University of Zurich[UZH]

EBPI Epidemiology, Biostatistics and Prevention Institute

Understanding and Applying Transformation Models

Torsten Hothorn

2016-10-13

Regression Models

Unconditional distribution

$$Y \sim \mathbb{P}_Y$$

Conditional distribution

$$Y | \mathbf{X} = \mathbf{x} \sim \mathbb{P}_{Y | \mathbf{X} = \mathbf{x}}$$

Aim: Obtain estimates $\hat{\mathbb{P}}_Y$ and $\hat{\mathbb{P}}_{Y | \mathbf{X} = \mathbf{x}}$

Unconditional Binary Response

$Y \in \{y_1, y_2\}$

$$\begin{aligned}\mathbb{P}(Y \le y_1) &= \pi_1 \\ \mathbb{P}(Y \le y_2) &= 1\end{aligned}$$

with $\pi_1 \in [0, 1]$ or, equivalently, with $\vartheta_1 \in \mathbb{R}$

$$\begin{aligned}\mathbb{P}(Y \le y_1) &= F_Z(\vartheta_1) \\ \mathbb{P}(Y \le y_2) &= 1 = F_Z(\infty)\end{aligned}$$

$F_Z : \mathbb{R} \to [0, 1]$ is cdf of some continuous rv Z

F_Z

$$\begin{aligned} F_Z(z) &= \Phi(z) \\ F_Z(z) &= F_{\text{SL}}(z) = (1 + \exp(-z))^{-1} \\ F_Z(z) &= F_{\text{MEV}}(z) = 1 - \exp(-\exp(z)) \\ &\vdots \end{aligned}$$

$\vartheta_1 = \log(\pi/(1 - \pi))$ for $F_Z = F_{\text{SL}}$

Conditional Binary Response

$$\begin{aligned}\mathbb{P}(Y \le y_1 \mid \mathbf{X} = \mathbf{x}) &= \pi_1(\mathbf{x}^\top \boldsymbol{\beta}) = F_Z(\vartheta_1 + \mathbf{x}^\top \boldsymbol{\beta}) \\ \mathbb{P}(Y \le y_2 \mid \mathbf{X} = \mathbf{x}) &= 1 = F_Z(\infty + \mathbf{x}^\top \boldsymbol{\beta})\end{aligned}$$

Probit regression: $F_Z = \Phi$
Logistic regression: $F_Z = F_{\text{SL}}$
Complementary log-log regression: $F_Z = F_{\text{MEV}}$

Unconditional Ordered Categorical Response

$Y \in \{y_1, y_2, \dots, y_K\}$

$$\begin{aligned}
\mathbb{P}(Y \le y_1) &= F_Z(\vartheta_1) \\
\mathbb{P}(Y \le y_2) &= F_Z(\vartheta_2) \\
&\vdots \\
\mathbb{P}(Y \le y_{K-1}) &= F_Z(\vartheta_{K-1}) \\
\mathbb{P}(Y \le y_K) &= F_Z(\infty)
\end{aligned}$$

st $\vartheta_k < \vartheta_{k+1}$ for $k = 1, \dots, K-1$

aka multinomial model

Conditional Ordered Categorical Response (Simple)

$$
\begin{aligned}
\mathbb{P}(Y \le y_1 \mid \mathbf{X} = \mathbf{x}) &= F_Z(\vartheta_1 + \mathbf{x}^\top \boldsymbol{\beta}) \\
\mathbb{P}(Y \le y_2 \mid \mathbf{X} = \mathbf{x}) &= F_Z(\vartheta_2 + \mathbf{x}^\top \boldsymbol{\beta}) \\
&\vdots \\
\mathbb{P}(Y \le y_{K-1} \mid \mathbf{X} = \mathbf{x}) &= F_Z(\vartheta_{K-1} + \mathbf{x}^\top \boldsymbol{\beta}) \\
\mathbb{P}(Y \le y_K | \mathbf{X} = \mathbf{x}) &= F_Z(\infty + \mathbf{x}^\top \boldsymbol{\beta})
\end{aligned}
$$

st $\vartheta_k < \vartheta_{k+1}$ for $k = 1, \dots, K-1$

Proportional odds ($F_Z = F_{\text{SL}}$) and proportional hazards ($F_Z = F_{\text{MEV}}$) cumulative models

