

Nonparametric Mode Hunting

Metodi non parametrici per riconoscere le mode di una densità

Marco Perone Pacifico

Abstract We propose a procedure for detecting the modes of a density estimate and test their significance. We use a data-splitting approach: potential modes are identified using the first half of the data and their significance is tested with the second half of the data. The mode test is based on nonparametric confidence intervals for the eigenvalues of the Hessian. In order to get valid bootstrap confidence sets even in presence of multiplicity of the eigenvalues, we use a bootstrap based on an elementary-symmetric-polynomial transformation.

Abstract *Si propone un metodo per identificare le mode di una stima di densità e verificarne la significatività. L'approccio prevede la suddivisione in due del campione: la prima metà dei dati viene usata per individuare le possibili mode, la cui significatività viene poi verificata sfruttando la seconda metà del campione. Il test è basato su intervalli di confidenza non parametrici per gli autovalori dell'hessiana. Al fine di ottenere intervalli validi anche in presenza di autovalori non tutti distinti, la procedura di bootstrap proposta prevede la trasformazione degli autovalori in polinomi simmetrici elementari.*

Key words: bootstrap, density estimation, elementary symmetric polynomials, modes

1 Introduction and main ideas

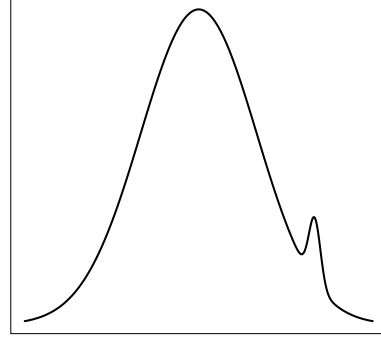
We present here some results of a paper in collaboration with Chris Genovese, Isabella Verdinelli and Larry Wasserman [8].

Figure 1 shows a one-dimensional density estimate with two modes. The leftmost mode is likely to correspond to a real mode in the true density, but the second

Marco Perone Pacifico
Sapienza University of Rome, Italy e-mail: marco.peronepacifico@uniroma1.it

smaller mode on the right may be due to random fluctuation. How can we recognize a real mode from random fluctuation? In this paper, we provide a test to answer this question that is easy to implement, even in multivariate problems.

Fig. 1 The mode on the left appears to be real. The mode on the right might be due to random fluctuation.



There are many reasons for mode hunting and many methods to find modes; see, for example, [9, 10, 6]. In particular, modes can be used as the basis of nonparametric clustering [3, 4, 5, 7, 10]. Before finding clusters, it is important to find out which modes are significant and which are explainable as random fluctuations. Even if based on similar ideas, our procedure is simpler, both theoretically and computationally, than the method in [2].

A mode is a maximum of the density, hence at each mode the gradient of the density is null and all the eigenvalues of the Hessian are negative. Consider a point $x \in \mathbf{R}^d$ and suppose we want to test

$$H_0 : x \text{ is not a mode of } p \quad \text{versus} \quad H_1 : x \text{ is a mode of } p.$$

First, testing the null hypothesis of “no mode” raises problems because the alternative forms a measure zero set. More precisely, if $g(x) = (g_1(x), \dots, g_d(x))^T$ is the gradient of p at x and $\lambda_1(x) \geq \dots \geq \lambda_d(x)$ are the eigenvalues of the Hessian of the density, then the alternative hypothesis

$$H_1 : \left\{ (\lambda_1, g) \in \mathbf{R} \times \mathbf{R}^d : \lambda_1 < 0, g = (0, \dots, 0)^T \right\}. \quad (1)$$

is a measure zero subset of $\mathbf{R}^{d+1}$. No meaningful test can be constructed of such a “reverse null hypothesis.” The second problem is that a mode can occur at uncountably many possible locations, leading potentially to a difficult multiple testing problem. Finally, the eigenvalues are not continuously differentiable functions of the Hessian which makes methods like the bootstrap and the delta method invalid.

We overcome these problems by combining several ideas:

1. We split data in two subsamples to separate the process of finding candidate modes from the process of hypothesis testing. This ameliorates the multiplicity problem and simplifies the hypothesis test as well.

