Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

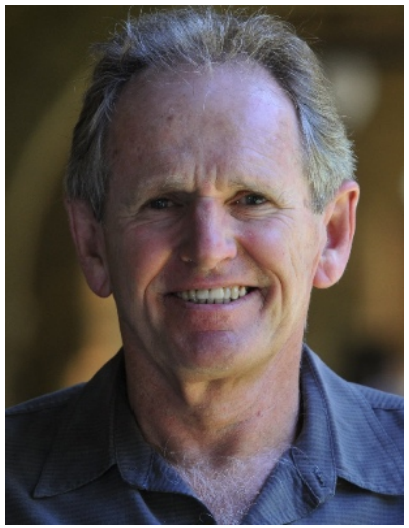
IFCS 2015, Bologna



Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

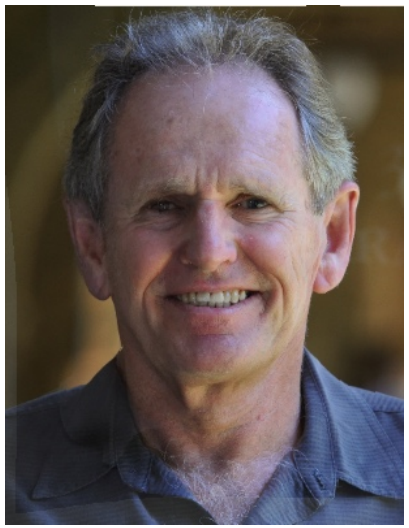
IFCS 2015, Bologna



Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

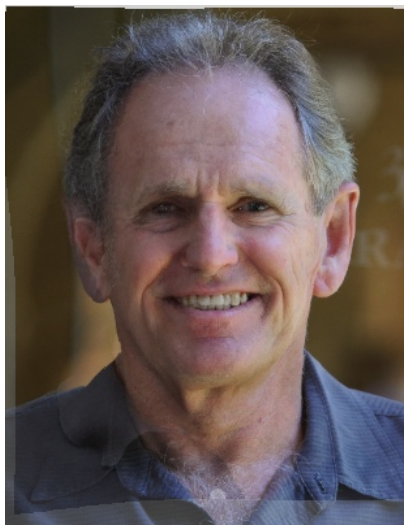
IFCS 2015, Bologna



Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

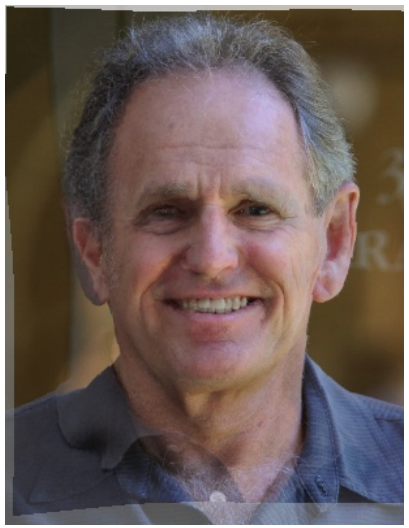
IFCS 2015, Bologna



Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

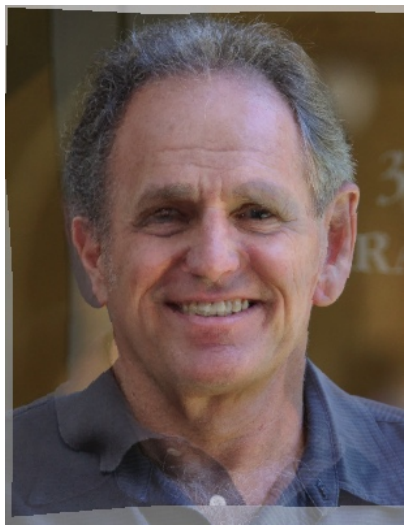
IFCS 2015, Bologna



Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

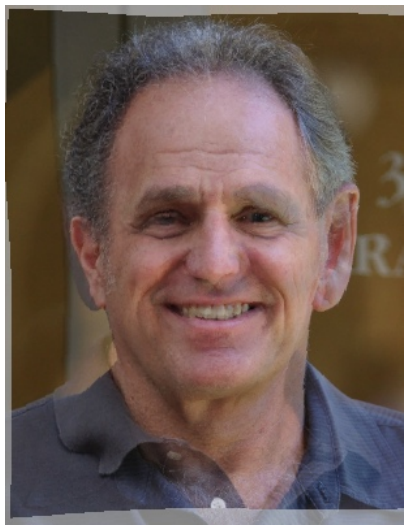
IFCS 2015, Bologna



Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

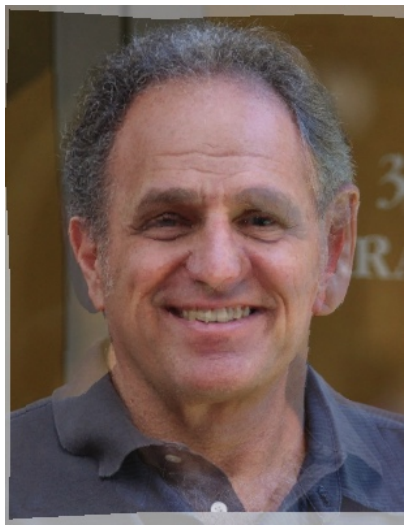
IFCS 2015, Bologna



Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

