

A Note on Efficient Density Estimators of Convolutions[☆]

Soutir Bandyopadhyay

Department of Mathematics, Lehigh University, Bethlehem, PA 18015

Abstract

It is already known that the convolution of a bounded density with itself can be estimated at the root- n rate using the two asymptotically equivalent kernel estimators: (i) Frees estimator (Frees (1994)) and (ii) Saavedra and Cao estimator (Saavedra and Cao (2000)). In this work, we investigate the efficiency of these estimators of the convolution of a bounded density. The efficiency criterion used in this work is that of a *least dispersed regular estimator* described in Begun et al. (1983). This concept is based on the Hájek-Le Cam convolution theorem for locally asymptotically normal (LAN) families.

Keywords: Convolution, Kernel density, Locally asymptotically normal, Hellinger derivative.

1. Introduction

Let X_1, \dots, X_n be independent and identically distributed random variables with distribution function F and bounded density f . Let p denote the convolution $f*f$ of f with itself,

$$p(x) = \int f(y)f(x-y)dy, \quad x \in \mathbb{R}. \quad (1.1)$$

In the field of statistics, density estimates play an important role. However in many instances, instead of estimating the density of the observations, one is interested in estimating the convolution density $p(\cdot)$ as in (1.1). Frees (1994) describes motivating examples of practical applications of $p(x)$ in the insurance, as well as reliability, areas. One can also use estimators of $p(\cdot)$ to test whether f belongs to a given family of densities that is closed under convolutions, like the Gaussian family of densities (see, Schick and Wefelmeyer (2007)). Moreover, if the density function is symmetric then $p(0) = \int f^2(y)dy$. Estimation of $p(0)$ is of extreme importance in the study of rank-based nonparametric inference problems because it is a basic quantity involved in the expressions for asymptotic efficiency of rank tests for problems in location shift, analysis of variance, etc.; see for example, Prakasa Rao (1983), Hall and Marron (1987), Bickel and Ritov (1988) and the references therein.

Frees (1994) first proposed a kernel estimator for the convolution as follows

$$\hat{p}(x) = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} K_{h_n}(x - X_i - X_j), \quad x \in \mathbb{R}, \quad (1.2)$$

where $K_{h_n}(x) = h_n^{-1}K(x/h_n)$ for some kernel density $K(\cdot)$ and for some bandwidth h_n . Later, Saavedra and Cao (2000) introduced the natural estimator of $p(x)$, defined by plugging in the kernel density estimator \hat{f}

[☆]AMS (2000) subject classification: Primary 62G07, 62G20 Secondary 62H12.

Email address: sob210@lehigh.edu (Soutir Bandyopadhyay)

Preprint submitted to *Journal of Statistical Planning and Inference*

April 19, 2012

in place of f in (1.1) and proved the asymptotic normality of the proposed estimator. However Schick and Wefelmeyer (2007) proved that the two estimators proposed by Frees (1994) and Saavedra and Cao (2000) were asymptotically equivalent. Hence, we restrict our attention to estimators (1.2). Frees (1994) showed that under some smoothness assumptions on the density, the convolution function $p(\cdot)$ can be estimated using (1.2) at the parametric rate $n^{-1/2}$. Schick and Wefelmeyer (2004) generalized the result and proved \sqrt{n} -consistency of the Frees estimator of the convolutions in weighted \mathcal{L}_1 norms under some additional moment conditions. They showed that $n^{1/2}(\hat{p} - p)$ converges in distribution to a centered Gaussian process in \mathcal{L}_1 norm whose covariance structure matches that of $2f(\cdot - X_1)$. Schick and Wefelmeyer (2007) further proved \sqrt{n} -consistency of the Frees estimator under weaker smoothness assumptions on the density f and they showed that the same asymptotic normality could still be achieved under those weaker assumptions. Finally Giné and Mason (2007) derived a functional central limit theorems in \mathcal{L}_p , $1 \leq p \leq \infty$, in the general setting of Frees (1994), and uniformly in the bandwidth.

