbon-14 has a half life of 5,730 years. By measuring the amount of Carbon-14 in a sample and comparing it to a fresh sample, scientists can determine its age.
beta decay
proton + Carbon-14
Nitrogen-14
© Pass My Exams

Nonparametric regression models - The problem

Radiocarbon data

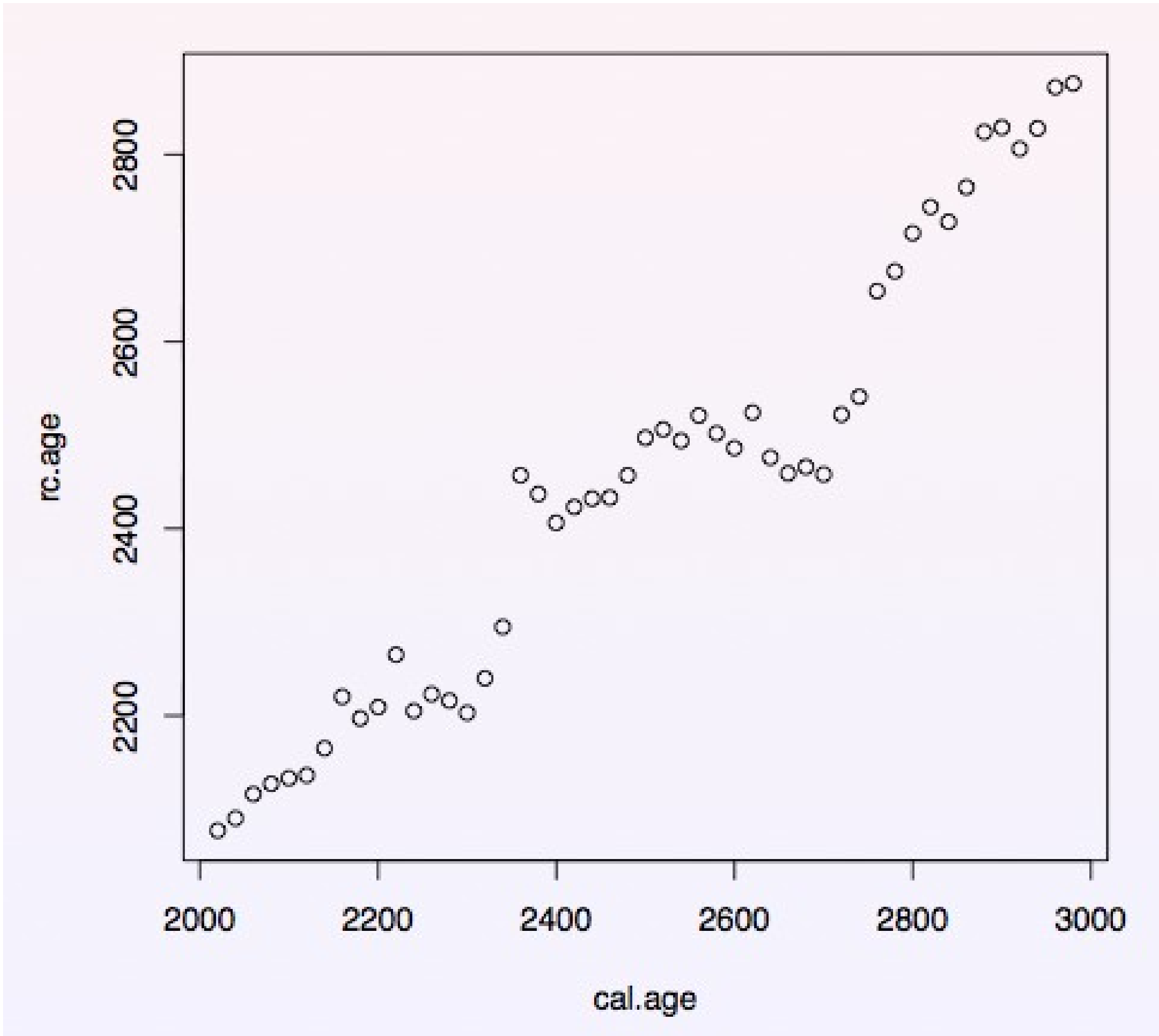
- high precision measurements of radiocarbon on Irish oak
- used to construct a calibration curve.

