IMS Lecture Notes–Monograph Series
Asymptotics: Particles, Processes and Inverse Problems
Vol. 55 (2007) 101–107

DOI: 10.1214/074921707000000292

Marshall's lemma for convex density estimation

Lutz Dümbgen[1], Kaspar Rufibach[1] and Jon A. Wellner[2]

University of Bern and University of Washington

Abstract: Marshall's [*Nonparametric Techniques in Statistical Inference* (1970) 174–176] lemma is an analytical result which implies $\sqrt{n}$–consistency of the distribution function corresponding to the Grenander [*Skand. Aktuarietidskr.* **39** (1956) 125–153] estimator of a non-decreasing probability density. The present paper derives analogous results for the setting of convex densities on $[0, \infty)$.

1. Introduction

Let $\mathbb{F}$ be the empirical distribution function of independent random variables X_1, $X_2, \ldots, X_n$ with distribution function F and density f on the halfline $[0, \infty)$. Various shape restrictions on f enable consistent nonparametric estimation of it without any tuning parameters (e.g. bandwidths for kernel estimators).

The oldest and most famous example is the Grenander estimator $\hat{f}$ of f under the assumption that f is non-increasing. Denoting the family of all such densities by $\mathcal{F}$, the Grenander estimator may be viewed as the maximum likelihood estimator,

$$\hat{f} = \operatorname{argmax}\Big\{ \int \log h \, d\mathbb{F} : h \in \mathcal{F} \Big\},$$

or as a least squares estimator,

$$\hat{f} = \operatorname{argmin}\Big\{ \int_0^\infty h(x)^2 dx - 2 \int h \, d\mathbb{F} : h \in \mathcal{F} \Big\};$$

cf. Robertson et al. [5]. Note that if $\mathbb{F}$ had a square-integrable density $\mathbb{F}'$, then the preceding argmin would be identical with the minimizer of $\int_0^\infty (h - \mathbb{F}')(x)^2 \, dx$ over all non-increasing probability densities h on $[0, \infty)$.

A nice property of $\hat{f}$ is that the corresponding distribution function $\hat{F}$,

$$\hat{F}(r) \ := \ \int_0^r \hat{f}(x) \, dx,$$

is automatically $\sqrt{n}$–consistent. More precisely, since $\hat{F}$ is the least concave majorant of $\mathbb{F}$, it follows from Marshall's [4] lemma that

$$\|\hat{F} - F\|_\infty \ \le \ \|\mathbb{F} - F\|_\infty.$$

A more refined asymptotic analysis of $\hat{F} - \mathbb{F}$ has been provided by Kiefer and Wolfowitz [3].

[1]Institute of Math. Statistics and Actuarial Science, University of Bern, Switzerland, e-mail: lutz.duembgen@stat.unibe.ch; kaspar.rufibach@stat.unibe.ch

[2]Dept. of Statistics, University of Washington, Seattle, USA, e-mail: jaw@stat.washington.edu

AMS 2000 subject classifications: 62G05, 62G07, 62G20.

Keywords and phrases: empirical distribution function, inequality, least squares, maximum likelihood, shape constraint, supremum norm.

2. Convex densities

Now we switch to the estimation of a convex probability density f on $[0,\infty)$. As pointed out by Groeneboom et al. [2], the nonparametric maximum likelihood estimator $\hat{f}_{ml}$ and the least squares estimator $\hat{f}_{ls}$ are both well-defined and unique, but they are not identical in general. Let $\mathcal{K}$ denote the convex cone of all convex and integrable functions g on $[0,\infty)$. (All functions within $\mathcal{K}$ are necessarily nonnegative and non-increasing.) Then

$$\begin{aligned}\hat{f}_{ml} &= \underset{h\in\mathcal{K}}{\operatorname{argmax}}\Big(\int \log h\, d\mathbb{F} - \int_0^\infty h(x)\,dx\Big),\\ \hat{f}_{ls} &= \underset{h\in\mathcal{K}}{\operatorname{argmin}}\Big(\int_0^\infty h(x)^2 dx - 2\int h\, d\mathbb{F}\Big).\end{aligned}$$

Both estimators have the following property:

Proposition 1. *Let $\hat{f}$ be either $\hat{f}_{ml}$ or $\hat{f}_{ls}$. Then $\hat{f}$ is piecewise linear with*

- *at most one knot in each of the intervals $(X_{(i)}, X_{(i+1)})$, $1 \le i < n$,*
- *no knot at any observation X_i, and*
- *precisely one knot within $(X_{(n)}, \infty)$.*

The estimators $\hat{f}_{ml}$, $\hat{f}_{ls}$ and their distribution functions $\hat{F}_{ml}$, $\hat{F}_{ls}$ are completely characterized by Proposition 1 and the next proposition.