The goal of this paper is to study the efficiency of the Frees estimator as in (1.2) for the convolution of a bounded density f with itself, for a fixed $x \in \mathbb{R}$. The previous works discuss the consistency and asymptotic properties of (1.2). However the efficiency of this estimator is still not known. In this work we have shown that the Frees estimator is actually an efficient estimator. This will also imply the efficiency of the Saavedra and Cao estimator since it is asymptotically equivalent to the Frees estimator. The efficiency criterion used in this work is that of a *least dispersed regular estimator* described in Begun et al. (1983), Pfanzagl and Wefelmeyer (1982) and Schick (1996); see also the monograph by Bickel et al. (1993). This concept is based on the Hájek-Le Cam convolution theorem for locally asymptotically normal (LAN) families.

The rest of the paper is organized as follows. In Section 2 we state the main result of this work followed by the proof of the result in Section 3.

2. Result

Let $\Theta \equiv$ the set of all bounded Lebesgue densities and X_1, X_2, \dots be measurable functions which are independently and identically distributed with distribution Q_ϑ for each $\vartheta \in \Theta$. $dQ_\vartheta = \vartheta d\lambda$ where λ denote the Lebesgue measure and for a fixed $x \in \mathbb{R}$, let us define $\kappa_x : \Theta \rightarrow \mathbb{R}$ by $\kappa_x(\vartheta) = \vartheta * \vartheta(x)$.

Let h be a measurable function from \mathbb{R} to \mathbb{R}^m , for a fixed $m \geq 1$ such that $\int h dF = 0$, $\int \|h\|^2 dF < \infty$, $H = \int h h^T dF$ is positive definite and $\Delta = \{\delta \in \mathbb{R}^m : \|\delta\| < 1\}$. Now let us define a model $\{f_{\delta,h} : \delta \in \Delta\}$ where $f_{\delta,h} = f(1 + \delta^T \tilde{h}_\delta)$ with $\tilde{h}_\delta = h_\delta - \int h_\delta dF$ where $h_\delta = h I(\|h\| < \|\delta\|^{-1/2}/2)$ for $\delta \in \mathbb{R}^m$. Now before stating the main theorem, let us briefly go over few definitions.

Definition 2.1. *By a path through f we mean a function π from an open neighborhood Δ_π of the origin into the function space Θ such that $\pi(0) = f$.*

Remark By construction of $f_{\delta,h}$ as above, we have a path $\pi(\delta) = f_{\delta,h}$ which defines the model $\{f_{\delta,h} : \delta \in \Delta\}$.

Definition 2.2. *We say that the path π defined by $\pi(\delta) = f_{\delta,h}$ is κ_x -smooth if the following two conditions are met.*

- (i) *The model $\{f_{\delta,h} : \delta \in \Delta\}$ is Hellinger differentiable at 0 with a positive definite information.*
- (ii) *The function $\kappa_x \circ \pi$ is differentiable at 0.*

Suppose now Π denote the collection of all such paths. Then we have the following definition.

Definition 2.3. *The closed linear span of the tangent set (see, Bickel et al. (1993)) is called the tangent space generated by Π and is denoted by $\dot{\Pi}$. We call Π proper if for every finite-dimensional linear subspace L of $\dot{\Pi}$ there exists a path $\pi \in \Pi$ with the tangent set containing L .*

Equipped with the above definitions (2.1) - (2.3), we propose the following theorem concerning the model $\{f_{\delta,h} : \delta \in \Delta\}$.

Theorem 2.4. *The model $\{f_{\delta,h} : \delta \in \Delta\}$ has the following properties.*

- (i) *For every $\delta \in \Delta$, $f_{\delta,h}$ is a bounded density.*
- (ii) *The model is Hellinger differentiable at 0 with Hellinger derivative h and invertible information H .*
- (iii) *The function $\kappa_x(f_{\delta,h})$ is differentiable at 0 with derivative $2f^*(hf)(x)$.*

In the light of above definitions (2.1) and (2.2), Theorem 2.4 shows that the path π defined by $\pi(\delta) = f_{\delta,h}$ is κ_x -smooth.

