Electronic Journal of Statistics
Vol. 13 (2019) 3243–3253
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1594

Distribution-free properties of isotonic regression

Jake A. Soloff, Adityanand Guntuboyina[*] and Jim Pitman

Department of Statistics
University of California, Berkeley
Berkeley, CA, 94720
e-mail: soloff@berkeley.edu; aditya@stat.berkeley.edu; pitman@stat.berkeley.edu

Abstract: It is well known that the isotonic least squares estimator is characterized as the derivative of the greatest convex minorant of a random walk. Provided the walk has exchangeable increments, we prove that the slopes of the greatest convex minorant are distributed as order statistics of the running averages. This result implies an exact non-asymptotic formula for the squared error risk of least squares in homoscedastic isotonic regression when the true sequence is constant that holds for every exchangeable error distribution.

Keywords and phrases: Shape-constrained regression, statistical dimension, convex minorant, fluctuation theory, quantiles of stochastic processes.

Received February 2019.

1. Introduction

Isotonic regression with homoscedastic errors refers to the problem of estimating a monotone sequence $\theta_1^* \leq \cdots \leq \theta_n^*$ based on a noisy observation vector Y which is assumed to be an additive perturbation of $\theta^* = (\theta_1^*, \ldots, \theta_n^*)$,

$$Y = \theta^* + \sigma Z,$$

where the components $Z_1, \ldots, Z_n$ of Z are assumed to have zero mean and unit variance. It is commonly assumed that $Z_1, \ldots, Z_n$ are independent and identically distributed (i.i.d.) but we work with the more general assumption of exchangeability in this paper. A natural estimator for θ^* in this setting is the isotonic Least Squares Estimator (LSE), defined as

$$\hat{\theta} := \Pi_{\mathcal{M}^n}(Y) := \operatorname{argmin}_{\theta \in \mathcal{M}^n} \|Y - \theta\|_2^2,$$

where $\|\cdot\|_2$ denotes the usual Euclidean norm on $\mathbb{R}^n$ and $\mathcal{M}^n := \{\theta \in \mathbb{R}^n : \theta_1 \leq \cdots \leq \theta_n\}$ is the monotone cone of length n non-decreasing sequences. As $\mathcal{M}^n$ is a closed convex cone, $\hat{\theta}$ as defined above exists uniquely; it can also be computed in $O(n)$ time by the pool adjacent violators algorithm [4, 11].

One approach to evaluating the statistical properties of $\hat{\theta}$ is to measure the risk, or expected deviation of $\hat{\theta}$ from θ^*. Indeed, the risk provides a convenient

[*]Supported by NSF CAREER Grant DMS-16-54589.

summary of the accuracy of $\hat{\theta}$ and many papers on isotonic regression have focussed on obtaining bounds for the risk of $\hat{\theta}$ (see e.g., [3, 12, 20]). In this paper, we primarily consider the normalized mean squared error:

$$R(\hat{\theta}, \theta^*) := \frac{1}{n}\mathbb{E}_{\theta^*}\|\hat{\theta} - \theta^*\|_2^2.$$

A key quantity in understanding $R(\hat{\theta}, \theta^*)$ is

$$\delta_n(\mu) := \mathbb{E}_{Z\sim\mu}\|\Pi_{\mathcal{M}^n}(Z)\|_2^2,$$

where μ denotes the law of the noise vector Z. Indeed, it is clear that

$$\frac{n}{\sigma^2}R(\hat{\theta}, \theta^*) = \delta_n(\mu) \qquad\qquad \text{when } \theta_1^* = \cdots = \theta_n^*.$$

When $\theta_1^* \leq \cdots \leq \theta_n^*$ are not all equal, let $(A_1, \ldots, A_k)$ be the coarsest partition of $\{1, \ldots, n\}$ such that θ^* is constant on each A_i. It has been shown [16, 9, 3] that

$$\frac{n}{\sigma^2}R(\hat{\theta}, \theta^*) \begin{cases} \leq \delta_{n_1}(\mu_{A_1}) + \cdots + \delta_{n_k}(\mu_{A_k}) & \text{for every } \sigma > 0 \\ \to \delta_{n_1}(\mu_{A_1}) + \cdots + \delta_{n_k}(\mu_{A_k}) & \text{as } \sigma \downarrow 0 \end{cases}, \tag{1.1}$$

where μ_{A_i} denotes the marginal distribution of $(Z_j)_{j\in A_i}$ and $n_i = |A_i|$ is the length of the i^{th} block for all $i = 1, \ldots, k$. We emphasize that (1.1) holds for arbitrarily dependent $Z_1, \ldots, Z_n$ with zero mean and finite variance. It was also shown in Bellec [3] that $\delta_n(\mu)$ also bounds the risk of the isotonic LSE in misspecified settings where θ^* does not lie in $\mathcal{M}^n$.

