Nonparametric penalized likelihood for density estimation

Stima della densità non-parametrica basata su verosimiglianza penalizzata

Federico Ferraccioli, Laura M. Sangalli and Livio Finos

Abstract In this work we consider a nonparametric likelihood approach to multivariate density estimation with a regularization based on the Laplace operator. The complexity of the estimation problem is tackled by means of a finite element formulation, that allows great flexibility and computational tractability. The model is suitable for any type of bounded planar domain and can be generalized to the non-Euclidean settings. Within this framework, we as well discuss a new approach to clustering based on the concept of diffusion in a potential field, and a permutation-based procedure for one and two samples hypothesis testing.

Abstract *In questo lavoro viene considerato un metodo di stima della densità tramite verosimiglianza non parametrica con un termine di regolarizzazione basato sull'operatore di Laplace. Il problema di stima è risolto attraverso l'uso di una formulazione ad elementi finiti, che assicura elevata flessibilità e trattabilità computazionale. Il modello è adatto per qualsivoglia tipo di dominio planare chiuso e può essere generalizzato al caso non Euclideo. Si propone un approccio al clustering basato sul concetto di diffusione, insieme ad una procedura di test di ipotesi per uno e due campioni basata su permutazione.*

Key words: finite elements, Laplace operator, mode clustering.

Federico Ferraccioli
Dipartimento di Scienze Statistiche, Via Cesare Battisti, 241, 35121 Padova (Italy)
e-mail: ferraccioli@stat.unipd.it

Laura M. Sangalli
MOX-Dipartimento di Matematica, Piazza L. da Vinci, 32, 20133 Milano (Italy)
e-mail: laura.sangalli@polimi.it

Livio Finos
Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Via Venezia, 8, 35131 Padova (Italy) e-mail: livio.finos@unipd.it

1 Introduction

The problem of density estimation plays a central role in statistics. It is a fundamental tool for the visualization of structure in exploratory data analysis, and it may be used as intermediate procedure in classification and custering problems. The last decades have seen enormous amount of research focused on kernel density estimation [10]. Simplicity of use and elegant analytic results are the key features of the success of the kernel approach. However, bandwidth selection remains a crucial problem for this method. Moreover, despite recent progress on the asymptotic convergence of the errors, good finite sample performance is by no means guaranteed and many practical challenges remain. The problem get even worse in the multidimensional setting, where the specification of a symmetric, positive definite bandwith matrix is needed, although it's common practice to use diagonal matrices.

Beside the class of kernel density estimators, many other smoothing methods for density estimation have been proposed. All these estimators are based on the idea of reducing the complexity of the problem with some type of approximations or some form of constraint on the space of solutions. In the former case, the approximation is given by basis expasion such as wavelets [4] or splines [9]. In the latter case, the two most prominent approaches are based on regularization of the likelihood functional [8] or shape constraints on the density, e.g. log-concavity [3]. The latter is a parameter-free method but at the cost of a severe restriction on model flexibility. Regularized likelihood methods are extremely flexible but because of the computational complexity they never reached popularity.

In this work we present a new nonparametric likelihood approach to density estimation. The model is based on a finite element formulation and can deal with data distributed over non-regular planar domains. We also briefly discuss a density based clustering method and a permutation based procedure for goodness of fit and for two-samples hypothesis testing.

2 Methodology

2.1 Classical approach

The problem of nonparametric maximum likelihood estimation, in the univariate case, has been considered for the first time in [6]. Let $X_1, \dots, X_n$ i.i.d. observations with distribution function F and density f on a bounded domain $\Omega \in \mathbb{R}$. Without further assumptions, the maximum likelihood estimator for f is not well defined. The likelihood function is unbounded above and the maximization procedure returns the trivial solution of sum of delta functions at the observations. Unlike classical parametric likelihood estimation, where the parameter space is finite, the estimator belongs to an infinite class of functions and some type of regularization becomes necessary to obtain a non-degenerate solution. The basic approach is to maximize a

score ω, depending on f and on the observations, defined by

$$\omega = \omega(f) = L - \alpha R(f)\,, \tag{1}$$

where $L = \sum_i \log f(x_i)$ is the log-likelihood, $R(f)$ is the roughness penalty, and the parameter $\alpha > 0$ controls the amount of smoothness. The authors consider, as measure of the roughness or complexity, the functional $R(f) = ||(\sqrt{f})^{(1)}||_2^2$, where the square root permits to avoid the positive contraints on the density. Further developments of this model are presented in [8], where the author considers a regularization functional of the form $R(f) = ||(\log f)^{(3)}||_2^2$. In this case the limiting estimate, as α tends to infinity, is the normal density with the same mean and variance as the data. Note that in this case the positive constraint is avoided by means of the logarithm transformation. Although both models could be generalized to the multivariate setting, consistency results and implementation are given only in the univariate case.