Proposition 2.5. *For the path defined by $\pi(\delta) = f_{\delta,h}$, the tangent space $\dot{\Pi} = \mathcal{L}_{2,0}(F) = \{\psi : \mathbb{R} \rightarrow \mathbb{R} : \int \psi dF = 0, \int \psi^2 dF < \infty\}$ and Π is proper.*

Now $2f^*(hf)(x) = \int 2f(x-y)h(y)f(y)dy = \int gh dF$ with $g(y) = 2f(x-y)$, for all $y \in \mathbb{R}$. Hence κ_x has Π -gradient g . A canonical gradient is given by

$$g_* = g - \int g dF = 2(f(x-y) - f^*f(x)),$$

which is also the influence function of $\hat{p}(x)$ from the representation given below in (2.3). For the proof of this representation we refer the reader to the paper Schick and Wefelmeyer (2007). They have shown that,

$$\hat{p}(x) - p(x) = \frac{2}{n} \sum_{j=1}^n [f(x - X_j) - p(x)] + o_p(n^{-1/2}). \quad (2.3)$$

Hence following the efficiency criterion described in Begun et al. (1983), Pfanzagl and Wefelmeyer (1982), Bickel et al. (1993) and Schick (1996) we say that, since Π is proper and g_* is a canonical gradient, then $\hat{p}(x)$ is *least dispersed regular* for κ_x and Π , since it has the influence function g_* from (2.3).

3. Proof

Before giving the proof of Theorem 2.4 let us describe the following lemma in a more general setup which will be needed to prove the theorem.

Lemma 3.1. *Let $(\Omega, \mathbb{A}, \{P_{\vartheta} : \vartheta \in \Theta\})$ and $(S, \mathbb{S}, \{Q_{\vartheta} : \vartheta \in \Theta\})$ be two experiments and X_1, X_2, \dots be measurable functions from Ω into S which are independently and identically distributed with distribution Q_{ϑ}*

under P_ϑ for each $\vartheta \in \Theta$. Let G be a probability measure on \mathbb{S} and h be a measurable function from S to \mathbb{R}^k such that $\int \|h\|^2 dG$ is finite. Suppose ψ be another measurable function from S to \mathbb{R}^p such that $\int \psi \psi^T dG$ is well defined and positive definite and the span $\{a^T \psi : a \in \mathbb{R}^p\}$ contains 1. Then there is a model $\{G_t : t \in \Delta\}$ with Δ an open neighborhood of the origin in \mathbb{R}^k with the following properties:

1. $G_0 = G$.
2. For each $t \in \Delta$, $\int \psi dG_t = \int \psi dG$.
3. The model $\{G_t : t \in \Delta\}$ is Hellinger differentiable at 0 with Hellinger derivative

$$\tau = h - \int h \psi^T dG \left(\int \psi \psi^T dG \right)^{-1} \psi.$$

4. For every G -integrable function ϕ , $\int \phi dG_t \rightarrow \int \phi dG$ as $t \rightarrow 0$.

Proof Let $C = \int h \psi^T dG$, $\Psi = \int \psi \psi^T dG$ and $M = C\Psi^{-1}$. Let $A = \|M\|$ denote the euclidean norm of M . For $t \in \mathbb{R}^k$ with $\|t\| < 1$, let h_t and ψ_t be measurable functions such that $\|h_t\| \leq \|t\|^{-1/2}/2$ and $\|\psi_t\| \leq \|t\|^{-1/2}/(2A+2)$ and that $\int \|h_t - h\|^2 dG \rightarrow 0$ and $\int \|\psi_t - \psi\|^2 dG \rightarrow 0$ as $t \rightarrow 0$. The choices $h_t = hI(\|h\| < \|t\|^{-1/2}/2)$ and $\psi_t = \psi I(\|\psi\| < \|t\|^{-1/2}/(2A+2))$ work. Then $C_t = \int h_t \psi_t^T dG$ converges to C and $\Psi_t = \int \psi_t \psi_t^T dG$ converges to Ψ as $t \rightarrow 0$. Thus there is a neighborhood Δ of the origin such that Ψ_t is invertible for $t \in \Delta$, and the matrix $M_t = C_t \Psi_t^{-1}$, which converges to M , has norm less than $A+1$. For $t \in \Delta$ set $\tilde{h}_t = h_t - M_t \psi_t$ and $g_t = 1 + t^T \tilde{h}_t$. Then $\|\tilde{h}_t\| \leq \|t\|^{-1/2}$ and

$$|g_t - 1| \leq \|t\|^{1/2}. \quad (3.4)$$

Thus g_t is positive as $\|t\| < 1$. Moreover,

$$\int g_t \psi^T a dG = \int \psi^T a dG + t^T [C_t - M_t \Psi_t] a = \int \psi^T a dG \quad (3.5)$$