The quantity $\delta_n(\mu)$ therefore crucially controls the risk of the isotonic LSE. The goal of this paper is to explicitly determine $\delta_n(\mu)$ for every $n \geq 1$ under the additional assumption that Z is exchangeable. Specifically, under the assumption of exchangeability, we show in Corollary 3.3 that, for all n,

$$\delta_n(\mu) = \rho n + (1 - \rho)H_n, \tag{1.2}$$

where $H_n := 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ is the n^{th} harmonic number and $\rho = \text{Cor}(Z_1, Z_2)$ is the pairwise correlation. Combined with (1.1), our result provides a sharp non-asymptotic bound on the risk of isotonic regression for *any* exchangeable noise vector. In the special case when $Z_1, \ldots, Z_n$ are i.i.d. with zero mean and unit variance, $\rho = 0$ and thus (1.2) gives:

$$\delta_n(\otimes_{i=1}^n \eta) = H_n \qquad \text{for every probability measure } \eta. \tag{1.3}$$

Here η is the common distribution of the independent variables $Z_1, \ldots, Z_n$.

Previously, the formula (1.3) was known when η is the standard Gaussian probability measure on $\mathbb{R}^n$. This was observed by Amelunxen et al. [2] who proved it by observing first that when $\mu = \otimes_{i=1}^n \eta$ and η is the standard Gaussian measure, the formula

$$\mathbb{E}\|\Pi_K(Z)\|_2^2 = \sum_{k=0}^{n} k\,\nu_k(K) \tag{1.4}$$

holds for every closed convex cone $K \subseteq \mathbb{R}^n$ where $\nu_k(K)$ is the k^{th} intrinsic volume of K. When $K = \mathcal{M}^n$ is the monotone cone, the right hand side in equation (1.4) can be shown to be equal to H_n by using the fact that the generating function $s \mapsto \sum_{k=0}^n s^k \nu_k(\mathcal{M}^n)$ can be computed in closed form. Amelunxen et al. [2] used the theory of finite reflection groups [7] to obtain the exact expression for this generating function. However, the exact expression for $\sum_{k=0}^n s^k \nu_k(\mathcal{M}^n)$ can already be found in the classical literature on isotonic regression (see Theorem 2.4.2 in Roberston et al. [17] and references therein).

The above proof does not work for non-Gaussian η mainly because the expression (1.4) does not hold for general η. In fact, the best available result on $\delta_n(\otimes_{i=1}^n \eta)$ for non-Gaussian η is in equation (2.11) of Zhang [20], who proved the asymptotic result:

$$\delta_n(\otimes_{i=1}^n \eta) = (1 + o(1))(1 + \log n) \qquad \text{as } n \to \infty.$$

This bound gives the right behavior as the right hand side of equation (1.3) but only as $n \to \infty$. We improve this result by proving for every $n \geq 1$ that $\delta_n(\otimes_{i=1}^n \eta)$ is always equal to the n^{th} harmonic number H_n for every probability measure η having mean 0 and variance 1.

We prove (1.2) by developing a precise characterization of the marginal distribution of each individual component $(\Pi_{\mathcal{M}^n}(Z))_k$ of $\Pi_{\mathcal{M}^n}(Z)$. Specifically, as long as Z is exchangeable, we show in Theorem 2.2 that $(\Pi_{\mathcal{M}^n}(Z))_k$ has the same distribution as $\bar{Z}_{(k)}$, the k^{th} order statistic of the running averages $\bar{Z}_j = \frac{Z_1 + \cdots + Z_j}{j}$. We prove Theorem 2.2 in Section 2, using a characterization of the components of the isotonic LSE as the left-hand slopes of the greatest convex minorant of the random walk with increments $Z_1, \ldots, Z_n$. This result and its continuous-time analogue may be of independent interest outside the study of isotonic regression, so in Section 2 we also address consequences for the greatest convex minorant of a stochastic process with exchangeable increments. The order statistics of the running averages $\{\bar{Z}_k\}_{k=1}^n$ can be fairly complicated even when Z is Gaussian; however, Theorem 2.2 easily implies results such as (1.2). In Section 3, we detail some risk calculations for isotonic regression and its variants which all follow from Theorem 2.2.

