MR4779847 62G05 62G07 62G20 62H05

Ameijeiras-Alonso, Jose (E-SACO-MRT); **Gijbels, Irène** (B-KUL)

Smoothed circulas: nonparametric estimation of circular cumulative distribution functions and circulas. (English. English summary)

Bernoulli **30** (2024), *no.* 4, 2747–2769.

In this article, the authors are concerned with nonparametric estimation of a multivariate cumulative distribution function (cdf) for the d-dimensional random angles $\boldsymbol{\Theta} = (\Theta_1, \dots, \Theta_d)$ on the torus $\mathbb{T}^d = [-\pi, \pi)^d$. To handle the dependencies among components of $\boldsymbol{\Theta}$, they specialize in kernel estimation of a circula $C\colon \mathbb{T}^d \mapsto [0,1]$ under a more general setup than in the previous study. A circula, an analogue to a copula on the Euclidean space, is given by a multivariate toroidal distribution whose marginals are circular uniform distributions.

Recognizing that if $\Theta \in [-\pi, \pi)$ is a circular random variable with a cdf F, then $2\pi F(\Theta) - \pi$ obeys a circular uniform distribution, the authors propose to estimate the circula C nonparametrically in two steps. First, the marginal cdf F_s of each component Θ_s, $s = 1, \dots, d$, is estimated nonparametrically. For a univariate second-order circular density kernel $k_{s;\rho_s}$ with a concentration parameter $\rho_s \in [0,1]$ (an analogue to the bandwidth on the Euclidean space), the circular marginal kernel is defined as $K_{s;\rho_s}(\theta_s) := \int_{-\pi}^{\theta_s} k_{s;\rho_s}(\phi)d\phi$. Then, given n observations $\{\Theta_{i,s}\}_{i=1}^n$, the cdf F_s can be estimated as

$$\widehat{F}_{s;\rho_s}(\theta_s) := \frac{1}{n} \sum_{i=1}^{n} \left(K_{s;\rho_s}(\theta_s - \Theta_{i,s}) - K_{s;\rho_s}(-\pi - \Theta_{i,s}) \right).$$

Second, from the marginals estimators $\{(\widehat{F}_{1,\rho_1}(\Theta_{i,1}), \dots, \widehat{F}_{d,\rho_d}(\Theta_{i,d}))\}_{i=1}^n$ obtained in the first step, the estimator of the circula C is defined as

$$\widehat{C}_{\boldsymbol{\nu},\boldsymbol{\rho}}(\boldsymbol{\theta}) := \frac{1}{n} \sum_{i=1}^{n} \prod_{s=1}^{d} \left(K_{s;\nu_s}\left(\theta_s - 2\pi \widehat{F}_{s;\rho_s}(\Theta_{i,s}) + \pi\right) - K_{s;\nu_s}\left(-2\pi \widehat{F}_{s;\rho_s}(\Theta_{i,s})\right)\right).$$

Such kernel estimator is referred to as the *smoothed circula*.

The authors also explore asymptotic properties of the d-dimensional cdf estimator $\widehat{\boldsymbol{F}}_{\boldsymbol{\rho}}(\boldsymbol{\theta}) := \prod_{s=1}^{d} \widehat{F}_{s;\rho_s}(\theta_s)$ and the smoothed circula $\widehat{C}_{\boldsymbol{\nu},\boldsymbol{\rho}}(\boldsymbol{\theta})$ under a set of regularity conditions. Focusing on the von Mises and several other univariate second-order circular density kernels, they derive asymptotic bias, variance and mean integrated squared error (AMISE) of each of $\widehat{\boldsymbol{F}}_{\boldsymbol{\rho}}(\boldsymbol{\theta})$ and $\widehat{C}_{\boldsymbol{\nu},\boldsymbol{\rho}}(\boldsymbol{\theta})$. There are two interesting findings. First, the optimal AMISE rates for the cdf estimator differ, depending on the specific second-order circular kernel chosen, unlike the cases of second-order kernels on the Euclidean space. Second, as long as the same kernel is employed in both steps, the AMISE for the smoothed circula can be simplified only through undersmoothing the first-step cdf estimation. However, if two different kernels are employed in each step, then their AMISE-optimal concentration parameters in the cdf estimation have different shrinkage rates, and the resulting plug-in rule can circumvent undersmoothing in the first step.

Masayuki Hirukawa

[References]

1. Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York. MR0415962
2. Ameijeiras-Alonso, J. and Gijbels, I. (2024). Supplement to “Smoothed circulas: Nonparametric estimation of circular cumulative distribution functions and circulas.” https://doi.org/10.3150/23-BEJ1693SUPPB

