Journal of Data Science and Modeling, Vol. 1, No. 1, 1-9, December 2022

Research Manuscript

A Berry-Esseen Type Bound for a Smoothed Version of Grenander Estimator

Zohre Lotfi Colvani*[1], Raheleh Zamini[2]

1. Assistant professor of statistics, Kharazmi University, Thehran, Iran.
2. M.Sc. student in statistics, Kharazmi University, Thehran, Iran.

Received: 28/3/2017 Accepted: 23/7/2017

Abstract:

In various statistical models, such as density estimation and estimation of regression curves or hazard rates, monotonicity constraints can arise naturally. A frequently encountered problem in nonparametric statistics is to estimate a monotone density function f on a compact interval. A known estimator for the density function of f under the restriction that f is decreasing, is Grenander estimator, where is the left derivative of the least concave majorant of the empirical distribution function of the data. Many authors worked on this estimator and obtained beneficial properties for this estimator. Grenander estimator is a step function, and hence, it is not smooth. In this paper, we discuss the estimation of a decreasing density function by the kernel smoothing method. Many works have been done due to the importance and applicability of the Berry-Esseen bound for the density estimator. In this paper, we study a Berry- Esseen type bound for a smoothed version of Grenander estimator.

Keywords: Berry-Esseen, Grenander estimator, Kernel, Least concave majorant.
Mathematics Subject Classification (2010): 62G05, 62G07,62G20.

*Corresponding author: lotfi-asal@ymail.com

1. Introduction

In various statistical models, such as density estimation and estimation of regression curves or hazard rates, monotonicity constraints can arise naturally. Let F_n be the empirical distribution function of a sample $X_1, \dots, X_n$ and suppose that the distribution F of the $X_i{}'s$ is continuous and concentrated on $[a, b]$. A famous estimator for the density function $f = F'$ under the restriction that f is decreasing is the Grenander $\tilde{f}_n$, defined on $(a, b]$ as the left derivative of the least concave majorant $\hat{F}_n$ of F_n, with $\tilde{f}_n(a) = \lim_{s\downarrow a} \tilde{f}_n(s)$. (See e.g. Durot and Lopuhaä (2014)). This estimator is a step function, and as a consequence, it is not smooth. In this paper, we are interested in an estimator that is both decreasing and smooth. We study a Berry-Esseen type bound for this estimator.

Many works have been done in different data sampling models due to the importance applicability of density estimation and Berry- Esseen bounds for the density estimator. See e.g., Chang and Rao (1989), Dewan and Prakasa Rao (2007), Birkel (1988), Isogai (1994), Sun and Zhu (1999), Zhou et al. (2006), Liang and Uña-Álvarez (2009), Liang and Baek (2008), Huang et al. (2011), Yang et al. (2012) among others.

Durot and Lopuhaä (2014) defined a smoothed version of the $\tilde{f}_n$ by

$$\hat{f}_n(t) = \begin{cases} \hat{f}_n(a+h_n) + \hat{f}_n'(a+h_n)(t-a-h_n), & a \le t \le a+h_n \\ \frac{1}{h_n}\int_{\mathbb{R}} K\big(\frac{t-x}{h_n}\big) d\hat{F}_n(x) & a+h_n \le t \le b-h_n \\ \hat{f}_n(b-h_n) + \hat{f}_n'(b-h_n)(t-b+h_n) & b-h_n \le t \le b, \end{cases}$$

where h_n is a sequence of positive bandwidths tending to zero as $n \to \infty$ and the kernel function $K : [-1, 1] \to [0, \infty)$ satisfies $\int_{\mathbb{R}} K(t)dt = 1$.

