Proceedings of the Royal Society of Edinburgh, **151**, 464–472, *2021*
DOI:10.1017/prm.2020.27

Nonasymptotic bounds for the quadratic risk of the Grenander estimator

Malkhaz Shashiashvili
Department of Mathematics, Faculty of Exact and Natural Sciences, Iv. Javakhishvili Tbilisi State University, 3 University Str., 0143 Tbilisi, Georgia (malkhaz.shashiashvili@tsu.ge)

(MS received 30 September 2019; accepted 9 March 2020)

This paper is dedicated to Professor Estate Khmaladze on the occasion of his 75th birthday.

There is an enormous literature on the so-called Grenander estimator, which is merely the nonparametric maximum likelihood estimator of a nonincreasing probability density on $[0,1]$ (see, for instance, Grenander (1981)), but unfortunately, there is no nonasymptotic (i.e. for arbitrary finite sample size n) explicit upper bound for the quadratic risk of the Grenander estimator readily applicable in practice by statisticians. In this paper, we establish, for the first time, a simple explicit upper bound $2n^{-1/2}$ for the latter quadratic risk. It turns out to be a straightforward consequence of an inequality valid with probability one and bounding from above the integrated squared error of the Grenander estimator by the Kolmogorov–Smirnov statistic.

Keywords: Grenander estimator; nonasymptotic bounds; nonparametric estimation; quadratic risk bounds; Kolmogorov–Smirnov statistic

2010 Mathematics Subject Classification: Primary 62DXX
Secondary 62GXX

1. Introduction and the formulation of main results

Nonparametric density estimation has mainly been devoted for a long time, to estimation of smooth densities using linear methods such as kernel estimators with fixed bandwidth or projection estimators (truncated series expansions, with estimated coefficients).

Suppose we know only that $f(x)$, $0 \leqslant x \leqslant 1$, is a nonincreasing probability density, unbounded and discontinuous, in general. It can very well be steep at some places and flat elsewhere. In this case, a special estimator $\widehat{f}_n(x)$, $0 \leqslant x \leqslant 1$, has been known for a long time. It has been introduced by Grenander [4] as the left derivative of the least concave majorant $\widehat{F}_n(x)$, $0 \leqslant x \leqslant 1$, of the empirical distribution function $F_n(x)$, $0 \leqslant x \leqslant 1$, and which is exactly the nonparametric maximum likelihood estimator of $f(x)$, $0 \leqslant x \leqslant 1$, restricted to the class $\mathcal{F}$ of all nonincreasing densities on $[0,1]$ (for a proof see, for instance, Grenander [5]).

There is an enormous literature on the Grenander estimator and its asymptotic properties, in particular, Prakasa Rao [10], Kiefer and Wolfowitz [9], Groeneboom

© The Author(s), 2020. Published by Cambridge University Press on behalf of The Royal Society of Edinburgh

https://doi.org/10.1017/prm.2020.27 Published online by Cambridge University Press

and Pyke [7], Groeneboom [6] and van der Vaart [11]. These studies give very deep asymptotic analysis of the Grenander estimator and the corresponding limiting probability distributions.

Contrary to the latter, our point of view is nonasymptotic, that is, for arbitrary sample size n, moderate or even big, our objective consists in establishing the explicit and simply computable upper bound for the quadratic risk of the Grenander estimator of the type

$$E_f \int_0^1 \left(\widehat{f}_n(x) - f(x)\right)^2 \mathrm{d}x \leqslant c \cdot n^{-\delta}, \tag{1.1}$$

where $\delta > 0$ and the constant c, $c > 0$, should depend on the entire class $\mathcal{F}$ of all nonincreasing densities and, by no means, on the individual representative f, $f \in \mathcal{F}$.

Bound (1.1) should be readily applicable in practice by a statistician in the following pragmatic situation:

Suppose, we know that the unknown probability density $f(x)$, $0 \leqslant x \leqslant 1$, is nonincreasing, but nothing can be assumed about its smoothness. How small will be the quadratic risk of the Grenander estimator $\widehat{f}_n(x)$, $0 \leqslant x \leqslant 1$, if we decide to use it as an estimator of the unknown density $f(x)$, $0 \leqslant x \leqslant 1$?

Unfortunately, the class $\mathcal{F}$ contains densities which are not square integrable, in particular, the following ones

$$f_\beta(x) = (1-\beta)x^{-\beta}, \quad 0 \leqslant x \leqslant 1, \quad \text{with } \frac{1}{2} < \beta < 1, \tag{1.2}$$

hence for these types of densities classical quadratic risk (1.1) cannot be defined.