The variables are:

Rc.age: age from the radiocarbon dating process

Precision: a measure of precision of the dating process

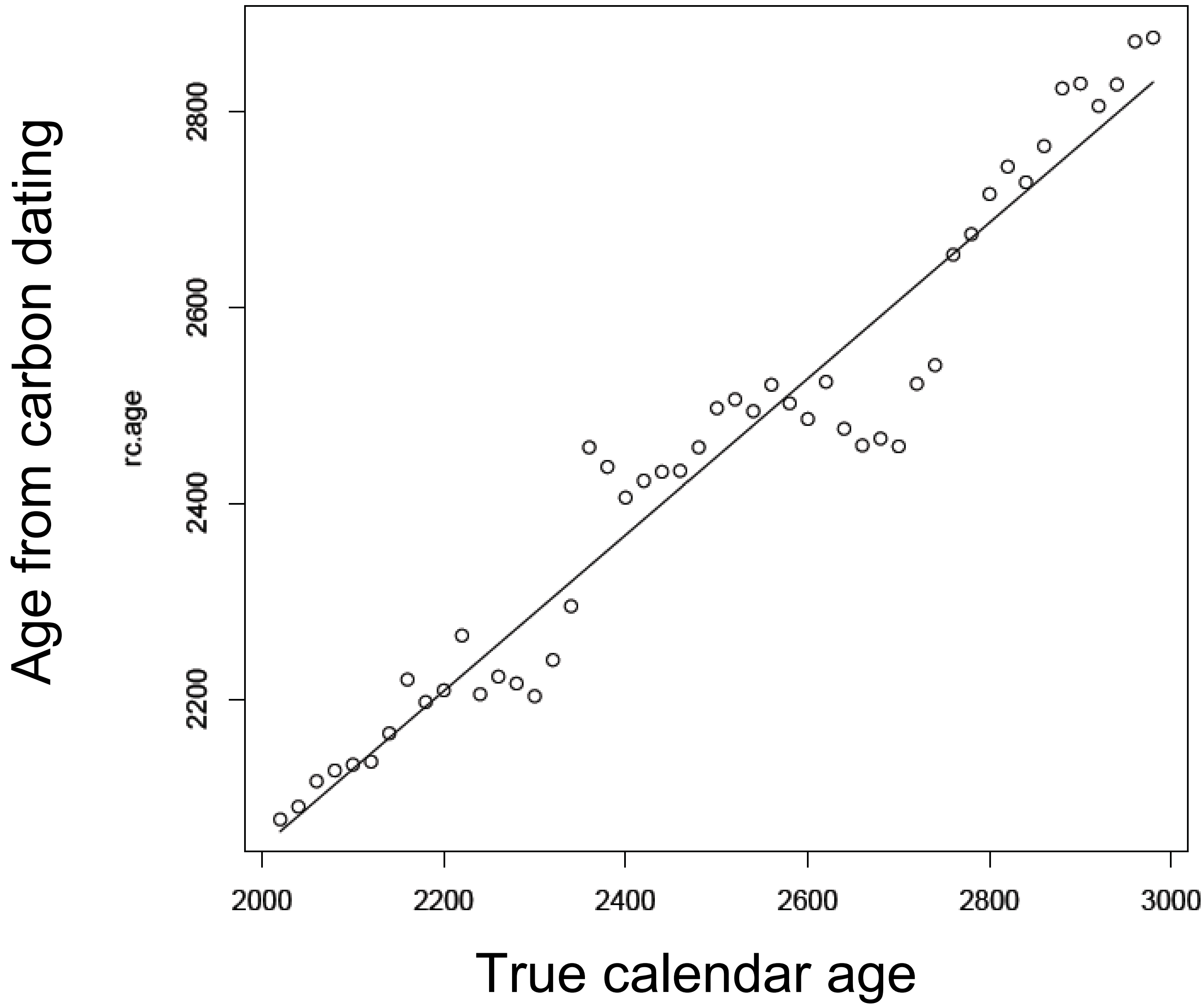
Cal.age: true calendar age

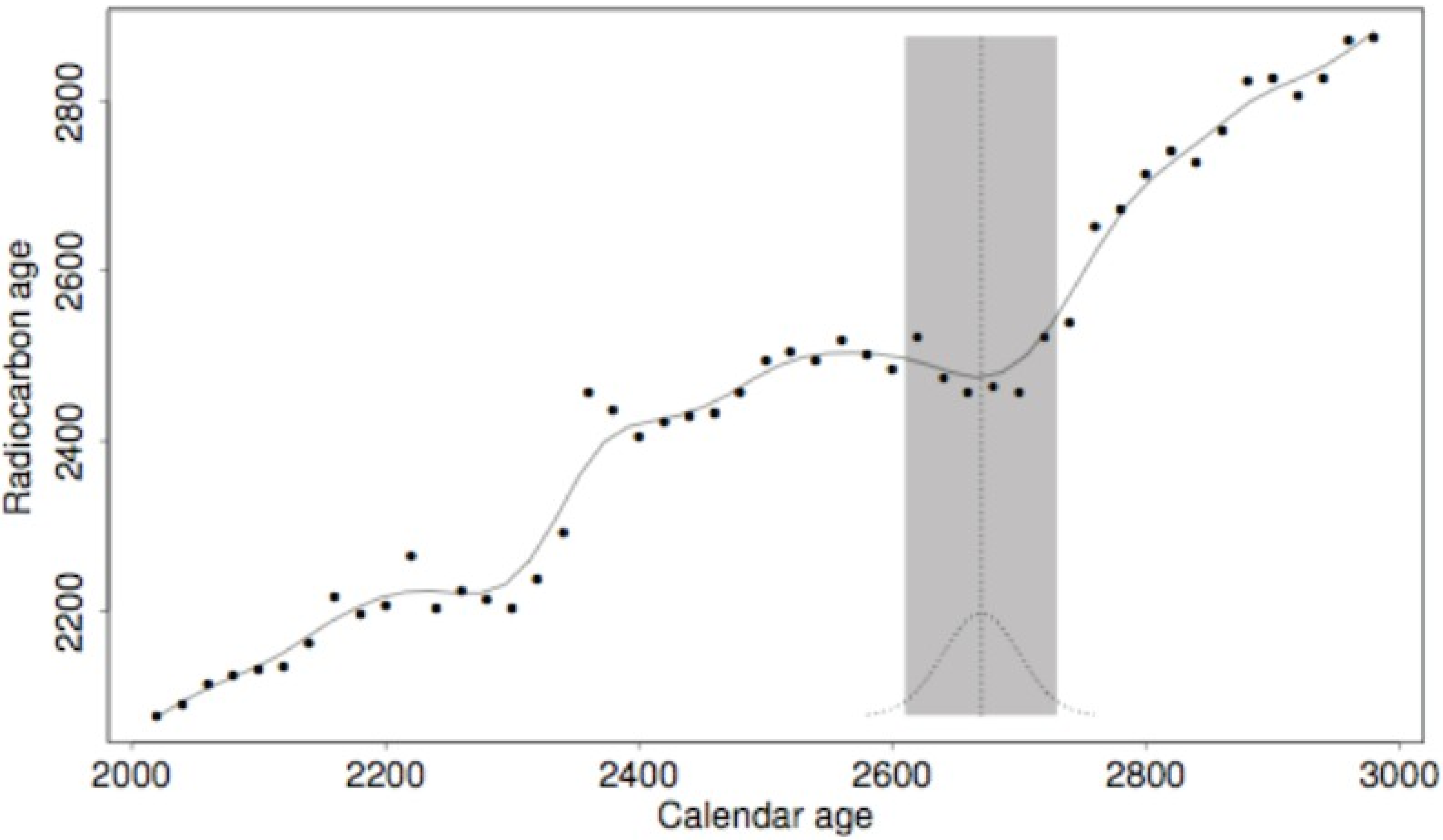


Age from carbon dating

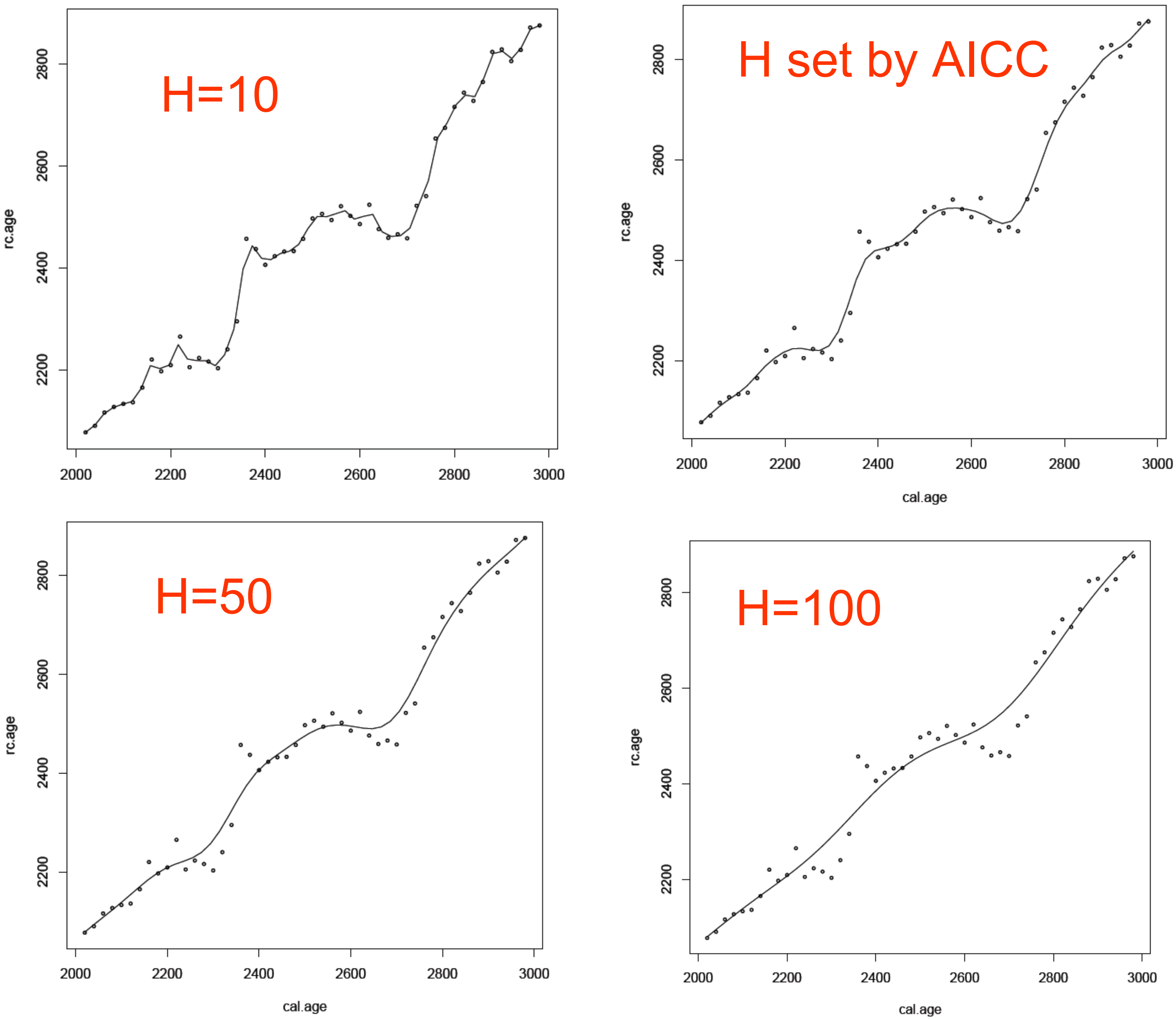
True calendar age

Linear interpolation





2200
2400
2600
2800
Radiocarbon age
2000
2200
2400
2600
2800
3000
Calendar age


H=10
H set by AICC
H=50
H=100
rc.age
cal.age
2000
2200
2400
2600
2800
3000

Embedding population dynamics models in inference

AIM

A generalized methodology for defining and fitting matrix population models that accommodates process variation (demographic and environmental stochasticity), observation error and model uncertainty