2. Main result

Let $S_k = \sum_{i=1}^k Z_i$ denote the partial sums for $k = 1, \ldots, n$, started at $S_0 = 0$. Identify the random walk $\{S_k\}_{k=0}^n$ with its *cumulative sum diagram* $S : [0, n] \to \mathbb{R}$, where $S(k) = S_k$ for integers $k = 0, \ldots, n$ and linearly interpolated between integers. Let $C : [0, n] \to \mathbb{R}$ denote the *greatest convex minorant* (GCM) of S, i.e. the greatest convex function that lies below S. See Figure 1 for a depiction of the GCM of S. With this notation, we now recall the graphical representation of the isotonic LSE as given in Theorem 1.2.1 of Roberston et al. [17].

Lemma 2.1. *For any vector Z, the isotonic LSE $\Pi_{\mathcal{M}^n}(Z)$ is given by the left-hand slopes of the greatest convex minorant of the cumulative sum diagram. For*

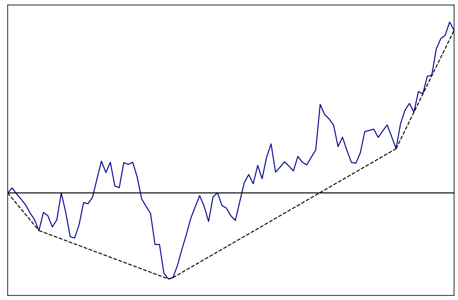

Fig 1. *Solid blue curve is the cumulative sum diagram S of increments $Z_1, \dots, Z_n$; dashed black curve is the greatest convex minorant C of S.*

all $k = 1, \dots, n$

$$(\Pi_{\mathcal{M}^n}(Z))_k = C(k) - C(k-1) = \partial_- C(k).$$

For the remainder of this section let

$$\Delta_k := \partial_- C(k) = \min_{k \le v \le n} \max_{0 \le u < k} \frac{S_v - S_u}{v - u} \tag{2.1}$$

denote the left-hand slope of the GCM at k, so $\Delta = (\Delta_1, \dots, \Delta_n)$ is equal to $\Pi_{\mathcal{M}^n}(Z)$ by the lemma. In particular, when $k = 1$ we have $\Delta_1 = \min_{1 \le v \le n} \frac{S_v}{v}$. When $k = n$, we have $\Delta_n = \max_{0 \le u < n} \frac{S_n - S_u}{n - u}$, and if $(Z_n, \dots, Z_1) \overset{d}{=} (Z_1, \dots, Z_n)$ then $\Delta_n \overset{d}{=} \max_{1 \le u \le n} \frac{S_u}{u}$. Our next result generalizes this observation, showing that the k^{th} slope Δ_k is equal in distribution to the k^{th} smallest running average if Z is exchangeable.

Theorem 2.2. *Suppose $Z = (Z_1, \dots, Z_n)$ is exchangeable. Let $\bar{Z}_k := \frac{1}{k} \sum_{i=1}^k Z_i$ denote the k^{th} running average for $k = 1, \dots, n$ and let $\bar{Z}_{(1)} \le \dots \le \bar{Z}_{(n)}$ denote their order statistics. Then*

$$\Delta_k \overset{d}{=} \bar{Z}_{(k)} \tag{2.2}$$

marginally for all $k = 1, \dots, n$.

Proof. As before, let S_k denote the k^{th} partial sum. Let M be the last argmin of the sequence $\{S_i\}_{i=0}^n$, and let N be the amount of time the walk is non-positive $N := \sum_{i=1}^n 1(S_i \le 0)$. We will use Corollary 11.14 of Kallenberg [13], due to Sparre-Andersen, which says $M \overset{d}{=} N$ as long as Z is exchangeable.