It turns out, surprisingly enough, that adjusting a bit definition of the quadratic risk by introducing a weight function $h(x) = x$, $0 \leqslant x \leqslant 1$, all nonincreasing densities become square integrable, i.e.

$$\int_0^1 f^2(x) x \, \mathrm{d}x < \infty \quad \text{for any } f \in \mathcal{F}. \tag{1.3}$$

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with values in $[0, 1]$ and the nonincreasing density function $f(x)$, $0 \leqslant x \leqslant 1$, and let $F(x)$, $0 \leqslant x \leqslant 1$, be the corresponding cumulative distribution function with $F(0) = 0$ and

$$F(x) = \int_0^x f(y) \, \mathrm{d}y, \quad 0 \leqslant x \leqslant 1. \tag{1.4}$$

We shall assume that $f(x)$, $0 \leqslant x \leqslant 1$, is a right-continuous version of the density function. Then it is evident that $f(x)$, $0 \leqslant x \leqslant 1$, is a right derivative of the concave function $F(x)$, $0 \leqslant x \leqslant 1$, and its left limit $f(x-)$, $0 < x \leqslant 1$, coincides with the left derivative of $F(x)$, $0 \leqslant x \leqslant 1$. It is a well-known mathematical fact that $f(x) = f(x-)$ everywhere except a countable set of points x, $0 \leqslant x \leqslant 1$.

Let $F_n(x)$, $0 \leqslant x \leqslant 1$, be the empirical distribution function, constructed from i.i.d. random variables $X_1, X_2, \ldots, X_n$. As the corresponding distribution function

https://doi.org/10.1017/prm.2020.27 Published online by Cambridge University Press

$F(x)$ is absolutely continuous with $F(0) = 0$, with probability one we have that the random variable

$$X_{\min} = \min(X_1, X_2, \ldots, X_n)$$

is strictly positive and hence

$$F_n(x) = 0 \quad \text{if } 0 \leqslant x < X_{\min} \ (P\text{-a.s.}). \tag{1.5}$$

Let $\widehat{F}_n(x)$, $0 \leqslant x \leqslant 1$, be the least concave majorant of $F_n(x)$, $0 \leqslant x \leqslant 1$, and let $\widehat{f}_n(x)$, $0 \leqslant x \leqslant 1$, denote the right derivative of the latter concave majorant.

Evidently, we have

$$\widehat{F}_n(x) = \int_0^x \widehat{f}_n(y)\,\mathrm{d}y, \quad 0 \leqslant x \leqslant 1, \quad \widehat{F}_n(0) = 0 \quad (P\text{-a.s.}). \tag{1.6}$$

As $F_n(x)$, $0 \leqslant x \leqslant 1$, is a nondecreasing function, the same property holds for its least concave majorant $\widehat{F}_n(x)$, $0 \leqslant x \leqslant 1$, and hence $\widehat{f}_n(x)$, $0 \leqslant x \leqslant 1$, is a nonnegative function. Moreover, the latter function is nonincreasing as the right derivative of the concave function. By its construction the function $\widehat{F}_n(x)$, $0 \leqslant x \leqslant 1$, is piecewise linear concave function and as a result we get that the function $\widehat{f}_n(x)$, $0 \leqslant x \leqslant 1$, is right-continuous step function, nonnegative and nonincreasing. The left limit $\widehat{f}_n(x-)$, $0 < x \leqslant 1$, of the function $\widehat{f}_n(x)$, $0 \leqslant x \leqslant 1$, coincides with the left derivative of the least concave majorant $\widehat{F}_n(x)$, $0 \leqslant x \leqslant 1$, and hence it is the celebrated Grenander estimator of the unknown nonincreasing density function $f(x)$, $0 \leqslant x \leqslant 1$.

