



A second Marshall inequality in convex estimation

Fadoua Balabdaoui, Kaspar Rufibach

► To cite this version:

Fadoua Balabdaoui, Kaspar Rufibach. A second Marshall inequality in convex estimation. 2007. hal-00363236

HAL Id: hal-00363236

<https://hal.science/hal-00363236>

Preprint submitted on 25 Feb 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A second Marshall inequality in convex estimation

Fadoua Balabdaoui^a and Kaspar Rufibach^{b, 1, 2}

^a*Centre de Recherche en Mathématiques de la Décision*

Université Paris-Dauphine, Paris, France

^b*Department of Statistics, Stanford University*

Stanford, CA 94305, United States

Abstract: We prove a second Marshall inequality for adaptive convex density estimation via Least Squares. The result completes the first inequality proved recently by Dümbgen, Rufibach and Wellner (2006), and is very similar to the original Marshall inequality in monotone estimation.

Key words and phrases: Convex density, Cubic polynomial, Kiefer–Wolfowitz theorem, Least Squares estimation, Marshall’s lemma.

1 Introduction and main result

1.1 Marshall’s lemma in monotone estimation

In monotone density estimation, the famous inequality of Marshall (1970) or better known as Marshall’s lemma can be stated as follows: If $\mathbb{G}_n$ is the empirical distribution function of $X_1, \dots, X_n$ from a common nonincreasing density g_0 on $[0, \infty)$ and $\widehat{G}_n$ is the least concave majorant of $\mathbb{G}_n$, then for any concave function G on $[0, \infty)$ the following inequality holds

$$\|\widehat{G}_n - G\|_\infty \leq \|\mathbb{G}_n - G\|_\infty \quad (1.1)$$

for $n \in \mathbb{N} \setminus \{0\}$, and where $\|\cdot\|_\infty$ denotes the supremum norm. Here, we recall that the concave majorant $\widehat{G}_n$ is equal to the cumulative distribution function corresponding to the Grenander estimator; that is the nonparametric Maximum Likelihood Estimator (MLE) of g_0 . Note that (1.1) is specifically true for $G_0(x) = \int_0^x g_0(t)dt$, $x \in [0, \infty)$.

In isotonic regression, the inequality above remains unchanged if $\mathbb{G}_n$ is replaced by the cumulative sum diagram of the data, $\widehat{G}_n$ by its greatest convex minorant and G is now convex rather than concave, as it is usually assumed that the regression curve is monotone increasing.

¹Supported by Swiss National Science foundation.

²Corresponding author. *E-mail address:* kaspar.rufibach@stanford.edu (K. Rufibach).

Note also that in this setting the compact interval $[0, 1]$ takes the place of the positive half line.

Marshall's lemma is a powerful result although its proof is rather simple (see Marshall (1970) and Robertson, Wright and Dykstra (1988), p. 329). It implies, in the sense of the supremum norm, that the least concave majorant or greatest convex minorant is a better estimator of the corresponding shape restricted functions than the empirical distribution function or the cumulative sum diagram respectively. In several contexts, this statement has been made sharper. In monotone density estimation, Kiefer and Wolfowitz (1976) used Marshall's lemma as a key argument to show that $(\log n)^{-1}n^{2/3}\|\widehat{G}_n - \mathbb{G}_n\|_{[0, \alpha_1]}$ converges to zero with probability equal to 1, where $\alpha_1 = \inf\{x : G_0(x) = 1\}$ is assumed to be finite, $\beta(G_0) = \inf_{0 < x < \alpha_1} |g'_0(x)|/g_0^2(x) > 0$, $\gamma(G_0) = \sup_{0 < x < \alpha_1} |g'_0(x)|/\inf_{0 < x < \alpha_1} g_0^2(x) < \infty$ and generally $\|f\|_I := \sup_{x \in I} |f(x)|$ for any function $f : I \rightarrow \mathbb{R}$ on any set $I \subset \mathbb{R}$. Here and throughout, G_0 is the distribution function corresponding to the true density g_0 .

Their proof implies also that $(\log n)^{-2/3}n^{2/3}\|\widehat{G}_n - \mathbb{G}_n\|_{[0, \alpha_1]}$ is almost surely bounded, although this was not explicitly stated by the authors. In their approach, Kiefer and Wolfowitz (1976) introduced an interpolated version of $\mathbb{G}_n$, $\mathbb{L}_n$ say, at deterministic points $a_j = a_{j,n}$ chosen such that the inter-distances tend uniformly to zero with a rate of order $(\log n/n)^{1/3}$ as $n \rightarrow \infty$. The piecewise linear function $\mathbb{L}_n$ turned out to be concave with probability tending to 1 (see their Lemma 4), and hence Marshall's lemma could be applied when the concavity event occurs yielding the inequality $\|\widehat{G}_n - \mathbb{G}_n\|_{[0, \alpha_1]} \leq 2\|\mathbb{G}_n - \mathbb{L}_n\|_{[0, \alpha_1]}$. The problem was then reduced to bounding the error $\|\mathbb{G}_n - \mathbb{L}_n\|_{[0, \alpha_1]}$, and this could be achieved by bounding $\|\mathbb{G}_n - \mathbb{L}_n - (G_0 - L_n)\|_{[0, \alpha_1]}$ and $\|G_0 - L_n\|_{[0, \alpha_1]}$, where L_n is the true counterpart of $\mathbb{L}_n$; i.e., the piecewise linear interpolant of G_0 at the same deterministic points a_j 's.

