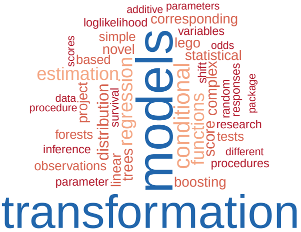
University of Zurich[UZH]

EBPI Epidemiology, Biostatistics and Prevention Institute

A Transformation Perspective on Marginal and Conditional Models

Torsten Hothorn and Luisa Barbanti



CEN 2023, 2023-09-04

Indeed, all statistical conclusions are based on marginality.

Lindsey and Lambert, SiM, 1998

Setup

- Clustered or longitudinal observations
- Interested in marginal effects for cluster elements or at specific times (for example, treatment effects)
- Possibly weird response distribution
- Possibly discrete or interval-censored data

Challenge

Obtain a simple analytic model for the marginal distribution.

What does “simple” mean?

$$\text{simple} := \text{transformation model}$$

Transformation Models ...

... describe whole distributions

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

not only means

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

Transformation Models ...

$$\mathbb{P}_Z \longrightarrow \mathbb{P}_{h(Y)}$$

The "Normal" Linear Regression Model

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

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

Adding a normal random effect $\boldsymbol{u}^\top \boldsymbol{R}$, $\boldsymbol{R} \sim \mathrm{N}_R(\mathbf{0}_R, \boldsymbol{G})$, still allows marginal inference.

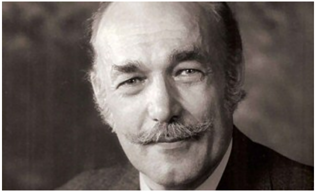
Generalisation I

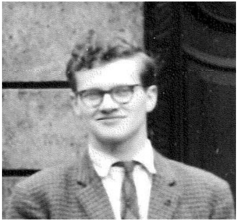
$$
\begin{aligned}
Y &= \alpha + \boldsymbol{x}^\top \gamma + \sigma Z, \quad Z \sim \mathrm{N}(0,1) \\
\iff \mathbb{E}(Y \mid \boldsymbol{X} = \boldsymbol{x}) &= \alpha + \boldsymbol{x}^\top \gamma \qquad , \quad Y \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{N}(,) \\
\hookrightarrow g(\mathbb{E}(Y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

Generalized Linear Models

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

Rothamsted Experimental Station, Harpenden, Herts





Generalisation II

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

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

$$\begin{aligned} h(Y) &= \boldsymbol{x}^\top \boldsymbol{\beta} + Z, \quad Z \sim F \\ \hookrightarrow \mathbb{P}(Y \leq y \mid \boldsymbol{X} = \boldsymbol{x}) &= F(h(y) - \boldsymbol{x}^\top \boldsymbol{\beta}) \end{aligned}$$

also allows discrete models via step-function h

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

aka Probabilistic index models:

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

Log-likelihoods

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

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

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

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

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

10.1111/sjos.12291

Correlated Observations

In both GLMs and transformation models, adding random effects defines conditional models

GLM:

$$g(\mathbb{E}(Y \mid \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{U} = \boldsymbol{u}, \boldsymbol{R})) \quad = \quad \alpha + \boldsymbol{x}^\top \gamma + \boldsymbol{u}^\top \boldsymbol{R}$$

Transformation model:

$$F^{-1}\left(\mathbb{P}\left(Y \leq y \mid \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{U} = \boldsymbol{u}, \boldsymbol{R}\right)\right) \quad = h(y) - \boldsymbol{x}^\top \beta - \boldsymbol{u}^\top \boldsymbol{R}$$

Except for random effects $\boldsymbol{R}$ from a bridge distribution to F or g^{-1}, β does not have a marginal interpretation.