One of the important problems to be investigated is to establish a Berry-Esseen-type result for the smooth estimate $\hat{f}_n(y)$, thereby providing asymptotic normality of $\sqrt{nh_n}\big(\hat{f}_n(y) - f(y)\big)$ with rates. In the present paper, and under the some basic assumptions, we discuss and resolve this problem. More precisely, it is shown that, under suitable regularity conditions:

$$\begin{aligned} \sup_{x\in\mathbb{R}} \Big| P\Big[\sqrt{nh_n}\Big(\hat{f}_n(y) - f(y)\Big) \le x\hat{\sigma}_n(y)\Big] - \Phi(x)\Big| &= O\big(\frac{(\log\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}\big) \\ &\quad + O\big((nh_n)^{-\frac{1}{2}}\big) + O\Big(\frac{(\log n)^{2/3}}{n^{1/6} h_n^{1/2}}\Big) \\ &\quad + O\big(h_n\big) + O\big(n^{1/2} h_n^{5/2}\big) \quad a.s. \end{aligned}$$

In Section 2, the main results are stated. Proofs of the results are given in the Appendix.

2. Main results

Before presenting the main results, we introduce some notations and mention some assumptions that are used in the following theorems and lemmas.

For every $y \in [a+h_n, b-h_n]$, consider an ordinary kernel estimator $f_n(y) = \frac{1}{h_n}\int K\big(\frac{y-x}{h_n}\big)dF_n(x)$. It is easy to see that

$$Var(f_n(y)) = \frac{1}{nh_n^2} var\big(K\big(\frac{y-X_1}{h_n}\big)\big)$$
$$= \frac{1}{nh_n}\int K^2(t)f(y-th_n)dt - \frac{h_n}{n}\Big(\int K(t)f(y-th_n)dt\Big)^2.$$

Let $\sigma_n^2(y) := nh_n Var\left(f_n\left(y\right)\right), \quad a+h_n \le y \le b-h_n$,
and $\sigma^2\left(y\right) := f(y)\int \left(K(u)\right)^2 du \quad a+h_n \le y \le b-h_n$.

Assumptions

$\mathbf{H_1}$ $\lim_{n\to\infty} nh_n = \infty$.
$\mathbf{H_2}$ $\lim_{n\to\infty} nh_n^5 = 0$.
$\mathbf{H_3}$ $\lim_{n\to\infty} \frac{nh_n^3}{(\log n)^4} = \infty$.
$\mathbf{K_1}$ $\int_{-1}^{+1} tK(t)dt = 0$
$\mathbf{K_2}$ $\int_{-1}^{+1} t^2|K(t)|dt < \infty$.
$\mathbf{K_3}$ $\int_{-1}^{+1} |K^m(t)|dt < \infty \quad m = 2, 3$.
F f, f' and f" are bounded.

In the following theorems, the Berry-Esseen type bounds for $\hat{f}_n(x)$ are presented.

Theorem 2.1. *Suppose that* $\mathbf{F}, \mathbf{K_3}$ *and* $\mathbf{H_1}$ *hold. Then for* $a+h_n \le y \le b-h_n$ *we have*

$$\sup_{x\in\mathbb{R}}\left|P\left(\sqrt{nh_n}[f_n(y) - Ef_n(y)] \le x\sigma_n(y)\right) - \Phi(x)\right| = O\left((nh_n)^{-\frac{1}{2}}\right),$$

where the letter O in the $h(n) = O(g(n))$ means that $|h|$ is bounded above by g asymptotically.

Theorem 2.2. *Suppose that fis strictly decreasing function. If* $\mathbf{F}, \mathbf{K_3}$, $\mathbf{H_1}$ *and* $\mathbf{H_3}$ *hold, then for* $a+h_n \le y \le b-h_n$, *we have*

$$\left|P\left(\sqrt{nh_n}[\hat{f}_n(y) - Ef_n(y)] \le x\sigma_n(y)\right) - \Phi(x)\right| = O\big((nh_n)^{-\frac{1}{2}}\big) + O\Big(\frac{(\log n)^{2/3}}{n^{1/6}h_n^{1/2}}\Big).$$