Other authors have adapted the approach of Kiefer and Wolfowitz (1976). In doing so, Marshall's lemma or some similar version of it was used. In monotone estimation, see e.g. Theorem 1 of Wang (1986) for the asymptotic equivalence of the empirical hazard rate and its greatest convex minorant in the problem of estimating a distribution function with an increasing hazard rate. In the context of isotonic regression, we refer to the recent work of Pal and Woodroffe (2006) who obtained a result analogous to that of Kiefer and Wolfowitz (1976). In their proof, they also use Marshall's lemma and consider linear interpolation of the cumulative sum diagram and its true counterpart (see their Proposition 2 and Lemma 2).

1.2 Convex estimation via Least Squares and the first Marshall inequality

In the smoother problem of estimating a convex decreasing density g_0 on $[0, \infty)$, the recent result of Balabdaoui and Wellner (2006) obtained for the Least Squares estimator (LSE) is analogous to the theorem of Kiefer–Wolfowitz in monotone density estimation. The LSE was defined and studied by Groeneboom, Jongbloed, and Wellner (2001) and its asymptotic distribution was derived at an interior point. Here and in the next section, we will use the following notation: $\hat{g}_n$ is the LSE based on a sample $X_1, \dots, X_n$ from g_0 , $\hat{G}_n(t) = \int_0^t \hat{g}_n(s) ds$, $\hat{H}_n(t) = \int_0^t \hat{G}_n(s) ds$, and $\mathbb{H}_n(t) = \int_0^t \mathbb{G}_n(s) ds$, where $\mathbb{G}_n$ is the empirical distribution function of $X_1, \dots, X_n$. By definition, $\hat{g}_n$ is the unique minimizer of the functional

$$\frac{1}{2} \int_0^\infty g^2(t) dt - \int_0^\infty g(t) d\mathbb{G}_n(t)$$

over the class of all convex functions on $[0, \infty)$. By Lemma 2.1 and Corollary 2.1 of Groeneboom, Jongbloed, and Wellner (2001), we know that $\hat{g}_n$ is piecewise linear and is a genuine density. The knots of $\hat{g}_n$, which are the jump locations of $\hat{g}'_n$, satisfy the following property:

$$a \text{ is a knot} \Rightarrow \hat{H}_n(a) = \mathbb{H}_n(a) \text{ and } \hat{H}'_n(a) = \mathbb{H}'_n(a). \quad (1.2)$$

This property follows immediately from the characterization of the LSE:

$$\hat{H}_n \geq \mathbb{H}_n \text{ and } \hat{H}_n(t) = \mathbb{H}_n(t) \text{ if } t \text{ is a knot of } \hat{g}_n, \quad (1.3)$$

see Lemma 2.2 of Groeneboom, Jongbloed, and Wellner (2001). Note that $\hat{G}_n(a) = \mathbb{G}_n(a)$ is an equivalent form of the second equality given in (1.2).

The work of Balabdaoui and Wellner (2006) appears to be the first attempt to deal with global asymptotic rates in adaptive convex density estimation. Their Theorem 2.2 implies that

$$\|\hat{G}_n - \mathbb{G}_n\|_{[0, \alpha_2]} = O((n^{-1} \log n)^{3/5}) \quad (1.4)$$

almost surely under specified assumptions on the restriction of g_0, g'_0 and g''_0 on the compact $[0, \alpha_2]$ where α_2 is any finite positive number such that the following conditions hold: $\beta_2(G_0) = \inf_{0 < t < \alpha_2} g''_0(t)/g_0^3(t) > 0$ and $\gamma_2(G_0) = \sup_{0 < t < \alpha_2} g''_0(t)/\inf_{0 < t < \alpha_2} g_0^3(t) < \infty$. The numbers $\beta_2(G_0)$ and $\gamma_2(G_0)$ appear to be generalized versions of $\beta(G_0)$ and $\gamma(G_0)$ defined by Kiefer and Wolfowitz (1976) in the monotone problem, and which we recalled in our Subsection 1.1.

Note that, as for $\beta(G_0)$ and $\gamma(G_0)$, $\beta_2(G_0)$ and $\gamma_2(G_0)$ are both invariant under scale and location changes of G_0 .