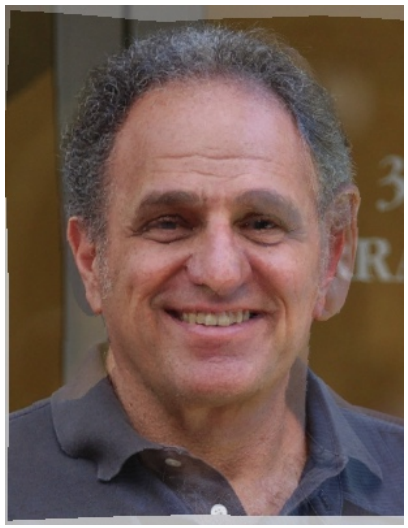
IFCS 2015, Bologna



Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

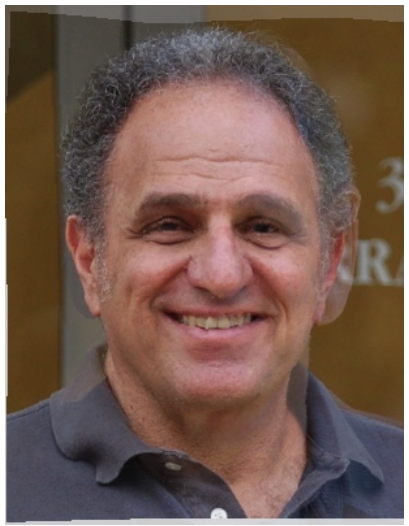
IFCS 2015, Bologna



Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

IFCS 2015, Bologna



Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

IFCS 2015, Bologna



Generalized additive models: a retrospective

Robert Tibshirani, Stanford University

IFCS 2015, Bologna



It all started around 1984



Background

- Hastie came from South Africa to Stanford in 1980 to do his PhD. He had worked at MRC in London in the interim
- Tibshirani came from Toronto to Stanford in 1981 to do his PhD. He had done a Masters in Stat at U of Toronto.
- Both had been strongly influenced by British Statistics- generalized linear models - Nelder, Wedderburn, McCullagh; the `glim` package
- For me, David Andrews from University of Toronto was a big influence.