Theorem 2.3. *Let f be strictly decreasing function. Under Assumptions* **F**, $\mathbf{K_1}$- $\mathbf{K_3}$, $\mathbf{H_1}$- $\mathbf{H_3}$, *for* $a+h_n \le y \le b-h_n$ *we can write*

$$\sup_{x\in\mathbb{R}}\left|P\left(\sqrt{nh_n}\Big[\hat{f}_n(y) - f(y)\Big] \le x\sigma(y) - \Phi(x) = O\big((nh_n)^{-\frac{1}{2}}\big)\right.\right.$$
$$+O\Big(\frac{(\log n)^{2/3}}{n^{1/6}h_n^{1/2}}\Big) + O\big(h_n\big) + O\big(n^{1/2}h_n^{5/2}\big).$$

Remark 2.4. *When f is unknown, Theorem 2.3 is not applicable. For example, it is not useful in finding a confidence interval for $f(y)$ or hypothesis testing. So we estimate $\sigma^2(y)$ by*

$$\hat{\sigma}_n^2(y) =: \hat{f}(y) \int (K(u))^2 du, \quad a + h_n \le y \le b - h_n. \tag{2.1}$$

In Lemma 2.5, the consistency of these proposed estimators is studied. Then we present another version of Theorem 2.3 using these estimators.

Lemma 2.5. *Let $\lim_{n\to\infty} \frac{nh_n^2}{\log\log n} = \infty$. We have*

$$\sup_{a+h_n \le y \le b-h_n} \left|\hat{\sigma}_n^2(y) - \sigma^2(y)\right| = O\big(h_n^2\big) + O\Big(\frac{(\log\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}\Big), \quad a.s.$$

Theorem 2.6. *Under assumptions of Theorem 2.3, for $a + h_n \le y \le b - h_n$, we have*

$$\begin{aligned}
\sup_{x\in\mathbb{R}} \left| P\left[\sqrt{nh_n}\left(\hat{f}_n(y) - f(y)\right) \le x\hat{\sigma}_n(y)\right] - \Phi(x)\right| &= O\Big(\frac{(\log\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}\Big) \\
&\quad + O\big((nh_n)^{-\frac{1}{2}}\big) + O\Big(\frac{(\log n)^{2/3}}{n^{1/6} h_n^{1/2}}\Big) \\
&\quad + O\big(h_n\big) + O\big(n^{1/2} h_n^{5/2}\big) \quad a.s.
\end{aligned}$$

Appendix

In order to prove the Berry-Esseen theorems, we need the following auxiliary result.

Lemma 2.7. *Suppose that* **F** *and* $\mathbf{K_3}$ *hold , then we have*

$$\sup_{a+h_n \le y \le b-h_n} \left|\sigma_n^2(y) - \sigma^2(y)\right| = O(h_n). \tag{2.2}$$

Proof. From the definition of $\sigma_n^2(y)$ and $\sigma^2(y)$ we can write

$$\begin{aligned}
\left|\sigma_n^2(y) - \sigma^2(y)\right| &\le \left|\int K^2(t) f(y - th_n) dt - f(y) \int K^2(t) dt\right| \\
&+ h_n^2 \left|\int K(t) f(y - th_n) dt\right|^2 \\
&=: I(y) + II(y).
\end{aligned} \tag{2.3}$$

Now, using the mean value theorem, we have

$$\begin{aligned}
\sup_y |I(y)| &\le h_n \sup_y |f'(y)| \int (K(u))^2 du \\
&= O(h_n).
\end{aligned} \tag{2.4}$$

Next, to deal with $II(x)$, we observe that

$$\lim_{n\to\infty}\Big(\int K(t)f(y-th_n)dt\Big)^2 = f^2(y)\Big(\int K(t)dt\Big)^2.$$

