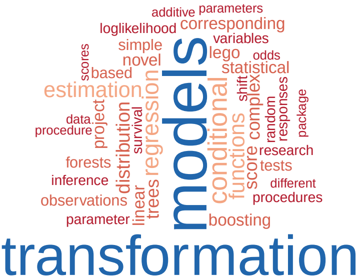
University of Zurich[UZH]

EBPI Epidemiology, Biostatistics and Prevention Institute

Transformation Models: Pushing the Boundaries

Torsten Hothorn



Inria Paris-Saclay, 2021-01-26

Transformation Models

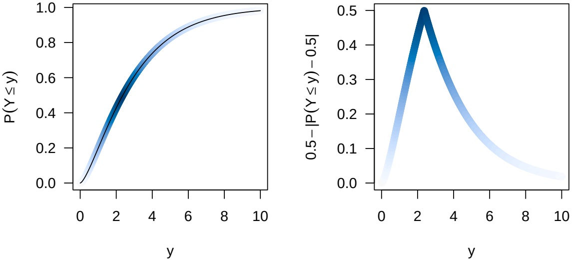
Models for Distributions, not Means

Regression:

$$\mathbb{P}(Y \le y \mid \boldsymbol{X} = \boldsymbol{x})$$

not only

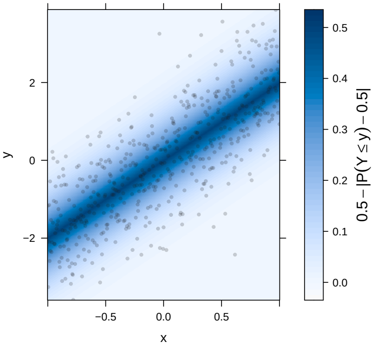
$$\mathbb{E}(Y \mid \boldsymbol{X} = \boldsymbol{x})$$



The Normal Linear Regression Model

$$Y = \alpha + \boldsymbol{x}^\top \boldsymbol{\gamma} + \sigma Z, \quad Z \sim \mathrm{N}(0, 1)$$

- everything but “normal”
- most special case
- $Y \mid \boldsymbol{x} \sim \mathrm{N}(\alpha + \boldsymbol{x}^\top \boldsymbol{\gamma}, \sigma^2)$
- no way escaping normal land



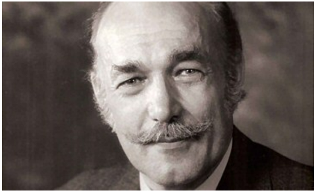
Generalisation I

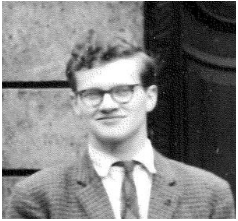
$$
\begin{aligned}
Y &= \alpha + \boldsymbol{x}^\top \gamma + \sigma Z, \quad & Z \sim \mathrm{N}(0, 1) \\
\iff \mathbb{E}(Y \mid \boldsymbol{X} = \boldsymbol{x}) &= \alpha + \boldsymbol{x}^\top \gamma \quad\quad\quad, & Y \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{N}(,) \\
\hookrightarrow g(\mathbb{E}(Y \mid \boldsymbol{X} = \boldsymbol{x})) &= \alpha + \boldsymbol{x}^\top \gamma \quad\quad\quad, & Y \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{ExpFam}(,)
\end{aligned}
$$

Generalized Linear Models

By J. A. Nelder and R. W. M. Wedderburn

Rothamsted Experimental Station, Harpenden, Herts





Generalisation II

$$
\begin{aligned}
Y &= \alpha + \boldsymbol{x}^\top \gamma + \sigma Z, \quad Z \sim \mathrm{N}(0,1) \\
\iff \frac{Y - \alpha}{\sigma} &= \boldsymbol{x}^\top \frac{\gamma}{\sigma} + Z, \quad Z \sim \mathrm{N}(0,1) \\
\iff \frac{Y - \alpha}{\sigma} &= \boldsymbol{x}^\top \boldsymbol{\beta} + Z, \quad Z \sim \mathrm{N}(0,1) \\
\hookrightarrow h(Y) &= \boldsymbol{x}^\top \boldsymbol{\beta} + Z, \quad Z \sim \mathrm{N}(0,1) \\
\hookrightarrow h(Y) &= \boldsymbol{x}^\top \boldsymbol{\beta} + Z, \quad Z \sim
\end{aligned}
$$