Trevor arrives at Stanford



Sequioa hall at Stanford



1927



1981

Sequioa hall today



Stanford Statistics faculty 1980-81

Ted Anderson, Charles Stein, Ingram Olkin, Brad Efron,

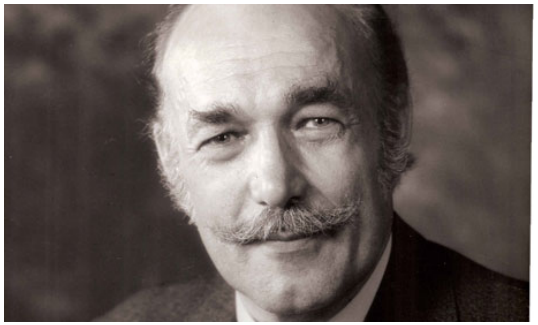
Thomas Cover, Rupert Miller, Paul Switzer, David Siegmund,

Lincoln Moses, Herb Solomon, Persi Diaconis, Jerome Friedman,

Iain Johnstone, Werner Stuetzle

Nelder and Wedderburn- GLMs

John Nelder 1924-2010





Robert William Maclagan Wedderburn 1947-1975

"His colleagues remember him as someone of engaging diffidence, who would nonetheless hold his own in argument when he was sure he was right (as he usually was)." - John Nelder

More on John Nelder

- While leading research at Rothamsted Experimental Station, Nelder developed and supervised the updating of the statistical software packages GLIM and GenStat: Both packages are flexible high-level programming languages that allow statisticians to formulate linear models concisely.
- GLIM influenced later environments for statistical computing such as S-PLUS and R
- In response-surface optimization, Nelder and Roger Mead proposed the Nelder-Mead simplex heuristic, widely used in engineering and statistics.
- He was responsible, with Max Nicholson and James Ferguson-Lees, for debunking the Hastings Rarities – sightings of a series of rare birds, preserved by a taxidermist and provided with bogus histories.

The GLM paper

J. R. Statist. Soc. A,
(1972), **135**, *Part* 3, p. 370

Generalized Linear Models

By J. A. NELDER and R. W. M. WEDDERBURN

Rothamsted Experimental Station, Harpenden, Herts

SUMMARY

The technique of iterative weighted linear regression can be used to obtain maximum likelihood estimates of the parameters with observations distributed according to some exponential family and systematic effects that can be made linear by a suitable transformation. A generalization of the analysis of variance is given for these models using log-likelihoods. These generalized linear models are illustrated by examples relating to four distributions; the Normal, Binomial (probit analysis, etc.), Poisson (contingency tables) and gamma (variance components).

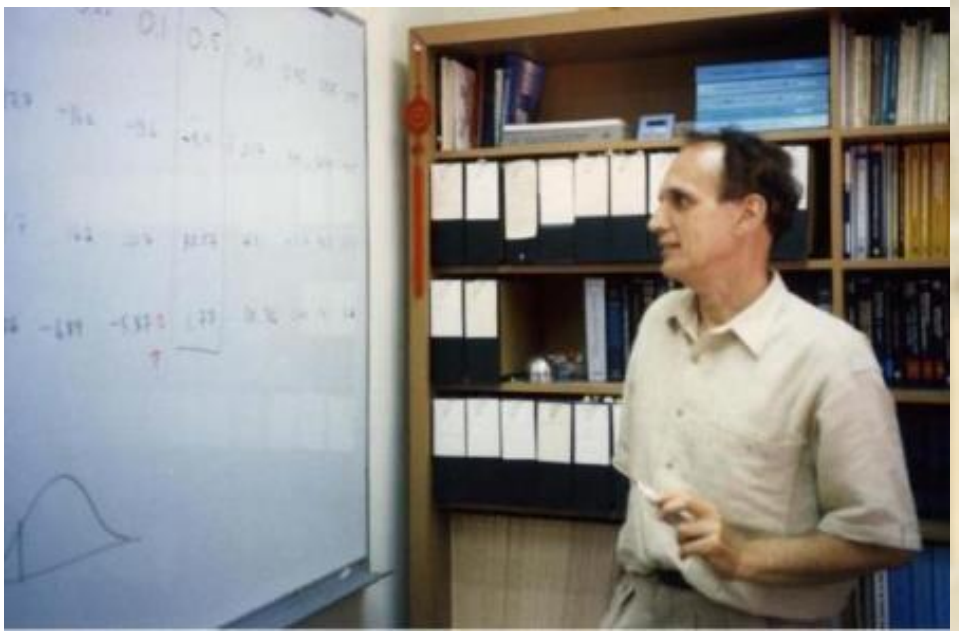
The implications of the approach in designing statistics courses are discussed.

