Using Statistical Models to Understand Child Eye Development

Julian Stander*, Luciana Dalla Valle*
Charlotte Taglioni† and Mario Cortina Borja‡

Centre for Mathematical Sciences, Plymouth University*, University of Padova† and UCL Institute of Child Health‡

The Eye Data

- Visual acuity is measured on 2810 children whose age is also recorded

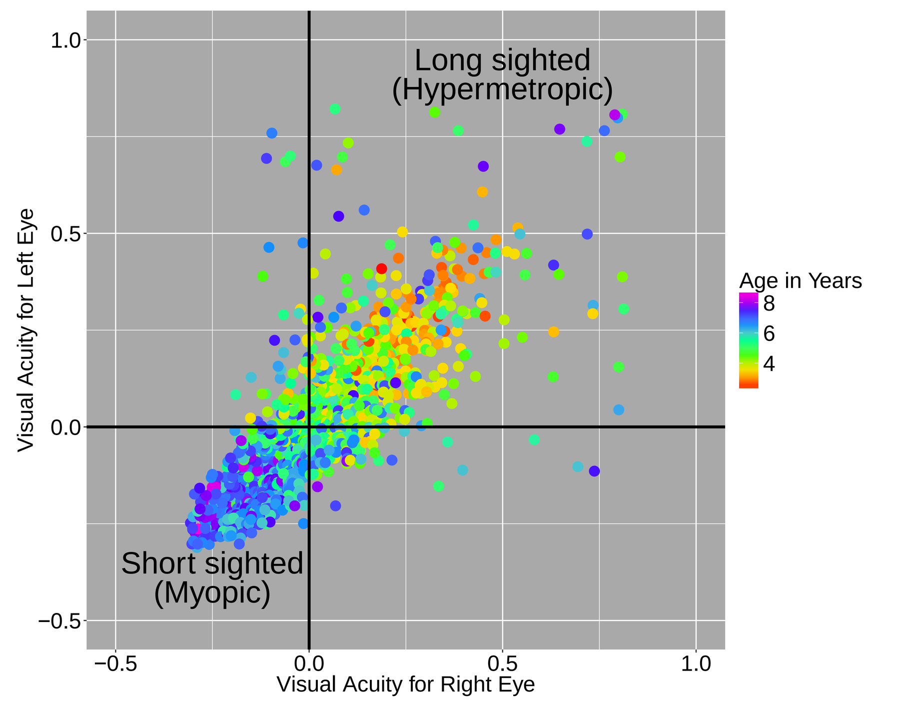

Fig. 1: The visual acuity of 2810 children with colour indicating age

- In general, as children grow, they become more short sighted
- We will present a statistical model for these data based on copulas and splines that helps us to quantify how sight changes with age
- Notation: for child $i = 1, \dots, n = 2810$, we have bivariate data (x_{1i}, x_{2i}) = (right visual acuity$_i$, left visual acuity$_i$) and a covariate $w_i = \text{age}_i$

What is a Copula?

- A bivariate copula is a probability density function with uniform marginals on $[0, 1]$
- An example of a copula density $c[u_1, u_2; \zeta]$, the shape of which is controlled by a parameter ζ, is shown here in the bottom left panel