Note that the slope of the GCM switches from non-positive to positive at time M, since the horizontal line with intercept S_M minorizes the GCM and touches it at time M. Hence, no matter the sequence of increments Z_i, there is the identity of events

$$(\Delta_k \le 0) = (M \ge k). \tag{2.3}$$

Also, for the time N that the walk is non-positive, since $S_i \leq 0$ if and only if $\bar{Z}_i \leq 0$, there is the identity of events

$$(\bar{Z}_{(k)} \leq 0) = (N \geq k).$$

The equality in distribution $M \overset{d}{=} N$ then implies

$$\mathbb{P}(\Delta_k \leq 0) = \mathbb{P}(\bar{Z}_{(k)} \leq 0).$$

If the sequence $\{Z_i\}$ is modified to $\{Z_i - z\}$ for some fixed z, the modified sequence is exchangeable, and the values of Δ_k and $\bar{Z}_{(k)}$ for the modified sequence are just $\Delta_k - z$ and $\bar{Z}_{(k)} - z$. Applying the above identity to the modified sequence gives

$$\mathbb{P}(\Delta_k \leq z) = \mathbb{P}(\Delta_k - z \leq 0) = \mathbb{P}(\bar{Z}_{(k)} - z \leq 0) = \mathbb{P}(\bar{Z}_{(k)} \leq z).$$

So Δ_k and $\bar{Z}_{(k)}$ have the same cumulative distribution function, hence the same distribution. □

The proof of Theorem 2.2 has a straightforward generalization to the setting where $S : [0,1] \to \mathbb{R}$ is a continuous-time stochastic process. Knight [14] showed that the analogous distributional identity $M \overset{d}{=} N$ holds when S has exchangeable increments and $S(0) = 0$. Hence, by a similar proof, we find that the slope $\Delta(p)$ of the greatest convex minorant of S at time $p \in [0,1]$ has the same distribution as the p^{th} percentile point of the occupation measure for the process $(\frac{S(t)}{t}, 0 \leq t \leq 1)$. We record this result as the following corollary.

Corollary 2.3. *Let S denote a real-valued càdlàg stochastic process on $[0,1]$ with exchangeable increments, such that $S(0) = 0$. Define $\Delta(t)$ as the slope of the greatest convex minorant of S at t, and let $F : \mathbb{R} \to [0,1]$ denote the (random) cdf associated with the occupation measure of $(\frac{S(t)}{t}, 0 \leq t \leq 1)$,*

$$F(x) = \lambda(\{t \in [0,1] : S(t) \leq tx\}), \tag{2.4}$$

where λ denotes Lebesgue measure. Then

$$\Delta(p) = \inf_{p \leq v \leq 1} \sup_{0 \leq u < p} \frac{S(v) - S(u)}{v - u} \overset{d}{=} F^{-1}(p) \tag{2.5}$$

marginally for all $p \in [0,1]$.

See Abramson et al. [1] for a general study of convex minorants of random walks and processes with exchangeable increments. In the special cases where S is a standard Brownian motion or Brownian bridge on the unit interval, Carolan & Dykstra [6] derive the distribution of the slope $\Delta(p)$, jointly with the process $S(p)$ and its convex minorant at p, for a fixed value $p \in [0,1]$. Given our corollary, their explicit formula for the slope $\Delta(p)$ provides the distribution of $F^{-1}(p)$, giving new information about the occupation measure of $(\frac{S(t)}{t}, 0 \leq t \leq 1)$ for Brownian motion and Brownian bridge. The distribution of the p^{th} percentile point of the occupation measure for $(S(t), 0 \leq t \leq 1)$ has been obtained under the same generality as Corollary 2.3: see the introduction of Dassios [8] and references therein.

3. Consequences for isotonic regression

Since the identity of Theorem 2.2 holds marginally, it allows us to simplify expectations of functions that are additive in the components of $\Pi_{\mathcal{M}^n}(Z)$. By Lemma 2.1, the k^{th} component $(\Pi_{\mathcal{M}^n}(Z))_k = \Delta_k$, which by Theorem 2.2 is equal in distribution to $\bar{Z}_{(k)}$. Hence, as long as Z is exchangeable,

$$\sum_{k=1}^{n} \mathbb{E}h((\Pi_{\mathcal{M}^n}(Z))_k) = \sum_{k=1}^{n} \mathbb{E}h(\bar{Z}_{(k)}) = \sum_{k=1}^{n} \mathbb{E}h(\bar{Z}_k). \tag{3.1}$$

Taking $h(x) = |x|^p$, we obtain our first corollary.

