Statistics and Probability Letters 99 (2015) 143–148

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Marshall lemma in discrete convex estimation

Fadoua Balabdaoui [a], Cécile Durot [b,*]

[a] CEREMADE, Université Paris-Dauphine, 75775, CEDEX 16, Paris, France

[b] UFR SEGMI, Université Paris Ouest Nanterre La Défense, F-92001, Nanterre, France

ARTICLE INFO

Article history:
Received 1 December 2014
Received in revised form 16 January 2015
Accepted 16 January 2015
Available online 25 January 2015

Keywords:
Convex
Nonparametric least squares
Marshall lemma
pmf
Shape constraints

ABSTRACT

We show that the supremum distance between the cumulative distribution of the convex LSE and an arbitrary distribution function F with a convex pmf on $\mathbb{N}$ is at most twice the supremum distance between the empirical distribution function and F.

1. Introduction

1.1. A brief overview

The first Marshall's inequality goes back to Marshall (1970). It states that if $\widehat{F}_n$ is the least concave majorant (LCM) of the empirical distribution function $\mathbb{F}_n$, then $\|\widehat{F}_n - F\|_\infty \le \|\mathbb{F}_n - F\|_\infty$ for an arbitrary concave distribution function F. Here, all the distributions involved are supported on $[0, \infty)$. The proof of this inequality is rather elementary and uses basic properties of the LCM. When F is the true distribution function with a decreasing density f supported on a compact real interval and assumed to be continuously differentiable with a strictly negative derivative, Kiefer and Wolfowitz (1976) showed that the global rate of convergence of $\|\widehat{F}_n - \mathbb{F}_n\|_\infty$ is of order $n^{-2/3}\log(n)$ almost surely. In convex density estimation, two Marshall-type of inequalities have been established for the convex least squares estimator (LSE) of a density on $[0, \infty)$ defined by Groeneboom et al. (2001). Let F be an arbitrary distribution function on $[0, \infty)$ such that F' is convex, $\mathbb{F}_n$ the empirical distribution function, $\widehat{F}_n$ the cumulative distribution of the LSE, $\widehat{H}_n = \int_0^{\cdot} \widehat{F}_n(s)ds$, $\mathbb{H}_n = \int_0^{\cdot} \mathbb{F}_n(s)ds$ and $H = \int_0^{\cdot} F(s)ds$ respectively. Then, Dümbgen et al. (2007) and Balabdaoui and Rufibach (2008) showed that $\|\widehat{F}_n - F\|_\infty \le 2\,\|\mathbb{F}_n - F\|_\infty$, and $\|\widehat{H}_n - H\|_\infty \le \|\mathbb{H}_n - H\|_\infty$ respectively. Those results were used by Balabdaoui and Wellner (2007) to show Kiefer–Wolfowitz-type of inequalities. Specifically, they show that when F is the true distribution such that $f = F'$ is twice continuously differentiable with $f'' > 0$, $\|\widehat{F}_n - \mathbb{F}_n\|_\infty$ and $\|\widehat{H}_n - \mathbb{H}_n\|_\infty$ are respectively of order $(n^{-1}\log(n))^{3/5}$ and $(n^{-1}\log(n))^{4/5}$.

1.2. Discrete versus continuous

In both the monotone and convex estimation problems, the Kiefer–Wolfowitz-type of inequalities recalled above clearly suggest that if one is ready to assume that the true density or its derivative does not admit any flat part on its support, then

* Corresponding author.
E-mail address: cecile.durot@gmail.com (C. Durot).

http://dx.doi.org/10.1016/j.spl.2015.01.016
0167-7152/

the empirical and constrained estimators for the unknown distribution function are equivalent for n large enough. In the discrete setting, this is not true anymore simply because the notion of a strict curvature does not exist in this case. In fact, with $\mathbb{F}_n$ the empirical distribution and $\widehat{F}_n$ the cumulative distribution of the LSE studied by Durot et al. (2013) and Balabdaoui et al. (2014), it can be shown with similar arguments as for the proof of Theorem 3.2 in Balabdaoui et al. (2014), that $(\widehat{F}_n, F_n)$ converges jointly, at the rate $\sqrt{n}$, to a quite complicated limiting distribution, so that $\sqrt{n}(\widehat{F}_n - F_n)$ converges to a distribution that we conjecture is not degenerate. Hence, it is somehow expected that the difference between $\widehat{F}_n$ and $\mathbb{F}_n$ converges exactly at the rate $1/\sqrt{n}$ in this case. Although we do not pursue this in this short note, an immediate consequence of our main theorem is that this difference is at most $O_p(1/\sqrt{n})$ in the supremum norm. This theorem gives the same form of the Marshall inequality proved by Dümbgen et al. (2007) for the convex LSE in the continuous setting. Although we re-adapt their idea of using concavity of the difference between the LSE and the density of F (with respect to the counting measure in our case) between two successive knot points $a < b$, our proof does not rely on the equalities $\mathbb{F}_n(a) = \widehat{F}_n(a)$ and $\mathbb{F}_n(b) = \widehat{F}_n(b)$ as they are no longer true for the discrete LSE. One has also to deal with discrete sums instead of integrals, which makes the final inequalities a bit less straightforward to obtain. One of the consequences of this note is to be able to assert that the convergence rate of $\widehat{F}_n$ to the truth F is the expected rate $1/\sqrt{n}$, whether F has a finite support or not. In Balabdaoui et al. (2014), the assumption that F is finitely supported on $\{0, \dots, S\}$ for some unknown integer $S > 0$ was made to be able to establish the limiting distribution of the LSE, the general case being much more complex to handle. However, this assumption was not at all necessary for Durot et al. (2013) to obtain that $\|\widehat{p}_n - p\|_k = O_p(1/\sqrt{n})$ for all $k \in [2, \infty]$, where $\widehat{p}_n$ and p are respectively the convex discrete LSE and the true probability mass function (pmf). However, for $k = \infty$, the transition from this result to showing that $\|\widehat{F}_n - F\|_\infty = O_p(1/\sqrt{n})$ is not immediate unless F admits a finite support. The Marshall lemma established in this note alleviates any existing doubt that the parametric convergence rate holds true independently of the nature of the support of the true convex pmf.

