Complex Regression for Complex Data

Rosa M. Crujeiras and María Alonso-Pena

Statistics and Operations Research
University of Santiago de Compostela (Spain)

New Bridges between Mathematics and Data Science

Motivation

- Complex is not always big (of course dimensionality poses challenges and it is a complexity)
- Complex by their intrinsic nature (even though they may be usual quantities)
- Complex by relations:

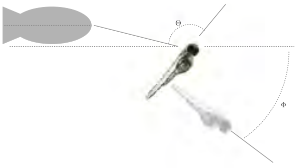
$$\mathbb{E}(Y|X=x) = m(x) \quad \textit{(mean regression)}$$

(the conditional mean approach may not be enough in some scenarios)

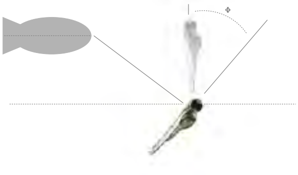
Goal of this talk: introduce a multimodal (complex) regression tool for circular (complex) data

Our example

- Escape behavior Φ (random circular variable)
- Stimulus: (robot) predator
- Experiment on larval zebrafish



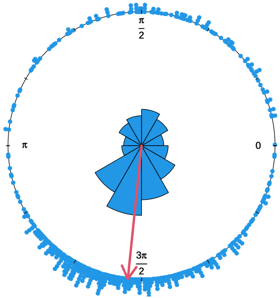
Contralateral

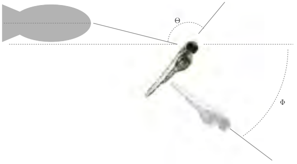


Ipsilateral

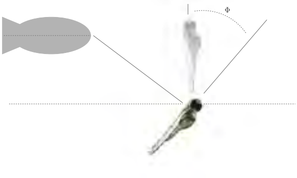
Nair, A., Changsing, K., Stewart, W.J. and McHenry, M.J. (2017)
Fish prey change strategy with the direction of a threat
Proceedings of the Royal Society B

Larvae escape direction: sample
$\Phi_i \in (-\pi, \pi]$, $i = 1, \dots, n$





Contralateral

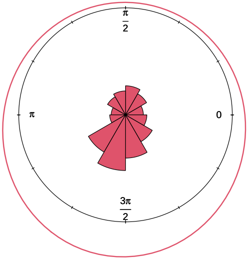


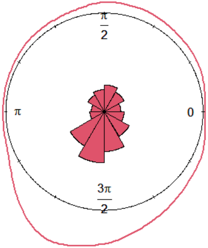
Ipsilateral

Circular mean

$$\hat{\mu} = \mathrm{atan2}\left(\frac{1}{n}\sum_{i=1}^{n}\sin\Phi_i, \frac{1}{n}\sum_{i=1}^{n}\cos\Phi_i\right)$$

Density estimation: parametric model vs. nonparametric density





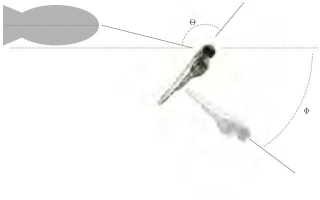
von Mises density

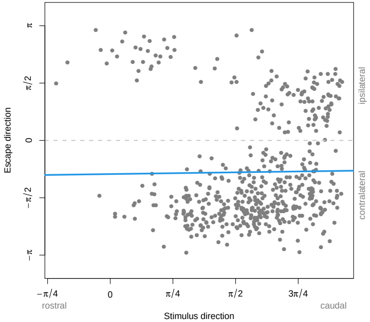
$$f(\phi;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa\cos(\phi-\mu)\}$$

kernel density estimator

$$\hat{f}(\phi;\kappa) = \frac{1}{n}\sum_{i=1}^{n} K_\kappa(\phi-\Phi_i)$$

In our example: (Θ_i, Φ_i), $i = 1, \ldots, n$, stimulus direction and escape direction

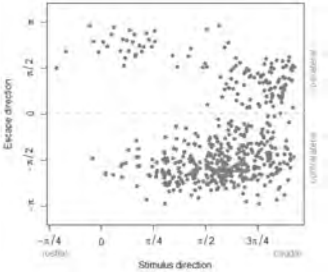




The naïve approach...