Corollary 3.1. *Suppose $Z = (Z_1, \dots, Z_n)$ is exchangeable. For $p > 0$,*

$$\mathbb{E}\|\Pi_{\mathcal{M}^n}(Z)\|_p^p = \sum_{k=1}^{n} \mathbb{E}\left|\frac{1}{k}\sum_{i=1}^{k} Z_i\right|^p, \tag{3.2}$$

provided $\mathbb{E}|Z_1|^p < \infty$.

Remark 3.2. *Viewed through its graphical representation, $\Delta_k = C(k) - C(k-1)$ is the left-derivative of the GCM C at k, so when the power $p = 1$, equation (3.2) yields the discrete arc-length formula*

$$\sum_{k=1}^{n} \mathbb{E}|C(k) - C(k-1)| = \mathbb{E}\|\Pi_{\mathcal{M}^n}(Z)\|_1 = \sum_{k=1}^{n} \frac{1}{k}\mathbb{E}|S_k| \tag{3.3}$$

Closely related to this formula is the identity of Spitzer & Widom [18], which takes $\tilde{Z}_1, \dots, \tilde{Z}_n$ to be a sequence of i.i.d. random variables in $\mathbb{R}^2$ (or the complex plane $\mathbb{C}$) with finite variance. If $\tilde{S}_k = \sum_{i=1}^{k} \tilde{Z}_i$ is the partial sum and $\tilde{L}_n$ is the length of the perimeter of the convex hull conv$(0, \tilde{S}_1, \dots, \tilde{S}_n)$, then

$$\mathbb{E}\tilde{L}_n = 2\sum_{k=1}^{n} \frac{1}{k}\mathbb{E}\|\tilde{S}_k\|. \tag{3.4}$$

These formulas connect the geometry of the convex hull of a random walk to the magnitudes of the running means.

Consider the case when $p = 2$. Since Z is exchangeable, every pair of components has the same correlation ρ. If we further assume Z_1 has zero mean and unit variance, the right hand side of equation (3.2) can be computed explicitly

$$\mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k} Z_i\right)^2 = \rho + \frac{1-\rho}{k}.$$

Summing over k yields our next result.

Corollary 3.3. *Suppose $Z \sim \mu$ is an exchangeable random vector with zero mean, unit variance, and pairwise correlation ρ. Then*

$$\delta_n(\mu) = \rho n + (1-\rho) H_n.$$

This result should be contrasted with other distribution-free identities, namely

$$\mathbb{E}\|Z\|_2^2 = n \text{ and } \mathbb{E}\|\bar{Z}_n \mathbf{1}_n\|_2^2 = 1,$$

provided Z has i.i.d. components with zero mean and unit variance. In particular, suppose we observe $Y = \theta^* + \sigma Z$ where Z has i.i.d. components with zero mean and unit variance, but it turns out that $\theta^* = c\mathbf{1}_n$ is constant. If we know θ^* is constant, we can estimate it by a constant sequence $\bar{Y}\mathbf{1}_n$ and pay a price of $\frac{\sigma^2}{n}$ in risk (normalized mean squared error). If we know nothing about the structure of θ^* and use $\hat{\theta} = Y$, the risk σ^2 is quite large by comparison. The monotone sequence estimate resides in the middle, with a much smaller risk of $\frac{H_n \sigma^2}{n}$ and knowledge only about the relative order.

Theorem 2.2 characterizes the distribution of a component of the isotonic LSE $\hat{\theta}$ when the underlying sequence θ^* is constant. When $\theta^* \in \mathcal{M}^n$ is not constant, Theorem 2.2 can be applied to characterize the distribution of a component $\hat{\theta}_i$ in the low noise limit $\sigma \downarrow 0$. In this limit, the distribution depends only on flat regions of θ^*:

Corollary 3.4. *Suppose $Y = \theta^* + \sigma Z$, for some $\theta^* \in \mathcal{M}^n$, and let $\hat{\theta} = \Pi_{\mathcal{M}^n}(Y)$ denote the isotonic LSE. Let $(A_1, \ldots, A_k)$ be the coarsest partition of $\{1, \ldots, n\}$ such that θ^* is constant on each A_j, and suppose Z is exchangeable on each of these blocks. If an index $i \in \{1, \ldots, n\}$ belongs to the j^{th} block, then letting $t_i = i + 1 - \min_{s \in A_j} s$ and $X = (Z_s)_{s \in A_j}$, we have*

$$\frac{\hat{\theta}_i - \theta_i^*}{\sigma} \xrightarrow{d} \bar{X}_{(t_i)} \text{ as } \sigma \downarrow 0. \tag{3.5}$$