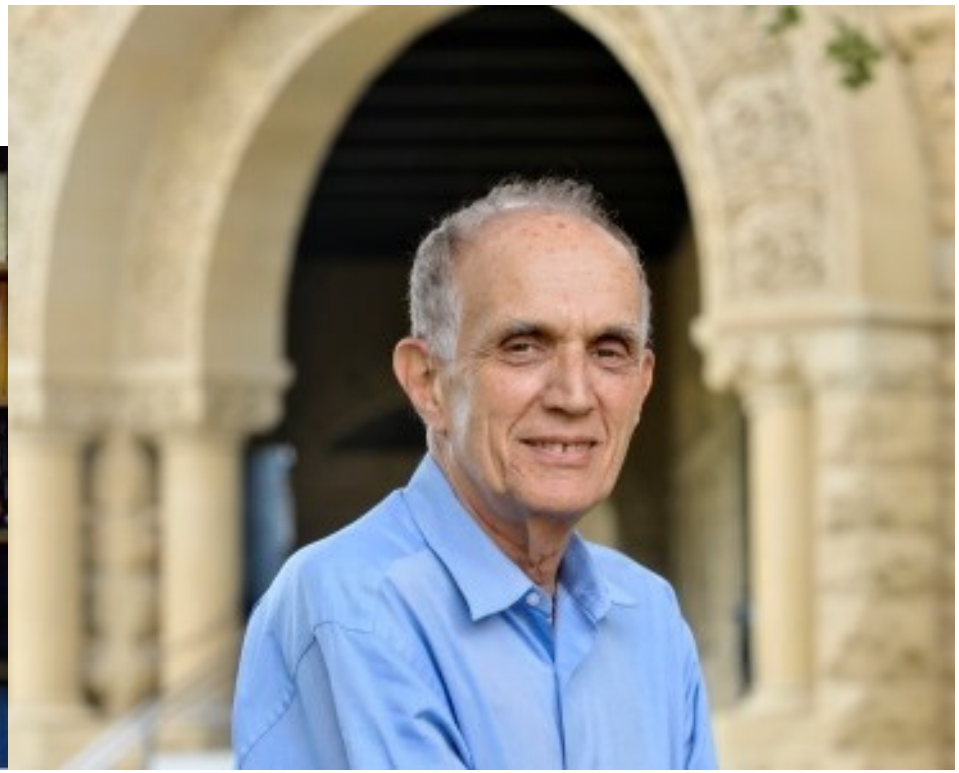
Keywords: ANALYSIS OF VARIANCE; CONTINGENCY TABLES; EXPONENTIAL FAMILIES; INVERSE POLYNOMIALS; LINEAR MODELS; MAXIMUM LIKELIHOOD; QUANTAL RESPONSE; REGRESSION; VARIANCE COMPONENTS; WEIGHTED LEAST SQUARES

Our experience at Stanford

- We were both taught and strongly influenced by Andreas Buja, Brad Efron, Jerome Friedman and Werner Stuetzle. Werner was Trevor's advisor; Brad was mine.
- Werner taught an applied statistics course to Trevor in 1980. Andreas did the same for me in 1981. These were amazing classes that changed he way that we thought about the subject! These class notes formed the kernel for our text "Elements of Statistical Learning" published almost 20 years later.