3. Bowman, A., Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika* **85** 799–808. MR1666695 https://doi.org/10.1093/biomet/85.4.799 MR1666695
4. Card, G. and Dickinson, M.H. (2008). Visually mediated motor planning in the escape response of Drosophila. *Curr. Biol.* **18** 1300–1307. https://doi.org/10.1016/j.cub.2008.07.094
5. Chen, S.X. and Huang, T.-M. (2007). Nonparametric estimation of copula functions for dependence modelling. *Canad. J. Statist.* **35** 265–282. MR2393609 https://doi.org/10.1002/cjs.5550350205 MR2393609
6. Di Marzio, M., Panzera, A. and Taylor, C.C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.* **79** 2066–2075. MR2571770 https://doi.org/10.1016/j.spl.2009.06.014 MR2571770
7. Di Marzio, M., Panzera, A. and Taylor, C.C. (2011). Kernel density estimation on the torus. *J. Statist. Plann. Inference* **141** 2156–2173. MR2772221 https://doi.org/10.1016/j.jspi.2011.01.002 MR2772221
8. Di Marzio, M., Panzera, A. and Taylor, C.C. (2012). Smooth estimation of circular cumulative distribution functions and quantiles. *J. Nonparametr. Stat.* **24** 935–949. MR2995485 https://doi.org/10.1080/10485252.2012.721517 MR2995485
9. Duong, T. (2016). Non-parametric smoothed estimation of multivariate cumulative distribution and survival functions, and receiver operating characteristic curves. *J. Korean Statist. Soc.* **45** 33–50. MR3456320 https://doi.org/10.1016/j.jkss.2015.06.002 MR3456320
10. Fernández-Durán, J.J. (2007). Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics* **63** 579–585. MR2370817 https://doi.org/10.1111/j.1541-0420.2006.00716.x MR2370817
11. Fisher, N.I. and Lee, A.J. (1983). A correlation coefficient for circular data. *Biometrika* **70** 327–332. MR0712021 https://doi.org/10.1093/biomet/70.2.327 MR0712021
12. Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 1096–1127. MR2469337 https://doi.org/10.1214/07-AIHP148 MR2469337
13. Genest, C., Rémillard, B. and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance Math. Econom.* **44** 199–213. MR2517885 https://doi.org/10.1016/j.insmatheco.2007.10.005 MR2517885
14. Heredia-Zavoni, E. and Montes-Iturrizaga, R. (2019). Modeling directional environmental contours using three dimensional vine copulas. *Ocean Eng.* **187** 106102.
15. Jammalamadaka, S.R. and Ramakrishna Sarma, Y. (1988). A correlation coefficient for angular variables. In *Statistical Theory and Data Analysis, II (Tokyo, 1986)* 349–364. Amsterdam: North-Holland. MR0999196 MR0999196
16. Johnson, R.A. and Wehrly, T. (1977). Measures and models for angular correlation and angular-linear correlation. *J. Roy. Statist. Soc. Ser. B* **39** 222–229. MR0494693 MR0494693
17. Jones, M.C. (1990). The performance of kernel density functions in kernel distribution function estimation. *Statist. Probab. Lett.* **9** 129–132. MR1045173 https://doi.org/10.1016/0167-7152(92)90006-Q MR1045173
18. Jones, M.C., Pewsey, A. and Kato, S. (2015). On a class of circulas: Copulas for circular distributions. *Ann. Inst. Statist. Math.* **67** 843–862. MR3390169 https://doi.org/10.1007/s10463-014-0493-6 MR3390169
19. Jupp, P.E. (2015). Copulae on products of compact Riemannian manifolds. *J. Multivariate Anal.* **140** 92–98. MR3372555 https://doi.org/10.1016/j.jmva.2015.04.

008 MR3372555
20. Jupp, P.E. and Mardia, K.V. (1980). A general correlation coefficient for directional data and related regression problems. *Biometrika* **67** 163–173. MR0570518 https://doi.org/10.1093/biomet/67.1.163 MR0570518
21. Justel, A., Peña, D. and Zamar, R. (1997). A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statist. Probab. Lett.* **35** 251–259. MR1484961 https://doi.org/10.1016/S0167-7152(97)00020-5 MR1484961
22. Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics. Wiley Series in Probability and Statistics*. Chichester: Wiley. MR1828667 MR1828667
23. Oliveira, M., Crujeiras, R.M. and Rodríguez-Casal, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Comput. Statist. Data Anal.* **56** 3898–3908. MR2957840 https://doi.org/10.1016/j.csda.2012.05.021 MR2957840
24. Omelka, M., Gijbels, I. and Veraverbeke, N. (2009). Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. *Ann. Statist.* **37** 3023–3058. MR2541454 https://doi.org/10.1214/08-AOS666 MR2541454
25. Pewsey, A., Lewis, T. and Jones, M.C. (2007). The wrapped t family of circular distributions. *Aust. N. Z. J. Stat.* **49** 79–91. MR2345412 https://doi.org/10.1111/j.1467-842X.2006.00465.x MR2345412
26. Rivest, L.-P. (1982). Some statistical methods for bivariate circular data. *J. Roy. Statist. Soc. Ser. B* **44** 81–90. MR0655377 MR0655377
27. Shieh, G.S., Zheng, S., Johnson, R.A., Chang, Y.-F., Shimizu, K., Wang, C.-C. and Tang, S.-L. (2011). Modeling and comparing the organization of circular genomes. *Bioinformatics* **27** 912–918.
28. Shimizu, K. and Iida, K. (2002). Pearson type VII distributions on spheres. *Comm. Statist. Theory Methods* **31** 513–526. MR1902308 https://doi.org/10.1081/STA-120003131 MR1902308
29. Taylor, C.C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Statist. Data Anal.* **52** 3493–3500. MR2427365 https://doi.org/10.1016/j.csda.2007.11.003 MR2427365
30. Tsuruta, Y. and Sagae, M. (2017). Asymptotic property of wrapped Cauchy kernel density estimation on the circle. *Bull. Inform. Cybernet.* **49** 1–10. MR3728316 MR3728316

Note: This list reflects references listed in the original paper as accurately as possible with no attempt to correct errors.