Conditional Ordered Categorical Response (Complex)

$$
\begin{aligned}
\mathbb{P}(Y \le y_1 \mid \mathbf{X} = \mathbf{x}) &= F_Z(\vartheta_1 + \mathbf{x}^\top \boldsymbol{\beta}_1) \\
\mathbb{P}(Y \le y_2 \mid \mathbf{X} = \mathbf{x}) &= F_Z(\vartheta_2 + \mathbf{x}^\top \boldsymbol{\beta}_2) \\
&\vdots \\
\mathbb{P}(Y \le y_{K-1} \mid \mathbf{X} = \mathbf{x}) &= F_Z(\vartheta_{K-1} + \mathbf{x}^\top \boldsymbol{\beta}_{K-1}) \\
\mathbb{P}(Y \le y_K | \mathbf{X} = \mathbf{x}) &= F_Z(\infty + \mathbf{x}^\top \boldsymbol{\beta}_K)
\end{aligned}
$$

st $\vartheta_k + \mathbf{x}^\top \boldsymbol{\beta}_k < \vartheta_{k+1} + \mathbf{x}^\top \boldsymbol{\beta}_{k+1}$ for $k = 1, \dots, K-1$ and all $\mathbf{x}$

Non-proportional odds ($F_Z = F_{\text{SL}}$), aka logistic multinomial regression, and non-proportional hazards ($F_Z = F_{\text{MEV}}$) cumulative models

Simplify (?) Notation

Unconditional

$$\mathbb{P}(Y \leq y) \quad = \quad F_Z(h_Y(y))$$

$$h_Y : \{y_1, \dots, y_K\} \to \mathbb{R} \text{ monotone, } h_Y(y_K) = \infty$$

Conditional

$$\mathbb{P}(Y \leq y \mid \mathbf{X} = \mathbf{x}) \quad = \quad F_Z(h_Y(y) + \mathbf{x}^\top \boldsymbol{\beta}(y))$$

st $h_Y(y_k) + \mathbf{x}^\top \boldsymbol{\beta}(y_k) < h_Y(y_{k+1}) + \mathbf{x}^\top \boldsymbol{\beta}(y_{k+1})$
for $k = 1, \dots, K-1$ and all $\mathbf{x}$

Unconditional Continuous Response

$y \in \mathbb{R}$

$$\mathbb{P}(Y \le y) \quad = \quad F_Y(y) = F_Z(h_Y(y))$$

$h_Y : \mathbb{R} \to \mathbb{R}$
st $h_Y(y) < h_Y(y + \delta)$ for all $\delta > 0$

Note: $h_Y = F_Z^{-1} \circ F_Y$ always exists and $Z = h_Y(Y)$

Conditional Continuous Response (Simple)

$y \in \mathbb{R}$

$$\mathbb{P}(Y \le y \mid \mathbf{X} = \mathbf{x}) \quad = \quad F_Z(h_Y(y) + \mathbf{x}^\top \boldsymbol{\beta})$$

st $h_Y(y) < h_Y(y + \delta)$ for all $\delta > 0$

Note: $Z = h_Y(Y) + \mathbf{x}^\top \boldsymbol{\beta}$ and thus $\mathbb{E}(h_Y(Y)) = \mathbb{E}(Z) - \mathbf{x}^\top \boldsymbol{\beta}$

Normal Linear Regression Model (NLRM)

$$Y|\mathbf{X} = \mathbf{x} \sim \mathcal{N}(\tilde{\alpha} + \mathbf{x}^\top \tilde{\boldsymbol{\beta}}, \sigma^2)$$

$$\begin{aligned}
\mathbb{P}(Y \leq y \mid \mathbf{X} = \mathbf{x}) &= \Phi\left(\frac{y - \tilde{\alpha} - \mathbf{x}^\top \tilde{\boldsymbol{\beta}}}{\sigma}\right) \\
&= \Phi(\vartheta_1 + \vartheta_2 y + \mathbf{x}^\top \boldsymbol{\beta}) \\
&= F_Z(h_Y(y) + \mathbf{x}^\top \boldsymbol{\beta})
\end{aligned}$$

$h_Y(y)$ is linear in y with positive slope $\vartheta_2 = \sigma^{-1}$

$$\mathbb{E}(h_Y(Y)) = \mathbb{E}(\vartheta_1 + \vartheta_2 Y) = \mathbf{x}^\top \boldsymbol{\beta}$$