Transformation models, $Z \in \mathbb{R}$ with absolute continuous log-concave density f_Z, $h : \mathbb{R} \to \mathbb{R}$ nondecreasing

An Analysis of Transformations (1964)

An Analysis of Transformations

By G. E. P. Box and D. R. Cox

University of Wisconsin — *Birkbeck College, University of London*

[Read at a RESEARCH METHODS MEETING of the SOCIETY, April 8th, 1964, Professor D. V. LINDLEY in the Chair]

Conceptually more powerful but, at the time, hard to compute and thus restricted to

$$h(y \mid \lambda) = \begin{cases} \frac{y^{\lambda}+1}{\lambda} & \lambda > 0 \\ \log(y) & \lambda = 0 \end{cases} \quad \text{with } Z \sim \mathrm{N}(0, \sigma^2)$$

“Box-Cox” power transformation

Conditional Distribution Functions

$$
\begin{aligned}
h(Y) &= \boldsymbol{x}^\top \beta + Z \\
\hookrightarrow \mathbb{P}(Y \le y \mid \boldsymbol{X} = \boldsymbol{x}) &= F_Z(h(y) - \boldsymbol{x}^\top \beta)
\end{aligned}
$$

also allows discrete models via step-function h

Linear transformation models: Proportional hazards, proportional odds, ...

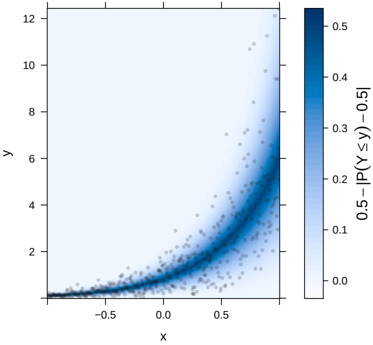
aka Probabilistic index models:

$$
\mathbb{P}(Y_1 < Y_2 \mid \boldsymbol{x}_1, \boldsymbol{x}_2) = m_Z((\boldsymbol{x}_1 - \boldsymbol{x}_2)^\top \beta)
$$

Cox Models

$$\mathbb{P}(Y \le y \mid \boldsymbol{X} = \boldsymbol{x}) \quad = \quad 1 - \exp(-\exp(h(y) + \boldsymbol{x}^\top \boldsymbol{\beta}))$$

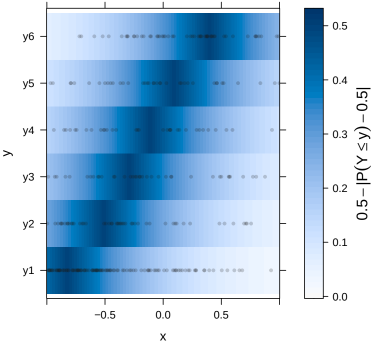
- the most prominent transformation model
- h is baseline log-cumulative hazard
- $\boldsymbol{x}^\top \boldsymbol{\beta}$ is log-hazard ratio
- partial likelihood profiles out h
- np score test: log-rank
- no censoring in *model*, only in *likelihood*



Proportional-odds Models

Ordinal outcome at categories $y_1 < y_2 < \cdots < y_K$
$h(y_k) = \vartheta_k, k = 1, \dots, K-1$ with $F_Z = \text{expit}$