States

We categorize animals by their state, and represent the population as numbers of animals by state.

Examples of factors that determine state: age; sex; size class; genotype; sub-population (metapopulations); species (e.g. predator-prey models, community models).

States

Suppose we have *m* states at the start of year *t*. Then numbers of animals by state are:

$$\mathbf{n}_t = \begin{bmatrix} n_{1,t} \\ n_{2,t} \\ n_{3,t} \\ \vdots \\ n_{m,t} \end{bmatrix}$$

NB: These numbers are unknown!

The BAS model

$$\mathrm{E}(\mathbf{n}_{t+1} \mid \mathbf{n}_t, \boldsymbol{\theta}) = \mathbf{BASn}_t$$

where

$$\mathbf{B} = \begin{bmatrix} \lambda_2 & \lambda_3 & \cdots & \lambda_m \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 1 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} \phi_1 & 0 & \cdots & 0 \\ 0 & \phi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_m \end{bmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\varphi} \end{pmatrix}$$

Leslie matrix

The product **BAS** is a Leslie projection matrix:

$$\mathbf{BAS} = \begin{bmatrix} \phi_1\lambda_2 & \phi_2\lambda_3 & \cdots & \phi_{m-1}\lambda_m & \phi_m\lambda_m \\ \phi_1 & 0 & \cdots & 0 & 0 \\ 0 & \phi_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \phi_{m-1} & \phi_m \end{bmatrix}$$

Observation equation

$$E(\mathbf{y}_t \mid \mathbf{n}_t, \boldsymbol{\theta}) = \mathbf{O}_t \mathbf{n}_t$$

e.g. metapopulation with two sub-populations,
each split into adults and young,
unbiased estimates of total abundance
of each sub-population available:

$$E(\mathbf{y}_t) = \begin{bmatrix} E(y_{1,t}) \\ E(y_{2,t}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} n_{01,t} \\ n_{11,t} \\ n_{02,t} \\ n_{12,t} \end{bmatrix}$$

Fitting models to time series of data

- Kalman filter
  Normal errors, linear models
  or linearizations of non-linear models

- Markov chain Monte Carlo

- Sequential Monte Carlo methods

Elements required for Bayesian inference

$g(\boldsymbol{\theta})$	Prior for parameters
$g_0(\mathbf{n}_0 \mid \boldsymbol{\theta})$	pdf (prior) for initial state
$g_t(\mathbf{n}_t \mid \mathbf{n}_{t-1},\ldots,\mathbf{n}_0,\boldsymbol{\theta})$	pdf for state at time *t* given earlier states
$f_t(\mathbf{y}_t \mid \mathbf{n}_t,\boldsymbol{\theta})$	Observation pdf

