SMALL BIASED CIRCULAR DENSITY ESTIMATION

Marco Di Marzio [1], Stefania Fensore[1], Agnese Panzera[2], and Charles C. Taylor[3]

[1] DSFPEQ, Università degli Studi di Chieti-Pescara,
(e-mail: mdimarzio@unich.it, stefania.fensore@unich.it)

[2] DiSIA, Università di Firenze,
(e-mail: a.panzera@disia.unifi.it)

[3] Departement of Statistics, Leeds University,
(e-mail: charles@maths.leeds.ac.uk)

ABSTRACT: We study higher order biased non-parametric estimators for circular densities. The idea is optimizing a local version of the log-likelihood function where the unknown log-density is replaced by a series expansion. It will be seen that the asymptotic bias will be reduced depending on the order of the expanding polynomial.

KEYWORDS: Circular data, Density estimation, Product kernels, Toroidal data.

1 Introduction

A circular (or directional) observation can be considered as a point on the circumference of the unit circle (or a unit vector in the plane) and measured (in radians) by an angle in $[-\pi, \pi)$ after both an origin and an orientation have been chosen. When dealing with circular data, the angle $\theta \in [-\pi, \pi)$ can be represented by any element in the set $\{2m\pi + \theta, m \in \mathbb{Z}\}$: this sets apart circular statistical analysis from standard real-line methods. Typical examples of circular data include flight direction of birds from a point of release, wind, and ocean current direction. A multidimensional directional observation lies in d-dimensional torus, i.e. the space $[-\pi, \pi)^d$, or in the hyper-sphere.

We propose a local likelihood approach for estimating toroidal densities.

To the best of author's knowledge, the only specific contribution on density estimation on the torus is due to Di Marzio *et al.*, 2011. Tibshirani & Hastie, 1987 introduced the concept of *local likelihood*. They proposed to fit a regression function using only the observations falling within a certain window around the estimation point. In the context of density estimation, local likelihood requires spatially weighting of the log-densities.

The local likelihood equations are obtained as an approximation of a counting process over the torus. A way to represent the *curse of dimensionality* in kernel density estimation is observing that, as the dimension increases, the classical bias-variance tradeoff is subject to failure since the optimal bandwidths must be large, and are generally too wide to avoid substantial bias. Therefore, higher order estimation should have the potential of improving the efficiency, especially in the regions where the bias is severe.

2 The toroidal likelihood model

Let f be a *toroidal* density, i.e. a 2π-periodic non-negative function defined on $[-\pi,\pi)^d$, $d \geq 1$, with $\int_{[-\pi,\pi)^d} f(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} = 1$. The objective is to estimate f at $\boldsymbol{\theta} \in [-\pi,\pi)^d$, using a random sample $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_n$ drawn from f.

Once the torus has been partitioned into S equal bins, say $C_1,\ldots,C_S$, let P_s and n_s denote the integral of f over C_s, $s \in (1,\ldots,S)$, and the number of sample observations in C_s, respectively. Then, with respect to C_s, each sample observation can be seen as a Bernoulli random variable with parameter P_s, and, due to the Mean Value Theorem, it results $P_s = f(\boldsymbol{\theta}_s)(2\pi)^d/S$, for some $\boldsymbol{\theta}_s \in C_s$, which is $O(S^{-1})$ for bounded f and finite d. Then, the likelihood is $L(f) := \prod P_s^{n_s}(1-P_s)^{n-n_s}$, and, letting $\ell_s := \log(P_s/(1-P_s))$, and the log-likelihood is $\mathcal{L}(f) := \sum\{n_s\ell_s - n\log\left(1+e^{\ell_s}\right)\}$.