Nair, A., Changsing, K., Stewart, W.J. and McHenry, M.J. (2017)
Fish prey change strategy with the direction of a threat
Proceedings of the Royal Society B

In our example: (Θ_i, Φ_i), $i = 1, \ldots, n$, stimulus direction and escape direction



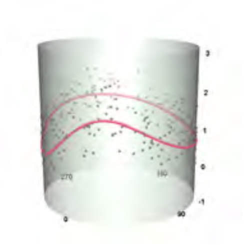
Nair, A., Changsing, K., Stewart, W.J. and McHenry, M.J. (2017)
Fish prey change strategy with the direction of a threat
Proceedings of the Royal Society B

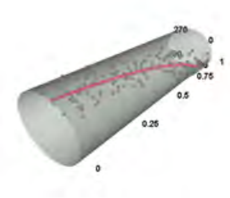
Circular regression

$Y = m(\Theta) + \varepsilon$

$\Phi = [m(X) + \varepsilon](\text{mod} 2\pi)$

$\Phi = [m(\Theta) + \varepsilon](\text{mod} 2\pi)$



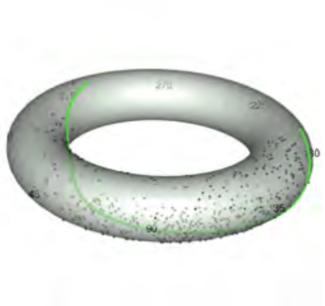


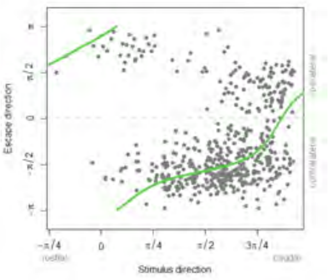
Circular predictor
Real-valued response
(Θ, Y)

Real-valued predictor
Circular response
(X, Φ)

Circular predictor
Circular response
(Θ, Φ)

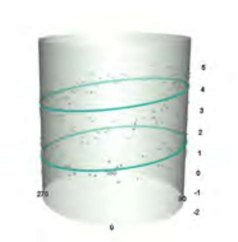
In our example: (Θ_i, Φ_i), $i = 1, \ldots, n$, stimulus direction and escape direction



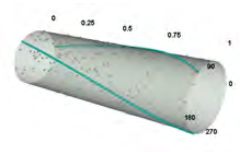


Di Marzio, Panzera and Taylor (2012)
Non-parametric regression for circular responses
Scandinavian Journal of Statistics

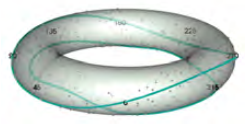
Multimodal circular regression



Circular predictor
Real-valued response
(Θ, Y)



Real-valued predictor
Circular response
(X, Φ)

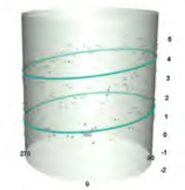


Circular predictor
Circular response
(Θ, Φ)

Multimodal circular regression: Circular predictor - Real-valued response

The regression multifunction is given by

$$M(\theta) = \left\{ y \in \mathbb{R} : \frac{\partial}{\partial y} f(y|\theta) = 0, \frac{\partial^2}{\partial y^2} f(y|\theta) < 0 \right\}$$



Estimation: indirect approach, estimating the conditional density

$$\hat{f}(y|\theta) = \frac{\sum_{i=1}^{n} K_\kappa(\theta - \Theta_i) L_h(y - Y_i)}{\sum_{i=1}^{n} K_\kappa(\theta - \Theta_i)}$$