Lemma 1.1. *We have for arbitrary nonincreasing density $f(x)$, $0 \leqslant x \leqslant 1$*

$$0 \leqslant f(x)x \leqslant F(x), \quad \lim_{x \downarrow 0}(f(x)x) = 0, \tag{1.7}$$

$$\int_0^1 f^2(x)x\,\mathrm{d}x \leqslant \frac{1}{2}. \tag{1.8}$$

Proof. As the density $f(x)$, $0 \leqslant x \leqslant 1$, is nonincreasing, we get

$$F(x) = \int_0^x f(y)\,\mathrm{d}y \geqslant \int_0^x f(x)\,\mathrm{d}y = f(x)x, \quad 0 \leqslant x \leqslant 1.$$

As $\lim_{x\downarrow 0} F(x) = F(0) = 0$, we come to relations (1.7). We use inequality (1.7) and write

$$\int_0^1 f(x)(f(x)x)\,\mathrm{d}x \leqslant \int_0^1 f(x)F(x)\,\mathrm{d}x = \int_0^1 \frac{1}{2}\,\mathrm{d}(F^2(x)) = \frac{1}{2}.$$

From now on we introduce the weighted quadratic risk of the Grenander estimator $\widehat{f}_n(x-)$, $0 < x \leqslant 1$,

$$E_f \int_0^1 \big(\widehat{f}_n(x-) - f(x)\big)^2 x\,\mathrm{d}x = E_f \int_0^1 \big(\widehat{f}_n(x) - f(x)\big)^2 x\,\mathrm{d}x. \tag{1.9}$$

https://doi.org/10.1017/prm.2020.27 Published online by Cambridge University Press

Our objective is to establish simple explicit upper bound for the following weighted quadratic risk

$$\sup_{f\in\mathcal{F}} E_f \int_0^1 \big(\widehat{f}_n(x) - f(x)\big)^2 x\,\mathrm{d}x. \tag{1.10}$$

□

In § 2, we shall prove the following main result of this paper.

THEOREM 1.2 Main result. (a) *The weighted integrated squared error of the Grenander estimator $\widehat{f}_n(x-)$, $0 < x \leqslant 1$, is bounded from above with probability one by the Kolmogorov–Smirnov statistic*

$$\int_0^1 \big(\widehat{f}_n(x) - f(x)\big)^2 x\,\mathrm{d}x \leqslant 2 \sup_{0\leqslant x\leqslant 1} |F_n(x) - F(x)| \quad (P\text{-}a.s.). \tag{1.11}$$

(b) *The weighted quadratic risk of the Grenander estimator $\widehat{f}_n(x-)$, $0 < x \leqslant 1$, admits the following simple upper bound for arbitrary finite sample size n*

$$\sup_{f\in\mathcal{F}} E_f \int_0^1 \big(\widehat{f}_n(x) - f(x)\big)^2 x\,\mathrm{d}x \leqslant 2n^{-1/2}. \tag{1.12}$$

Nonasymptotic L^1-risk bounds of the Grenander estimator were obtained by Birge [**1**, **2**]. His upper bound is of the following type

$$E_f \int_0^1 |\widehat{f}_n(x) - f(x)|\,\mathrm{d}x \leqslant c(f) n^{-1/3} \tag{1.13}$$

which is well-adapted to the well-known 'cube-root' asymptotic convergence of the Grenander estimator. Unfortunately Birge's constant factor $c(f)$ depends on the supremum of the unknown nonincreasing density function $f(x)$, $0 \leqslant x \leqslant 1$, and hence is not directly applicable in practice.

Theorem 24.6 in van der Vaart [**11**] states that the integrated squared error of the Grenander estimator is asymptotically of order $n^{-2/3}$ for bounded nonincreasing densities. This suggests that the following stronger than (1.12) nonasymptotic upper bound might be true

$$\sup_{f\in\mathcal{F}} E_f \int_0^1 \big(\widehat{f}_n(x) - f(x)\big)^2 x\,\mathrm{d}x \leqslant cn^{-2/3} \tag{1.14}$$

with a 'reasonable' constant factor c. We were unable to prove the hypothetical upper bound (1.14) and it is worth here to make an important remark: from the asymptotic point of view bound (1.14) is superior over bound (1.12) even for constants c being too large. But what is asymptotically justified can be far from reasonable with an ordinary (even large) amount of observations.

Indeed, let us take $c = 20$ in (1.14). If we equate the right-hand sides of (1.12) and (1.14), we shall have $n = 10^6$. Hence for the sample sizes $n < 1$ million bound (1.12) is better than bound (1.14). The conclusion is that in pragmatic cases the smaller

https://doi.org/10.1017/prm.2020.27 Published online by Cambridge University Press

value of constant factor c is crucial and the direct application of the asymptotical estimates of nonparametric statistics can be quite misleading.