2.2 Model and estimation procedure

In this work we propose a generalization to the estimation of density defined over bounded planar domains. Suppose we observe $X_1, \ldots, X_n$ i.i.d. observations drawn from a distribution F on a bounded planar domain $\Omega \in \mathbb{R}^2$. Instead of considering the density f, let us define the log density $g = \log f$, where g is a real function on Ω. This transformation is particularly convenient from the theoretical as well as the practical point of view.
We are interested in a penalized maximum likelihood estimation for g. As previously stated, some types of regularization are necessary, in order to restrict the class of possible solutions. More formally, we consider the estimator that is a solution of the optimization problem

$$\text{minimize} \qquad -\frac{1}{n}\sum_{i=1}^{n} g(X_i) + \int_\Omega \exp(g(x))\,dx + \lambda R(g) \tag{2}$$

$$\text{subject to} \qquad g \in \mathscr{H}^2(\Omega)\,, \tag{3}$$

where $\mathscr{H}^2(\Omega)$ is Sobolev space of functions with continuous weak derivatives up to the second order. As pointed out by [8], the second term of the functional ensures the unitary contraint on the density. We consider here the penalization functional $R(f) = \int_\Omega (\Delta \log f)^2\,dx$, where Δ is the Laplace operator. The Laplacian is a measure of local curvature that is invariant with respect to Euclidean transformations of spatial coordinates, and therefore ensures that the concept of smoothness does not depend on the orientation of the coordinate system.
A more complex prior knowledge concerning the domain, which can be translated into a partial differential operator, could be incorporated in the regularization term. We shall consider linear second order elliptic operators of the form

$$Lg = -\text{div}(K\nabla g) + b\nabla g + cf \tag{4}$$

The diffusion term $-\text{div}(K\nabla g)$ induces a smoothing with a preferential direction that corresponds to the first eigenvector of the diffusion tensor K. The degree of anisotropy is controlled by the ratio between its first and second eigenvalue. The transport term $b\nabla g$ induces a smoothing only in the direction specified by the transport vector b. Finally, the reaction term cf has instead a shrinkage effect towards a uniform density on the domain.

Likewise in [7] and [1], the estimation problem is tackled by means of finite element method (FEM), a methodology mainly developed and used in engineering applications, to solve partial differential equations. The strategy of finite element analysis is very similar in spirit to that of univariate splines, and consists of partitioning the problem domain into small disjoint sub-domains and defining polynomial functions on each of these sub-domains in such a way that the union of these pieces closely approximates the solution. Convenient domain partitions are given for instance by triangular meshes. The simplified problem is made computationally tractable by the choice of the basis functions for the space of piecewise polynomials on the domain partition. Each piece of the partition, equipped with the basis functions defined over it, is named a finite element.

Unlike kernel density estimation, the proposed approach admits a likelihood formulation and it's well defined on any domain Ω, without necessity for boundary correction. The absence of constraints on f allows the estimation of extremely complex structures, a fundamental feature in research areas such as density based clustering. Based on the proposed method, we shall in particular discuss a clustering procedure stimulated by Morse theory [2]. We shall also introduce one and two-sample nonparametric tests, based on a permutation approach.

3 Simulation study

Let us consider a complex domain, defined by the closed annulus $ann(a;r,R) = \{x \in \mathbb{R}^2 : r \leq ||x-a|| \leq R\}$, where a is the center and (r,R) the internal and external radii. In our case we consider the annulus centered at the origin, with internal radius 3 and external radius 5. The distribution we have defined on the annulus is the joint probability $\theta \sim \text{Unif}(0, 2\pi)$ and a truncated Gaussian distribution in the interval $[1,1]$ with zero mean and standard deviation $\sigma = 0.3$. The Gaussian defines a random distance from the circle with center the origin and radius 4, in the direction normal to the perimeter. This domain includes complex characteristics such as nonlinear boundaries and holes. Despite the presence of complex domains in real data, none of the standard methods in density estimation is appropriate for these problems.