To show the convergence in (1.4), the following new Marshall inequality proved by Dümbgen, Rufibach and Wellner (2006) in convex estimation was a key result:

Theorem 1.1. *If G is a function on $[0, \infty)$ such that G' is convex, then*

$$\|\widehat{G}_n - G\|_\infty \leq 2\|\mathbb{G}_n - G\|_\infty.$$

The factor 2 is shown to be sharp (see Dümbgen, Rufibach and Wellner (2006)). Theorem 1.1 follows easily from an intermediate inequality where the supremum is taken over $[a, b]$, and a and b are two consecutive knots of $\widehat{g}_n$. Here, we also start with proving our (second) Marshall inequality on a knot interval $[a, b]$, and we refer to the proof of Theorem 1.2 for more details on how the transition from knot intervals to the positive half line is achieved.

Beside the characterization (1.3), the proof of Theorem 1.1 relies substantially on convexity of the function G' by introducing a suitable auxiliary function. At the integral level, a function G in Theorem 1.1 is replaced by any of its primitives; i.e, $H = \int G$. The problem becomes more complicated as convexity is now a property of H'' , and one realizes quickly that direct use of this fact is no longer possible, requiring the development of a new approach. The class to which the function H belongs admits some specific properties. These properties are presented in detail and subsequently exploited to establish important intermediate results that will lead us to the proof of the main theorem.

1.3 Completing the proof for a Kiefer–Wolfowitz theorem in convex estimation

Choosing to extend the Kiefer–Wolfowitz theorem for the LSE instead of the MLE in the convex estimation problem was mainly due to the lack of an analogue of Theorem 1.1 for the MLE. On the other hand, note that the MLE in monotone estimation is actually equal to the LSE (see e.g. Robertson, Wright and Dykstra (1988)). Hence, from this point of view, it might be actually more natural to study the LSE in the convex problem.

To be able to use the idea of Kiefer and Wolfowitz (1976) in the convex problem, linear interpolants $\mathbb{L}_n$ and L_n of $\mathbb{G}_n$ and G_0 respectively had to be replaced by $\mathbb{S}_n$ and S_n , the cubic spline interpolants of $\mathbb{H}_n$ and $H_0(t) = \int_0^t G_0(s)ds = \int_0^t \int_0^s g_0(u)duds$. Balabdaoui and Wellner (2006) showed that $\mathbb{S}_n''$ is convex with increasing probability. This turned out to be

the most technical part in the proof. The knots of the cubic splines are chosen such that their inter-distances tend uniformly to zero at a rate $(\log n/n)^{1/5}$. Following the same idea of Kiefer and Wolfowitz (1976), Theorem 1.1 together with the triangle inequality were used to reduce the problem to bounding $\|\widehat{G}_n - S'_n - (G_0 - S'_n)\|_{[0, \alpha_2]}$ and $\|G_0 - S'_n\|_{[0, \alpha_2]}$. Balabdaoui and Wellner (2006) were then able to prove the (first) Kiefer–Wolfowitz theorem in convex density estimation; i.e., the result given in (1.4).

It is natural to ask whether an “integrated” version of (1.4) exists. It turns out that a scheme of proof very similar to the one described above for the first Kiefer–Wolfowitz result could be followed to show that

$$\|\widehat{H}_n - \mathbb{H}_n\|_{[0, \alpha_2]} = O((n^{-1} \log n)^{4/5}) \quad (1.5)$$

almost surely. Hence, under the same conditions on g_0 and its derivatives on $[0, \alpha_2]$, the argument can be constructed by approximating the empirical process $\mathbb{H}_n$ and its true counterpart H_0 via the random and deterministic cubic splines $\mathbb{S}_n$ and S_n introduced above. However, one crucial part is missing to complete the proof of this (second) Kiefer–Wolfowitz theorem: An analogue of Theorem 1.1 at the “integrated” level. In the next subsection, we present this analogue which is the main result of the paper. This completes the proof of (1.5) which was conjectured by Balabdaoui and Wellner (2006).

1.4 The second Marshall inequality in convex estimation

Our main result is the following theorem. The relevant results needed in the proof are deferred to Section 2.