Beyond Normality

Linear Transformation Model:

$$\mathbb{P}(Y \leq y \mid \mathbf{X} = \mathbf{x}) \quad = \quad F_Z(h_Y(y) + \mathbf{x}^\top \boldsymbol{\beta})$$

h_Y	Φ	F_{SL}	F_{MEV}
$\vartheta_1 + \vartheta_2 y$	NLRM		
$\vartheta_1 + \vartheta_2 \log(y)$	log-normal	log-logistic	exponential/Weibull
Box-Cox	Box-Cox		
monotone			Cox

Beyond Normality

Linear Transformation Model:

$$\mathbb{P}(Y \leq y \mid \mathbf{X} = \mathbf{x}) \quad = \quad F_Z(h_Y(y) + \mathbf{x}^\top \beta)$$

h_Y	Φ	F_{SL}	F_{MEV}
$\vartheta_1 + \vartheta_2 y$	NLRM	?	?
$\vartheta_1 + \vartheta_2 \log(y)$	log-normal	log-logistic	exponential/Weibull
Box-Cox	Box-Cox	?	?
monotone	!!!	!!!	Cox

Conditional Continuous Response (Complex)

$y \in \mathbb{R}$

$$\mathbb{P}(Y \le y \mid \mathbf{X} = \mathbf{x}) \quad = \quad F_Z(h_Y(y) + \mathbf{x}^\top \boldsymbol{\beta}(y))$$

st $h_Y(y) + \mathbf{x}^\top \boldsymbol{\beta}(y) < h_Y(y + \delta) + \mathbf{x}^\top \boldsymbol{\beta}(y + \delta)$ for all $\delta > 0$ and $\mathbf{x}$

Note: $Z = h_Y(Y) + \mathbf{x}^\top \boldsymbol{\beta}(Y)$

Time-varying Cox/AFT or non-proportional hazards models, distribution regression (talk Samantha Leorato Oct 28!)

Conditional Continuous Response (Too Complex?)

$y \in \mathbb{R}$

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

st $h(y|\mathbf{x}) < h(y + \delta|\mathbf{x})$ for all $\delta > 0$ and $\mathbf{x}$

Note: $Z = h(Y|\mathbf{x})$ instead of the usual

$$Y = h^{-1}(Z|\mathbf{x}) = g(\mathbf{x}) + \sigma Z$$

Unconditional Discrete Response

$y \in \mathbb{N}$

$$\mathbb{P}(Y \le y) \quad = \quad F_Z(h_Y(y))$$

$h_Y : \mathbb{N} \to \mathbb{R}$
st $h_Y(y) < h_Y(y+1)$

Conditional Discrete Response

$y \in \mathbb{N}$

$$\mathbb{P}(Y \le y \mid \mathbf{X} = \mathbf{x}) \quad = \quad F_Z(h(y|\mathbf{x}))$$

st $h(y|\mathbf{x}) < h(y+1|\mathbf{x})$ for all $\mathbf{x}$

Conditional Transformation Models

For all univariate Y

$$\mathbb{P}(Y \le y \mid \mathbf{X} = \mathbf{x}) \quad = \quad F_Z(h(y|\mathbf{x}))$$

st $h(y|\mathbf{x})$ monotone in y for all $\mathbf{x}$

h is called "conditional transformation function" by Hothorn, Kneib and Bühlmann (2014, JRSS B)

The Likelihood

Datum $(\underline{y}, \bar{y}] \subset \mathbb{R}$ (continuous) or $(\underline{y}, \bar{y}] = (y_{k-1}, y_k]$ (discrete)

Fisher's "exact" likelihood

$$
\begin{aligned}
\mathcal{L}(h | Y \in (\underline{y}, \bar{y}], \mathbf{X} = \mathbf{x}) \quad &:= \quad F_Z(h(\bar{y} \mid \mathbf{x})) - F_Z(h(\underline{y} \mid \mathbf{x})) \\
&= \quad 1 - F_Z(h(\underline{y} \mid \mathbf{x})) \quad \text{right-censored} \\
&= \quad F_Z(h(\bar{y} \mid \mathbf{x})) - 0 \quad \text{left-censored}
\end{aligned}
$$

