Parametric rates for nonparametric density estimates

Valentin Patillea [1] & Lionel Truquet [2]

[1] *CREST, Ensai, Campus de Ker-Lann, 35172, Bruz, France; patilea@ensai.fr*
[2] *IRMAR, Ensai, Campus de Ker-Lann, 35172, Bruz, France; lionel.truquet@ensai.fr*

Résumé. Nous étudions le problème d'estimation de la densité d'un vecteur aléatoire à partir des observations indépendantes de même loi. Dans certains modèles de régression linéaires ou non linéaires homoscédastiques, il est bien connu que la densité de la réponse peut-être estimée à une vitesse paramétrique, en utilisant la représentation de la densité objectif comme un produit de convolution entre la densité d'une fonctionnelle des covariables et celle du terme d'erreur. Nous considérons des généralisations de ce problème d'estimation non paramétrique pour des modèles de régression paramétriques conditionnellement hétéroscédastiques et non nécessairement additifs. Dans ce cas, la densité objectif est toujours une fonction lisse de la densité d'une transformation des covariables et de celle des erreurs et peut être estimée à partir d'une $U-$statistique construite à l'aide d'un estimateur à noyau de la densité des termes d'erreur. Cette approche demande au préalable une estimation consistante à la vitesse $n^{-1/2}$ des coefficients du modèle.

Mots-clés. Convolution, Estimateur à noyau, Modèles non separables, Vitesse de convergence paramétrique

Abstract. The problem of estimating a multivariate density of some random vector of interest using independent identically distributed observations is considered. For some linear or nonlinear homoscedastic models, it is well known that the density of the response variable could be estimated at the parametric rate. This is achieved using the representation of the response variable density as a convolution between a functional of the density of the covariates and the density of the error term. We generalize this nonparametric estimation setup to heteroscedastic, non necessarily additive regression models. In our case, the density of interest is still a smooth map of the density of a transformation of the covariates and of the density of the error term. Thus the density of interest could be estimated at a parametric rate using a $U-$statistic built from a kernel estimator of the error term density. A preliminary $n^{-1/2}-$consistent estimator of the parameter of the models is required.

Keywords. Convolution, Kernel estimation, Non séparable models, Parametric rate

1 Introduction

Density estimation received a lot of attention in the statistical literature. The nonparametric approaches have been very successful, especially when the estimation procedure could adapt to the regularity of the density and yields optimal rate estimates. In this paper we consider a broad, possibly nonparametric class of multivariate densities for which the density could be estimated at a parametric rate. To give a flavor on our approach, let us denote by Y the random vector with density f_Y defined on $\mathbb{R}^{d_Y}$, that we want to estimate. The vector Y is not necessarily observed, but instead let us suppose that one observed a random vector X. For the informal insight we want to provide at this stage, let us assume that X takes values in $\mathbb{R}^{d_X}$, it has a density f_X, and let $f_{Y,X}$ denote the joint density of Y and X, that for the moment is also assumed to exist. Then, one can write

$$f_Y(y) = \int_{\mathbb{R}^{d_X}} f_{Y|x}(y) f_X(x) dx = \mathbb{E}\left[f_{Y|X}(y)\right], \qquad y \in \mathbb{R}^{d_Y},$$

where $f_{Y|x}$ stands for the density of Y given that $X = x$. If the conditional densities $f_{Y|x}$, $x \in \mathbb{R}^{d_X}$, does not have any particular structure, in general one proceeds to a local estimation for each x, and thus the effective sample size is of smaller order than the sample size. This results in nonparametric rates.

In this paper, we propose a flexible structure that links Y and X and, in some sense, allows using the whole sample for estimation of the conditional density of Y given the value of X. This will result in parametric rates for simple density estimates defined by replacing the expectation with respect to the law of X by a sample mean. The model we propose include a large set of models studied in the statistical literature.