Estimated regression multifunction:

$$\hat{M}(\theta) = \left\{ y \in \mathbb{R} : \frac{\partial}{\partial y} \hat{f}(y|\theta) = 0, \frac{\partial^2}{\partial y^2} \hat{f}(y|\theta) < 0 \right\}$$

Computation of the modes: conditional mean shift algorithm

The critical point condition yields a fixed-point equation:

$$\frac{\partial}{\partial y}\hat{f}(y|\theta) = 0 \iff y = \frac{\sum_{i=1}^{n} K_\kappa(\theta - \Theta_i) G\left(\frac{y-Y_i}{h}\right) Y_i}{\sum_{i=1}^{n} K_\kappa(\theta - \Theta_i) G\left(\frac{y-Y_i}{h}\right)}$$

Fukunaga, K. and Hostetler, L. (1975)
The estimation of the gradient of a density function, with applications in pattern recognition
IEEE Transactions on Information Theory

Chen, Y.C. et al. (2016)
Nonparametric modal regression
The Annals of Statistics

Cheng, Y. (1995)
Mean shift, mode seeking and clustering
IEEE Transactions on Pattern Analysis and Machine Intelligence

Einbeck, J. and Tutz, G (2006)
Modelling beyond regression functions: an application of multimodal regression to speed-flow data
Applied Statistics

Algorithm: circular predictor - real-valued response

Sample $\{(\Theta_i, Y_i)\}_{i=1}^n$, smoothing parameters κ and h.

1. Initialize mesh points $\mathcal{T} \subset (-\pi, \pi]$.
2. For each $\theta \in \mathcal{T}$, select starting points $y_0^{(1)}(\theta), ..., y_0^{(p)}(\theta)$.
3. For each $\theta \in \mathcal{T}$ and for $k = 1, ..., p$ iterate until convergence:

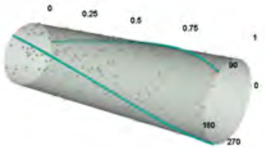
$$y_{l+1}^{(k)} = \frac{\sum_{i=1}^n K_\kappa(\theta - \Theta_i) G\left(\frac{y_l^{(k)} - Y_i}{h}\right) Y_i}{\sum_{i=1}^n K_\kappa(\theta - \Theta_i) G\left(\frac{y_l^{(k)} - Y_i}{h}\right)}, \quad \text{with} \quad l = 0, 1, ...$$

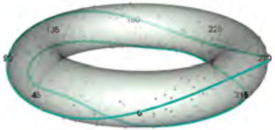
Multimodal circular regression: Circular response

The variable Δ denotes a general covariate ($\Delta = X$ or $\Delta = \Theta$)

The regression multifunction is given by

$$M(\delta) = \left\{ \phi \in \mathbb{S}^1 : \frac{\partial}{\partial \phi} f(\phi|\delta) = 0, \frac{\partial^2}{\partial \phi^2} f(\phi|\delta) < 0 \right\}$$





Zhang, Y. and Chen, Y.C. (To appear)
Kernel smoothing, mean shift and their learning theory with directional data
Journal of Machine Learning Research

Estimation of the regression multifunction (indirect approach)

- $\Delta = X$ (real-valued predictor)

$$\hat{f}(\phi|x)\frac{\sum_{i=1}^{n} L_h(x - X_i)K_\kappa(\phi - \Phi_i)}{\sum_{i=1}^{n} L_h(x - X_i)}$$

- $\Delta = \Theta$ (circular predictor)

$$\hat{f}(\phi|\theta)\frac{\sum_{i=1}^{n} K_\nu(\theta - \Theta_i)K_\kappa(\phi - \Phi_i)}{\sum_{i=1}^{n} K_\nu(\theta - \Theta_i)}$$

In short:

$$\hat{f}(\phi|\delta) = \frac{1}{n\hat{f}(\delta)}\sum_{i=1}^{n} w_\delta(\Delta_i)K_\kappa(\phi - \Phi_i), \quad \hat{f}(\delta) = \frac{1}{n}\sum_{i=1}^{n} w_\delta(\Delta_i)$$