Hence, assuming *sparse asymptotics*, i.e. both n and S go to infinity in such a way that just a few observations fall in each bin, we have that $n \approx S$, $n_s \to 1$, and $P_s \to 0$. Consequently, $\mathcal{L}(f) \approx \sum(\log P_s - nP_s)$. Using a spatial weight $\boldsymbol{K}_{\kappa_1,\ldots,\kappa_d}$, approximating the second sum with an integral, and ignoring constants, we motivate the following definition of local likelihood at $\boldsymbol{\theta} \in [-\pi,\pi)^d$,

$$\mathcal{L}_{\boldsymbol{\theta}}(f) := \sum_{i=1}^{n} \boldsymbol{K}_{\kappa_1,\ldots,\kappa_d}(\boldsymbol{\theta}_i - \boldsymbol{\theta})\log f(\boldsymbol{\theta}_i) - n \int\limits_{[-\pi,\pi)^d} \boldsymbol{K}_{\kappa_1,\ldots,\kappa_d}(\boldsymbol{\alpha}-\boldsymbol{\theta}) f(\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha},$$

where the weight function is a *toroidal kernel*, i.e. $\boldsymbol{K}_{\kappa_1,\ldots,\kappa_d}(\boldsymbol{\beta}) := \prod K_{\kappa_j}(\boldsymbol{\beta}^{(j)})$, $\boldsymbol{\beta} \in [-\pi,\pi)^d$, with K_{κ_j} being a *circular kernel* with zero mean direction and concentration parameter $\kappa_j \in (0,\infty)$, $j \in (1,\ldots,d)$ (see Di Marzio *et al.*, 2011). Then, for $\boldsymbol{\theta}_i$ belonging to a neighborhood of $\boldsymbol{\theta}$, the contribution of $\log f(\boldsymbol{\theta}_i)$ is weighted by $\boldsymbol{K}_{\kappa_1,\ldots,\kappa_d}(\boldsymbol{\theta}_i - \boldsymbol{\theta})$, with the neighborhood being as wide along the jth direction as small κ_j is. Clearly, here κ_j plays the role of the (inverse of the) smoothing factor along the jth dimension.

Consider the (2π-periodic) pth degree polynomial (Di Marzio *et al.*, 2009)

$$\mathcal{P}_p(\boldsymbol{\lambda}) := a_0 + \sum_{s=1}^{p} \frac{(\boldsymbol{S}'_{\boldsymbol{\lambda}})^{\otimes s} \boldsymbol{a}_s}{s!},$$

with $\boldsymbol{\lambda} \in [-\pi,\pi)^d$, $a_0 \in \mathbb{R}$, $\boldsymbol{a}_s \in \mathbb{R}^{d^s}$, $s \in (1,\ldots,p)$, $\boldsymbol{u}^{\otimes s}$ denoting the sth order Kronecker power of a vector $\boldsymbol{u}$ and $\boldsymbol{S}_{\boldsymbol{\lambda}} := (\sin(\boldsymbol{\lambda}^{(1)}),\ldots,\sin(\boldsymbol{\lambda}^{(d)}))'$.

Thus, for $\boldsymbol{a} := (a_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_p)'$, we see that $\mathcal{L}_{\boldsymbol{\theta}}(\boldsymbol{a})$ equals

$$\sum_{i=1}^{n} \boldsymbol{K}_{\kappa_1,\ldots,\kappa_d}(\boldsymbol{\theta}_i - \boldsymbol{\theta}) \mathcal{P}_p(\boldsymbol{\theta}_i - \boldsymbol{\theta}) - n \int\limits_{[-\pi,\pi)^d} \boldsymbol{K}_{\kappa_1,\ldots,\kappa_d}(\boldsymbol{\alpha} - \boldsymbol{\theta}) \exp(\mathcal{P}_p(\boldsymbol{\alpha} - \boldsymbol{\theta})) \mathrm{d}\boldsymbol{\alpha},$$

and equating to $\mathbf{0}$ the associated system of partial derivatives leads to

$$\frac{1}{n}\sum_{i=1}^{n} \mathcal{A}(\boldsymbol{\theta}_i - \boldsymbol{\theta}) \boldsymbol{K}_{\kappa_1,\ldots,\kappa_d}(\boldsymbol{\theta}_i - \boldsymbol{\theta}) = \int\limits_{[-\pi,\pi)^d} \mathcal{A}(\boldsymbol{\alpha} - \boldsymbol{\theta}) \boldsymbol{K}_{\kappa_1,\ldots,\kappa_d}(\boldsymbol{\alpha} - \boldsymbol{\theta}) \exp\left(\mathcal{P}_p(\boldsymbol{\alpha} - \boldsymbol{\theta})\right) \mathrm{d}\boldsymbol{\alpha}, \quad (1)$$