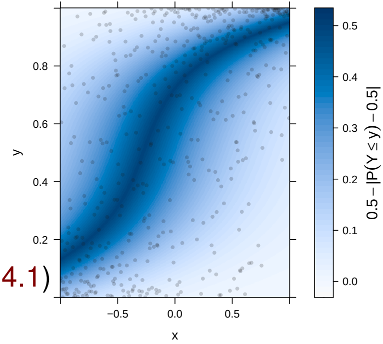
- Proportional-odds model
- Popular for ordinal data analysis
- $\boldsymbol{x}^\top \boldsymbol{\beta}$ is log-odds ratio
- *Simultaneous* estimation of h (via ϑ_k) and $\boldsymbol{\beta}$



Conditional Outcome Logistic Regression / Ordinal Regression Model

$$\mathbb{P}(Y \leq y \mid \boldsymbol{X} = \boldsymbol{x}) = \text{expit}(h(y) - \boldsymbol{x}^\top \beta)$$

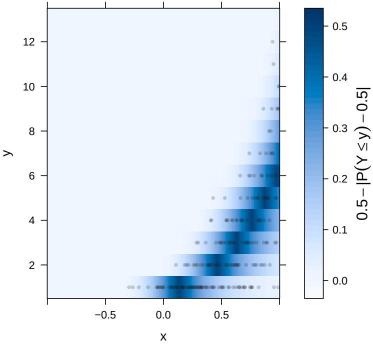
- Continuous proportional-odds model
- h is baseline log-odds function
- $\boldsymbol{x}^\top \beta$ is log-odds ratio
- np score test: Wilcoxon
- parametric: Colr (10.12688/f1000research.12934.1)
- nonparametric: orm (10.1002/sim.7433)



Count Transformation Models

Count outcome $Y \in \{0, 1, 2, \dots\}$
$h(k) := h(\lfloor y \rfloor) \quad \forall k \leq y < k + 1$ with $h : \mathbb{R} \to \mathbb{R}$

- Smooth $h(\cdot \mid \vartheta)$ evaluated discretely
- More flexible than Poisson/NB
- Discrete count likelihood
- 10.1111/2041-210X.13383



Distribution Regression

Quantile regression:

$$Q(\tau \mid \boldsymbol{X} = \boldsymbol{x}) \quad = \quad \alpha(\tau) + \boldsymbol{x}^\top \boldsymbol{\delta}(\tau)$$

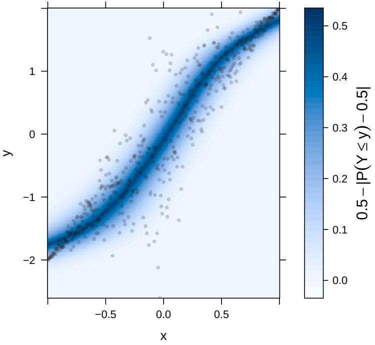
Distribution regression:

$$F_Z^{-1}(\mathbb{P}(Y \leq y \mid \boldsymbol{X} = \boldsymbol{x})) \quad = \quad h(y) - \boldsymbol{x}^\top \boldsymbol{\beta}(y)$$

i.e. quantile regression on the log-cumulative hazard / odds or probit scale

Distribution Regression

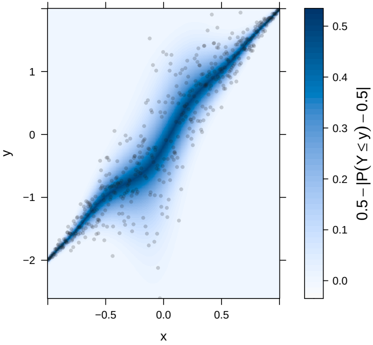
- Often (re-)discovered
- $\text{logit}(\mathbb{E}(I(Y \leq c)) = \alpha_c + \boldsymbol{x}^\top \boldsymbol{\beta}_c$
- Full likelihood possible
- Splines for $h(y)$ and $\beta(y)$



Conditional Transformation Models

$$\mathbb{P}(Y \leq y \mid \boldsymbol{X} = \boldsymbol{x}) = F_Z(h(y \mid \boldsymbol{x}))$$