Theorem 1.2. *For any $n \in \mathbb{N} \setminus \{0\}$ and a function H such that H'' is convex on $[0, \infty)$ we have*

$$\|\widehat{H}_n - H\|_\infty \leq \|\mathbb{H}_n - H\|_\infty.$$

Proof. It follows from the inequality $\widehat{H}_n \geq \mathbb{H}_n$ on $[0, \infty)$ that for any knot interval $[a, b]$ (see the characterization in (1.3))

$$\begin{aligned} - \inf_{t \in [a, b]} \{\widehat{H}_n(t) - H(t)\} &\leq - \inf_{t \in [a, b]} \{\mathbb{H}_n(t) - H(t)\} \\ &\leq \|\mathbb{H}_n - H\|_{[a, b]}. \end{aligned} \quad (1.6)$$

On the other hand, we have by Proposition 2.4

$$\begin{aligned} \sup_{t \in [a, b]} \{\widehat{H}_n(t) - H(t)\} &\leq \max\{|\mathbb{H}_n(a) - H(a)|, |\mathbb{H}_n(b) - H(b)|\} \\ &\leq \|\mathbb{H}_n - H\|_{[a, b]}. \end{aligned} \quad (1.7)$$

Combining (1.6) and (1.7) yields

$$\|\widehat{H}_n - H\|_{[a, b]} \leq \|\mathbb{H}_n - H\|_{[a, b]}. \quad (1.8)$$

The inequality in (1.8) can be directly compared with the one obtained in Theorem 1 of Dümbgen, Rufibach and Wellner (2006). It is its “integrated” version.

Let $0 = a_0 < \dots < a_m$ denote the sequence of jump points of $\widehat{g}'_n$ and define $a_{m+1} = \infty$. Note that for all $t \geq a_m$ we have $\widehat{H}_n(t) = \mathbb{H}_n(t)$. Indeed, we know by Corollary 2.1 (ii) and (iii) in Groeneboom, Jongbloed, and Wellner (2001) that $a_m > X_{(n)}$ and $\widehat{G}_n(t) = 1$ for all $t \geq a_m$, where $X_{(n)}$ denotes the n -th order statistic. On the other hand, we have for $t \geq a_m$

$$\begin{aligned} \widehat{H}_n(t) = \widehat{H}_n(a_m) + \int_{a_m}^t \widehat{G}_n(u) du &= \mathbb{H}_n(a_m) + \int_{a_m}^t du \\ &= \mathbb{H}_n(a_m) + \int_{a_m}^t \mathbb{G}_n(u) du = \mathbb{H}_n(t), \end{aligned}$$

since $\mathbb{G}_n(u) = 1$ for $u \geq a_m$. Hence, we can write

$$\begin{aligned} \|\widehat{H}_n - H\|_\infty &= \max_{0 \leq j \leq m} \|\widehat{H}_n - H\|_{[a_j, a_{j+1}]} \leq \max_{0 \leq j \leq m} \|\mathbb{H}_n - H\|_{[a_j, a_{j+1}]} \\ &= \|\mathbb{H}_n - H\|_\infty. \end{aligned} \quad \square$$

In a sense, this integrated version corresponds naturally to the original Marshall’s lemma in monotone estimation. An analogy between the monotone and convex problems is drawn in Table 1. From the characterization of the LSE $\widehat{g}_n$ given in (1.3), it follows that $\widehat{H}_n$ is an outer envelope for the integrated process $\mathbb{H}_n$, as the least concave majorant is for the empirical distribution function: They both share the property of staying above the empirical function $\mathbb{G}_n$ or $\mathbb{H}_n$ and touching it at the points where the adaptive estimator jumps or changes slope. However, it is important to note that a completely analogue geometric interpretation for $\widehat{H}_n$ is not available. Unlike for the class of concave functions, it is not possible to define a pointwise ordering on the set of functions whose second derivative is convex. Actually, the infimum of two such functions does not even have to be continuously differentiable.

On the other hand, we would like to note that the functions G in the original Marshall's lemma are assumed to be concave and hence it is clear that they belong to the same class as the true distribution function G_0 . In convex estimation, the functions G considered in the (first) Marshall inequality of Dümbgen, Rufibach and Wellner (2006) satisfy G' is convex, but G' is not assumed to be necessarily monotone nonincreasing (like the true density g_0). This also applies in our second inequality where we consider functions H such that H'' is convex without any further monotonicity constraints on H'' (see also Table 1). This might suggest that the Marshall inequalities in the convex problem hold for a wider class of functions. However, this is not exactly the case. In general, nonincreasing densities, and nonincreasing and convex densities are special cases of k -monotone densities g ; i.e., g is nonincreasing if $k = 1$, and $(-1)^k g^{(k-2)}$ is convex if $k \geq 2$ (see e.g. Balabdaoui (2004), Gneiting (1999) and the references therein). Hence, it is not surprising that the “monotonicity part” was not needed in the convex problem; i.e., $k = 2$. This also suggests that if a generalized Marshall's Lemma could be proved for the LSE estimator in the case $k > 2$ (see Balabdaoui (2004)), such a lemma will probably involve functions H such that $H^{(2k-2)}$ is convex if k is even and concave if k is odd.