2. First, we use the first subsample to find a finite set of candidate modes.
3. We then use the second half of the data to estimate the Hessian of the density at the candidate modes. We transform the eigenvalues of the Hessian using elementary symmetric polynomials (ESP). As noted in [1], the bootstrap leads to asymptotically valid confidence sets for the transformed eigenvalues. We then invert the mapping to get a valid confidence set for the eigenvalues. This provides useful shape information about the modes, which we call an eigenportrait.
4. The eigenportrait can be used to formulate a test for the importance of the mode. As a surrogate for testing whether a candidate mode is not really a mode, we instead test if x is an “approximate mode”. This requires reformulating the alternative to capture the idea of an approximate mode. There is no unique way to do this. One possibility is to consider

$$H'_0 : \mathbf{R}^{d+1} - H'_1 \quad \text{versus} \quad H'_1 : \left\{ (\lambda_1, g) \in \mathbf{R} \times \mathbf{R}^d : \lambda_1 < 0, \|g\| < \delta \right\} \quad (2)$$

where $\delta > 0$ is a small positive constant. We will show that, in practice, the constraint $\|g\| < \delta$ has no effect on the test so we can simplify matters by just testing $\lambda_1 \geq 0$ versus $\lambda_1 < 0$.

2 The procedure

For simplicity, assume that the sample size is even, say $2n$. Our testing procedure involves the following steps:

Data splitting: Randomly split the data into two subsamples X and Y of size n .

Stage 1: Using X , construct a density estimate $\hat{p}_{X,h}$ and find its modes $\hat{\mathcal{M}}$.

Stage 2: Use Y to construct another density estimate $\hat{p}_{Y,h}$.

At each candidate mode $\hat{m} \in \hat{\mathcal{M}}$ obtained in Stage 1, compute the Hessian of $\hat{p}_{Y,h}$ and its eigenvalues $\hat{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_d)$.

Construct a confidence rectangle G for λ using the ESP bootstrap procedure described later in Section 3. From G get a confidence interval $\mathcal{C}$ for the leading eigenvalue.

Test: If $\sup \{x \in \mathcal{C}\} < 0$, reject $H'_0 : \lambda_1 \geq 0$ and declare $\hat{m}$ to be a real mode.

Here are some details on the steps.

Data splitting: The purpose of the data splitting is to assure the validity of the confidence intervals. If we did not split the data, we could instead get a valid test by treating the estimated Hessian as a stochastic process over the whole space and then estimating the maximum fluctuations of this process. While this is possible, splitting the data and focusing on finitely many points is much simpler.

Stage 1: We use the kernel density estimate with Gaussian kernel K and bandwidth $h > 0$. To locate the modes we use the Mean Shift Algorithm (see [7, 5]):

1. Start with a mesh of points $\{a_1^{(0)}, \dots, a_N^{(0)}\}$.

2. For each mesh point $a_j^{(0)}$, iterate the following equation until convergence:

$$a_j^{(s+1)} \leftarrow \frac{\sum_{i=1}^n X_i K\left(\frac{\|a_j^{(s)} - X_i\|}{h}\right)}{\sum_{i=1}^n K\left(\frac{\|a_j^{(s)} - X_i\|}{h}\right)}.$$

3. Let $\widehat{\mathcal{M}} = \{\widehat{m}_1, \dots, \widehat{m}_k\}$ be the unique values of the set $\{a_1^{(\infty)}, \dots, a_N^{(\infty)}\}$.

Stage 2: We estimate the Hessian at $\widehat{m}$ by using the Hessian of the kernel density estimator from the second half of the data.

Using the method described later in Section 3, for each $\widehat{m} \in \widehat{\mathcal{M}}$ we construct $1 - \alpha/k$ confidence set G for λ , where k is the number of candidate modes.