- Practically no assumptions
- Tensor-product spline bases for y and $\boldsymbol{x}$
- Full likelihood possible
- 10.1111/rssb.12017



Transformation Likelihood

Log-likelihoods

Observed $(\underline{y}, \bar{y}] \subset \mathbb{R}$:

$$\log[F_Z\{h(\bar{y} \mid \boldsymbol{x})\} - F_Z\{h(\underline{y} \mid \boldsymbol{x})\}]$$

This includes discrete and censored observations and, via $(y_{(k)}, y_{(k+1)}]$, the nonparametric likelihood.

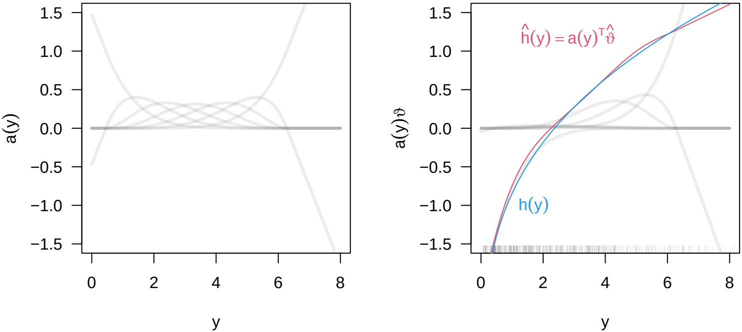
Observed $Y \in \mathbb{R}$:

$$\approx \log[f_Z\{h(y \mid \boldsymbol{x})\}] + \log\{h'(y \mid \boldsymbol{x})\}$$

10.1111/sjos.12291

Parameterisation

- h typically swept under the carpet
- More fun: parameterise $h(y \mid \boldsymbol{\vartheta}) = \boldsymbol{a}(y)^\top \boldsymbol{\vartheta}$ in terms of $\boldsymbol{\vartheta} \in \Theta$ (a la 10.1080/15598608.2013.772835)
- Estimate all parameters *simultaneously* (via ML)
- Does it hurt? Not really: 10.1002/sim.8425



Model Baking

- **Ingredients:** Take F_Z, parameterise $h(y \mid \boldsymbol{\vartheta})$, define impact of $\boldsymbol{x}$ via $\boldsymbol{x}^\top \boldsymbol{\beta}$, $\boldsymbol{x}^\top \boldsymbol{\beta}(y)$, $h(y \mid \boldsymbol{x})$
- **Mix:** Data defines likelihood function (handles discrete and continuous observations, censoring)
- **Oven:** Optimise, get $\hat{\boldsymbol{\vartheta}}, \hat{\boldsymbol{\beta}}$ + limiting distribution
- **Serve:** Interpret/interrogate fully specified model

And now for some tasty dishes...

Pushing the Boundaries

Correlated Observations

Mixed-effects transformation models via conditional distribution

$$\mathbb{P}\left(Y \le y \mid \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{U} = \boldsymbol{u}, \boldsymbol{\xi}\right) \quad = F_Z\left(h(y \mid \vartheta) - \boldsymbol{x}^\top \boldsymbol{\beta} - \boldsymbol{u}^\top \boldsymbol{\xi}\right)$$

with normal random effects $\boldsymbol{\xi} \sim \mathrm{N}_q(\mathbf{0}, \Sigma)$

Integrate conditional likelihood wrt random effects, using the fabulous TMB package

10.32614/RJ-2021-075 10.1093/biostatistics/kxab045

High Dimensions

$$\mathbb{P}(Y \leq y \mid \boldsymbol{X} = \boldsymbol{x}) \quad = \quad F_Z(h(y \mid \boldsymbol{\vartheta}) - \boldsymbol{x}^\top \boldsymbol{\beta})$$