Proposition 2. *Let Δ be any function on $[0,\infty)$ such that $\hat{f}_{ml} + t\Delta \in \mathcal{K}$ for some $t > 0$. Then*

$$\int \frac{\Delta}{\hat{f}_{ml}}\, d\mathbb{F} \;\le\; \int \Delta(x)\,dx.$$

Similarly, let Δ be any function on $[0,\infty)$ such that $\hat{f}_{ls} + t\Delta \in \mathcal{K}$ for some $t > 0$. Then

$$\int \Delta\, d\mathbb{F} \;\le\; \int \Delta\, d\hat{F}_{ls}.$$

In what follows we derive two inequalities relating $\hat{F} - F$ and $\mathbb{F} - F$, where $\hat{F}$ stands for $\hat{F}_{ml}$ or $\hat{F}_{ls}$:

Theorem 1.

$$\inf_{[0,\infty)} (\hat{F}_{ml} - F) \ge \frac{3}{2} \inf_{[0,\infty)} (\mathbb{F} - F) - \frac{1}{2} \sup_{[0,\infty)} (\mathbb{F} - F), \tag{1}$$

$$\big\|\hat{F}_{ls} - F\big\|_\infty \le 2\,\big\|\mathbb{F} - F\big\|_\infty. \tag{2}$$

Both results rely on the following lemma:

Lemma 1. *Let $F, \hat{F}$ be continuous functions on a compact interval $[a,b]$, and let $\mathbb{F}$ be a bounded, measurable function on $[a,b]$. Suppose that the following additional assumptions are satisfied:*

$$\hat{F}(a) = \mathbb{F}(a) \text{ and } \hat{F}(b) = \mathbb{F}(b), \tag{3}$$

$$\hat{F} \text{ has a linear derivative on } (a,b), \tag{4}$$

$$F \text{ has a convex derivative on } (a,b), \tag{5}$$

$$\int_r^b \hat{F}(y)\,dy \;\le\; \int_r^b \mathbb{F}(y)\,dy \quad \text{for all } r \in [a,b]. \tag{6}$$

Then

$$\sup_{[a,b]} (\hat{F} - F) \ \le\ \frac{3}{2} \sup_{[a,b]} (\mathbb{F} - F) - \frac{1}{2} (\mathbb{F} - F)(b).$$

If condition (6) is replaced with

$$\int_a^r \hat{F}(x)\, dx \ \ge\ \int_a^r \mathbb{F}(x)\, dx \quad \textit{for all } r \in [a,b], \tag{7}$$

then

$$\inf_{[a,b]} (\hat{F} - F) \ \ge\ \frac{3}{2} \inf_{[a,b]} (\mathbb{F} - F) - \frac{1}{2} (\mathbb{F} - F)(a).$$

The constants 3/2 and 1/2 are sharp. For let $[a,b] = [0,1]$ and define

$$F(x) := \begin{cases} x^2 - c & \text{for } x \ge \epsilon, \\ (x/\epsilon)(\epsilon^2 - c) & \text{for } x \le \epsilon, \end{cases}$$
$$\hat{F}(x) := 0,$$
$$\mathbb{F}(x) := 1\{0 < x < 1\}(x^2 - 1/3)$$

for some constant $c \ge 1$ and some small number $\epsilon \in (0, 1/2]$. One easily verifies conditions (3)–(6). Moreover,

$$\sup_{[0,1]} (\hat{F} - F) = c - \epsilon^2, \quad \sup_{[0,1]} (\mathbb{F} - F) = c - 1/3 \quad \text{and} \quad (\mathbb{F} - F)(1) = c - 1.$$

Hence the upper bound $(3/2) \sup(\mathbb{F} - F) - (1/2)(\mathbb{F} - F)(1)$ equals $\sup(\hat{F} - F) + \epsilon^2$ for any $c \ge 1$. Note the discontinuity of $\mathbb{F}$ at 0 and 1. However, by suitable approximation of $\mathbb{F}$ with continuous functions one can easily show that the constants remain optimal even under the additional constraint of $\mathbb{F}$ being continuous.