The Likelihood

Truncation to $(y_l, y_r]$:

$$\frac{\mathcal{L}(h|Y \in (\underline{y}, \bar{y}], \mathbf{X} = \mathbf{x})}{\mathcal{L}(h|Y \in (y_l, y_r], \mathbf{X} = \mathbf{x})}$$

Closed forms for scores and Fisher information available

For continuous datum $y \in \mathbb{R}$ approximate by density

$$f_Y(y \mid \mathbf{x}) = f_Z(h(y \mid \mathbf{x}))h'(y \mid \mathbf{x})$$

Parameterisation

With basis function $\mathbf{c}$

$$h(y \mid \mathbf{x}) = \mathbf{c}(y, \mathbf{x})^\top \boldsymbol{\vartheta}$$

$(F_Z, \mathbf{c}, \boldsymbol{\vartheta})$ is a fully specified parametric model

$\mathbf{c}(y, \mathbf{x})^\top \hat{\boldsymbol{\vartheta}}_{\text{ML}}$ is called most likely transformation (MLT)

$\hat{\boldsymbol{\vartheta}}_{\text{ML}}$ from constrained convex optimisation (augmented Lagrangian adaptive barrier minimization in **alabama** or spectral projected gradient in **BB**)

mlt Package

The **mlt** package (on CRAN) implements maximum-likelihood estimation for

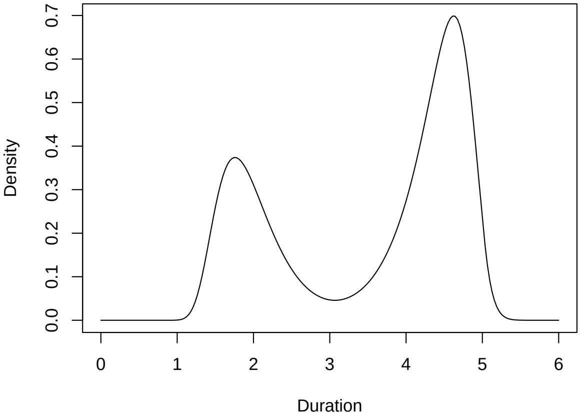
- unconditional and conditional transformation models, including all stratified linear transformation models
- for discrete (also counts) and continuous responses
- under all forms of random censoring and truncation,
- based on a variety of basis functions (log, polynomial, Bernstein, Legendre, ...) and combinations thereof,
- allowing specification, inference and the model-based bootstrap for unfitted (got no data yet) and fitted transformation models.

Old Faithful

```
> library("mlt")
> var_d <- numeric_var("duration", support = c(1.0, 5.0),
+                      add = c(-1, 1), bounds = c(0, Inf))
> B_d <- Bernstein_basis(var = var_d, order = 8, ui = "increasing")
> ctm_d <- ctm(response = B_d, todistr = "Normal")
> str(nd_d <- mkgrid(ctm_d, 200))

List of 1
 $ duration: num [1:200] 0 0.0302 0.0603 0.0905 0.1206 ...

> data("geyser", package = "TH.data")
> system.time(mlt_d <- mlt(ctm_d, data = geyser))

   user  system elapsed
  0.556   0.008   0.564

> logLik(mlt_d)

'log Lik.' -317.766 (df=9)
```

Old Faithful

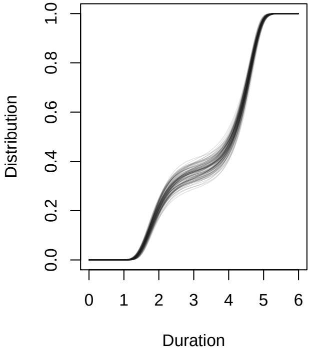
```
> nd_d$d <- predict(mlt_d, newdata = nd_d, type = "density")
> plot(d ~ duration, data = nd_d, type = "l", ylab = "Density",
+      xlab = "Duration")
```

