

*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?

simple := transformation model

Transformation Models ...

... describe whole distributions

$$\mathbb{P}(Y \leq y \mid \mathbf{X} = \mathbf{x})$$

not only means

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

Transformation Models ...

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



The “Normal” Linear Regression Model

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

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

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

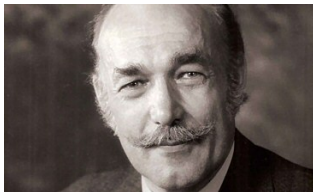
Generalisation I

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

Generalized Linear Models

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

Rothamsted Experimental Station, Harpenden, Herts



Generalisation II

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

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

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

“Box-Cox” power transformation

Conditional Distribution Functions

$$h(Y) = \mathbf{x}^\top \boldsymbol{\beta} + Z, \quad Z \sim F$$
$$\Leftrightarrow \mathbb{P}(Y \leq y \mid \mathbf{X} = \mathbf{x}) = F(h(y) - \mathbf{x}^\top \boldsymbol{\beta})$$

also allows discrete models via step-function h

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

aka Probabilistic index models:

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

Log-likelihoods

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

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

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

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

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

10.1111/sjos.12291

Correlated Observations

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

GLM:

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

Transformation model:

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

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

Mixed-effects Transformation Models

Mixed-effects transformation models via conditional distribution

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

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

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

10.32614/RJ-2021-075

10.1093/biostatistics/kxab045

tramME package

Joint Model for Clustered or Longitudinal Data

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

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

with unstructured Cholesky factor $\boldsymbol{\Lambda}(\boldsymbol{\gamma}) \in \mathbb{R}^{R \times R}$

- $\mathbf{D}_i(\boldsymbol{\gamma}) = \text{diag}(\boldsymbol{\Sigma}_i(\boldsymbol{\gamma}))^{1/2} \cdot \mathbf{I}_{N_i}$

And the joint distribution function is

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

Joint Model for Clustered or Longitudinal Data

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

Ouch!

Probit-type Models

With $F = \Phi$, we get

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

with marginal distribution of some Y in \mathbf{Y} ,

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

General Marginal Distribution

For some Y in \mathbf{Y} ,

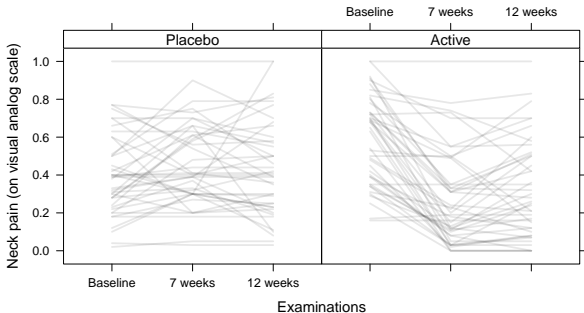
$$\mathbb{P}(Y \leq y \mid \mathbf{x}, \mathbf{u}) = F \left(\frac{h(y) - \mathbf{x}^\top \boldsymbol{\beta}}{\sqrt{\mathbf{u}^\top \boldsymbol{\Lambda}(\boldsymbol{\gamma}) \boldsymbol{\Lambda}(\boldsymbol{\gamma})^\top \mathbf{u} + 1}} \right).$$

- 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_j -dimensional normal
- Simultaneous *exact* maximum likelihood inference for h (suitably parameterised), β , 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 :

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

Interpretation:

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

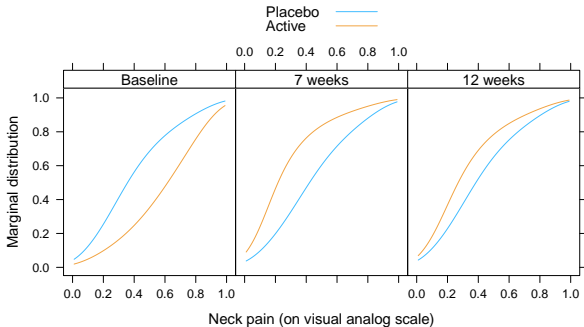
is marginal odds-ratio at time t for all cut-off points $y \in [0, 1]$.

Can be re-formulated as probabilistic index.

Fitting the Model

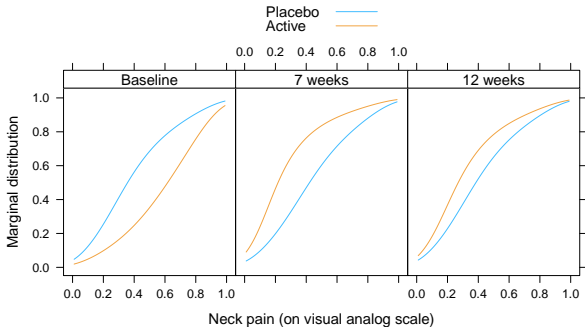
```
R> library("tram")
R> m <- Colr(vas ~ laser * time, data = pain_df,
+           bounds = c(0, 1), support = c(0, 1))
R> mtram(m, ~ (1 | id), data = pain_df)
```

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

- **tram**: Implementation
- [10.1093/biostatistics/kxac048](https://doi.org/10.1093/biostatistics/kxac048)
- **mtram** package vignette: Empirical evaluation and worked examples

Parameterisation

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