for all $a \in \mathbb{R}^p$. For a such that $a^T \psi = 1$, we then get $\int g_t dG = 1$. This shows that g_t is a probability density function with respect to G . Let now G_t denote the probability measure with density g_t with respect to G . Then $G_0 = G$; by (3.5), $\int \psi dG_t = \int \psi dG$; and by (3.4), $|\int \phi dG_t - \int \phi dG| \leq \int |g_t - 1| |\phi| dG \leq \|t\|^{1/2} \int |\phi| dG$. Thus we are left to show the Hellinger differentiability of the model. For this we use the inequality

$$|\sqrt{1+x} - 1 - \frac{1}{2}x| \leq |x|^2, \quad |x| < 1/2,$$

which is derived via a Taylor expansion. Then, for small enough t , we have $|t^T \tilde{h}_t|^2 \leq \|t\| < 1/2$ and thus

$$\int \left(\sqrt{g_t} - 1 - \frac{1}{2} t^T \tilde{h}_t \right)^2 dG \leq \int |t^T \tilde{h}_t|^4 dG \leq \|t\|^3 \int \|\tilde{h}_t\|^2 dG.$$

Thus we have

$$\int \left(\sqrt{g_t} - 1 - \frac{1}{2} t^T \tilde{h}_t \right)^2 dG = o(\|t\|^2).$$

It follows from what we have already shown that $\int \|\tilde{h}_t - \tau\|^2 dG \rightarrow 0$ as $t \rightarrow 0$. Thus

$$\int \left(\sqrt{g_t} - 1 - \frac{1}{2} t^T \tau \right)^2 dG \leq 2 \int \left(\sqrt{g_t} - 1 - \frac{1}{2} t^T \tilde{h}_t \right)^2 dG + 2\|t\|^2 \int \|\tilde{h}_t - \tau\|^2 dG = o(\|t\|^2).$$

This shows that the model is Hellinger differentiable with the desired Hellinger derivative τ .

We shall now use Lemma 3.1 to prove Theorem 2.4.

3.1. Proof of Theorem 2.4

(i) Fix a $\delta \in \Delta$. By the construction of the functions, h_δ and \tilde{h}_δ are measurable functions such that $\|h_\delta\| \leq \|\delta\|^{-1/2}/2$ and hence $\|\tilde{h}_\delta\| \leq \|\delta\|^{-1/2}$. Then,

$$\|f_{\delta,h} - f\| = \|f\| \|\delta^T \tilde{h}_\delta\| \leq \|f\| \|\delta\|^{1/2} \leq \|f\|.$$

This implies $\|f_{\delta,h}\| = \|f_{\delta,h} - f + f\| \leq 2\|f\|$. Hence, $f_{\delta,h}$ is bounded. Also $\int f_{\delta,h} d\lambda = \int f d\lambda + \int \delta^T \tilde{h}_\delta f d\lambda = 1$ since \tilde{h}_δ is centered. Therefore, $f_{\delta,h}$ is a bounded density function.

(ii) We can prove this by the Lemma 3.1 applied with $\psi(x) = 1$.

(iii) To prove this part, we will use the following properties of the convolution function.

- a. $f * g = g * f$ for all bounded density functions f and g .
- b. $(f + g) * (f + g) = f * f + 2f * g + g * g$.
- c. For any $x \in \mathbb{R}$, $|f * g|(x) \leq \|f\|_p \|g\|_q$, where $\frac{1}{p} + \frac{1}{q} = 1$. In particular, we will use the cases when $p = 1, q = \infty$ and $p = 2, q = 2$.

Also, since the model is Hellinger differentiable, so using Proposition 3 of Appendix 5 in Bickel et al. (1993) we get $\|f_{\delta,h} - f - \delta^T h f\|_1 = o(\|\delta\|)$ as $\delta \rightarrow 0$. Hence,

$$\begin{aligned} D(x) &\equiv [f_{\delta,h} * f_{\delta,h} - f * f - 2\delta^T f * (h f)](x) \\ &= [2f * (f_{\delta,h} - f - \delta^T h f) + (f_{\delta,h} - f) * (f_{\delta,h} - f)](x). \end{aligned}$$

Hence,

$$\begin{aligned} |D(x)| &\leq [2|f * (f_{\delta,h} - f - \delta^T h f)| + |(f_{\delta,h} - f) * (f_{\delta,h} - f)|](x) \\ &\leq 2\|f\|_\infty \|f_{\delta,h} - f - \delta^T h f\|_1 + \|f_{\delta,h} - f\|_2^2 \\ &\leq 2O(1)o(\|\delta\|) + \left(\|\sqrt{f_{\delta,h}}\|_\infty + \|\sqrt{f}\|_\infty\right)^2 \|\sqrt{f_{\delta,h}} - \sqrt{f}\|_2^2 \\ &\leq o(\|\delta\|). \end{aligned}$$

Therefore,

$$\kappa_x(f_{\delta,h}) - \kappa_x(f) - 2\delta^T f * (h f)(x) = o(\|\delta\|).$$

Hence the function $\kappa_x(f_{\delta,h})$ is differentiable at 0 with derivative $2f * (h f)(x)$.