Proof. As $\sigma \downarrow 0$, the ratio $\frac{\hat{\theta} - \theta^*}{\sigma}$ tends to the directional derivative $D_Z \Pi_{\mathcal{M}^n}(\theta^*)$. Lemma 4.6 in Zarantonello [19] shows that this derivative exists and equals the projection of Z onto the tangent cone $T_{\mathcal{M}^n}(\theta^*)$. Hence

$$\frac{\hat{\theta} - \theta^*}{\sigma} \to \Pi_{T_{\mathcal{M}^n}(\theta^*)}(Z) \text{ as } \sigma \downarrow 0. \tag{3.6}$$

From the tangent cone computation in Bellec [3], we have

$$\left(\Pi_{T_{\mathcal{M}^n}(\theta^*)}(Z)\right)_i = \left(\Pi_{\mathcal{M}^{n_i}}(X)\right)_{t_i},$$

where $n_i = |A_{j_i}|$. Finally, by Theorem 2.2, $\left(\Pi_{\mathcal{M}^{n_i}}(X)\right)_{t_i} \stackrel{d}{=} \bar{X}_{(t_i)}$. □

We explained in Section 1 how risk calculations when $\theta^* = 0$ generalize to MSE bounds that are sharp in the low noise limit for arbitrary θ^*. For example, when $\theta^* \in \mathcal{M}^n$ has k constant pieces, then (1.1), Corollary 3.3 and the fact that $H_l \leq \log(el)$ for every $l \geq 1$ imply that

$$R(\hat{\theta}, \theta^*) \leq \frac{k\sigma^2}{n} \log\left(\frac{en}{k}\right) \tag{3.7}$$

whenever $Z_1, \ldots, Z_n$ are i.i.d. with mean zero and unit variance. The bound (3.7) should be compared with the risk of the structure-respecting estimator that averages over the constant blocks and achieves a risk of exactly $\frac{k\sigma^2}{n}$ when the blocks are all of size $\frac{n}{k}$. If $\theta^* \in \mathbb{R}^n$ is not necessarily in $\mathcal{M}^n$, then Corollary 3.3, together with the results of [3], implies that

$$R(\hat{\theta}, \theta^*) \leq \inf_{\theta \in \mathcal{M}^n} \left(\frac{1}{n} \|\theta - \theta^*\|^2 + \sigma^2 \frac{k(\theta)}{n} \log\left(\frac{en}{k(\theta)}\right)\right),$$

where $k(\theta)$ is the number of constant pieces of the vector θ. These formulae (with the leading constant of 1 in front of the $\frac{k\sigma^2}{n} \log \frac{en}{k}$ term on the right hand side) were previously only known when the distribution of $Z_1, \ldots, Z_n$ was standard Gaussian.

Define the L^p-risk of the isotonic LSE

$$R^{(p)}(\hat{\theta}, \theta^*) = \frac{1}{n} \mathbb{E}\|\hat{\theta} - \theta^*\|_p^p$$

so that $R(\hat{\theta}, \theta^*) = R^{(2)}(\hat{\theta}, \theta^*)$. We can similarly employ Theorem 2.2 to explicitly calculate the L^p-risk of the isotonic LSE $\hat{\theta}$ when θ^* is constant and Z is Gaussian:

Corollary 3.5. *Suppose $Z \sim \mathcal{N}(0, I_n)$. Then for any $p > 0$,*

$$\mathbb{E}\|\Pi_{\mathcal{M}^n}(Z)\|_p^p = H_{n,p/2} \mathbb{E}|Z_1|^p = H_{n,p/2} \sqrt{\frac{2^p}{\pi}} \Gamma\left(\frac{p+1}{2}\right),$$

where $H_{n,m} = \sum_{k=1}^n \frac{1}{k^m}$.

Proof. Note $\mathbb{E}\left|\frac{1}{k}\sum_{i=1}^k Z_i\right|^p = \left(\frac{2}{k}\right)^{p/2} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}$ and apply the theorem. □

Corollary 3.5 should similarly be contrasted with the following identities when $Z \sim \mathcal{N}(0, I_n)$:

$$\mathbb{E}\|Z\|_p^p = n\mathbb{E}|Z_1|^p \text{ and } \mathbb{E}\|\bar{Z}\mathbf{1}_n\|_p^p = n^{1-p/2}\mathbb{E}|Z_1|^p$$

respectively. In particular, when $p > 2$, the bound $H_{n,p/2} < \sum_{k=1}^\infty \frac{1}{k^{p/2}} < \infty$ holds for all n, which is to say $\mathbb{E}\|\Pi_{\mathcal{M}^n}(Z)\|_p^p$ is bounded when $p > 2$ whereas $\mathbb{E}\|Z\|_p^p$ grows without bound as n grows.