Connections to m -convexity. In the literature on generalized convexity and shape constrained approximation, a function f defined on some real interval I and admitting a convex $(m - 2)$ -th derivative on I is called m -convex. See e.g. Bullen (1971), Roberts and Varberg (1973) and Zwick (1986). Using this terminology, the functions G and H considered in the first and second Marshall inequalities in the convex problem are 3- and 4-convex on $[0, \infty)$, respectively.

Marshall inequality in monotone estimation revisited. Our approach can be applied in monotone estimation, offering a third way of proving the original Marshall's lemma (see Marshall (1970) and Robertson, Wright and Dykstra (1988), p. 329 for the two already available proofs). In fact, the least concave majorant $\hat{G}_n$ is the linear interpolant of $\mathbb{G}_n$ on $[a, b]$ where a and b are successive jump points of the Grenander estimator (see also Table 1). A property similar to the one shown in Proposition 2.3 could be easily proved for an arbitrary concave function G . The Hermite operator Q should then be replaced by the linear operator L ; i.e., $L(f, t) = f(a) + (t - a)(f(b) - f(a))/(b - a)$. This yields $\sup_{t \in [a, b]} [\hat{G}_n(t) - G(t)] \leq \max\{\mathbb{G}_n(a) - G(a), \mathbb{G}_n(b) - G(b)\}$. Marshall's lemma follows using $\hat{G}_n \geq \mathbb{G}_n$ and the fact that

Table 1: Analogy between monotone and convex density estimation.

Monotone	Convex
$\ \widehat{G}_n - G\ _\infty \leq \ \mathbb{G}_n - G\ _\infty$	$\ \widehat{H}_n - H\ \leq \ \mathbb{H}_n - H\ _\infty$
True density g_0 is monotone $\searrow$	True density g_0 is convex $\searrow$
G' is monotone $\searrow$ (G is concave)	H'' is convex (H is 4-convex)
$\mathbb{G}_n$: Empirical distribution function	$\mathbb{H}_n$: Integral of the empirical distribution function
$\widehat{G}_n$: Least concave majorant of $\mathbb{G}_n$	$\widehat{H}_n$: Outer envelope of $\mathbb{H}_n$
$\widehat{G}_n$: Concave spline of degree 1 (piecewise linear and concave)	$\widehat{H}_n$: Spline of degree 3 with convex second derivative ($\widehat{H}_n$ is 4-convex)
$\widehat{G}_n$ touches $\mathbb{G}_n$ at the jump points of the Grenander estimator	$\widehat{H}_n$ touches $\mathbb{H}_n$ at the break points of the LSE
$\widehat{G}_n$: The linear interpolant of $\mathbb{G}_n$ on a knot interval	$\widehat{H}_n$: The cubic Hermite polynomial interpolant of $\mathbb{H}_n$ on a knot interval

$$\widehat{G}_n(t) = \mathbb{G}_n(t) \text{ for } t \geq X_{(n)}.$$

Convex regression. In convex regression where the true regression curve r_0 is compactly supported, e.g. on $[0, 1]$, Theorem 1.2 has a natural analogue if the estimator, $\widehat{r}_n$ say, is the minimizer of the functional

$$\frac{1}{2} \int_0^1 r^2(t) dt - \frac{1}{n} \sum_{i=1}^n r(x_{n,i}) Y_{n,i}$$

over the class of convex functions r on $[0, 1]$. Above, $0 < x_{n,1} < \dots < x_{n,n} < 1$ are the design points and $Y_{n,i} = r_0(x_{n,i}) + \epsilon_{n,i}$ are the observations where $\{\epsilon_{n,i}, i = 1, \dots, n\}$ is a triangular array of i.i.d. random variables.

If $\mathbb{R}_n(t) = n^{-1} \sum_{i=1}^n Y_{n,i}(t - x_{n,i})_+$; that is, $\mathbb{R}_n$ is the integrated cumulative sum diagram, and $\widehat{R}_n(t) = \int_0^t \int_0^s \widehat{r}_n(u) du ds$, then for any 4-convex function R

$$\|\widehat{R}_n - R\|_\infty \leq \|\mathbb{R}_n - R\|_\infty. \quad (1.9)$$

The functions $\mathbb{R}_n$ and $\widehat{R}_n$ play the same roles as $\mathbb{H}_n$ and $\widehat{H}_n$ in convex density estimation, and the proof of the Marshall inequality in (1.9) is exactly similar to that of Theorem 1.2. However, we should note that this result does not hold *a priori* for the “classical” regression estimator, $\tilde{r}_n$ say, defined as the minimizer of

$$\frac{1}{2} \sum_{i=1}^n (Y_{n,i} - r(x_{n,i}))^2$$

over the class of convex functions r which are piecewise linear and change slopes only at the design points $x_{n,i}$ (see Groeneboom, Jongbloed, and Wellner (2001), pages 1666–1667). The characterization is different and involves a function that is less smooth than a second integral of $\tilde{r}_n$. This is mainly due to the completely discrete nature of the criterion function to be minimized. Hence, unlike in monotone regression, where it is easy to check that both the minimization problems yield the same solution (the slopes of the Greatest Convex Minorant of the cumulative sum diagram), the estimators $\tilde{r}_n$ and $\widehat{r}_n$ are in principal different in convex estimation.