It was noticed long ago (see, e.g. Groeneboom and Pyke [7]) that for the uniform distribution with density $f(x) = 1$, $0 \leqslant x \leqslant 1$, the following asymptotic result does hold

$$E \int_0^1 (\widehat{f}_n(x) - 1)^2 \, \mathrm{d}x \approx (\log n) n^{-1}, \tag{1.15}$$

that is for flat density the order of approximation is almost classical n^{-1}.

A result similar to bound (1.15) holds for arbitrary finite sample size n and for nonincreasing piecewise constant densities on the subset U of $[0, 1]$. This is the subject of theorem 1.3 stated below. Its proof is deferred to § 2.

THEOREM 1.3. *Let $f(x)$, $0 \leqslant x \leqslant 1$, be the nonincreasing density, which is known to be flat, that is, piecewise constant on the union U of the subintervals*

$$U = \bigcup_{i=1}^{m} [a_i, b_i), \tag{1.16}$$

where $0 \leqslant a_1 < b_1 \leqslant a_2 < b_2 \leqslant \cdots \leqslant a_i < b_i \leqslant \cdots \leqslant a_m < b_m \leqslant 1$. Then

(a) *the weighted integrated squared error of the Grenander estimator $\widehat{f}_n(x-)$, $0 < x \leqslant 1$, is bounded from above with probability one by the squared Kolmogorov–Smirnov statistic*

$$\sum_{i=1}^{m} \int_{a_i}^{b_i} \left(\widehat{f}_n(x) - f(x)\right)^2 (x - a_i)(b_i - x) \, \mathrm{d}x \leqslant 4 \left(\sup_{0 \leqslant x \leqslant 1} |F_n(x) - F(x)| \right)^2 \quad (P\text{-}a.s.) \tag{1.17}$$

(b) *the weighted quadratic risk of the Grenander estimator $\widehat{f}_n(x-)$, $0 < x \leqslant 1$, admits the following simple upper bound for arbitrary finite sample size n*

$$E_f \sum_{i=1}^{m} \int_{a_i}^{b_i} \left(\widehat{f}_n(x) - f(x)\right)^2 (x - a_i)(b_i - x) \, \mathrm{d}x \leqslant 4n^{-1}. \tag{1.18}$$

2. Proof of theorems 1.2 and 1.3

We prove at first our main result.

Proof of theorem 1.2. Introduce the following notations

$$\widehat{G}_n(x) = \widehat{F}_n(x) - F(x), \quad 0 \leqslant x \leqslant 1, \tag{2.1}$$

$$\widehat{g}_n(x) = \widehat{f}_n(x) - f(x), \quad \widehat{g}_n(x-) = \widehat{f}_n(x-) - f(x-), \quad 0 \leqslant x \leqslant 1. \tag{2.2}$$

Note that

$$\widehat{G}_n(x) = \int_0^x \widehat{g}_n(y) \, \mathrm{d}y = \int_0^x \widehat{g}_n(y-) \, \mathrm{d}y, \quad 0 \leqslant x \leqslant 1, \tag{2.3}$$

https://doi.org/10.1017/prm.2020.27 Published online by Cambridge University Press

Take $\delta > 0$ arbitrary small, $0 < \delta < 1$. The functions $\widehat{G}_n(x)$ and $\widehat{g}_n(x)$, $0 \leqslant x \leqslant 1$, are right-continuous functions of bounded variation on the interval $[\delta, 1]$ (we remind that the $\lim_{x\downarrow 0} f(x)$ can be equal to $+\infty$).

The important formula of the integration by parts is valid for the functions of bounded variation $\widehat{G}_n(x)$ and $\widehat{g}_n(x)$ on the interval $[\delta, 1]$ and has the following form (see Hewitt and Stromberg [8, theorem 21.67])

$$d(\widehat{G}_n \cdot \widehat{g}_n) = \widehat{G}_n \, d\widehat{g}_n + \widehat{g}_n(-) \, d\widehat{G}_n, \quad \text{or} \quad \widehat{g}_n(-) \, d\widehat{G}_n = d(\widehat{G}_n \cdot \widehat{g}_n) - \widehat{G}_n \, d\widehat{g}_n, \tag{2.4}$$

which after multiplication by x, and the subsequent integration, becomes

$$\int_\delta^1 x\widehat{g}_n(x-) \, d\widehat{G}_n(x) = \int_\delta^1 x \, d\big(\widehat{G}_n(x)\widehat{g}_n(x)\big) - \int_\delta^1 x\widehat{G}_n(x) \, d\widehat{g}_n(x). \tag{2.5}$$