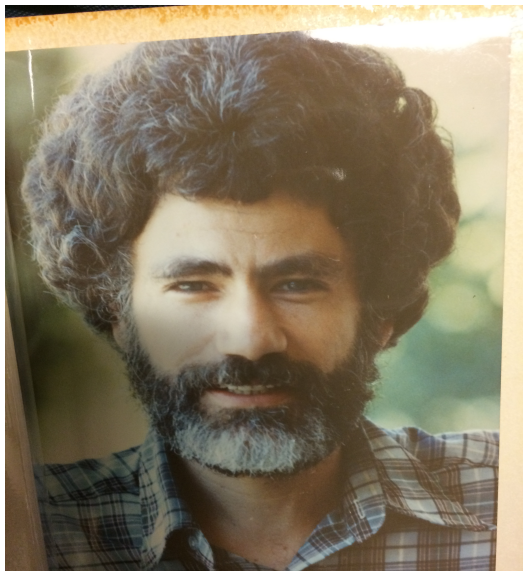
Our mentors





Brad Efron

Our mentors- continued





Jerry Friedman

Our mentors- continued





Andreas Buja and Werner Stuetzle

Smoothing meets GLMs

- We knew about GLMs from our past (few others at Stanford were familiar with them). Buja, Friedman, Stuetzle taught us about smoothing of scatterplots. How could we combine the two to obtain a more flexible model?
- I tried one approach in my PhD dissertation : "Local likelihood estimation"— in which a GLM is fit in a local neighborhood of each x. Developed with Trevor, published in JASA
- Approach was somewhat clumsy — there was no overall objective function. We sought something better

What is a Generalized Additive Model?

- Response variable Y from an exponential family distribution with mean μ. Predictors $X_1, X_2, \ldots X_p$.
- model $g(\mu) = \beta_0 + \sum_j f_j(X_j)$ where g is a known link function. Functions f_j are unspecified and may be modeled e.g. as cubic splines.

Smoothing meets GLMs- continued

- We were writing out the local scoring (Newton-Raphson) equations for GLMs and realized that one could replace the weighted-least-squares step by weighted smoothers.
- Excitement! We coded it up (in Mortran?) and it seems to work. We rushed to write it up

GENERALIZED ADDITIVE MODELS

BY

TREVOR J. HASTIE
AND
ROBERT J. TIBSHIRANI

TECHNICAL REPORT NO. 98
OCTOBER 1984

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT 5 R01 GM21215-10

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



GLM conference 1985

We travelled to "beautiful" Lancaster England for the 1985 GLM conference. We met many luminaries- John Nelder, Peter Diggle, Peter Green, Bernard Silverman, Murray Aitken, Bent Jorgensen, Joe Whittaker. Afraid we were going to get scooped ... but survived.

Lecture Notes in Statistics

Edited by D. Brillinger, S. Fienberg, J. Gani,
J. Hartigan, and K. Krickeberg

32

Generalized Linear Models

Proceedings of the GLIM 85 Conference,
held in Lancaster, UK, Sept. 16–19, 1985

Edited by R. Gilchrist, B. Francis and J. Whittaker



Springer-Verlag
Berlin Heidelberg New York Tokyo

GLIM 85: CONTENTS

INTRODUCTION: GLIM AND GENERALIZED LINEAR MODELS 1
R Gilchrist.

GLIM4 - DIRECTIONS FOR DEVELOPMENT 6
M Aitkin.

SIMULTANEOUS EQUATION SYSTEMS WITH CATEGORICAL OBSERVED VARIABLES 15
G Arminger and U Kusters.

THE NULL EXPECTED DEVIANCE FOR AN EXTENDED CLASS OF GENERALIZED LINEAR MODELS 27
G M Cordeiro.

COMPARING ESTIMATED SPECTRAL DENSITIES USING GLIM 35
P J Diggle.

SEMI-PARAMETRIC GENERALIZED LINEAR MODELS 44
P J Green and B S Yandell.

AN ALGORITHM FOR DEGREE OF FREEDOM CALCULATIONS IN SPARSE COMPLETE CONTINGENCY TABLES 56
S Haslett.

GENERALIZED ADDITIVE MODELS; SOME APPLICATIONS 66
T Hastie and R Tibshirani.

"Generalized additive models"

- the name (whose idea?) was really important.
- Etymology: From Italian gamba (?leg?) (slang) A person's leg, especially an attractive woman's leg.