Estimation of the regression multifunction

- Estimated regression multifunction

$$\hat{M}(\delta) = \left\{ \phi \in \mathbb{S}^1 : \frac{\partial}{\partial \phi} \hat{f}(\phi|\delta) = 0, \frac{\partial^2}{\partial \phi^2} \hat{f}(\phi|x) < 0 \right\}$$

- For the critical point condition:

$$\frac{\partial}{\partial \phi} \hat{f}(\phi|\delta) = \frac{\kappa c_\kappa}{n \hat{f}(\delta)} \sum_{i=1}^{n} w_\delta(\Delta_i) K'[\kappa(1 - \cos(\phi - \Phi_i))] \sin(\phi - \Phi_i)$$

Estimation of the regression multifunction

Expanding the last expression and equating to zero, we get:

$$\phi = \mathsf{atan2}\left(S_\delta(\phi), C_\delta(\phi)\right)$$

where we denote

$$S_\delta(\phi) = \sum_{i=1}^{n} w_\delta(\Delta_i) T(\phi - \Phi_i) \sin \Phi_i$$

$$C_\delta(\phi) = \sum_{i=1}^{n} w_\delta(\Delta_i) T(\phi - \Phi_i) \cos \Phi_i$$

and $T(\cdot)$ is proportional to $K'[\kappa(1 - \cos(\cdot))]$.

Algorithm: circular response

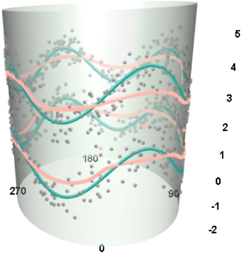
Sample $\{(\Delta_i, \Phi_i)\}_{i=1}^n$, smoothing parameters κ and h/ν.

1. Initialize mesh points $\mathcal{S} \subset \mathbb{R}$ if $\Delta = X$ or $\mathcal{T} \subset (-\pi, \pi]$ if $\Delta = \Theta$.
2. For each $\delta \in \mathcal{S}$ (or $\delta \in \mathcal{T}$), select starting points $\phi_0^{(1)}(\delta), ..., \phi_0^{(p)}(\delta)$
3. For $k = 1, ..., p$ iterate until convergence:

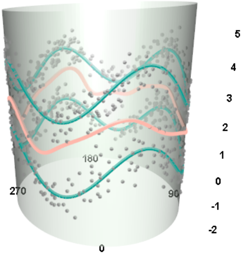
$$\phi_{l+1}^{(k)} = \text{atan2}\left(\sum_{i=1}^{n} w_\delta(\Delta_i) T(\phi_l^{(k)} - \Phi_i) \sin\Phi_i, \sum_{i=1}^{n} w_\delta(\Delta_i) T(\phi_l^{(k)} - \Phi_i) \cos\Phi_i\right)$$

with $l = 0, 1, ...$

Behavior of the smoothing parameters



Fixed h, varying κ



Fixed κ, varying h

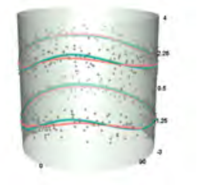
- The parameter associated to the predictor controls the smoothing
- The parameter associated to the response affects the number of estimated branches