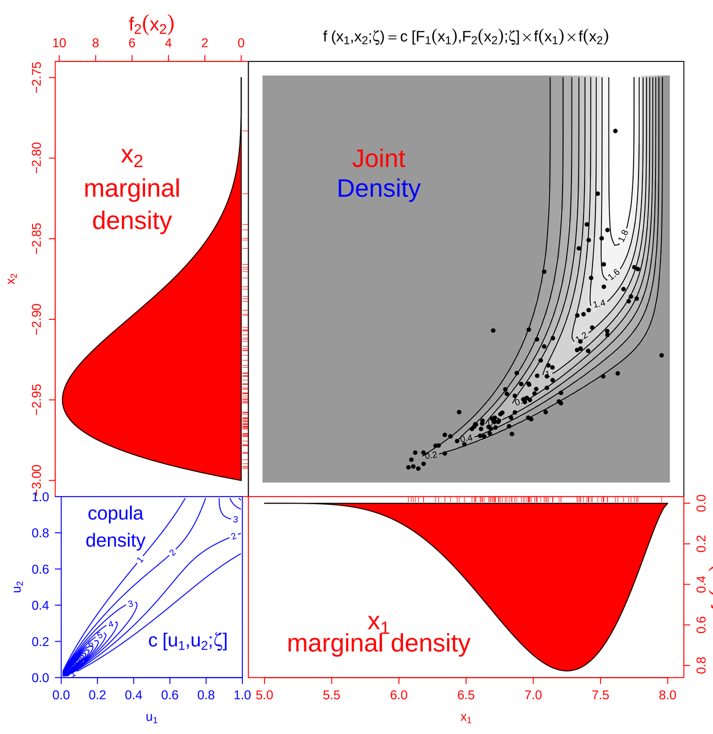

Fig. 2: The general bivariate density copula construction

- Flexible probability density functions can be created by combining c with marginal densities $f_1(x_1)$ and $f_2(x_2)$ (bottom right, top left) to produce a joint density $f(x_1, x_2; \zeta)$ (top right)
- This construction conceptually separates marginal from dependence modelling

Example of Copula Families

- There are many different families of copula densities $c[u_1, u_2; \zeta]$
- Here we see the 0°–90° double Gumbel family

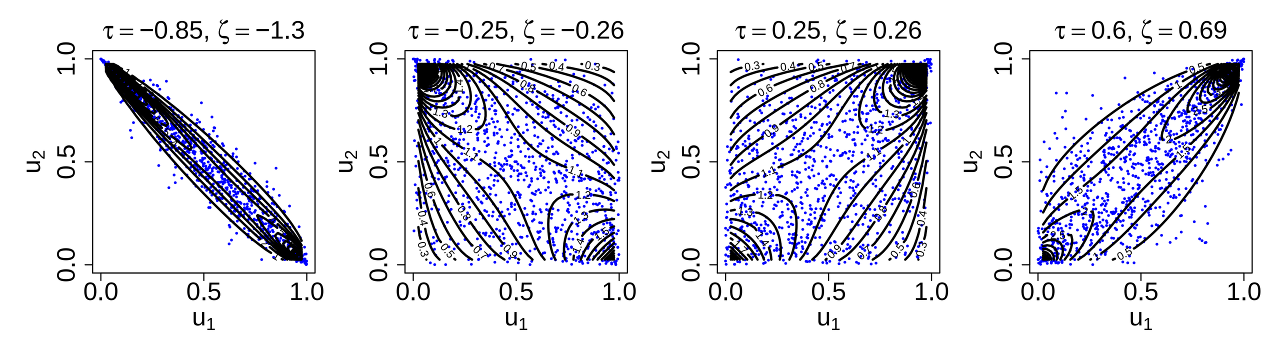

Fig. 3: The 0°–90° double Gumbel family

- We parametrize our copulas using $\zeta \in \mathbb{R}$, Fisher's transform of Kendall's correlation $\tau \in [-1, 1]$
- The first/second (third/fourth) copulas give negative (positive) correlation τ between u_1 and u_2 and hence have negative (positive) ζ
- Rotating 0°–90° double copulas by 180° gives the 180°–270° family
- We consider the normal, the 0°–90° & 180°–270° double Clayton and Gumbel, and the Frank copula families, all parameterized by ζ
- Following Thompson *et al.* (2010) and Craiu and Sabeti (2012), we aim to understand how the copula density shape, controlled by ζ, depends on a covariate w such as age by estimating the relationship g between ζ and w: $\zeta = g(w)$
- Like these authors we perform inference in the Bayesian framework
- We also consider two parameter copulas such as the BB7 so that we can control the shape of the upper and lower tails separately

What is a Spline?