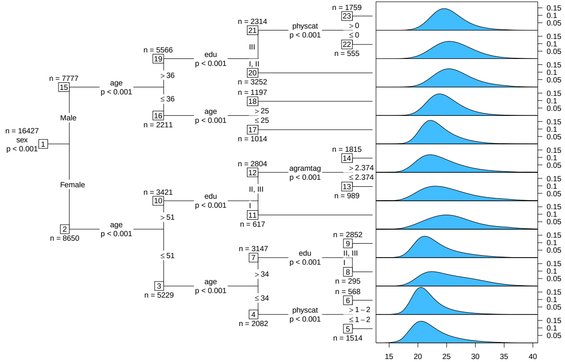
with $\boldsymbol{\beta} \in \mathbb{R}^p$, with p being large

Add L_1, L_2, ... penalty for $\boldsymbol{\beta}$ to likelihood, using the fabulous CVXR package

Transformation ridge, lasso, elastic net, etc. 10.32614/RJ-2021-054

Trees and Forests

Model-based recursive partitioning (MOB) based on transformation models, trees and forests for distributional regression



10.1515/ijb-2019-0063, 10.1177/0962280219862586

Additive Models

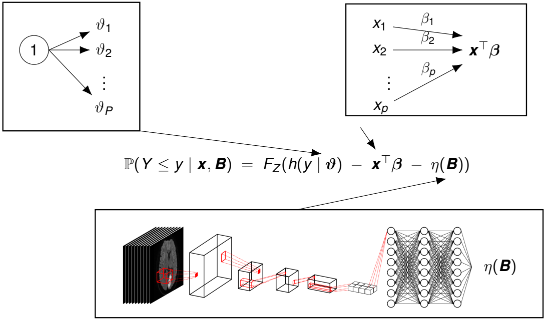
$$\mathbb{P}(Y \le y \mid \boldsymbol{X} = \boldsymbol{x}) \quad = \quad F_Z\left(h(y \mid \vartheta) - \sum_{j=1}^{J} f_i(\boldsymbol{x})\right)$$

aka “transform-both-sides”

Use connection to mixed models and leverage mgcv infrastructure; in the making

Alternative: Boosting via mboost, 10.1007/s11222-019-09870-4

Unstructured Information



"Deep" transformations, 10.1016/j.patcog.2021.108263
10.1007/978-3-030-86523-8_1

Multivariate Transformation Models

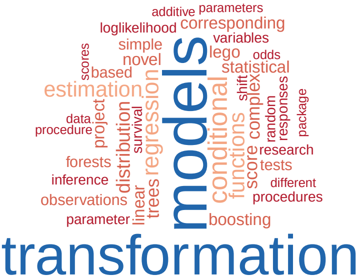
$$\mathbb{P}\left(\bigcap_{j=1}^{J} Y_j \le y_j \mid \boldsymbol{X} = \boldsymbol{x}\right) = F_{\boldsymbol{Z}}\Big(h_1(y_1 \mid \boldsymbol{x}) + h_2(y_2 \mid \boldsymbol{x}) + \lambda_{21}(\boldsymbol{x}) h_1(y_1 \mid \boldsymbol{x}) + \dots h_J(y_J \mid \boldsymbol{x}) + \sum_{j=1}^{J-1} \lambda_{Jj}(\boldsymbol{x}) h_j(y_j \mid \boldsymbol{x})\Big)$$

with $\boldsymbol{Z} \sim \mathrm{N}_J(\mathbf{0}, \boldsymbol{I})$. Correlations depend on $\boldsymbol{x}$ through $\lambda_{jj'}(\boldsymbol{x})$

Connection to Gaussian copulas and normalising flows, 10.1111/sjos.12501

R Add-on Packages

- mlt: Basic infrastructure
- tram: Model interfaces, multivariate models
- cotram: Count models
- tramnet: Penalisation
- tramME: Mixed-effects
- tbm: Boosting
- trtf: Trees and Forests



https://ctm.R-forge.R-project.org/