2 The framework

Let ε be a random vector with values in $\mathbb{R}^{d_Y}$. Let f_ε denote the density of ε, that is supposed to exist. Assume that

$$\varepsilon \quad \text{and} \quad X \quad \text{are independent.}$$

Let $Y \in \mathbb{R}^{d_Y}$ be the random vector defined by some transformation

$$Y = T(X, \varepsilon).$$

The map $(x, e) \mapsto T(x, e) \in \mathbb{R}^{d_Y}$ is such that, for any $x \in \mathbb{R}^{d_X}$, the function $e \mapsto T(x, e)$ is one-to-one. Let $S(x, \cdot)$ be the inverse, so that

$$\varepsilon = S(X, Y).$$

Assuming that for any x, the application $S(x, \cdot)$ has continuous partial derivatives of order 1, and the expectations are well-defined, one could write the density of Y under the form

$$f_Y(y) = \mathbb{E}\left[|J(X, y)|\, f_\varepsilon\left(S(X, y)\right)\right], \tag{1}$$

where

$$J(X, y) = \det \nabla_y S(X, y)$$

is the Jacobian matrix. The aim is to estimate $f_Y(y)$ assuming an independent sample (X_i, ε_i), $1 \le i \le n$, from (X, ε), where only the X_i's are observed.

Concerning the realizations of Y, two cases are allowed: all the Y_i's are available and none of the Y_i's are available. When the $Y_i = T(X_i, \varepsilon_i)$, $1 \le i \le n$, are observed the inverse transformation S could be some unknown element of a given parametric family of one-to-one transformations

$$\{S_\theta : \theta \in \Theta\},$$

where Θ is the some parameter space, typically subset of $\mathbb{R}^p$. Let θ_0 be the unknown value of the parameter that corresponds to transformation S, *i.e.* $\varepsilon_i = S_{\theta_0}(X_i, Y_i)$. Let $J_{\theta_0}(X, y) = \det \nabla_y S_{\theta_0}(X, y)$ be the corresponding Jacobian. When the sample of Y is not observed, we will necessarily assume that the transformation S is given.

Concerning the density f_ε, we will not necessarily assume that it is known, and we will use a kernel estimate to approximate it. This will be possible only when the ε_i's could be estimated from the data, that is when the Y_i's are observed together with the X_i's.

The framework we consider is quite general and includes many examples of statistical models, such as Berkson error model, parametric location-scale regression models and econometric non separable models.

3 The estimators

Inspired by the identity (1), one can define an estimator for the density f_Y for each of the four situations where the transformation S is given or not, and the density f_ε is known or has to be estimated.

If the transformation S, the Jacobian J and the density f_ε are given, a simple estimator of $f_Y(y)$ would be

$$\widetilde{f}_Y(y) = \frac{1}{n} \sum_{i=1}^{n} |J(X_i, y)| \, f_\varepsilon(S(X_i, y)).$$

Under mild conditions, this estimator converges at the $\sqrt{n}$ parametric rate. For instance, this result is well known in the case of Berkson error model with one-dimensional Y; see, for instance, Delaigle (2007).

If the transformation S and the Jacobian J depend on some parameter θ, we could define the density estimator

$$\widetilde{f}_Y(y; \widehat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left| J_{\widehat{\theta}}(X_i, y) \right| f_\varepsilon(S_{\widehat{\theta}}(X_i, y)),$$

where $\widehat{\theta}$ is some estimator of the true unknown value θ_0 of the parameter.

Next, let us consider the case where f_ε is unknown. Let us suppose for the moment that the transformation S and the Jacobian J are given. The idea we propose for building and estimator of f_Y is to use a leave-one-out Parzen-Rosenblatt density estimate of $f_\varepsilon(S(X_i, y))$ based on the sample $\varepsilon_j = S(X_j, Y_j)$, $j \neq i$. More precisely, consider the estimator

$$\widehat{f}_Y(y) = \frac{1}{n(n-1)b^d} \sum_{1 \leq i \neq j \leq n} |J(X_i, y)| \, K_b(S(X_i, y) - S(X_j, Y_j)),$$

with $K_b(t) = K(t/b)$ with $K(\cdot)$ a kernel and b a bandwidth.