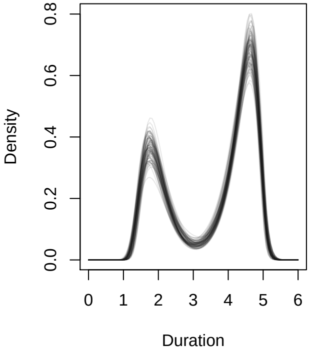


Parametric Bootstrap Old Faithful

```
> new_d <- simulate(mlt_d, nsim = 100)
> llr <- numeric(length(new_d))
> pdist <- vector(mode = "list", length = length(new_d))
> pdens <- vector(mode = "list", length = length(new_d))
> ngeyser <- geyser
> q <- mkgrid(var_d, 100)[[1]]
> for (i in 1:length(new_d)) {
+     ngeyser$duration <- new_d[[i]]
+     mlt_i <- mlt(ctm_d, data = ngeyser, scale = TRUE,
+                  theta = coef(mlt_d))
+     llr[[i]] <- logLik(mlt_i) - logLik(mlt_i, parm = coef(mlt_d))
+     pdist[[i]] <- predict(mlt_i, newdata = data.frame(1),
+                           type = "distribution", q = q)
+     pdens[[i]] <- predict(mlt_i, newdata = data.frame(1),
+                           type = "density", q = q)
+ }
```

Parametric Bootstrap Old Faithful





Boston Housing: Normal Linear Regression

$$\mathsf{medv} | \mathbf{X} = \mathbf{x} \sim \mathsf{N}(\alpha + \mathbf{x}^\top \boldsymbol{\beta}, \sigma^2)$$

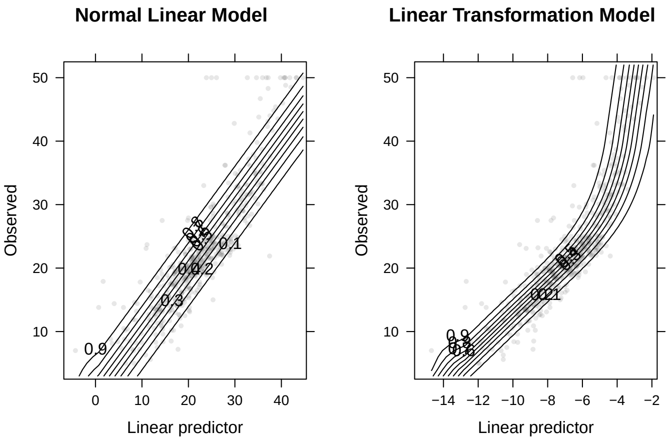
```
> data("BostonHousing2", package = "mlbench")
> lm_BH <- lm(cmedv ~ crim + zn + indus + chas + nox + rm + age +
+                      dis + rad + tax + ptratio + b + lstat,
+             data = BostonHousing2)
> logLik(lm_BH)

'log Lik.' -1494.245 (df=15)
```

Boston Housing: Linear Transformation Model

$$\begin{aligned}\mathbb{P}(\text{medv} \leq y \mid \mathbf{X} = \mathbf{x}) &= \Phi(h_{\text{medv}}(y) + \mathbf{x}^\top \beta) \\ &= \Phi(\mathbf{a}_{\text{Bs},6}(y)^\top \vartheta + \mathbf{x}^\top \beta)\end{aligned}$$

```
> BostonHousing2$medvc <- with(BostonHousing2,
+                              Surv(cmedv, cmedv < 50))
> var_m <- numeric_var("medvc", support = c(10.0, 40.0),
+                      bounds = c(0, Inf))
> fm_BH <- medvc ~ crim + zn + indus + chas + nox + rm + age +
+                  dis + rad + tax + ptratio + b + lstat
> B_m <- Bernstein_basis(var_m, order = 6, ui = "increasing")
> ctm_BH <- ctm(B_m, shift = fm_BH[-2L], data = BostonHousing2,
+               todistr = "Normal")
> system.time(mlt_BH <- mlt(ctm_BH, data = BostonHousing2,
+                           scale = TRUE))

   user  system elapsed
  0.328   0.004   0.332

> logLik(mlt_BH)

'log Lik.' -1324.698 (df=20)
```