Proof of Lemma 1. We define $G := \hat{F} - F$ with derivative $g := G'$ on (a,b). It follows from (3) that

$$\max_{\{a,b\}} G \ =\ \max_{\{a,b\}} (\mathbb{F} - F) \ \le\ \frac{3}{2} \sup_{[a,b]} (\mathbb{F} - F) - \frac{1}{2} (\mathbb{F} - F)(b).$$

Therefore it suffices to consider the case that G attains its maximum at some point $r \in (a,b)$. In particular, $g(r) = 0$. We introduce an auxiliary linear function $\bar{g}$ on $[r,b]$ such that $\bar{g}(r) = 0$ and

$$\int_r^b \bar{g}(y)\, dy \ =\ \int_r^b g(y)\, dy \ =\ G(b) - G(r).$$

Note that g is concave on (a,b) by (4)–(5). Hence there exists a number $y_o \in (r,b)$ such that

$$g - \bar{g} \ \begin{cases} \ge 0 \text{ on } [r, y_o], \\ \le 0 \text{ on } [y_o, b). \end{cases}$$

This entails that

$$\int_r^y (g - \bar{g})(u)\, du \ =\ -\int_y^b (g - \bar{g})(u)\, du \ \ge\ 0 \quad \text{for any } y \in [r,b].$$

Consequently,

$$\begin{aligned} G(y) &= G(r) + \int_r^y g(u)\,du \\ &\geq G(r) + \int_r^y \bar{g}(u)\,du \\ &= G(r) + \frac{(y-r)^2}{(b-r)^2}[G(b) - G(r)], \end{aligned}$$

so that

$$\begin{aligned} \int_r^b G(y)\,dy &\geq (b-r)G(r) + \frac{G(b)-G(r)}{(b-r)^2}\int_r^b (y-r)^2\,dy \\ &= (b-r)\Big[\frac{2}{3}G(r) + \frac{1}{3}G(b)\Big] \\ &= (b-r)\Big[\frac{2}{3}G(r) + \frac{1}{3}(\mathbb{F}-F)(b)\Big]. \end{aligned}$$

On the other hand, by assumption (6),

$$\int_r^b G(y)\,dy \ \leq \ \int_r^b (\mathbb{F}-F)(y)\,dy \ \leq \ (b-r)\sup_{[a,b]}(\mathbb{F}-F).$$

This entails that

$$G(r) \ \leq \ \frac{3}{2}\sup_{[a,b]}(\mathbb{F}-F) - \frac{1}{2}(\mathbb{F}-F)(b).$$

If (6) is replaced with (7), then note first that

$$\min_{\{a,b\}} G \ = \ \min_{\{a,b\}}(\mathbb{F}-F) \ \geq \ \frac{3}{2}\min_{\{a,b\}}(\mathbb{F}-F) - \frac{1}{2}(\mathbb{F}-F)(a).$$

Therefore it suffices to consider the case that G attains its minimum at some point $r \in (a,b)$. Now we consider a linear function $\bar{g}$ on $[a,r]$ such that $\bar{g}(r) = 0$ and

$$\int_a^r \bar{g}(x)\,dx \ = \ \int_a^r g(x)\,dx \ = \ G(r) - G(a).$$

Here concavity of g on (a,b) entails that

$$\int_a^x (g-\bar{g})(u)\,du \ = \ -\int_x^r (g-\bar{g})(u)\,du \ \leq \ 0 \quad \text{for any } x \in [a,r],$$

so that

$$\begin{aligned} G(x) &= G(r) - \int_x^r g(u)\,du \\ &\leq G(r) - \int_x^r \bar{g}(u)\,du \\ &= G(r) - \frac{(r-x)^2}{(r-a)^2}[G(r) - G(a)]. \end{aligned}$$