"Generalized additive models"

- published in Statistical Science in 1986
- Hastie move to Bell Labs in 1986, I moved to University of Toronto the year before. We met back at Stanford for the summer of 1990 wiring internet in the unfinished basement of Sequioa Hall
- We remained friends and collaborators: I visited Bell Labs for a week or two every year
- This new thing called "email" became quite helpful in our collaboration around 1986 (it would arrive overnight)

Linear smoother and additive models

- 1989 paper with Buja laid out some "theory" for additive models, including conditions for convergence of the backfitting algorithm
- Some of these ideas were around earlier in the 1981 "Gifi" book from University of Leiden.
- With this 1989 paper in place, we felt that the topic had enough "meat" to warrant a monograph on the subject. We published "Generalized Additive Models" in 1990.

Progress since that time

- Mostly by others: we published varying coefficient models in 1993; Bayesian backfitting in 2000
- Software!!— Splus GLM and GAM functions (Chambers and Hastie). Later ported to R.
- we made the `gamfit` program for PCs in the late 1980's, shipping it by mail on a floppy.
- Notable work since then: *P-splines* (1996) Eilers and Marx; Generalized Additive Models: An Introduction with R (2006) Simon Wood — `mgcv` R package.
- Lots of related ideas: Friedman's MARS algorithm, which incorporates interactions; AdaBoost, gradient boosting...

[BOOK] Generalized additive models

TJ Hastie, RJ Tibshirani - 1990 - books.google.com
This book describes an array of power tools for data analysis that are based on nonparametric regression and smoothing techniques. These methods relax the linear assumption of many standard models and allow analysts to uncover structure in the data ...
Cited by 11076 Related articles All 19 versions Cite Save More

Generalized additive models

T Hastie, R Tibshirani - Statistical science, 1986 - JSTOR
Likelihood-based regression models such as the normal linear regression model and the linear logistic model, assume a linear (or some other parametric) form for the covariates X 1, X 2,⋯, X p. We introduce the class of
Cited by 1440 Related articles All 22 versions Cite Save More

[BOOK] Generalized additive models: an introduction with R

S Wood - 2006 - books.google.com
Now in widespread use, **generalized additive models** (GAMs) have evolved into a standard statistical methodology of considerable flexibility. While Hastie and Tibshirani's outstanding 1990 research monograph on GAMs is largely responsible for this, there has been a long- **...**
Cited by 3950 Related articles All 13 versions Cite Save More

Generalized linear and generalized additive models in studies of species distributions: setting the scene

A Guisan, TC Edwards, T Hastie - Ecological modelling, 2002 - Elsevier
An important statistical development of the last 30 years has been the advance in regression analysis provided by **generalized** linear **models** (GLMs) and **generalized additive models**
Cited by 974 Related articles All 17 versions Cite Save

On the use of generalized additive models in time-series studies of air pollution and health

F Dominici, A McDermott, SL Zeger… - American journal of …, 2002 - Oxford Univ Press
Abstract The widely used **generalized additive models** (GAM) method is a flexible and effective technique for conducting nonlinear regression analysis in time-series studies of the
Cited by 553 Related articles All 8 versions Cite Save

Stable and efficient multiple smoothing parameter estimation for generalized additive models

SN Wood - Journal of the American Statistical Association, 2004 - amstat.tandfonline.com
Representation of **generalized additive models** (GAM's) using penalized regression splines allows GAM's to be employed in a straightforward manner using penalized regression

Incorporating random effects

- Lin and Zhang (1999). Generalized additive mixed models. JRSSB
- Rigby and Stasinopoulos 2005). Generalized additive models for location, scale and shape. JRSSC (with discussion)
- Simon Wood's GAMM R package

What have we been doing since then?

- in large part, our focus has been on sparsity, beginning with the lasso paper in 1996
- There are a lot of parallels with GAMs: an additive model assumes a kind of sparsity, in an appropriate function space. Trevor will give more details in his talk
- Some proposals like COSSO and SpAM combine additive models and sparsity. Again, Trevor will give details
- The glmnet package for sparse modeling uses coordinate decent as its bread and butter. The backfitting algorithm for GAMs is just blockwise coordinate descent (we didn't realize this at the time)

Wrapping up

- Thanks to everyone for their interest in GAMs
- Thanks to Angela Montanari and the conference organizers for inviting us