

## Functional approach for excess mass estimation in the density model

Cristina Butucea  
Mathilde Mougeot  
Karine Tribouley, *Universite PX France*

### Abstract

We consider a multivariate density model where we estimate the excess mass of the unknown probability density  $f$  at a given level  $\nu > 0$  from  $n$  i.i.d. observed random variables. This problem has several applications such as multimodality testing, density contour clustering, anomaly detection, classification and so on. For the first time in the literature we estimate the excess mass as an integrated functional of the unknown density  $f$ . We suggest an estimator and evaluate its rate of convergence, when  $f$  belongs to general Besov smoothness classes, for several risk measures. A particular care is devoted to implementation and numerical study of the studied procedure. It appears that our procedure improves dramatically on the plug-in estimator of the excess mass.

AMS 2000 subject classifications: Primary 62G05, 62G20, 62H12; secondary 62C20.

Keywords: excess mass, functional estimation, multivariate probability density, rates of convergence, upper bounds of the risk error.



Full Text: [PDF](#)

Butucea, Cristina, Mougeot, Mathilde, Tribouley, Karine, Functional approach for excess mass estimation in the density model, *Electronic Journal of Statistics*, 1, (2007), 449-472 (electronic). DOI: 10.1214/07-EJS079.

## References

- [1] Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhy* Ser. A 50, 3, 381–393. [MR1065550 \(91e:62079\)](#)
- [2] Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* 23, 1, 11–29. [MR1331653](#)
- [3] Cai, T. T. and Low, M. G. Lower bounds for excess mass estimation. Manuscript.
- [4] Chandler, G. and Polonik, W. (2006). Discrimination of locally stationary time series based on the excess mass functional. *J. Amer. Statist. Assoc.* 101, 473, 240–253. [MR2268042](#)
- [5] Cheng, M.-Y. and Hall, P. (1999). Mode testing in difficult cases. *Ann. Statist.* 27, 4, 1294–1315. [MR1740110 \(2001a:62044\)](#)
- [6] Daubechies, I. (1992). Ten lectures on wavelets. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 61. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. [MR1162107](#)

- [7] Fisher, N. I. and Marron, J. S. (2001). Mode testing via the excess mass estimate. *Biometrika* 88, 2, 499–517. [MR1844848](#)
- [8] Gayraud, G. and Rousseau, J. (2005). Rates of convergence for a Bayesian level set estimation. *Scand. J. Statist.* 32, 4, 639–660. [MR2232347](#)
- [9] Hall, P. (1982). Rates of convergence in the central limit theorem. *Research Notes in Mathematics*, Vol. 62. Pitman (Advanced Publishing Program), Boston, Mass. [MR0668197 \(84k:60032\)](#)
- [10] Härdle, W., Kerkycharian, G., Picard, D., and Tsybakov, A. (1998). Wavelets, approximation, and statistical applications. *Lecture Notes in Statistics*, Vol. 129. Springer-Verlag, New York. [MR1618204 \(99f:42065\)](#)
- [11] Hartigan, J. A. and Hartigan, P. M. (1985). The dip test of unimodality. *Ann. Statist.* 13, 1, 70–84. [MR0773153 \(86h:62068\)](#)
- [12] Ibragimov, I. A. and Khasminskiĭ, R. Z. (1991). Asymptotically normal families of distributions and efficient estimation. *Ann. Statist.* 19, 4, 1681–1724. [MR1135145 \(93d:62093\)](#)
- [13] Kerkycharian, G. and Picard, D. (1996). Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.* 24, 2, 485–507. [MR1394973 \(97e:62062\)](#)
- [14] Klemelä, J. (2004). Complexity penalized support estimation. *J. Multivariate Anal.* 88, 2, 274–297. [MR2025614](#)
- [15] Korostelev, A. P. (1991). On the accuracy of the estimation of nonsmooth functionals of regression. *Theory Probab. Appl.* 35, 4, 784–787. [MR1090505 \(92e:62074\)](#)
- [16] Lepski, O., Nemirovski, A., and Spokoiny, V. (1999). On estimation of the  $L_p$  norm of a regression function. *Probab. Theory Related Fields* 113, 2, 221–253. [MR1670867](#)
- [17] Mammen, E. and Tsybakov, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* 23, 2, 502–524. [MR1332579](#)
- [18] Meyer, Y. (1992). Wavelets and operators. *Cambridge Studies in Advanced Mathematics*, Vol. 37. Cambridge University Press, Cambridge. Translated from the 1990 French original by D. H. Salinger. [MR1228209](#)
- [19] Müller, D. W. and Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* 86, 415, 738–746. [MR1147099](#)
- [20] Nemirovski, A. (2000). Topics in non-parametric statistics. In *Lectures on probability theory and statistics (Saint-Flour, 1998)*. *Lecture Notes in Math.*, Vol. 1738. Springer, Berlin, 85–277. [MR1775638 \(2001c:60003\)](#)
- [21] Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters—an excess mass approach. *Ann. Statist.* 23, 3, 855–881. [MR1345204](#)
- [22] Polonik, W. and Wang, Z. (2005). Estimation of regression contour clusters—an application of the excess mass approach to regression. *J. Multivariate Anal.* 94, 2, 227–249. [MR2167913 \(2006i:62055\)](#)
- [23] Rigollet, P. and Vert, R. (2006). Fast rates for plug-in estimators of density level-sets. hal-00114180.
- [24] Tribouley, K. (2000). Adaptive estimation of integrated functionals. *Math. Methods Statist.* 9, 1, 19–38. [MR1772223](#)
- [25] Tribouley, K. (2003). A wavelet method for confidence intervals in the density model. *Scand. J. of Stat* 30, 4, 739–756. [MR1447735](#)
- [26] Tsybakov, A. B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* 25, 3, 948–969. [MR1447735](#)