We have

$$\int_\delta^1 x\widehat{g}_n(x-) \, d\widehat{G}_n(x) = \int_\delta^1 (\widehat{g}_n(x-))^2 x \, \mathrm{d}x. \tag{2.6}$$

We have also

$$\begin{aligned} \int_\delta^1 x \, d\big(\widehat{G}_n(x)\widehat{g}_n(x)\big) &= x\widehat{G}_n(x)\widehat{g}_n(x)\Big|_\delta^1 - \int_\delta^1 \widehat{G}_n(x)\widehat{g}_n(x) \, \mathrm{d}x \\ &= -\delta\widehat{G}_n(\delta)\widehat{g}_n(\delta) - \int_\delta^1 \frac{1}{2} \, d(\widehat{G}_n(x))^2 \\ &= -\delta\widehat{G}_n(\delta)\widehat{g}_n(\delta) + \frac{1}{2} \, (\widehat{G}_n(\delta))^2, \quad \text{as } \widehat{G}_n(1) = 0. \end{aligned}$$

We should note that here and throughout the paper the integral $\int_a^b$ stands for $\int_{(a,b]}$, that is, including only right end point.

Thus we get

$$\int_\delta^1 (\widehat{g}_n(x-))^2 x \, \mathrm{d}x = -\delta\widehat{G}_n(\delta)\widehat{g}_n(\delta) + \frac{1}{2} \, (\widehat{G}_n(\delta))^2 - \int_\delta^1 x\widehat{G}_n(x) \, d\widehat{g}_n(x). \tag{2.7}$$

Let us bound the last term of the latter equality (2.7)

$$\begin{aligned} \left| - \int_\delta^1 x\widehat{G}_n(x) \, d\widehat{g}_n(x) \right| &\leqslant \int_\delta^1 x|\widehat{G}_n(x)| \, d(\operatorname{var} \widehat{g}_n(x)) \\ &\leqslant \sup_{0\leqslant x\leqslant 1} |\widehat{G}_n(x)| \int_\delta^1 x \, d\big((-\widehat{f}_n(x)) + (-f(x))\big). \end{aligned}$$

We have

$$\begin{aligned} \int_\delta^1 x \, d\big((-\widehat{f}_n(x)) + (-f(x))\big) &= x\big((-\widehat{f}_n(x)) + (-f(x))\big)\Big|_\delta^1 + \int_\delta^1 \big(\widehat{f}_n(x) + f(x)\big) \, \mathrm{d}x \\ &\leqslant \delta\big(\widehat{f}_n(\delta)) + f(\delta)\big) + \int_\delta^1 \big(\widehat{f}_n(x) + f(x)\big) \, \mathrm{d}x, \end{aligned}$$

https://doi.org/10.1017/prm.2020.27 Published online by Cambridge University Press

hence we get the bound

$$\left| -\int_\delta^1 x\widehat{G}_n(x)\, d\widehat{g}_n(x) \right| \leqslant \sup_{0\leqslant x\leqslant 1} |\widehat{G}_n(x)| \left[\delta\big(\widehat{f}_n(\delta) + f(\delta)\big) + \int_\delta^1 \big(\widehat{f}_n(x) + f(x)\big)\, \mathrm{d}x \right]. \tag{2.8}$$

From equality (2.7) and bound (2.8) we come to the inequality

$$\begin{aligned} &\int_\delta^1 (\widehat{g}_n(x-))^2 x\, \mathrm{d}x \\ &\quad \leqslant \delta\big(\widehat{f}_n(\delta) + f(\delta)\big)|\widehat{G}_n(\delta)| + \frac{1}{2}\,(\widehat{G}_n(\delta))^2 \\ &\qquad + \sup_{0\leqslant x\leqslant 1} |F_n(x) - F(x)| \left[\delta\big(\widehat{f}_n(\delta) + f(\delta)\big) + \int_\delta^1 \big(\widehat{f}_n(x) + f(x)\big)\, \mathrm{d}x \right], \end{aligned} \tag{2.9}$$

where we have used the well-known Marshall's inequality

$$\sup_{0\leqslant x\leqslant 1} |\widehat{F}_n(x) - F(x)| \leqslant \sup_{0\leqslant x\leqslant 1} |F_n(x) - F(x)|. \tag{2.10}$$

Now we pass to limit $\delta \downarrow 0$ in inequality (2.9) taking into account limit relation (1.7) of lemma 1.1 and get bound (1.11) of theorem 1.2.