Bayesian inference

Joint prior for $\boldsymbol{\theta}$ and the $\mathbf{n}_t$:

$$g(\boldsymbol{\theta}) \times g_0(\mathbf{n}_0 \mid \boldsymbol{\theta}) \times \prod_{t=1}^{T} g_t(\mathbf{n}_t \mid \mathbf{n}_{t-1}, \ldots, \mathbf{n}_0, \boldsymbol{\theta})$$

Likelihood:

$$\prod_{t=1}^{T} f_t(\mathbf{y}_t \mid \mathbf{n}_t, \boldsymbol{\theta})$$

Posterior:

$$g(\mathbf{n}_0, \ldots, \mathbf{n}_T, \boldsymbol{\theta} \mid \mathbf{y}_1, \ldots, \mathbf{y}_T) = \frac{g(\boldsymbol{\theta}) \times g_0(\mathbf{n}_0 \mid \boldsymbol{\theta}) \times \prod_{t=1}^{T} \left\{ g_t(\mathbf{n}_t \mid \mathbf{n}_{t-1}, \ldots, \mathbf{n}_0, \boldsymbol{\theta}) \times f_t(\mathbf{y}_t \mid \mathbf{n}_t, \boldsymbol{\theta}) \right\}}{f(\mathbf{y}_1, \ldots, \mathbf{y}_T)}$$

Generalizing the framework

$g(\mathbf{M})$	Model prior
$g(\boldsymbol{\theta} \mid \mathbf{M})$	Prior for parameters
$g_0(\mathbf{n}_0 \mid \boldsymbol{\theta}, \mathbf{M})$	pdf (prior) for initial state
$g_t(\mathbf{n}_t \mid \mathbf{n}_{t-1}, \ldots, \mathbf{n}_0, \boldsymbol{\theta}, \mathbf{M})$	pdf for state at time *t* given earlier states
$f_t(\mathbf{y}_t \mid \mathbf{n}_t, \boldsymbol{\theta}, \mathbf{M})$	Observation pdf

Generalizing the framework

Replace $$\mathrm{E}(\mathbf{n}_{t+1} \mid \mathbf{n}_t, \boldsymbol{\theta}) = \mathbf{P}_t \mathbf{n}_t$$

by $$\mathbf{n}_{t+1} = \mathbf{P}_t(\mathbf{n}_t)$$

where $$\mathbf{P}_t(\mathbf{n}_t) = \mathbf{P}_{K,t}(\mathbf{P}_{K-1,t}(\cdots \mathbf{P}_{1,t}(\mathbf{n}_t)\cdots))$$

and $\mathbf{P}_{k,t}(\cdot)$ is a possibly random operator

The Scottish Mathematical Sciences Training Centre

Statistics stream

Aims

Introduce some key topics which lie at the heart of research in statistical methods.

The intention is not to be a comprehensive study of all the most advanced statistical techniques available (that would be somewhat difficult in approximately 40 hours!) but to present some key concepts that form a basis for more advanced and sophisticated ideas.

NB: for statistics students, more advanced and more specific skills are taught in APTS workshops:
https://www2.warwick.ac.uk/fac/sci/statistics/apts/

Aims

Develop good computational skills using R.

R is a very powerful statistical computing environment with an extensive suite of libraries/packages and one of the main platforms for statistical research both in academia and industry throughout the world. Knowledge of R is likely to be very useful whetever your PhD topic may be.

The Scottish Mathematical Sciences Training Centre

Regression and Simulation Methods

Prerequisites

Basic concepts in:

(i) probability (elementary probability distributions);

(ii) statistics (ideas of estimation, confidence intervals, hypothesis tests); and

(iii) calculus.

The level required in these areas would usually be provided in a first undergraduate course.