Boston Housing

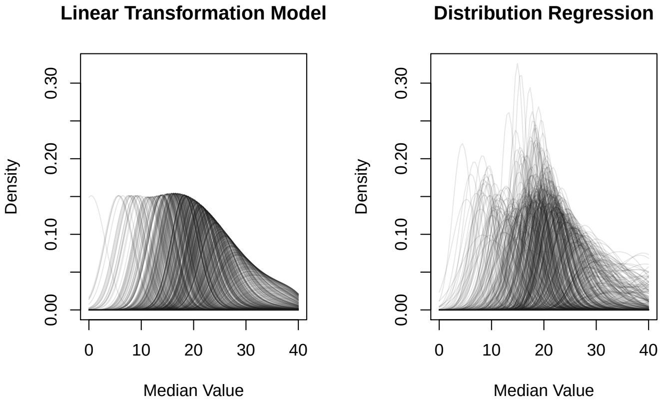


Boston Housing: Distribution Regression

$$
\begin{aligned}
\mathbb{P}(\text{medv} \leq y | \mathbf{X} = \mathbf{x}) &= \Phi\left(h_Y(y) + \sum_{j=1}^{J} \beta_j(y)\mathbf{x}_j\right) \\
&= \Phi\left(\mathbf{a}_{\text{Bs},6}(y)^\top \boldsymbol{\vartheta}_1 + \sum_{j=1}^{J} \mathbf{a}_{\text{Bs},6}(y)^\top \boldsymbol{\vartheta}_{j+1}\mathbf{x}_j\right)
\end{aligned}
$$

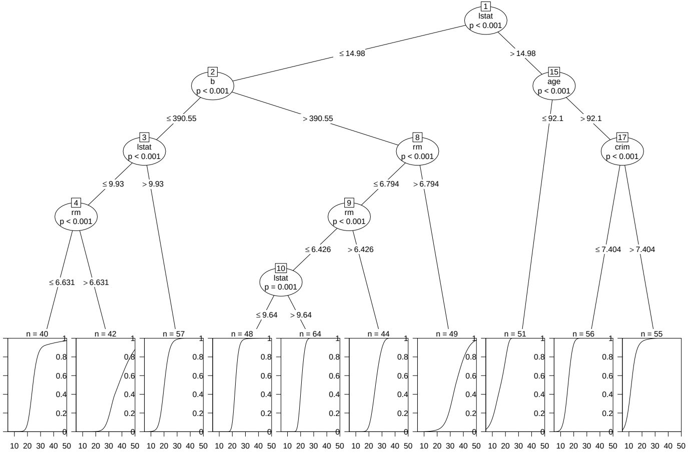
```
> b_BH_s <- as.basis(fm_BH[-2L], data = BostonHousing2, scale = TRUE)
> ctm_BHi <- ctm(B_m, interacting = b_BH_s, sumconstr = FALSE)
> system.time(mlt_BHi <- mlt(ctm_BHi, data = BostonHousing2,
+              scale = TRUE))

   user  system elapsed
  6.908   0.004   6.912

> logLik(mlt_BHi)

'log Lik.' -1274.367 (df=98)
```

Boston Housing



Boston Housing: Transformation Tree

$$\mathbb{P}(\text{medv} \le y | \mathbf{X} = \mathbf{x}) = \Phi(\mathbf{a}_{\text{Bs},4}(y)^\top \boldsymbol{\vartheta}(\mathbf{x}))$$

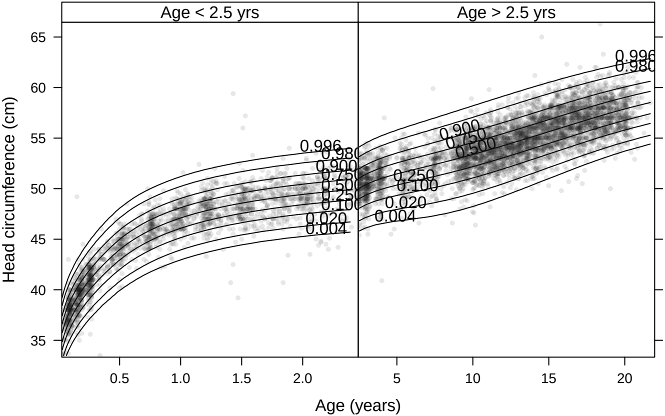


Growth Curves: Head Circumference (HC)

$$\mathbb{P}(\text{HC} \leq y \mid \text{age} = a) = \Phi((\mathbf{a}_{\text{Bs},3}(y)^\top \otimes \mathbf{b}_{\text{Bs},3}(a^{1/3})^\top)\boldsymbol{\vartheta})$$

```
> data("db", package = "gamlss.data")
> db$lage <- with(db, age^(1/3))
> var_head <- numeric_var("head", bounds = range(db$head),
+                         support = quantile(db$head, c(.1, .9)))
> B_head <- Bernstein_basis(var_head, order = 3, ui = "increasing")
> var_lage <- numeric_var("lage", bounds = range(db$lage),
+                         support = quantile(db$lage, c(.1, .9)))
> B_age <- Bernstein_basis(var_lage, order = 3, ui = "none")
> ctm_head <- ctm(B_head, interacting = B_age)
> system.time(mlt_head <- mlt(ctm_head, data = db, scale = TRUE))

   user  system elapsed 
  2.608   0.004   2.616 
```