Consequently,

$$\begin{aligned} \int_a^r G(x)\,dx &\leq (r-a)G(r) - \frac{G(r)-G(a)}{(r-a)^2}\int_a^r (r-x)^2\,dx \\ &= (r-a)\Big[\frac{2}{3}G(r) + \frac{1}{3}(\mathbb{F}-F)(a)\Big], \end{aligned}$$

whereas

$$\int_a^r G(x)\,dx \;\geq\; \int_a^r (\mathbb{F}-F)(x)\,dx \;\geq\; (r-a)\inf_{[a,b]}(\mathbb{F}-F),$$

by assumption (7). This leads to

$$G(r) \;\geq\; \frac{3}{2}\inf_{[a,b]}(\mathbb{F}-F) - \frac{1}{2}(\mathbb{F}-F)(a). \qquad \square$$

$\square$

Proof of Theorem 1. Let $0 =: t_0 < t_1 < \cdots < t_m$ be the knots of $\hat{f}$, including the origin. In what follows we derive conditions (3)–(5) and (6/7) of Lemma 1 for any interval $[a,b] = [t_k, t_{k+1}]$ with $0 \leq k < m$. For the reader's convenience we rely entirely on Proposition 2. In case of the least squares estimator, similar inequalities and arguments may be found in Groeneboom et al. [2].

Let $0 < \epsilon < \min_{1\leq i\leq m}(t_i - t_{i-1})/2$. For a fixed $k \in \{1,\ldots,m\}$ we define Δ_1 to be continuous and piecewise linear with knots $t_{k-1}-\epsilon$ (if $k>1$), t_{k-1}, t_k and $t_k+\epsilon$. Namely, let $\Delta_1(x) = 0$ for $x \notin (t_{k-1}-\epsilon, t_k+\epsilon)$ and

$$\Delta_1(x) \;:=\; \begin{Bmatrix} \hat{f}_{ml}(x) \text{ if } \hat{f} = \hat{f}_{ml} \\ 1 \quad \text{ if } \hat{f} = \hat{f}_{ls} \end{Bmatrix} \quad \text{for } x \in [t_{k-1}, t_k].$$

This function Δ_1 satisfies the requirements of Proposition 2. Letting $\epsilon \searrow 0$, the function $\Delta_1(x)$ converges pointwise to

$$\begin{cases} 1\{t_{k-1} \leq x \leq t_k\}\hat{f}_{ml}(x) \text{ if } \hat{f} = \hat{f}_{ml}, \\ \quad 1\{t_{k-1} \leq x \leq t_k\} \quad\;\; \text{if } \hat{f} = \hat{f}_{ls}, \end{cases}$$

and the latter proposition yields the inequality

$$\mathbb{F}(t_k) - \mathbb{F}(t_{k-1}) \;\leq\; \hat{F}(t_k) - \hat{F}(t_{k-1}).$$

Similarly let Δ_2 be continuous and piecewise linear with knots at t_{k-1}, $t_{k-1}+\epsilon$, $t_k-\epsilon$ and t_k. Precisely, let $\Delta_2(x) := 0$ for $x \notin (t_{k-1}, t_k)$ and

$$\Delta_2(x) \;:=\; \begin{Bmatrix} -\hat{f}_{ml}(x) \text{ if } \hat{f} = \hat{f}_{ml} \\ -1 \quad \text{ if } \hat{f} = \hat{f}_{ls} \end{Bmatrix} \quad \text{for } x \in [t_{k-1}+\epsilon, t_k-\epsilon].$$

The limit of $\Delta_2(x)$ as $\epsilon \searrow 0$ equals

$$\begin{cases} -1\{t_{k-1} < x < t_k\}\hat{f}_{ml}(x) \text{ if } \hat{f} = \hat{f}_{ml}, \\ \quad -1\{t_{k-1} < x < t_k\} \quad\;\; \text{if } \hat{f} = \hat{f}_{ls}, \end{cases}$$

and it follows from Proposition 2 that

$$\mathbb{F}(t_k) - \mathbb{F}(t_{k-1}) \;\geq\; \hat{F}(t_k) - \hat{F}(t_{k-1}).$$

This shows that $\mathbb{F}(t_k) - \mathbb{F}(t_{k-1}) = \hat{F}(t_k) - \hat{F}(t_{k-1})$ for $k = 1,\ldots,m$. Since $\hat{F}(0) = 0$, one can rewrite this as

$$\mathbb{F}(t_k) \;=\; \hat{F}(t_k) \quad \text{for } k = 0,1,\ldots,m. \tag{8}$$