The prerequisite for Modern Regression and Bayesian Methods is Regression and Simulation Methods (or equivalent)

The Scottish Mathematical Sciences Training Centre

Statistics stream

Method of delivery

As standard each lecture will be a total of 2 hours (including a tea/coffee break!) on Tuesday from 1.00-3.00.

Questions are encouraged during lectures and there will also be opportunities to discuss particular issues that arise within lectures and/or associated exercises.

The first half of semester 1 (Regression and Simulation Methods) will be run as an online video course. It covers what for most will be revision. We ask you to check the material covered. If any of it is unfamiliar, you can view the relevant lectures, and attempt the related tutorial questions. Tutorial help will be arranged locally.

Assessments

Short projects after each block of 5 sessions.

These will be marked and individual feedback provided.

The Scottish Mathematical Sciences Training Centre

Statistics stream

Stream outline

Regression and simulation methods

Introduction to R
Review of linear models
Likelihood and optimisation

Review of generalised linear models (GLMs)
Simulation and bootstrapping
Case study

Modern regression and Bayesian methods

Random effects models
Modern regression
Case study

An Introduction to Markov chain Monte Carlo (MCMC) Methods

Regression and simulation methods

Syllabus

Introduction to R

1 Data structures and types; standard plotting facilities; elementary statistical functions; distributions within R; simple control structures; simple example of writing a function; a taster of more sophisticated facilities.

The Scottish Mathematical Sciences Training Centre

Regression and simulation methods

Syllabus

Review of linear models

2 Basic results on estimation, confidence intervals and tests within the linear model; model checking; the use of factors; fitting linear models in R.

3 The analysis of simple designed experiments; case studies of linear models

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j$$

Likelihood and optimisation

4 Likelihood principles and key distributional results; examples of likelihood fitting and analysis; Newton's method for optimisation.

5 Plotting and inspection of two-parameter likelihoods; more general methods of optimisation of multiparameter functions; implementation in R.

The Scottish Mathematical Sciences Training Centre

Regression and simulation methods

Syllabus

Review of GLMs

6 Exponential family, with examples for standard distributions (e.g. normal, gamma, Binomial, Poisson, Negative Binomial); link functions; examples.

7 Iteratively weighted least squares; model fitting within R, including function `glm`; case studies.

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j$$

Nonlinear transformation

Regression and simulation methods

Syllabus

Simulation and bootstrapping

8 Non-parametric bootstrap for calculating standard errors; confidence intervals (percentile intervals); implementing the bootstrap within R.

9 Parametric bootstrap, simulation methods and implementation in R; examples (e.g. linear regression).

Case study

10 This session will be constructed around real scientific studies where statistical methods were central to the solution of the problem of interest. The methods required will involve some of the techniques discussed earlier in the course. However, some further techniques will be introduced, as required by the analysis.

Random effects models

11 A summary of methods for linear mixed effects models as in Pinheiro & Bates; case studies.

12 A summary of methods for non-linear mixed effects models as in Pinheiro & Bates; case studies.

Syllabus

Modern regression

13 Density estimation; different methods of nonparametric regression with one and two covariates; bandwidth selection; examples of use.

14 Additive models; the backfitting algorithm; generalized additive models; examples.

An Introduction to MCMC Methods

16 Introduction to Bayesian methods, prior specification, posterior distribution, summary statistics, prior sensitivity, marginal distributions; underlying idea behind Markov chain Monte Carlo.

17 Metropolis-Hastings algorithm; Gibbs sampler; issues of convergence; length of burn-in; mixing properties; tuning parameters.

18 Introduction to WinBUGS; basic examples to demonstrate previous principles.

19 Coding MCMC simulations within R; further examples.

20 Introduction to advanced topics, for example, the use of auxiliary variables (e.g. random effects), missing data, model selection; WinBUGS/R.

Case study

This session will be constructed around real scientific studies where statistical methods were central to the solution of the problem of interest. The methods required will involve some of the techniques discussed earlier in the course. However, some further techniques will be introduced, as required by the analysis.