When θ^* is constant and $Z \sim \mathcal{N}(0, I_n)$, the L^p risk of isotonic regression is

$$R^{(p)}(\hat{\theta}, \theta^*) = \frac{H_{n,p/2}}{n} \sigma^p \mathbb{E}|Z_1|^p. \tag{3.8}$$

When $1 \le p \le 2$, Theorem 2.3 of Zhang [20] shows an asymptotic result for the L^p risk on constant θ^* that agrees with equation (3.8).

The continuous-time distributional identity in Corollary 2.3 applies to the asymptotic distribution of the isotonic least squares estimator. A standard model for studying the asymptotic behavior of isotonic regression is

$$\theta_k^* = f^*\left(\frac{k}{n}\right)$$

where $f^* : [0,1] \to \mathbb{R}$ is non-decreasing. We observe Y, a noisy version of θ^*, and calculate $\hat{\theta}$ by projecting Y onto the monotone cone. The function estimate $\hat{f}$ is defined by $\hat{f}\left(\frac{k}{n}\right) = \hat{\theta}_k$ and linearly interpolated between design points. Here, as before, the dependence on n in $\theta^* \in \mathcal{M}^n$ is suppressed, but now we are interested in the behavior of isotonic least squares $\hat{f}(t_0)$ at a fixed point $t_0 \in [0,1]$ as $n \to \infty$.

Define the partial sum process $S^{(n)} : [0,1] \to \mathbb{R}$ by $S^{(n)}(k/n) = \frac{Y_1 + \cdots + Y_k}{\sqrt{n}}$, linearly interpolated between design points. When the function $f^* \equiv c$ is constant, the quantity

$$\sqrt{n}(\hat{f}(t_0) - f^*(t_0))$$

is given by the left-derivative of the greatest convex minorant of $S^{(n)}$ at t_0. By the invariance principle, this converges in distribution to the left-derivative of the greatest convex minorant of standard Brownian motion $B = (B(t), 0 \le t \le 1)$ at t_0. This asymptotic result is well known and a similar result was noted for the Grenander estimator in Carolan & Dykstra [5], where Brownian motion is replaced with a Brownian bridge. Corollary 2.3 relates this asymptotic distribution to the percentile points of the occupation measure for $(\frac{B(t)}{t}, 0 \le t \le 1)$.

Finally, Corollary 3.3 on the projection onto $\mathcal{M}^n$ extends over to that of the set of non-negative monotone sequences $\mathcal{M}_+^n = \mathcal{M}^n \cap \mathbb{R}_+^n$. Theorem 1 of Németh & Németh [15] observes that the projection of Z onto $\mathcal{M}_+^n$ is given by $\Pi_{\mathcal{M}_+^n}(Z) = \Pi_{\mathcal{M}^n}(Z)_+$, the element-wise positive part of the projection onto $\mathcal{M}^n$. Hence the distributional identity Theorem 2.2 yields a similar set of identities for non-negative isotonic regression.

Corollary 3.6. *For any exchangeable noise vector Z,*

$$(\Pi_{\mathcal{M}_+^n}(Z))_k \stackrel{d}{=} (\bar{Z}_{(k)})_+ \tag{3.9}$$

Provided $\mathbb{E}|Z_i|^p < \infty$,

$$\mathbb{E}\|\Pi_{\mathcal{M}_+^n}(Z)\|_p^p = \sum_{k=1}^n \mathbb{E}\left(\frac{1}{k}\sum_{i=1}^k Z_i\right)_+^p . \tag{3.10}$$

Furthermore, if Z is symmetric with unit variance, the generalized statistical dimension of the monotone cone is

$$\mathbb{E}\|\Pi_{\mathcal{M}_+^n}(Z)\|_2^2 = \frac{\rho n + (1-\rho)H_n}{2}, \tag{3.11}$$

where ρ is the pairwise correlation.