2 Intermediate results and proofs

In the following, we present some preparatory results and sketch the line of arguments used to show Theorem 1.2.

2.1 The results

We start with the following definition.

Definition 2.1. *For a function f defined on a real interval I , the m -th divided difference of the function f at $(m+1)$ distinct points $x_0 < \dots < x_m \in I$, denoted by $[x_0, \dots, x_m]f$, is defined as*

$$[x_0]f = f(x_0), \quad [x_0, \dots, x_m]f = \frac{[x_1, \dots, x_m]f - [x_0, \dots, x_{m-1}]f}{x_m - x_0} \quad \text{for } m \geq 1.$$

See e.g. Roberts and Varberg (1973), p. 237.

The definition of divided differences is recursive. When $m = 0$, the 0-th divided difference is simply the value of the function at a point. When $m = 1$, it reduces to the slope of f between two points x_0 and x_1 . For $m \geq 2$, it takes the form of differences of slopes. Now, note that

monotonicity of the function f on I is equivalent to having the divided difference $[x_0, x_1]f$ nonnegative for all possible choices of $x_0 \neq x_1 \in I$. Also, it is easy to check that f is convex on I if and only if $[x_0, x_1, x_2]f \geq 0$ for all distinct points $x_0, x_1, x_2 \in I$. The following proposition, due to Bullen (1971) (see Corollary 8), gives a natural extension to this equivalence in the case of functions whose second derivative is convex. See also Lemma 3.1 in Kopotun and Shadrin (2003).

Proposition 2.1. *Let f be a twice differentiable function on a real interval I . Then, the second derivative f'' is convex if and only if the 4-th divided difference $[x_0, x_1, x_2, x_3, x_4]f$ is nonnegative for all choices of five distinct points $x_0, \dots, x_4$ in I .*

We need the above proposition to prove a general property of functions with convex second derivative (see Proposition 2.3). This property will turn out to be very useful for bounding $\sup_{t \in [a, b]} [\hat{H}_n(t) - H(t)]$, where a and b are any consecutive knots of the LSE $\hat{g}_n$ and H'' is convex on $[a, b]$. Now, for such knots a and b it follows from the characterization of $\hat{g}_n$, that $\hat{H}_n$ is the cubic Hermite polynomial interpolant of the empirical function $\mathbb{H}_n$ on $[a, b]$; that is, the unique cubic polynomial matching $\mathbb{H}_n$ (the interpolated function here) and its first derivative at the end points a and b . In the following, we give the known formula for the cubic Hermite polynomial interpolating some function f at two distinct points a and b (see e.g. DeVore and Lorentz (1993)).

Proposition 2.2. *Let f be a function defined on a real interval $[a, b]$ such that $f'(a)$ and $f'(b)$ exist. For all $t \in [a, b]$, the cubic Hermite polynomial $Q(f, t)$ is given by the formula*

$$\begin{aligned} Q(f, t) = & f(a)\phi\left(\frac{t-a}{b-a}\right) + f'(a)(b-a)\psi\left(\frac{t-a}{b-a}\right) \\ & + f(b)\phi\left(\frac{b-t}{b-a}\right) - f'(b)(b-a)\psi\left(\frac{b-t}{b-a}\right), \end{aligned} \quad (2.10)$$

where $\phi(x) = 1 + x^2(2x - 3)$ and $\psi(x) = x(1 - x)^2$, for all $x \in [0, 1]$.

The above formula gives a convenient representation of $\hat{H}_n$ as a function of $\mathbb{H}_n$ on any interval $[a, b]$ such that a and b are successive knots of the LSE $\hat{g}_n$: $\hat{H}_n$ is the cubic Hermite polynomial interpolant of $\mathbb{H}_n$ on $[a, b]$; that is

$$\hat{H}_n(t) = Q(\mathbb{H}_n, t) \quad \text{for } t \in [a, b].$$

This follows easily from the properties of the LSE $\hat{g}_n$ given in (1.2).

Now we are ready to state a general property satisfied by any function with convex second derivative on some finite real interval.

Proposition 2.3. *For any function f on $[a, b]$ such that f'' is convex and f is differentiable at a and b , we have for any $t \in [a, b]$*

$$Q(f, t) - f(t) \leq 0$$

where $Q(f, t)$ is the cubic Hermite polynomial interpolant defined in (2.10).