Finally, in the case where f_ε is unknown, and the transformation S and the Jacobian J are given up to some unknown parameter, we propose the density estimator

$$\widehat{f}_Y(y; \widehat{\theta}) = \frac{1}{n(n-1)b^d} \sum_{1 \leq i \neq j \leq n} \left| J_{\widehat{\theta}}(X_i, y) \right| K_b(S_{\widehat{\theta}}(X_i, y) - S_{\widehat{\theta}}(X_j, Y_j)), \qquad (2)$$

where $\widehat{\theta}$ is some estimator of θ_0. The estimator $\widehat{\theta}$ could be build inside the specific model using standard approaches. A general approach for non separable models was proposed by Brown & Wegkamp (2002).

4 The results

The problem we consider was investigated, under different assumptions, by Escanciano & Jacho-Chávez (2012), Müller *et al.* (2013), Støve & Tjøstheim (2012), Schick & Wefelmeyer (2013), among many others.

The asymptotic behavior of $\widetilde{f}_Y(y)$ follows directly from quite standard results on the behavior of the $U-$statistics of order 2. Convergence, uniform convergence with respect to y in some suitable, and $n^{-1/2}-$asymptotic normality could be derived. For the estimator $\widehat{f}_Y(y; \widehat{\theta})$, one has to control the effect induced by the fact that the true ε_i are not available. This could be a difficult issue and was usually handled in a regression setup using a suitable decomposition of the difference $\widehat{f}_Y(y; \widehat{\theta}) - \widehat{f}_Y(y; \theta_0)$ and i.i.d. representations of the residual empirical process in the parametric model to control this difference. See Khmaladze and Koul (2009) for recent reference on such representations of the residual process. We use a different decomposition of $\widehat{f}_Y(y; \widehat{\theta}) - \widehat{f}_Y(y; \theta_0)$ and uniform results on the behavior of $U-$processes to control this difference. Our approach seems more convenient for studying multivariate densities. Under suitable conditions, we derive the $n^{-1/2}-$convergence for $\widehat{f}_Y(y; \widehat{\theta})$. The uniformity with respect to y is also studied. Next, we derive the asymptotic normality $\widehat{f}_Y(y; \widehat{\theta})$ and study its efficiency.

Bibliographie

[1] Brown, D.J. & Wegkamp, M.H. (2002). Weighted Minimum Mean-Square Distance from Independence Estimation. *Econometrica*, 70, 2035–2051.

[2] Delaigle, A. (2007). Nonparametric density estimation from data with a mixture of Berkson and classical errors. *Canad. J. Statist.*, 35, 89–104.
[3] Escanciano, J.C. & Jacho-Chávez, D.T. (2012). $\sqrt{n}-$ Uniformly Consistent Density Estimation in Nonparametric Regression Models. *Journal of Econometrics*, 167, 305–316.
[4] Khmaladze, E.V. & Koul, H.L. (2009). Goodness-of-fit problem for errors in nonparametric regression: Distribution free approach. *Ann. Statist.*, 37, 3165–3185.
[5] Müller, U.U., Schick A. & Wefelmeyer W. (2013). Non-standard behavior of density estimators for functions of independent observations. *Comm. Statist. Theory Methods*, 42, 16, 2291–2300.
[6] Støve, B. & Tjøstheim, D. (2012). A Convolution Estimator for the Density of Nonlinear Regression Observations, *Scandinavian Journal of Statistics*, 39, 282–304.
[7] Schick, A. & Wefelmeyer, W. (2013). Uniform convergence of convolution estimators for the response density in nonparametric regression. *Bernoulli*, 19, 2250–2276.