Pointwise error for a real-valued response

$$\Lambda(\theta) = \mathsf{Haus}(M(\theta), \hat{M}(\theta))$$

$$\mathsf{Haus}(A, B) = \max\left\{\sup_{x\in A} d(x, B), \sup_{x\in B} d(x, A)\right\}$$

$$d(x, A) = \inf_{z\in A} |x - z|$$

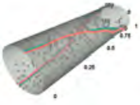


Pointwise error for a circular response

$$\tilde{\Lambda}(\delta) = \widetilde{\mathsf{Haus}}(M(\delta), \hat{M}(\delta))$$

$$\widetilde{\mathsf{Haus}}(A, B) = \max\left\{\sup_{x\in A} \tilde{d}(x, B), \sup_{x\in B} \tilde{d}(x, A)\right\}$$

$$\tilde{d}(x, A) = \inf_{z\in A} 1 - \cos(x - z)$$



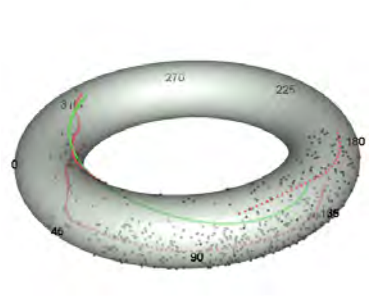
		$\tau = 6$		$\tau = 8$		$\tau = 10$	
Model	(n_1, n_2)	B	CV	B	CV	B	CV
LC-1	$(100, 100)$	0.021	0.034	0.016	0.023	0.013	0.018
	$(100, 200)$	0.016	0.039	0.012	0.024	0.011	0.018
	$(200, 200)$	0.011	0.019	0.008	0.012	0.007	0.010
	$(200, 300)$	0.010	0.016	0.007	0.011	0.006	0.009
	$(300, 300)$	0.007	0.012	0.006	0.008	0.005	0.006
CC-1	$(100, 100)$	0.144	0.395	0.117	0.213	0.102	0.148
	$(100, 200)$	0.120	0.182	0.097	0.147	0.087	0.120
	$(200, 200)$	0.066	0.089	0.054	0.066	0.047	0.057
	$(200, 300)$	0.058	0.068	0.046	0.054	0.040	0.045
	$(300, 300)$	0.042	0.056	0.035	0.045	0.030	0.038

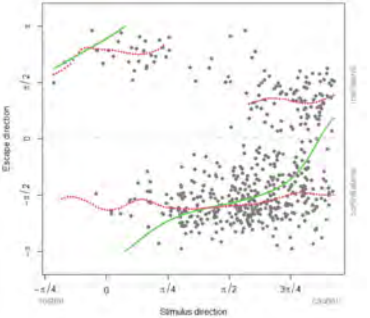
$\widetilde{\mathsf{CMIE}}_m(\hat{M}) = \mathbb{E}\left[\int \tilde{\Lambda}(\delta) d\delta\right] \to 0$ as n grows.

B: benchmark (optimal) CV: modal cross-validation selector.

τ: controls the data concentration (from less to more concentrated).

Back to the example: escape direction vs. predator direction approach

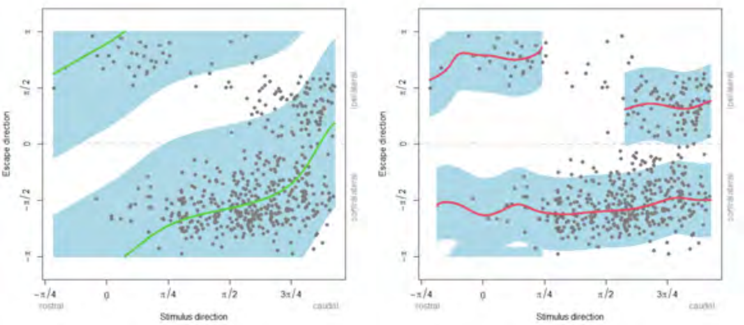




Mean regression: ipsilateral escape (from rostral approach); contralateral escape (from caudal approach).

Multimodal regression: two trends, ipsilateral/contralateral when stimulus appears from peripherical vision (rostral/caudal); third trend, contralateral, if stimulus is not peripherical.

Prediction sets



Thanks!

Grant PID2020-116587GB-I00 funded by
MCIN/AEI/10.13039/501100011033 and the European Union.

Alonso-Pena, M. and Crujeiras, R.M. (2021)
Nonparametric multimodal regression for circular data
arXiV (under 2nd revision)