Proposition 2.4. *Let a and b be two consecutive jump points of $\hat{g}'_n$. Then for any function H on $[a, b]$ such that H'' is convex we have*

$$\sup_{t \in [a, b]} [\hat{H}_n(t) - H(t)] \leq \max \{ \mathbb{H}_n(a) - H(a), \mathbb{H}_n(b) - H(b) \}.$$

2.2 Proofs

Proof of Proposition 2.3. For an arbitrary function g on $[a, b]$ such that g'' is convex, the maximal number of zeros of g in $[a, b]$ is four (counting multiplicities). Indeed, g'' is convex on $[a, b]$, and hence it admits at most 2 zeros in that interval. By application of Rolle's theorem to g' and g , the claim follows.

Now, let $g = f - Q(f, \cdot)$, where f'' is convex on $[a, b]$ and f is differentiable at a and b . Since $Q(f, \cdot)$ is a polynomial of degree 3, g'' is convex. On the other hand, by definition of $Q(f, \cdot)$ we have $g(a) = g(b) = g'(a) = g'(b) = 0$. The function g has then 2 zeros, at a and b . Therefore, we must have either $g \geq 0$ or $g \leq 0$ on $[a, b]$. To find out the sign, we calculate g at the midpoint $(a + b)/2$. Without loss of generality, we take $a = 0$ and $b = 1$. By Proposition 2.1, we have $[0, \delta, 1/2, 1 - \delta, 1]f \geq 0$ for any $\delta \in (0, 1/2)$. Since f is differentiable at 0 and 1, we obtain by letting δ tend to 0

$$0 \leq [0, 0, 1/2, 1, 1]f = 2(8f(1/2) - 4(f(0) + f(1)) + f'(1) - f'(0)).$$

Hence,

$$f(1/2) \geq \frac{1}{2}(f(0) + f(1)) + \frac{1}{8}(f'(0) - f'(1)).$$

Now, note that the right side of the previous inequality is exactly equal to $Q(f, 1/2)$, since $\phi(1/2) = 1/2$ and $\psi(1/2) = 1/8$ (ϕ and ψ are cubic polynomials defined in Proposition 2.2). We conclude that $g(1/2) \geq 0$. Hence, $g(t) \geq 0$ on $[a, b]$ or equivalently $Q(f, t) \leq f(t)$, $t \in [a, b]$.

□

Proof of Proposition 2.4. We assume first that $H'(a) = \mathbb{H}'_n(a)$ and $H'(b) = \mathbb{H}'_n(b)$. From the characterization of the LSE $\widehat{g}_n$, it follows that $\widehat{H}_n$ is the cubic Hermite polynomial of $\mathbb{H}_n$ on $[a, b]$. By Proposition 2.2, we can write

$$\begin{aligned}
\widehat{H}_n(t) - H(t) &= Q(\mathbb{H}_n, t) - Q(H, t) + Q(H, t) - H(t) \\
&\leq Q(\mathbb{H}_n, t) - Q(H, t), \quad \text{by Proposition 2.3} \\
&= (\mathbb{H}_n - H)(a)\phi\left(\frac{t-a}{b-a}\right) + (b-a)(\mathbb{H}'_n - H')(a)\psi\left(\frac{t-a}{b-a}\right) \\
&\quad + (\mathbb{H}_n - H)(b)\phi\left(\frac{b-t}{b-a}\right) - (b-a)(\mathbb{H}'_n - H')(b)\psi\left(\frac{b-t}{b-a}\right) \\
&= (\mathbb{H}_n - H)(a)\phi\left(\frac{t-a}{b-a}\right) + (\mathbb{H}_n - H)(b)\phi\left(\frac{b-t}{b-a}\right) \\
&\leq \max\{(\mathbb{H}_n - H)(a), (\mathbb{H}_n - H)(b)\}. \tag{2.11}
\end{aligned}$$

The last inequality follows easily from the fact that $\phi(1-u) = 1 - \phi(u)$ and $0 \leq \phi(u) \leq 1$ for all $u \in [0, 1]$.

Now, we consider the general case where H' does not necessarily match $\mathbb{H}'_n$ at a and b . Let $r = \operatorname{argmax}_{t \in [a, b]} \{\widehat{H}_n(t) - H(t)\}$, and consider the polynomial $P(t) = \alpha_3(t-a)^3 + \alpha_2(t-a)^2 + \alpha_1(t-a)$ satisfying the conditions

$$P'(a) = H'(a) - \widehat{H}'_n(a) \tag{2.12}$$

$$P'(b) = H'(b) - \widehat{H}'_n(b). \tag{2.13}$$

The function $H - P$ admits a second derivative that is convex on $[a, b]$ since P is a polynomial of degree 3. Furthermore, by (2.12) and (2.13) together with the fact that $\widehat{H}'_n(a) = \mathbb{H}'_n(a)$ and $\widehat{H}'_n(b) = \mathbb{H}'_n(b)$, we have $(H - P)'(a) = \mathbb{H}'_n(a)$ and $(H - P)'(b) = \mathbb{H}'_n(b)$. Hence, we are now in a position to use the result obtained in (2.11) to conclude that