Therefore

$$\sup_x \left|II(x)\right| = O(h_n^2). \tag{2.5}$$

(2.3),(2.4) and (2.5) complete the proof. □

Proof of Theorem 2.1. Set

$$\begin{aligned} f_n(y)-E(f_n(y)) &= \frac{1}{nh_n}\sum_{i=1}^{n}\Big\{K\Big(\frac{y-X_i}{h_n}\Big)-E\Big(\frac{y-X_i}{h_n}\Big)\Big\} \\ &=: \sum_{i=1}^{n} L_{ni}. \end{aligned}$$

By Theorem 5.7 of Petrov (1995), it can be written that

$$\begin{aligned} &\sup_{x\in\mathbb{R}}\left|\mathrm{P}\left[\sqrt{nh_n}\left(f_n\left(y\right)-Ef_n\left(y\right)\right)\le x\sigma_n\left(y\right)\right]-\Phi\left(x\right)\right| \\ &\le \frac{Cn^{5/2}h_n^{3/2}}{\sigma_n^3(y)}E|L_{n1}|^3 \\ &\le \frac{2Cn^{-1/2}h_n^{-3/2}}{\sigma_n^3(y)}E\left|K\Big(\frac{y-X_1}{h_n}\Big)\right|^3 \\ &= \frac{Ch_n^{-1/2}n^{-1/2}}{\sigma_n^3(y)}\int_{\mathbb{R}}\left|K(t)\right|^3 dtf(y-th_n)dt. \end{aligned} \tag{2.6}$$

Sine f is bounded in a neighbourhood of y, Lemma 2.7 yields the result. □

Proof of Theorem 2.2. By using Lemma 2 of Chang and Rao (1989) for $a = \frac{\sqrt{nh_n}}{\sigma_n(y)}\left|\hat{f}_n\left(y\right)-f_n\left(y\right)\right|$, we can see that

$$\begin{aligned} \sup\nolimits_{x\in\mathbb{R}} \quad &\left|P\Big[\sqrt{nh_n}\Big(\hat{f}_n(y)-E(f_n(y))\Big)\le x\sigma_n(y)\Big]-\Phi(x)\right| \\ \le \quad &\sup_{x\in\mathbb{R}}\left|P\Big[\sqrt{nh_n}\Big(f_n(y)-E(f_n(y))\Big)\le x\sigma_n(y)\Big]-\Phi(x)\right| + \frac{a}{\sqrt{2\pi}}. \end{aligned} \tag{2.7}$$

Kiefer and Wolfowitz (1976) proved that when f is strictly decreasing function,

$$\sup_{x\in\mathbb{R}}\left|\hat{F}_n(x)-F_n(x)\right| = O\Big(\big(\frac{\log n}{n}\big)^{2/3}\Big) \quad a.s. \tag{2.8}$$

This last result and Lemma 2.7 conclude that

$$\begin{aligned}
\sup_{a+h_n \le y \le b-h_n} \left|\frac{a}{\sqrt{2\pi}}\right| &\le \frac{n^{1/2}}{\sqrt{2\pi}\sigma_n(y)h_n^{1/2}} \sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F_n(x)\right| \int_{\mathbb{R}} \left|dK(u)\right| \\
&= O\Big(n^{-1/6} h_n^{-1/2} (\log n)^{2/3}\Big) \quad a.s. \qquad (2.9)
\end{aligned}$$

Using (2.7), (2.9) and Theorem 2.1, we get the result. □

Proof of Theorem 2.3. Using Lemma 2 of Chang and Rao (1989), it is not difficult to show that

$$\begin{aligned}
&\sup\nolimits_{x\in\mathbb{R}} \quad \left|P\Big[\sqrt{nh_n}\Big(\hat{f}_n(y) - f(y)\Big) \le x\sigma(y)\Big] - \Phi(x)\right| \\
&\le \quad \sup_{x\in\mathbb{R}} \left|P\Big[\frac{\sqrt{nh_n}}{\sigma_n(y)}\Big(\hat{f}_n(y) - E(f_n(y))\Big) \le x\frac{\sigma(y)}{\sigma_n(y)}\Big] - \Phi\Big(\frac{\sigma(y)}{\sigma_n(y)}x\Big)\right| \\
&+ \quad \sup_{x\in\mathbb{R}} \left|\Phi\Big(\frac{\sigma(y)}{\sigma_n(y)}x\Big) - \Phi(x)\right| + \frac{\sqrt{nh_n}}{\sqrt{2\pi}\sigma(y)}\Big|E(f_n(y)) - f(y)\Big|. \qquad (2.10)
\end{aligned}$$