Mixed-effects Transformation Models

Mixed-effects transformation models via conditional distribution

$$\mathbb{P}(Y \leq y \mid \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{U} = \boldsymbol{u}, \boldsymbol{R}) \quad = F\left(h(y) - \boldsymbol{x}^\top \beta - \boldsymbol{u}^\top \boldsymbol{R}\right)$$

with normal random effects $\boldsymbol{R} \sim \mathrm{N}_R(\mathbf{0}_R, \boldsymbol{G})$

$\rightarrow$ next talk by Bálint Tamási

Joint Model for Clustered or Longitudinal Data

- $i = 1, \dots, N$ independent observational units, each consisting of N_i correlated observations of the response $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{iN_i})^\top \in \Xi^{N_i}$
- "fixed" effects design matrix $\boldsymbol{X}_i$
- "random" effects design matrix $\boldsymbol{U}_i$
- variance parameters $\gamma \in \mathbb{R}^{R(R+1)/2}$ such that

$$\Sigma_i(\gamma) := \boldsymbol{U}_i \Lambda(\gamma) \Lambda(\gamma)^\top \boldsymbol{U}_i^\top + \boldsymbol{I}_{N_i}$$

  with unstructured Cholesky factor $\Lambda(\gamma) \in \mathbb{R}^{R \times R}$
- $\boldsymbol{D}_i(\gamma) = \text{diag}(\Sigma_i(\gamma))^{1/2} \cdot \boldsymbol{I}_{N_i}$

And the joint distribution function is

$$\mathbb{P}(\boldsymbol{Y}_i \le \boldsymbol{y} \mid \boldsymbol{X}_i, \boldsymbol{U}_i) =$$

Joint Model for Clustered or Longitudinal Data

$$\mathbb{P}(\boldsymbol{Y}_i \le \boldsymbol{y} \mid \boldsymbol{X}_i, \boldsymbol{U}_i) = \underbrace{\Phi_{\boldsymbol{0}_{N_i}, \Sigma_i(\gamma)}\left(\underbrace{\boldsymbol{D}_i(\gamma)\,\underbrace{\Phi_{N_i}^{-1}\left(\underbrace{F_{N_i}\{\underbrace{\boldsymbol{D}_i(\gamma)^{-1}[\underbrace{h_{N_i}(\boldsymbol{y}) - \boldsymbol{X}_i\beta}_{\text{element-wise trafo}}]}_{\text{standardise}}\}}_{\text{into}[0,1]}\right)}_{\text{make normal}}}_{\text{de-standardise}}\right)}_{\text{evaluate}}$$

Ouch!

Probit-type Models

With $F = \Phi$, we get

$$\mathbb{P}(\boldsymbol{Y}_i \leq \boldsymbol{y} \mid \boldsymbol{X}_i, \boldsymbol{U}_i) = \Phi_{\mathbf{0}_{N_i}, \Sigma_i(\gamma)}\left(h_{N_i}(\boldsymbol{y}) - \boldsymbol{X}_i \beta\right)$$

with marginal distribution of some Y in $\boldsymbol{Y}$,

$$\mathbb{P}(Y \leq y \mid \boldsymbol{x}, \boldsymbol{u}) = \Phi\left(\frac{h(y) - \boldsymbol{x}^\top \beta}{\sqrt{\boldsymbol{u}^\top \Lambda(\gamma)\Lambda(\gamma)^\top \boldsymbol{u} + 1}}\right).$$

General Marginal Distribution

For some Y in $\boldsymbol{Y}$,

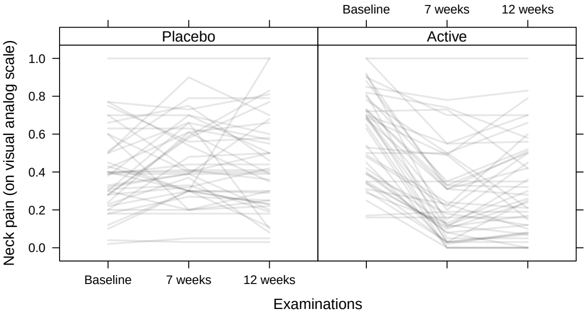
$$\mathbb{P}(Y \leq y \mid \boldsymbol{x}, \boldsymbol{u}) = F\left(\frac{h(y) - \boldsymbol{x}^\top \beta}{\sqrt{\boldsymbol{u}^\top \Lambda(\gamma)\Lambda(\gamma)^\top \boldsymbol{u} + 1}}\right).$$

- This is a linear transformation model
- Describes whole marginal distribution
- Parameter interpretation defined by F
- This is almost “simple” (the scaling is a bit nasty)