Proof. Equation (3.10) follows from equation (3.1) by taking $h(x) = (x)_+^p$. When $Z_i \stackrel{d}{=} -Z_i$ is symmetric with unit variance,

$$\mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k} Z_i\right)_+^2 = \frac{1}{2}\mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k} Z_i\right)^2 = \frac{1}{2}\left(\rho + \frac{1-\rho}{k}\right).$$

Summing over k yields equation (3.11). □

Equation (3.11) is also shown in Amelunxen et al. [2] in the special case $Z \sim \mathcal{N}(0, I_n)$ using the theory of finite reflection groups. The identity (3.10) allows us to show equation (3.11) for a much wider variety of noise vectors, and as before also allows us to obtain relations for the expected L^p norms of the projection of the noise vector. All of our exact formulae follow from the distributional identity in Theorem 2.2, which exploits the geometric characterization of the isotonic LSE in Lemma 2.1. An interesting open question is whether similar characterizations—such as for convex regression [10]—may yield exact non-asymptotic risk calculations in other shape-constrained estimation problems.

References

[1] J. Abramson, J. Pitman, N. Ross, and G. U. Bravo. Convex minorants of random walks and Lévy processes. *Electronic Communications in Probability*, 16:423–434, 2011. MR2831081

[2] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: Phase transitions in convex programs with random data. *Information and Inference: A Journal of the IMA*, 3(3):224–294, 2014. MR3311453

[3] P. C. Bellec. Sharp oracle inequalities for least squares estimators in shape restricted regression. *The Annals of Statistics*, 46(2):745–780, 2018. MR3782383

[4] H. D. Brunk, R. E. Barlow, D. J. Bartholomew, and J. M. Bremner. *Statistical inference under order restrictions; the theory and application of isotonic regression*. Wiley, New York, 1972. MR0326887

[5] C. Carolan and R. Dykstra. Asymptotic behavior of the grenander estimator at density flat regions. *Canadian Journal of Statistics*, 27(3):557–566, 1999. MR1745821

[6] C. Carolan and R. Dykstra. Marginal densities of the least concave majorant of Brownian motion. *Ann. Statist.*, 29(6):1732–1750, 2001. MR1891744

[7] H. S. M. Coxeter and W. O. J. Moser. *Generators and relations for discrete groups*, volume 14. Springer Science & Business Media, 2013. MR0349820

[8] A. Dassios. On the quantiles of Brownian motion and their hitting times. *Bernoulli*, 11(1):29–36, 2005. MR2121453

[9] B. Fang and A. Guntuboyina. On the risk of convex-constrained least squares estimators under misspecification. *Bernoulli*, 25(3):2206–2244, 2019. MR3961246

[10] P. Groeneboom, G. Jongbloed, and J. A. Wellner. Estimation of a convex function: characterizations and asymptotic theory. *The Annals of Statistics*, 29(6):1653–1698, 2001. MR1891742

[11] S. J. Grotzinger and C. Witzgall. Projections onto order simplexes. *Applied mathematics and Optimization*, 12(1):247–270, 1984. MR0768632

[12] A. Guntuboyina and B. Sen. Nonparametric shape-restricted regression. *Statistical Science*, 33(4):568–594, 2018. MR3881209

[13] O. Kallenberg. *Foundations of modern probability*. Springer Science & Business Media, 2006. MR1464694

[14] F. B. Knight. The uniform law for exchangeable and Lévy process bridges. *Astérisque*, (236):171–188, 1996. MR1417982

[15] A. B. Németh and S. Z. Németh. How to project onto the monotone nonnegative cone using pool adjacent violators type algorithms. *arXiv preprint arXiv:1201.2343*, 2012.

[16] S. Oymak and B. Hassibi. Sharp mse bounds for proximal denoising. *Foundations of Computational Mathematics*, 16(4):965–1029, 2016. MR3529131

[17] T. Robertson, F. T. Wright, and R. L. Dysktra. *Order restricted statistical inference*. 1988. MR0961262

[18] F. Spitzer and H. Widom. The circumference of a convex polygon. *Proceedings of the American Mathematical Society*, 12(3):506–509, 1961. MR0130616

[19] E. H. Zarantonello. Projections on convex sets in Hilbert space and spectral theory. In *Contributions to Nonlinear Functional Analysis*, 237–424, 1971. MR0388177

[20] C.-H. Zhang. Risk bounds in isotonic regression. *The Annals of Statistics*, 30(2):528–555, 2002. MR1902898