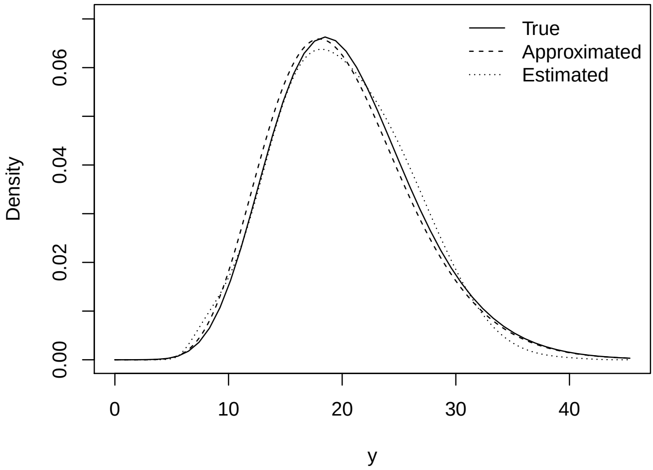
Growth Curves: Head Circumference



Computing on Models

```
> pY <- function(x) pchisq(x, df = 20)
> dY <- function(x) dchisq(x, df = 20)
> qY <- function(p) qchisq(p, df = 20)
> yvar <- numeric_var("y", support = qY(c(.001, 1 - .001)),
+                      bounds = c(0, Inf))
> By <- Bernstein_basis(yvar, order = ord <- 15, ui = "increasing")
> mod <- ctm(By)
> h <- function(x) qnorm(pY(x))
> x <- seq(from = support(yvar)[["y"]][1],
+          to = support(yvar)[["y"]][2],
+          length.out = ord + 1)
> mlt::coef(mod) <- h(x)
> d <- as.data.frame(mkgrid(yvar, n = 500))
> d$grid <- d$y
> d$y <- simulate(mod, newdata = d)
> fmod <- mlt(mod, data = d, scale = TRUE)
> logLik(fmod)

'log Lik.' -1597.168 (df=16)

> logLik(fmod, parm = coef(mod))

'log Lik.' -1603.067 (df=16)
```

Computing on Models



Where to?

- understanding and teaching: Distributions, not means
- rethink parametric vs. non-parametric statistics
- top-down model diagnostics and checking
- penalise $\boldsymbol{\beta}$ in linear transformation models
- random-effect models:

$$\begin{aligned}
\mathbb{P}(Y \le y \mid \mathbf{X} = \mathbf{x}, i) &= F_Z(h_Y(y) + \gamma_i + \mathbf{x}^\top \boldsymbol{\beta}) \quad \text{or} \\
\mathbb{P}(Y \le y \mid \mathbf{X} = \mathbf{x}, i) &= F_Z(h_Y(y) + \gamma_i + \sigma_i y + \mathbf{x}^\top \boldsymbol{\beta}) \quad \text{or} \\
\mathbb{P}(Y \le y \mid \mathbf{X} = \mathbf{x}, i) &= F_Z(h_Y(y) + h_i(y) + \mathbf{x}^\top \boldsymbol{\beta}) \quad ?
\end{aligned}$$

- transformation GAM
  $\mathbb{P}(Y \le y \mid \mathbf{X} = \mathbf{x}) = F_Z(h_Y(y) + g(\mathbf{x}))$
- transformation trees and forests (inverse quantile regression forests)
- ...

Resources

- CRAN packages **mlt.docreg**, **mlt**, **basefun**, **variables**
- package vignette **mlt.docreg**
- DOI 10.1111/rssb.12017 (with T. Kneib and P. Bühlmann)
- arXiv 1508.06749 (with L. Möst and P. Bühlmann)
- `torsten.hothorn@R-project.org`