Now we consider first the maximum likelihood estimator $\hat{f}_{ml}$. For $0 \leq k < m$ and $r \in (t_k, t_{k+1}]$ let $\Delta(x) := 0$ for $x \notin (t_k - \epsilon, r)$, let Δ be linear on $[t_k-\epsilon, t_k]$, and let $\Delta(x) := (r - x)\hat{f}_{ml}(x)$ for $x \in [t_k, r]$. One easily verifies, that this function Δ satisfies the conditions of Proposition 2, too, and with $\epsilon \searrow 0$ we obtain the inequality

$$\int_{t_k}^{r} (r - x)\,\mathbb{F}(dx) \ \le \ \int_{t_k}^{r} (r - x)\,\hat{F}(dx).$$

Integration by parts (or Fubini's theorem) shows that the latter inequality is equivalent to

$$\int_{t_k}^{r} (\mathbb{F}(x) - \mathbb{F}(t_k))\,dx \ \le \ \int_{t_k}^{r} (\hat{F}(x) - \hat{F}(t_k))\,dx.$$

Since $\mathbb{F}(t_k) = \hat{F}(t_k)$, we end up with

$$\int_{t_k}^{r} \mathbb{F}(x)\,dx \ \le \ \int_{t_k}^{r} \hat{F}(x)\,dx \quad \text{for } k = 0, 1, \ldots, m-1 \text{ and } r \in (t_k, t_{k+1}].$$

Hence we may apply Lemma 1 and obtain (1).

Finally, let us consider the least squares estimator $\hat{f}_{ls}$. For $0 \le k < m$ and $r \in (t_k, t_{k+1}]$ let $\Delta(x) := 0$ for $x \notin (t_k - \epsilon, r)$, let Δ be linear on $[t_k - \epsilon, t_k]$ as well as on $[t_k, r]$ with $\Delta(t_k) := r - t_k$. Then applying Proposition 2 and letting $\epsilon \searrow 0$ yields

$$\int_{t_k}^{r} (r - x)\,\mathbb{F}(dx) \ \le \ \int_{t_k}^{r} (r - x)\,\hat{F}(dx),$$

so that

$$\int_{t_k}^{r} \mathbb{F}(x)\,dx \ \le \ \int_{t_k}^{r} \hat{F}(x)\,dx \quad \text{for } k = 0, 1, \ldots, m-1 \text{ and } r \in (t_k, t_{k+1}].$$

Thus it follows from Lemma 1 that

$$\inf_{[0,\infty)} (\hat{F} - F) \ \ge \ \frac{3}{2} \inf_{[0,\infty)} (\mathbb{F} - F) - \frac{1}{2} \sup_{[0,\infty)} (\mathbb{F} - F) \ \ge \ -2\,\big\|\mathbb{F} - F\big\|_\infty.$$

Alternatively, for $1 \le k \le m$ and $r \in [t_{k-1}, t_k)$ let $\Delta(x) := 0$ for $x \notin (r, t_k + \epsilon)$, let Δ be linear on $[r, t_k]$ as well as on $[t_k, t_k + \epsilon]$ with $\Delta(t_k) := -(t_k - r)$. Then applying Proposition 2 and letting $\epsilon \searrow 0$ yields

$$\int_{r}^{t_k} (t_k - x)\,\mathbb{F}(dx) \ \ge \ \int_{r}^{t_k} (t_k - x)\,\hat{F}(dx),$$

so that

$$\int_{r}^{t_k} \mathbb{F}(x)\,dx \ \ge \ \int_{t_k}^{r} \hat{F}(x)\,dx \quad \text{for } k = 1, 2, \ldots, m \text{ and } r \in [t_{k-1}, t_k).$$

Hence it follows from Lemma 1 that

$$\sup_{[0,\infty)} (\hat{F} - F) \ \le \ \frac{3}{2} \sup_{[0,\infty)} (\mathbb{F} - F) - \frac{1}{2} \inf_{[0,\infty)} (\mathbb{F} - F) \ \le \ 2\,\big\|\mathbb{F} - F\big\|_\infty. \qquad \square$$

Acknowledgment. The authors are grateful to Geurt Jongbloed for constructive comments and careful reading.

References

[1] Grenander, U. (1956). On the theory of mortality measurement, part II. *Skand. Aktuarietidskr.* **39** 125–153.

[2] Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001). Estimation of a convex function: Characterization and asymptotic theory. *Ann. Statist.* **29** 1653–1698.

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

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

[5] Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* Wiley, New York.