Alternatively, a little calculation and Lemma 2.7 yield

$$\begin{aligned}
\sup_{x\in\mathbb{R}} \left|\Phi\left(\frac{\sigma(y)}{\sigma_n(y)}x\right) - \Phi(x)\right| &= O\left(\left|\sigma_n^2(y) - \sigma^2(y)\right|\right) \\
&= O(h_n). \qquad (2.11)
\end{aligned}$$

On the other hand, by a two term Taylor expansion and condition $\int_{-1}^{+1} tK(t)dt = 0$, we obtain

$$E(f_n(y)) - f(y) = \frac{1}{2}h_n^2 f''(\theta_n(y)) \int_{\mathbb{R}} u^2|K(u)|du$$

with

$$y - h_n \le \theta_n(y) \le y + h_n. \qquad (2.12)$$

Condition $\sup_{y\in\mathbb{R}} \left|f''(y)\right| < \infty$ and (2.12) imply that there is a constant C such that $\left|f''(\theta_n(y))\right| \le C$. Hence

$$\sup_y \Big|E(f_n(y)) - f(y)\Big| = O(h_n^2). \qquad (2.13)$$

Now Theorem 2.3 follows from (2.10), (2.11), (2.13) and Theorem 2.2.

□

Proof of Lemma 2.5. By using the definition of $\hat{\sigma}_n^2(y)$ and $\sigma^2(y)$ for $a+h_n \leq y \leq b-h_n$, we have

$$\begin{aligned} \left|\hat{\sigma}_n^2(y)-\sigma^2(y)\right| &\leq \int K^2(t)dt\left|\hat{f}(y)-f(y)\right| \\ &\leq \left(\int K^2(t)dt\right)\left(\left|\hat{f}(y)-E(f_n(y))\right|+\left|E(f_n(y))-f(y)\right|\right) \\ &\leq \left(\int K^2(t)dt\right)\Big(h_n^{-1}\sup_x\left|\hat{F}(x)-F(x)\right|\int\left|dK(u)\right| \\ &\quad +\left|E(f_n(y))-f(y)\right|\Big). \qquad (2.14) \end{aligned}$$

Now, from Marshall's lemma (Marshall (1970)) and Chung's law of the iterated logarithm for $\sup_x\left|F_n(x)-F(x)\right|$ (see e.g. Shorack and Wellner (1986), Page 505), we know that with $b_n=(2\log\log n)^{1/2}$,

$$\begin{aligned} \limsup_{n\to} n^{1/2}\frac{\sup_x\left|\hat{F}(x)-F(x)\right|}{b_n} &\leq \limsup_{n\to} n^{1/2}\frac{\sup_x\left|F_n(x)-F(x)\right|}{b_n} \\ &= \frac{1}{2} \quad a.s. \qquad (2.15) \end{aligned}$$

(2.14),(2.15) and (2.13) follow the result.