The validity of the bootstrap, together with the independence from sample splitting, ensures that

$$\liminf_{n \rightarrow \infty} \mathbf{P}(\lambda \in G, \text{ for all } \widehat{m}) \geq 1 - \alpha.$$

Test: In principle, for all $\widehat{m} \in \widehat{\mathcal{M}}$, we would like to test the null hypothesis $\widehat{m}$ is *not a mode* versus the alternative $\widehat{m}$ is *a mode* formalized in (1). But, as we explained earlier, it is not possible to construct a non-trivial test for this hypothesis since the alternative has measure 0. Instead we could replace the alternative with the statement: $\widehat{m}$ is *an approximate mode*. This suggests testing the hypotheses in (2). However, thanks to the data-splitting, testing H'_0 versus H'_1 is asymptotically equivalent to testing

$$H''_0 : \lambda_1 \geq 0 \quad \text{versus} \quad H''_1 : \lambda_1 < 0.$$

This follows since $\widehat{m}$ is a mode of $\widehat{p}_{X,h}$, hence $\|\widehat{g}_{X,h}(\widehat{m})\| = 0$ and

$$\|\widehat{g}_{Y,h}(\widehat{m})\| = \|\widehat{g}_{X,h}(\widehat{m})\| + O_P\left(\frac{1}{nh^{d+2}}\right) = 0 + O_P\left(\frac{1}{nh^{d+2}}\right) = o_P(1).$$

Hence, with probability tending to 1, $\|\widehat{g}_{Y,h}(\widehat{m})\| < \delta$ and, asymptotically, we reject H'_0 if and only if we reject H''_0 . In summary, we interpret the rejection of H''_0 to mean that $\widehat{m}$ is an approximate mode.

3 The ESP Bootstrap

We construct confidence regions for the eigenvalues using the bootstrap. Bootstrapping the eigenvalues poses some problems since $\lambda = (\lambda_1, \dots, \lambda_d)$ is not a continuously differentiable function of the Hessian of the density. As a result, standard bootstrapping applied to the Hessian will not produce valid confidence sets for the eigenvalues. However, Beran and Srivastava (1985) note that if the eigenvalues are

transformed using elementary symmetric polynomials, then the confidence set obtained has the correct asymptotic coverage.

Given ordered, not necessarily distinct, eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$, the elementary symmetric polynomials (ESP) is a vector $s = (s_1, \dots, s_d)$ with generic entry

$$s_t = \sum_{i_1=1}^d \sum_{i_2=i_1+1}^d \dots \sum_{i_t=i_{t-1}+1}^d \lambda_{i_1} \cdot \lambda_{i_2} \dots \lambda_{i_t}(x). \quad (3)$$

The ESP is a continuously differentiable function of the Hessian, hence bootstrap confidence sets for the vector s have the right coverage. Since the map from λ to s is one-to-one, a confidence set for s can be turned into a confidence set for λ .

The steps in the bootstrap, at a particular candidate mode $\hat{m}$ are as follows:

1. Let $\hat{\lambda}$ be the eigenvalues of the Hessian of $\hat{p}_{Y,h}$ and let $\hat{s}$ be its ESP.
2. Draw $Y_1^*, \dots, Y_n^* \sim P_n$ where P_n is the empirical distribution of $Y_1, \dots, Y_n$.
3. Using $Y_1^*, \dots, Y_n^*$ compute the density estimate, the Hessian and the estimated eigenvalues $\lambda^* = (\lambda_1^*, \dots, \lambda_d^*)$. Use (3) to get the ESP-transformed eigenvalues $s^* = (s_1^*, \dots, s_d^*)$.
4. Repeat steps 2 and 3 B times yielding B vectors $s^{*1}, \dots, s^{*B}$.
5. Compute the $1 - \alpha/k$ bootstrap confidence set $\mathcal{S}$ for s and the corresponding confidence rectangles G for the eigenvalues.

4 Concluding remarks

The numerical example reported in Figure 2 shows promising results. The data have three well-separated modes and a ring. The three modes are detected in Stage 1 together with three other modes on the ring. In Stage 2, modes on the ring are declared to be non significant, since the corresponding confidence interval for λ_1 lies across 0. The eigen portrait in the two bottom panels distinguishes the difference in shape between the different significant modes.