$$\begin{aligned}
\widehat{H}_n(t) - (H(t) - P(t)) &\leq \\
&\max\{\mathbb{H}_n(a) - (H(a) - P(a)), \mathbb{H}_n(b) - (H(b) - P(b))\} \tag{2.14}
\end{aligned}$$

for all $t \in [a, b]$. Suppose without loss of generality that $\max\{\mathbb{H}_n(a) - H(a) + P(a), \mathbb{H}_n(b) - H(b) + P(b)\} = \mathbb{H}_n(a) - H(a) + P(a)$. We impose now our third condition on P ; i.e., $P(r) = P(a)$. When $r \neq a$, such a polynomial is unique and it can be easily checked that

$$\begin{aligned}
\alpha_1 &= H'(a) - \widehat{H}'_n(a), \quad \alpha_2 = -\frac{3\alpha_1(b-a)^2 + \Delta(r-a)^2}{D}, \quad \text{and} \\
\alpha_3 &= \frac{\Delta(r-a) + 2\alpha_1(b-a)}{D}
\end{aligned}$$

where

$$\begin{aligned}\Delta &= H'(b) - \widehat{H}'_n(b) - (H'(a) - \widehat{H}'_n(a)), \text{ and} \\ D &= (b-a)(r-a)(3(b-a) - 2(r-a)) \neq 0, \text{ since } 3(b-a) > 2(r-a).\end{aligned}$$

Now, for $t = r$ the inequality in (2.14) yields

$$\widehat{H}_n(r) - H(r) + P(r) \leq \mathbb{H}_n(a) - H(a) + P(a)$$

or equivalently $\widehat{H}_n(r) - H(r) \leq \mathbb{H}_n(a) - H(a)$. Note that when $r = a$, the third condition is not at all needed, and one could take any of the solutions P satisfying the derivative conditions in (2.12) and (2.13). The case $\max\{\mathbb{H}_n(a) - H(a) + P(a), \mathbb{H}_n(b) - H(b) + P(b)\} = \mathbb{H}_n(b) - H(b) + P(b)$ can be handled similarly, where the third condition should be changed into $P(r) = P(b)$. $\square$

Acknowledgments. We thank a referee for helpful remarks.

References

- Balabdaoui, F. (2004), *Estimation of a k -monotone density: A new asymptotic distribution theory* (Unpublished Ph.D. dissertation, University of Washington).
- Balabdaoui, F. and Wellner, J. A. (2006) A Kiefer–Wolfowitz theorem for convex densities. *Submitted to IMS Lecture Notes–Monograph Series*. Preprint available at <http://www.stochastik.math.uni-goettingen.de/preprints/KWCvx28.pdf>.
- Bullen, P. S. (1971), A criterion for n -convexity, *Pacific J. Math.* **36**, 81–98.
- DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation* (Springer Verlag, New York).
- Dümbgen, L., Rufibach, K. and Wellner, J. A. (2006). Marshall’s lemma for convex density estimation. *Submitted to IMS Lecture Notes–Monograph Series*. Preprint available at <http://arxiv.org/abs/math.ST/0609277>.
- Gneiting, T. (1999), Radial positive definite functions generated by Euclid’s hat, *J. Multivariate Analysis* **69**, 88–119.
- Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001), Estimation of convex functions: characterizations and asymptotic theory, *Ann. Statist.* **29**, 1653–1698.

- Kiefer, J. and Wolfowitz, J. (1976), Asymptotically minimax estimation of concave and convex distribution functions, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **34**, 73–85.
- Kopotun, K. and Shadrin A. (2003), On k -monotone approximation by free knot splines, *Siam. J. Math. Anal.* **34**, 901–924.
- Marshall, A. W. (1970). Discussion on Barlow and Van Zwet’s paper, in M. L. Puri, ed. *Nonparametric techniques in statistical inference* (Cambridge University Press, London) 174–176.
- Pal, J. K. and Woodroffe, M. (2006), On the distance between cumulative sum diagram and its greatest convex minorant for unequally design points, *Scan. J. Statist.* **33**, 279–291.
- Roberston, T., Wright, F. and Dykstra, R. L. (1988), *Order Restricted Statistical Inference* (Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Ltd., Chichester).
- Roberts, W. A. and Varberg, D. A. (1973), *Convex Functions* (Academic Press, New York).
- Wang, J. L. (1986), Asymptotically minimax estimation for distributions with increasing failure rate, *Ann. Statist.* **14**, 1113–1131.
- Zwick, D. (1986), Existence of best n -convex approximations, *Proc. Amer. Math. Soc.* **97**, 267–273.