□

Proof of Theorem 2.6.

$$\begin{aligned} &\sup_{x\in\mathbb{R}}\left|P\left[\sqrt{nh_n}\left(\hat{f}_n(y)-f(y)\right)\leq x\hat{\sigma}_n(y)\right]-\Phi(x)\right| \\ &\leq \sup_{x\in\mathbb{R}}\left|P\left[\frac{\sqrt{nh_n}}{\sigma(y)}\left(\hat{f}_n(y)-f(y)\right)\leq x\frac{\hat{\sigma}_n(y)}{\sigma(y)}\right]-\Phi\left(x\frac{\hat{\sigma}_n(y)}{\sigma(y)}\right)\right| \\ &\quad +\sup_{x\in\mathbb{R}}\left|\Phi\left(x\frac{\hat{\sigma}_n(y)}{\sigma(y)}\right)-\Phi(x)\right|. \end{aligned}$$

(2.16)

Now, Lemma 2.5 follows that

$$\begin{aligned} \sup_{x\in\mathbb{R}}\left|\Phi\left(x\frac{\hat{\sigma}_n(y)}{\sigma(y)}\right)-\Phi(x)\right| &= O\left(\left|\hat{\sigma}_n^2(y)-\sigma^2(y)\right|\right) \\ &= O\left(\frac{(\log\log n)^{1/2}}{n^{1/2}h_n}\right)+O(h_n^2) \quad a.s. \quad (2.17) \end{aligned}$$

This last result, (2.16) and Theorem 2.3 give the result. □

Conclusions

In this paper, we study a Berry- Esseen type bound for a smoothed version of the Grenander estimator.

References

Birkel T. (1988), On the convergence rate in the central limit theorem for associated processes, *The Annals of Probability,* **16,** 1685–1698.

Chang M. N. and Rao P. V. (1989), Berry-Esseen bound for the Kaplan-Meier estimator, *Communications in Statistics-Theory and Methods*, **18(12)**, 4647–4664.

Dewan I. and Prakasa Rao B. L. S. (2007), Non-uniform and uniform Berry-Esseen type bounds for stationary associated sequences, *Journal of Nonparametric Statistics,* **17(2)**, 217–235.

Durot C. and Lopuhaä H. (2014), A Kiefer-Wolfowitz type of result in a general setting, with an application to smooth monotone estimation, *Electronic Journal of Statistics, Institute of Mathematical Statistics and Bernoulli Society*, **8(2)**, 2479-2513.

Huang C., Wang H. and Zhang L. (2011), Berry-Esseen bounds for kernel estimates of stationary processes, *Journal of Statistical Planning and Inference,* **141**, 1290–1296.

Isogai E. (1994), A Berry-Esseen-type bound for recursive estimators of a density and its derivatives, *Journal of Statistical Planning and Inference*, **40**, 1–14.

Kiefer J. and Wolfowitz J. (1976), Asymptotically minimax estimation of concave and convex distribution functions, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **34**, 73-85.

Liang H. Y. and Uña-Álvarez J. (2009), A Berry-Esseen type bound in kernel density estimation for strong mixing censored samples, *Journal of Multivariate Analysis*, **100**, 1219–1231.

Liang H. and Baek J. (2008), Berry-Esseen bounds for density estimates under NA assumption, *Metrika,* **68,** 305–322.

Marshall A. W. (1970), Discussion on Barlow and van Zwet's paper. In *Nonparametric Techniques in Statistical Inference,* (M. L. Puri, ed.).*Proceedings of the First International Symposium on Nonparametric Techniques held at Indiana University, June* 1969 174–176. Cambridge University Press, London.

Petrov V. V. (1995), *Limit theorems of probability theory*, Oxford University Press Inc., New York.

Sun L. and Zhu L. (1999), A Berry-Esseen type bound for kernel density estimators under Random Censorship, *Acta Mathematica Sinica,* **42,** 627–636.

Shorack G. R. and Wellner J. A. (1986), *Empirical processes with applications to statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics,* John Wiley and Sons, Inc., New York.

Yang W., Hu S., Wang X. and Ling N. (2012), The Berry-Esseen type bound of sample quantiles for strong mixing sequence, *Journal of Statistical Planning and Inference,* **142**, 660–672.

Zhou Y., Wu G. and Li D. (2006), A kernel-type estimator of a quantile function under randomly truncated data, *Acta Mathematica Scientia,* **26B**, 585–594.