3.2. Proof of Proposition 2.5

For any h , $T_{\pi_h} = \{\sum_{i=1}^m a_i h_i : a \in \mathbb{R}^m\} \subseteq \mathcal{L}_{2,0}(F)$. Hence the *tangent set* $T_\Pi = \bigcup_{\pi_h \in \Pi} T_{\pi_h} \subseteq \mathcal{L}_{2,0}(F)$. Conversely, suppose $\psi \in \mathcal{L}_{2,0}(F)$. If $\psi = 0$, then $\psi \in T_{\pi_h}$ for any h (Take $a = 0$). If $\psi \neq 0$, then using Theorem 2.4, we can construct a κ_x -smooth path π_ψ such that $\psi \in T_{\pi_\psi}$ implying $\psi \in T_\Pi$. Hence $\mathcal{L}_{2,0}(F) \subseteq T_\Pi$. Therefore $T_\Pi = \mathcal{L}_{2,0}(F)$ and hence $\dot{\Pi} = \mathcal{L}_{2,0}(F)$.

Now suppose L be any finite-dimensional linear subspace of $\dot{\Pi}$ with an orthonormal basis $[\psi_1, \psi_2, \dots, \psi_m]^T$ and let us define $h = [\psi_1, \psi_2, \dots, \psi_m]^T$. Then $\int h dF = 0$ and $\int h h^T dF = I$ and

$$T_{\pi_h} = \left\{ a^T h = \sum_{i=1}^m a_i \psi_i : a \in \mathbb{R}^m \right\} = L.$$

Hence Π is proper.

Acknowledgment The author would like to thank Anton Schick and Wei-Min Huang whose detailed comments were very helpful; in particular the present proof of Lemma 3.1 belongs to Anton Schick. The author would also like to thank an anonymous referee for some constructive suggestions that improved the presentation of the paper.

References

- Begun, J., Hall, W. J., Huang, W. M., and Wellner, J. (1983). Information and Asymptotic Efficiency in Parametric-nonparametric Models. *Annals of Statistics II* 432–452.
- Bickel, P. J., and Ritov, Y. (1988). Estimating Integrated Squared Density Derivatives: Sharp Best Order Convergence Estimates. *Sankhya, Series A* **50** 381–393.
- Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore.
- Frees, E. W. (1994). Estimating Densities of Functions of Observations. *Journal of American Statistical Association* **89** 517–525.
- Giné, E., and Mason, D. M. (2007). On Local U -statistic Process and the Estimation of Densities of Functions of Several Sample Variables. *Annals of Statistics* **35** 1105–1145.
- Hall, P., and Marron, J. S. (1987). Estimation of Integrated Squared Density Derivatives. *Statistics and Probability Letters* **6** 109–115.
- Pfanzagl, J., and Wefelmeyer, W. (1982). Contributions to a General Asymptotic Statistical Theory. *Lecture Notes in Statistics* **13** Springer, Berlin.
- Prakasa Rao, B. L. S. (1983). *Nonparametric Functional Estimation*. Academic Press, New York.
- Saavedra, A., and Cao, R. (2000). On the Estimation of the Marginal Density of a Moving Average Process. *Canadian Journal of Statistics* **28** 799–815.
- Schick, A. (1996). Efficient Estimation in a Semiparametric Additive Regression Model with Autoregressive Errors. *Stochastic Processes and their Applications* **61** 339–361.
- Schick, A., and Wefelmeyer, W. (2004). Root-n Consistent Density Estimators for Sums of Independent Random Variables. *Journal of Nonparametric Statistics* **16** 925–935.
- Schick, A., and Wefelmeyer, W. (2007). Root-n Consistent Density Estimators of Convolutions in Weighted \mathcal{L}_1 Norms. *Journal of Statistical Planning and Inference* **137** 1765–1774.