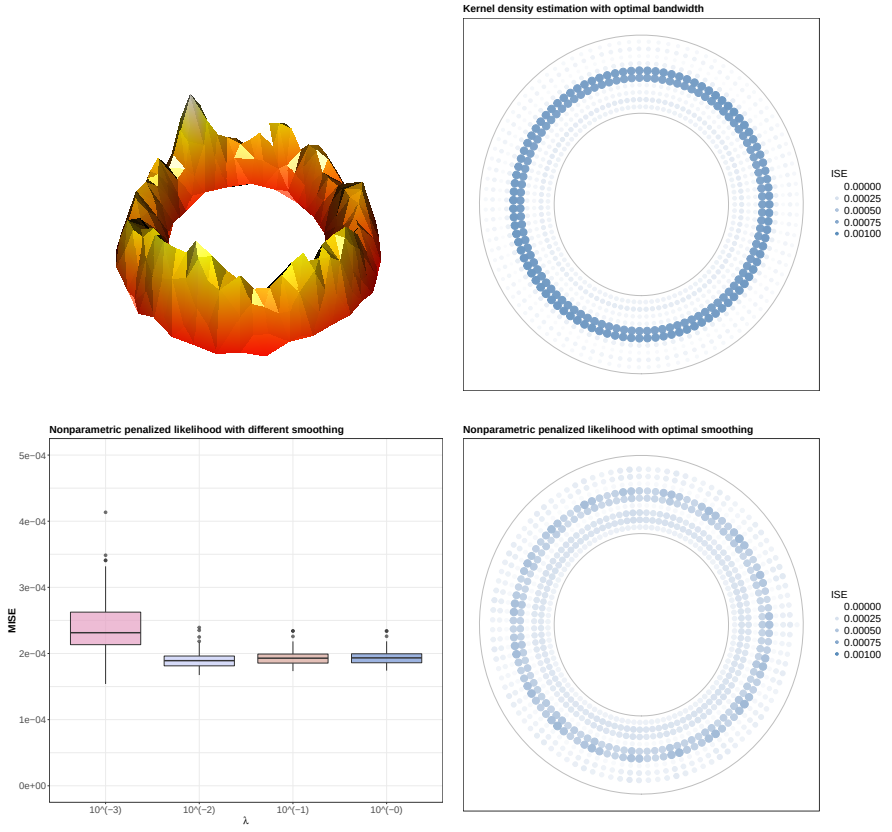

Fig. 1 On the left, an example of estimated density (top) and the distrubution of the MISE of 1000 simulation for different values of the smoothing parameters (bottom). On the right, the MSE surface of 1000 simulation for the nonparametric likelihood (top) and the KDE estimator (bottom), respectively.

4 Model extensions and conclusions

The proposed model performs well with respect to the state of the art of density estimators, while reducing the number of parameters to be selected. The estimator is also well defined on every bounded planar domain. The model can be generalized to non-euclidean setting, e.g. manifolds [5], and a time dependency can be included. These two features are extremely important in applications such as the study of brain activity, where the distribution of the signals over an highly convoluted domain, the cerebral cortex, changes over time. To the best of the authors knowledge, none of the existing methods is appropriate for this type of problems. Clustering procedures and two-samples testing based on the proposed estimator are also presented. Future

works will consider convergence of the estimators, consistency in the multivariate case and time-dependent generalizations.

References

[1] Laura Azzimonti et al. "Mixed finite elements for spatial regression with PDE penalization". In: *SIAM/ASA Journal on Uncertainty Quantification* 2.1 (2014), pp. 305–335.

[2] José E Chacón et al. "A population background for nonparametric density-based clustering". In: *Statistical Science* 30.4 (2015), pp. 518–532.

[3] Madeleine Cule, Richard Samworth, and Michael Stewart. "Maximum likelihood estimation of a multi-dimensional log-concave density". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.5 (2010), pp. 545–607.

[4] David L Donoho et al. "Density estimation by wavelet thresholding". In: *The Annals of Statistics* (1996), pp. 508–539.

[5] Bree Ettinger, Simona Perotto, and Laura M Sangalli. "Spatial regression models over two-dimensional manifolds". In: *Biometrika* 103.1 (2016), pp. 71–88.

[6] IJ Good and RA Gaskins. "Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data". In: *Journal of the American Statistical Association* 75.369 (1980), pp. 42–56.

[7] Laura M Sangalli, James O Ramsay, and Timothy O Ramsay. "Spatial spline regression models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75.4 (2013), pp. 681–703.

[8] Bernard W Silverman. "On the estimation of a probability density function by the maximum penalized likelihood method". In: *The Annals of Statistics* (1982), pp. 795–810.

[9] Grace Wahba. *Spline models for observational data*. Vol. 59. Siam, 1990.

[10] Matt P Wand and M Chris Jones. *Kernel smoothing*. Crc Press, 1994.