- We assume that the curve g such that $\zeta = g(w)$ is smooth by taking it to be a natural cubic spline with N fixed knots at $\tau_1, \dots, \tau_N$:

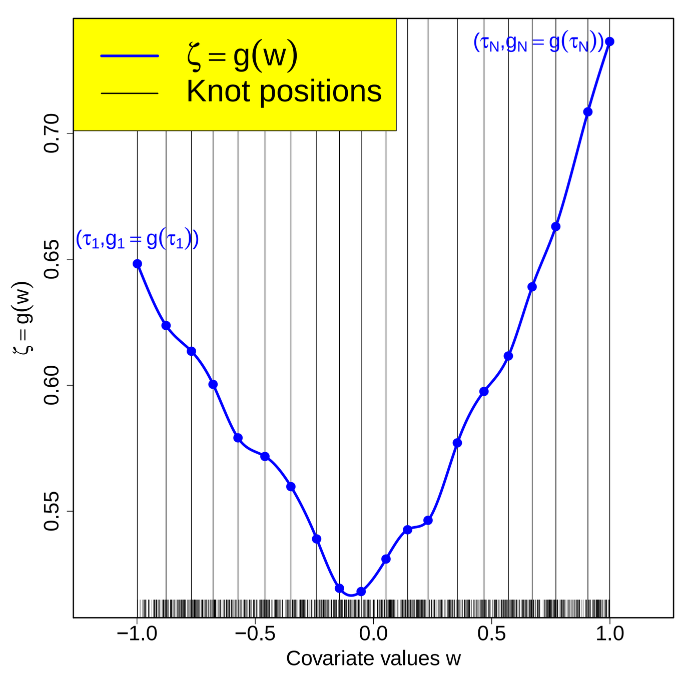

Fig. 4: Example of a natural cubic spline

- A natural cubic spline curve g is cubic between and linear beyond the knots, with continuous g'', a consequence of which is that g is uniquely defined by its knot values $\mathbf{g} = (g_1, \dots, g_N)^T$

A Statistical Model and Bayesian Inference

- Data model
  - Data (x_{1i}, x_{2i}) from child i (after marginal transformation to $U[0, 1]$) arises from a copula c with shape parameter ζ_i:

$$(x_{1i}, x_{2i}) \sim c\left[\cdot; \zeta_i\right], \; i = 1, \dots, n, \text{ independently}$$

  - ζ_i is related to the age covariate w_i through the natural cubic spline g: $\zeta_i = g(w_i)$
- We believe prior to seeing the data that g should be smooth in the sense of not wiggly, so we define

$$\pi(g \mid \lambda) \propto \lambda^{(N-2)/2} \exp\left(\frac{1}{2}\,\lambda \text{ smoothness of } g\right) = \lambda^{(N-2)/2} \exp\left(-\frac{1}{2}\lambda\, \mathbf{g}^T K \mathbf{g}\right) \text{ for a known matrix } K$$

  - Larger values of λ result in more probability density being given to smoother curves g
- Finally, we adopt a gamma hyper-prior for λ: $\pi(\lambda) \sim \Gamma(a, b)$
- Our Bayesian inference is based on the posterior distribution

$$\pi(g, \lambda \mid \mathbf{x}) \propto \pi(\mathbf{x} \mid g)\, \pi(g \mid \lambda)\, \pi(\lambda) = \left\{\prod_{i=1}^{n} c\left[x_{1i}, x_{2i} \mid \zeta_i = g(w_i)\right]\right\} \pi(g \mid \lambda)\, \pi(\lambda)$$

  in which $\mathbf{x} = ((x_{1i}, x_{2i}),\, i = 1, \dots, n)$ is all the data
- We simulate from $\pi(g, \lambda \mid \mathbf{x}) = \pi(\mathbf{g}, \lambda \mid \mathbf{x})$ using a Markov chain Monte Carlo algorithm

Results

- Of all the one and two parameter copula families considered, the one with the lowest Deviance Information Criterion (Spiegelhalter, *et al.* 2014) is the 180°–270° double Gumbel, which corresponds well to the data when divided into age groups
- Here is our Bayesian inference about $\zeta = g(w)$:

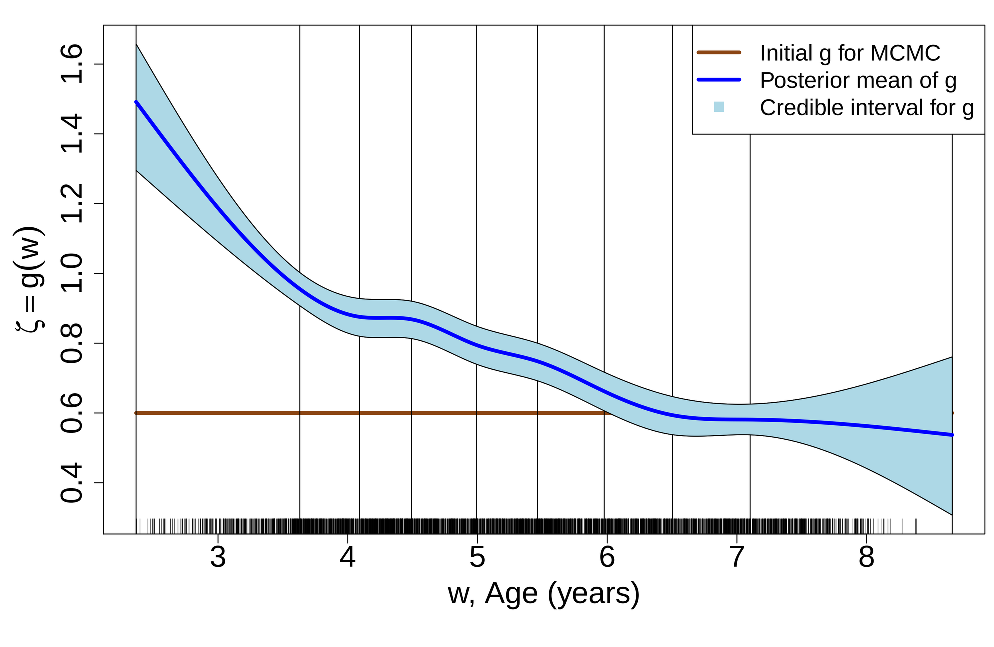

Fig. 5: The dependence on age of Fisher's transformation ζ of Kendall's correlation τ: $\zeta = g(\text{age})$

- Right and left eye visual acuity **correlation** drops from 0.9 for younger to 0.5 for older children
- Roughly speaking, eyes become less similar as children grow. More precisely, the lower tail of the visual acuity distribution is less concentrated for older children

Further Work

We plan to:

- fully include the modelling of the marginal densities in the Bayesian framework
- extend the model to more than one covariate
- extend the model from two to many dimensions using a vine based approach

This future work will be motivated by another application, this time from forensic science: we aim to understand how measurements of long bone structures such as overall height, shoulder and pelvis width depend on measurements taken from the skull

References

[1] Craiu, R. V. and Sabeti, A. (2012). In mixed company: Bayesian inference for conditional copulas models with discrete and continuous outcomes. *Journal of Multivariate Analysis* **110**, 106–120.

[2] Salt, A. T., Wade, A. M., Proffitt, R., Heavens, S. and Sonksen, P. M. (2007). The Sonksen logMAR Test of Visual Acuity: I. Testability and reliability. *Journal of American Association of Pediatric Ophthalmology and Strabismus* **11**, 589–596.

[3] Spiegelhalter, D. J., Best, N. G., Carlin, B. P and Van der Linde, A. (2014). The deviance information criterion: 12 years on. *J. R. Statist. Soc.* B **76**, 485–493.

[4] Thompson, P., Cai, Y., Moyeed, R., Reeve, D. and Stander, J. (2010). Bayesian nonparametric quantile regression using splines. *Computational Statistics and Data Analysis* **54**, 1138–1150.

LaTeX TikZposter