Denote

$$z_n = \sup_{0\leqslant x\leqslant 1} |F_n(x) - F(x)|. \tag{2.11}$$

Then from theorem 3.3 in Devroye and Lugosi [3] we know that

$$E_f z_n^2 \leqslant \frac{1}{n}, \quad E_f z_n \leqslant \frac{1}{\sqrt{n}}. \tag{2.12}$$

Let us take the expectation in both sides of inequality (1.11), then from the latter estimate (2.12) we come easily to the desired bound (1.12). □

Next we prove theorem 1.3.

Proof of theorem 1.3. Let us introduce the following notation

$$h(x; a_i, b_i) = \begin{cases} (x - a_i)(b_i - x), & \text{if } a_i \leqslant x \leqslant b_i, \\ 0, & \text{elsewhere}, \end{cases} \tag{2.13}$$

$$0 \leqslant x \leqslant 1, \quad i = 1, \ldots, m.$$

We have from Marshall's inequality (2.10) and (2.11)

$$\widehat{G}_n(x) - z_n \leqslant 0, \quad 0 \leqslant x \leqslant 1, \tag{2.14}$$

where

$$\widehat{G}_n(x) = \widehat{F}_n(x) - F(x) = \int_0^x \widehat{g}_n(y)\, \mathrm{d}y, \quad 0 \leqslant x \leqslant 1. \tag{2.15}$$

https://doi.org/10.1017/prm.2020.27 Published online by Cambridge University Press

The integration by parts formula (2.4) gives us

$$\widehat{g}_n(-)\,\mathrm{d}(\widehat{G}_n - z_n) = \mathrm{d}\big[(\widehat{G}_n - z_n)\widehat{g}_n\big] - (\widehat{G}_n - z_n)\,\mathrm{d}\widehat{g}_n. \tag{2.16}$$

Multiplying the latter equality by $h(x; a_i, b_i)$ and integrating it we get

$$\int_{a_i}^{b_i} h(x; a_i, b_i)\widehat{g}_n(-)\,\mathrm{d}(\widehat{G}_n - z_n) = \int_{a_i}^{b_i} h(x; a_i, b_i)\,\mathrm{d}\big[(\widehat{G}_n - z_n)\widehat{g}_n\big] - \int_{a_i}^{b_i} h(x; a_i, b_i)(\widehat{G}_n - z_n)\,\mathrm{d}\widehat{g}_n. \tag{2.17}$$

We have

$$\int_{a_i}^{b_i} h(x; a_i, b_i)\widehat{g}_n(-)\,\mathrm{d}(\widehat{G}_n - z_n) = \int_{a_i}^{b_i} (\widehat{g}_n)^2 h(x; a_i, b_i)\,\mathrm{d}x. \tag{2.18}$$

From the definition of the Lebesgue–Stieltjes integral we can write

$$\int_{a_i}^{b_i} h(x; a_i, b_i)(\widehat{G}_n - z_n)\,\mathrm{d}\widehat{g}_n = \int_{(a_i, b_i)} h(x; a_i, b_i)(\widehat{G}_n - z_n)\,\mathrm{d}\widehat{g}_n + h(b_i; a_i, b_i)(\widehat{G}_n(b_i) - z_n)\big(\widehat{g}_n(b_i) - \widehat{g}_n(b_i-)\big). \tag{2.19}$$

But $h(b_i; a_i, b_i) = 0$, $\widehat{G}_n(x) - z_n \leqslant 0$, and as the density $f(x)$ is constant on $[a_i, b_i)$, then $\widehat{g}_n(x) = \widehat{f}_n(x) - f(x)$ is nonincreasing on $[a_i, b_i)$, hence we come to the crucial inequality

$$\int_{a_i}^{b_i} h(x; a_i, b_i)(\widehat{G}_n - z_n)\,\mathrm{d}\widehat{g}_n \geqslant 0. \tag{2.20}$$

Thus from equality (2.17) taking into account the latter inequality (2.20) we get

$$\int_{a_i}^{b_i} (\widehat{g}_n(x))^2 h(x; a_i, b_i)\,\mathrm{d}x \leqslant \int_{a_i}^{b_i} h(x; a_i, b_i)\,\mathrm{d}\big[(\widehat{G}_n - z_n)\widehat{g}_n\big]. \tag{2.21}$$