An accurate bandwidth selection is crucial in mode detection: too much smoothness could hide true modes (and, consequently, some existing cluster), too little smoothness could highlight fake modes. The fake modes removal in Stage 2 makes it possible to use our procedure for bandwidth selection, choosing the bandwidth to maximize the number of significant modes.

References

1. Beran, R. and Srivastava, M.S.: Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics*, **13**, 95–115 (1985)
2. Burman, P. and Polonik, W.: Multivariate mode hunting: Data analytic tools with measures of significance. *Journal of Multivariate Analysis*, **100**, 1198–1218 (2009)

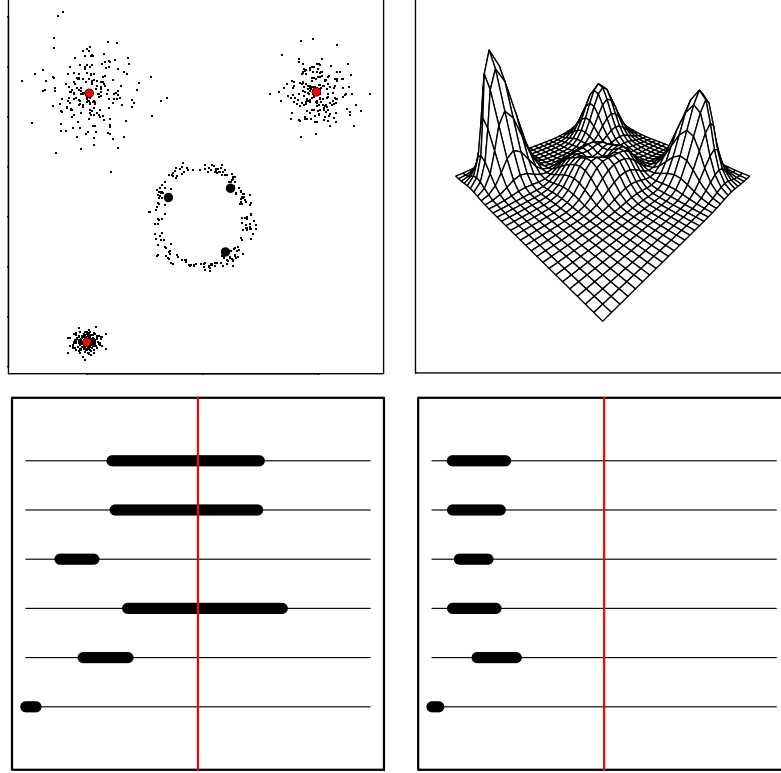


Fig. 2 TOP: scatterplot (with 6 detected modes) and estimated density. BOTTOM: confidence intervals for λ_1 (left) and λ_2 (right) at each detected mode.

3. Chacón, J.: Clusters and water flows: a novel approach to modal clustering through Morse Theory. *arXiv:1212.1384* (2012)
4. Chazal, F., Guibas, L.J., Oudot, S.Y. and Skraba, P.: Persistence-based clustering in Riemannian manifolds. *Proceedings of the 27th annual ACM symposium on Computational geometry*, 97–106 (2011)
5. Comaniciu, D. and Meer, P.: Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 603–619 (2002)
6. Dümbgen, L. and Walther, G.: Multiscale Inference about a density. *The Annals of Statistics*, **36**, 1758–1785 (2008)
7. Fukunaga, K. and Hostetler, L.D.: The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition. *IEEE Transactions on Information Theory*, **21**, 32–40 (1975)
8. Genovese, C., Perone Pacifico, M., Verdinielli, I. and Wasserman, L.: Nonparametric Inference for Density Modes. *arXiv:1312.7567* (2013)
9. Klemelä, J.: Smoothing of Multivariate Data: Density Estimation and Visualization. Wiley, New York (2009)
10. Li, J., Ray, S. and Lindsay, B.G.: A nonparametric statistical approach to clustering via mode identification. *Journal of Machine Learning Research*, **8**, 1687–1723 (2007)