where $\mathcal{A}(\boldsymbol{\lambda}) := \mathrm{vec}(1, \boldsymbol{S}'_{\boldsymbol{\lambda}}, \ldots, (\boldsymbol{S}'_{\boldsymbol{\lambda}})^{\otimes p})$. Under the assumption that $\log f$ is smooth enough at $\boldsymbol{\theta}$, solving for $\boldsymbol{a}$ gives the estimates $\hat{\boldsymbol{a}} := (\hat{a}_0, \hat{\boldsymbol{a}}_1, \cdots, \hat{\boldsymbol{a}}_p)'$ of $\tilde{\boldsymbol{a}} := (\tilde{a}_0, \tilde{\boldsymbol{a}}_1 \ldots, \tilde{\boldsymbol{a}}_p)'$, where, for $\boldsymbol{\theta} \in [-\pi,\pi)^d$, $\tilde{a}_0 := \log f(\boldsymbol{\theta})$ and $\tilde{\boldsymbol{a}}_s$ is the vector of the mixed partial derivatives of total order s of $\log f$ at $\boldsymbol{\theta}$. For example, $\tilde{\boldsymbol{a}}_1$ is the gradient vector, and $\tilde{\boldsymbol{a}}_2 = \mathrm{vec}(\mathrm{H})$, where H denotes the Hessian matrix. Clearly, the local-likelihood density estimator of $f(\boldsymbol{\theta})$ is $\hat{f}(\boldsymbol{\theta}) := \exp(\hat{a}_0)$. Notice that, when $p = 0$, system (1) reduces to a single equation whose solution coincides with the kernel estimator of a toroidal density considered by Di Marzio *et al.*, 2011; when $p > 0$ the estimate is always non-negative, whereas it is not guaranteed that it is a proper density, and a normalization step becomes necessary after estimation. Loader, 1996 introduced a likelihood approach that could be considered as an euclidean counterpart of ours. If $p = 0$ the estimator has a closed form solution. This allows us to calculate the accuracy measures explicitly. For the circle case we get a bias equal to $1/2 f''(\theta)/f(\theta) \int \sin^2 K_\kappa$, and a variance equal to $1/(nf(\theta)) \int K_\kappa^2$. A comparison between local constant and local linear fit of a_0 shows that, when $d = 1$, this latter has bias $1/2(f''(\theta)/f(\theta) - f'(\theta)^2/f(\theta)^2) \int \sin^2 K_\kappa$, whereas the variance is identical for both. We can observe that the respective convergence rates are the same as well. In Figure 1 we show a few population examples and highlight

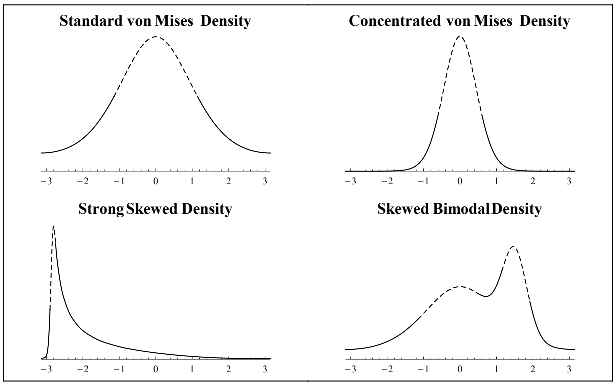

Figure 1: Population examples. Continuous line indicates where local linear is more accurate, dashed line indicates where local constant is superior.

which estimator is better. Around the modes, where $f'' < 0$ we have smaller bias for the local constant fit, while the opposite situation happens at the tail regions. Notice that at the stationary points, where $f' = 0$, biases are identical.

References

DI MARZIO, M., PANZERA, A., & TAYLOR, C.C. 2009. Local polynomial regression for circular predictors. *Statistics & Probability Letters.*, **79**, 2066–2075.

DI MARZIO, M., PANZERA, A., & TAYLOR, C.C. 2011. Kernel density estimation on the torus. *Journal of Statistical Planning and Inference.*, **141**, 2156–2173.

LOADER, C.R. 1996. Local likelihood density estimation. *The Annals of Statistics.*, **24**, 1602–1618.

TIBSHIRANI, R., & HASTIE, T. 1987. Local likelihood estimation. *Journal of the American Statistical Association.*, **82**, 559–567.