Let us bound the right-hand side of (2.21). We have

$$\begin{aligned}
&\int_{a_i}^{b_i} h(x; a_i, b_i)\,\mathrm{d}\big[(\widehat{G}_n - z_n)\widehat{g}_n\big] \\
&= h(x; a_i, b_i)(\widehat{G}_n - z_n)\widehat{g}_n\Big|_{a_i}^{b_i} - \int_{a_i}^{b_i} (\widehat{G}_n - z_n)\widehat{g}_n h'_x(x; a_i, b_i)\,\mathrm{d}x \\
&= -\int_{a_i}^{b_i} h'_x(x; a_i, b_i)\, d\left(\frac{1}{2}\,(\widehat{G}_n - z_n)^2\right) \\
&= -h'_x(x; a_i, b_i)\,\frac{1}{2}\,(\widehat{G}_n - z_n)^2\Big|_{a_i}^{b_i} + \int_{a_i}^{b_i} \frac{1}{2}\,(\widehat{G}_n - z_n)^2 h''_{xx}(x; a_i, b_i)\,\mathrm{d}x
\end{aligned}$$

https://doi.org/10.1017/prm.2020.27 Published online by Cambridge University Press

$$= (b_i - a_i) \cdot \frac{1}{2} \, (\widehat{G}_n(b_i) - z_n)^2 + (b_i - a_i) \cdot \frac{1}{2} \, (\widehat{G}_n(a_i) - a_n)^2$$

$$- \int_{a_i}^{b_i} (\widehat{G}_n - z_n)^2 \, \mathrm{d}x \leqslant 4(b_i - a_i) z_n^2.$$

Hence we get the estimate

$$\int_{a_i}^{b_i} (\widehat{g}_n(x))^2 h(x; a_i, b_i) \, \mathrm{d}x \leqslant 4(b_i - a_i) z_n^2, \quad i = 1, \ldots, m. \tag{2.22}$$

Summing up the latter estimate through i, $i = 1, \ldots, m$, we come to the inequality

$$\sum_{i=1}^{m} \int_{a_i}^{b_i} \big(\widehat{f}_n(x) - f(x)\big)^2 (x - a_i)(b_i - x) \, \mathrm{d}x \leqslant 4 z_n^2. \tag{2.23}$$

Taking the expectation in both sides of inequality (2.23) together with the well-known estimate (2.12) we get the bound

$$E_f \sum_{i=1}^{m} \int_{a_i}^{b_i} \big(\widehat{f}_n(x) - f(x)\big)^2 (x - a_i)(b_i - x) \, \mathrm{d}x \leqslant 4 n^{-1}. \tag{2.24}$$

□

Acknowledgments

The author would like to thank professor Estate Khmaladze for his suggestions which greatly improved a former version of this paper.

References

1 L. Birgé. The Grenander estimator: a nonasymptotic approach. *Ann. Statist.* **17** (1989), 1532–1549.

2 L. Birgé. Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* **25** (1997), 970–981.

3 L. Devroye and G. Lugosi. *Combinatorial methods in density estimation*. Springer Series in Statistics (New York: Springer-Verlag, 2001).

4 U. Grenander. On the theory of mortality measurement. Part II. *Skand. Aktuarietidskr.* **1956** (1957), 125–153.

5 U. Grenander. *Abstract inference*. Wiley Series in Probability and Mathematical Statistics (New York: John Wiley & Sons, Inc., 1981).

6 P. Groeneboom, Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, vol. II (Berkeley, Calif., 1983)*. Wadsworth Statist./Probab. Ser., pp. 539–555 (Belmont, CA: Wadsworth, 1985).

7 P. Groeneboom and R. Pyke. Asymptotic normality of statistics based on the convex minorants of empirical distribution functions. *Ann. Probab.* **11** (1983), 328–345.

8 E. Hewitt and K. Stromberg. *Real and abstract analysis. A modern treatment of the theory of functions of a real variable*. Graduate Texts in Mathematics, vol. 25 (New York, Heidelberg: Springer-Verlag, 1975).

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

10 B. L. S. Prakasa Rao. Estimation of a unimodal density. *Sankhyā Ser. A* **31** (1969), 23–36.

11 A. W. van der Vaart. *Estimation of a unimodal density*. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3 (Cambridge: Cambridge University Press, 1998).

https://doi.org/10.1017/prm.2020.27 Published online by Cambridge University Press