2. Marshall inequality

Based on a n-sample from a convex discrete probability mass function (pmf) on $\mathbb{N}$, let $\widehat{p}_n$ denote the discrete convex LSE of the pmf as defined by Durot et al. (2013). We recall that the points in $\mathbb{N}\backslash\{0\}$ where $\widehat{p}_n$ changes its slopes are called knots, and that the characterization of $\widehat{p}_n$ is given by

$$\sum_{x=0}^{z-1} \widehat{F}_n(x) \begin{cases} \geq \sum_{x=0}^{z-1} \mathbb{F}_n(x), & z \in \mathbb{N}\backslash\{0\} \\ = \sum_{x=0}^{z-1} \mathbb{F}_n(x), & \text{if } z \text{ is a knot of } \widehat{p}_n, \end{cases} \tag{2.1}$$

where $\mathbb{F}_n$ is the empirical distribution function and $\widehat{F}_n$ is the distribution function corresponding to $\widehat{p}_n$, see Proposition 2.1 in Balabdaoui et al. (2014). In the sequel, F is a distribution function on $\mathbb{N}$ with corresponding pmf, p, that is decreasing and convex on $\mathbb{N}$. We denote by $\mathcal{K}_n$ the set consisting of the point 0 and all knots of $\widehat{p}_n$ and by $\bar{p}_n$ the empirical pmf, that is the pmf corresponding to $\mathbb{F}_n$. We start with the following result.

Lemma 2.1. *For $\tau \in \mathcal{K}_n$, we have $\widehat{F}_n(\tau - 1) \leq \mathbb{F}_n(\tau - 1)$ and $0 \leq \widehat{F}_n(\tau) - \mathbb{F}_n(\tau) \leq \widehat{p}_n(\tau) - \bar{p}_n(\tau)$.*

Proof. First, suppose that $\tau \geq 2$. From the characterization in (2.1), it follows that

$$\sum_{k=0}^{\tau-1} \widehat{F}_n(k) = \sum_{k=0}^{\tau-1} \mathbb{F}_n(k) \tag{2.2}$$

$$\sum_{k=0}^{\tau-2} \widehat{F}_n(k) \geq \sum_{k=0}^{\tau-2} \mathbb{F}_n(k). \tag{2.3}$$

Then, (2.2)–(2.3) yields $\widehat{F}_n(\tau - 1) \leq \mathbb{F}_n(\tau - 1)$. Likewise, (2.2) combined to (2.1) with $z = \tau + 1$ yields $\mathbb{F}_n(\tau) \leq \widehat{F}_n(\tau)$. As the inequality $\widehat{F}_n(\tau - 1) \leq \mathbb{F}_n(\tau - 1)$ can be rewritten under the alternative form $\widehat{F}_n(\tau) \leq \mathbb{F}_n(\tau) + \widehat{p}_n(\tau) - \bar{p}_n(\tau)$, this gives the result for $\tau \geq 2$.

If $\tau = 1$, then the equality in (2.2) gives the first claimed inequality of the lemma $\widehat{F}_n(0) = \mathbb{F}_n(0)$. Combined with the inequality in (2.1) with $z = 2$, this yields the second claimed inequality of the lemma, that is, $\mathbb{F}_n(1) \leq \widehat{F}_n(1) = \mathbb{F}_n(1) + \widehat{p}_n(1) - \bar{p}_n(1)$.

Finally, if $\tau = 0$, then $\widehat{F}_n(-1) = \mathbb{F}_n(-1) = 0$ and (2.1) with $z = 1$ yields $0 \leq \widehat{F}_n(0) - \mathbb{F}_n(0) = \widehat{p}_n(0) - \bar{p}_n(0)$ which completes the proof of the Lemma. □

In what follows, we will establish two important inequalities linking the extrema of $\widehat{F}_n - F$ to those of $\mathbb{F}_n - F$.

Proposition 2.2. *We have*

$$\max_{x \in \mathbb{N}} \left(\widehat{F}_n(x) - F(x)\right) \leq 2 \max_{x \in \mathbb{N}} \left|\mathbb{F}_n(x) - F(x)\right| \tag{2.4}$$