Good News

- Analytic log-likelihood and score functions for absolute continuous responses available
- Log-likelihood for discrete and censored observations only requires evaluation of R-dimensional normal, not N_i-dimensional normal
- Simultaneous *exact* maximum likelihood inference for h (suitably parameterised), $\boldsymbol{\beta}$, and γ possible and computationally efficient
- For $F = \Phi$ and linear h, fit is identical to probit mixed-effects models
- For $F = \text{logit}^{-1}$ and binary responses, inference for marginal log-odds ratio practically identical to GEE inference
- Analytic marginal distribution close to approximate marginal distribution from **tramME**

Longitudinal Pain Assessment



Longitudinal Pain Assessment

Joint model with marginal distribution function for pain $y \in [0, 1]$ at time t:

$$
\begin{aligned}
\operatorname{logit}(\mathbb{P}(\text{Pain} \le y \mid t, \text{Placebo})) &= \frac{h(y) + \beta(t)}{\sqrt{1+\gamma^2}} \\
\operatorname{logit}(\mathbb{P}(\text{Pain} \le y \mid t, \text{Active})) &= \frac{h(y) + \beta_\text{A} + \beta(t) + \beta_\text{A}(t)}{\sqrt{1+\gamma^2}}
\end{aligned}
$$

Interpretation:

$$
\Rightarrow \exp\left(\frac{\beta_\text{A} + \beta_\text{A}(t)}{\sqrt{1+\gamma^2}}\right)
$$

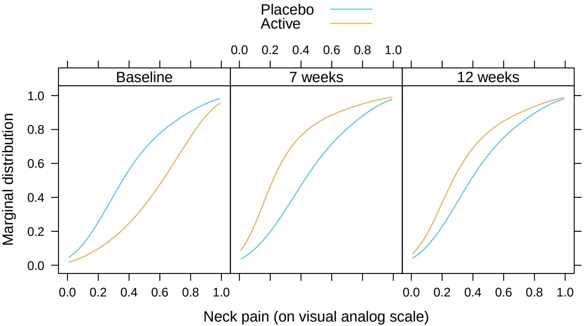
is marginal odds-ratio at time t for all cut-off points $y \in [0, 1]$.

Can be re-formulated as probabilistic index.

Fitting the Model

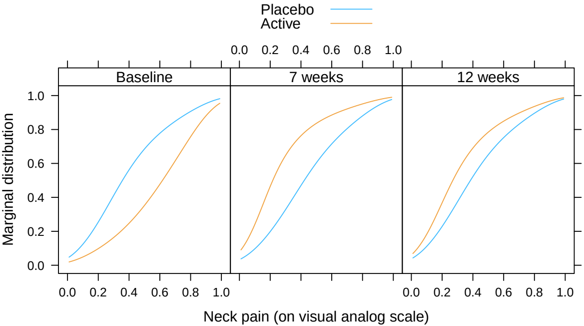
```
R> library("tram")
R> m <- Colr(vas ~ laser * time, data = pain_df,
+             bounds = c(0, 1), support = c(0, 1))
R> mtram(m, ~ (1 | id), data = pain_df)
```

Longitudinal Pain Assessment



Time	Probabilistic Index	95% CI
Baseline	0.72	[0.58; 0.83]
7 weeks	0.29	[0.17; 0.43]
12 weeks	0.38	[0.24; 0.54]

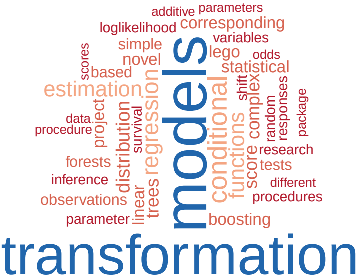
Longitudinal Pain Assessment



Time	Probabilistic Index	95% CI	
Baseline	0.72	[0.58; 0.83]	hmmm...
7 weeks	0.29	[0.17; 0.43]	strike!
12 weeks	0.38	[0.24; 0.54]	well...

Further Information

- 10.1093/biostatistics/kxac048
- tram: Implementation
- mtram package vignette: Empirical evaluation and worked examples



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