# A nonasymptotic law of iterated logarithm for general M-estimators

**Nicolas Schreuder** **Victor-Emmanuel Brunel** **Arnak Dalalyan**
CREST, ENSAE, IP Paris

## Abstract

$M$-estimators are ubiquitous in machine learning and statistical learning theory. They are used both for defining prediction strategies and for evaluating their precision. In this paper, we propose the first non-asymptotic "any-time" deviation bounds for general $M$-estimators, where "any-time" means that the bound holds with a prescribed probability for *every* sample size. These bounds are non-asymptotic versions of the law of iterated logarithm. They are established under general assumptions such as Lipschitz continuity of the loss function and (local) curvature of the population risk. These conditions are satisfied for most examples used in machine learning, including those ensuring robustness to outliers and to heavy-tailed distributions. As an example of application, we consider the problem of best arm identification in a stochastic multi-armed bandit setting. We show that the established bound can be converted into a new algorithm, with provably optimal theoretical guarantees. Numerical experiments illustrating the validity of the algorithm are reported.

## 1 Introduction

Perhaps the most fundamental theorems in statistics are the law of large numbers (LLN) and the central limit theorem (CLT). Morally, they state that a sample average converges almost surely or in probability to the population average, and if one zooms in by multiplying by a square root factor, a much weaker form of stochastic convergence still holds, namely, convergence in distribution towards a Gaussian law. A fine intermediate result shows what happens in between the two scales: the law of iterated logarithm (LIL). By zooming in slightly less than in the CLT, *i.e.*, by rescaling the sample average with a slightly smaller factor than in the CLT, it is possible to gain a guarantee for infinitely many sample sizes, almost surely. In practice, however, the LIL has limited applicability, since it does not specify for which sample sizes the guarantee holds. The goals of the present work are to lift this limitation, by proving a LIL valid for every sample size, and holding for general $M$-estimators, rather than for the sample mean only.

The precise statement of the LIL, discovered by Khintchine (1924); Kolmogoroff (1929) almost a century ago, is as follows: for a sequence of i.i.d. random variables $\{Y_i\}_{i\in\mathbb{N}}$ with mean $\theta$ and variance $\sigma^2 < \infty$, the sample averages $\bar{Y}_n = (Y_1 + \ldots + Y_n)/n$ satisfy the relations

$$\liminf_{n\to\infty} \frac{\sqrt{n}\,(\bar{Y}_n - \theta)}{\sigma\sqrt{2\ln\ln n}} = -1,$$
$$\limsup_{n\to\infty} \frac{\sqrt{n}\,(\bar{Y}_n - \theta)}{\sigma\sqrt{2\ln\ln n}} = 1,$$

almost surely. This provides a guarantee on the deviations of the sample average as an estimator of the mean $\theta$ since it yields that, with probability one, for any constant $c > 1$, there exists an integer $n_0 \in \mathbb{N}$ such that $|\bar{Y}_n - \theta| \le c\sigma(2\ln\ln n/n)^{1/2}$ for every $n \ge n_0$. As compared to the deviation guarantees provided by the central limit theorem, the one of the last sentence has the advantage of being valid for any sample size large enough. This advantage is gained at the expense of a factor $(\ln\ln n)^{1/2}$. Akin for the classic version of the CLT, the applicability of the LIL is limited by the fact that it is hard to get any workable expression of $n_0$.

In the case of the CLT and its use in statistical learning, the drawback related to $n_0$ was lifted by exploiting concentration inequalities, such as the Hoeffding or the Bernstein inequalities, that can be seen as non-asymptotic versions of the CLT. For bounded random variables, the aforementioned concentration inequalities imply that for a prescribed tolerance level $\delta \in (0,1)$, for every $n \in \mathbb{N}$, the event[1] $\mathcal{A}_n = \{|\bar{Y}_n - \theta| \le$

Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108. 

[1] Here $C$ is a universal constant.

$C(\ln(1/\delta)/n)^{1/2}\}$ holds with probability at least $1-\delta$. Such a deviation bound is satisfactory in a batch setting, when all the data are available in advance. In contrast, when data points are observed sequentially, as in online learning, or when the number of acquired data points depends on the actual values of the data points, the event of interest is $\bar{\mathcal{A}}_N = \mathcal{A}_1 \cap \ldots \cap \mathcal{A}_N$ or even a version of it in which $N$ can be replaced by $\infty$. One can use the union bound to ensure that $\bar{\mathcal{A}}_N$ has a probability at least $1 - N\delta$ but this is too crude. Furthermore, replacing in $\mathcal{A}_n$ the confidence $\delta$ by $\delta/n^2$, we get coverage $1-\pi^2\delta/6$, valid for any sample size $n$, for an interval of length $O((\ln n/n)^{1/2})$. This result, obtained by a straightforward application of the union bound, is sub-optimal. A remedy to such a sub-optimality—in the form of a nonasymptotic version of the LIL—was proposed by Jamieson et al. (2014) and further used by Howard et al. (2018); Kaufmann and Koolen (2018); Kaufmann et al. (2016). In addition, its relevance for online learning was demonstrated by deriving guarantees for the best arm selection in a multi-armed bandit setting. Note that these recent results apply exclusively to the sample mean in the one-dimensional setting; there is no equivalent of these bounds for other types of (possibly multivariate) estimators.

In this work, we establish a non-asymptotic LIL in a general setting encompassing many estimators, far beyond the sample mean. More precisely, we focus on the class of (penalized) $M$-estimators comprising the sample mean but also the sample median, the quantiles, the least-squares estimator, etc. Of particular interest to us are estimators that are robust to outliers and/or to heavy-tailed distributions. This is the case of the median, the quantiles, the Huber estimator, etc. (Huber, 1964; Huber and Ronchetti, 2009). It is well known that under mild assumptions, $M$-estimators are both consistent and asymptotically normal, *i.e.*, suitably adapted versions of the LLN and the CLT apply to them (Collins, 1977; Portnoy, 1984; van der Vaart, 1998). Moreover, some versions of the LIL were also shown for $M$-estimators (Arcones, 1994; He and Wang, 1995). They suffer, however, from the same limitations as those explained above for the standard LIL. Our contributions allow to circumvent these limitations by providing a general non-asymptotic LIL for $M$-estimators both in one dimensional and in multivariate cases.

We apply the developed methodology to the problem of multi-armed bandits when the rewards are heavy-tailed or contaminated by outliers. In such a context, Altschuler et al. (2018) tackled the problem of best median arm identification; this corresponds to replacing the average regret by the median regret. The relevance of this approach relies on the fact that even a small number of contaminated samples obtained from each arm may make the corresponding means arbitrarily large. In that setup, would it be possible to improve the upper bounds on the sample complexity of their algorithm—similarly to Jamieson et al. (2014)—by using some finite-sample any-time version of the LIL for empirical medians or, more generally, for robust estimators? Our main results yield a positive answer to this question and establish rate-optimality of the proposed algorithm.

The rest of the paper is organized as follows. The next section contains the statement of the LIL in a univariate setting and provides some examples satisfying the required conditions. A multivariate version of the LIL for penalized $M$-estimators is presented in Section 3.3. An application to online learning is carried out in Section 4.2, while a summary of the main contributions and some future directions of research are outlined in Section 6. Detailed proofs are deferred to the supplementary material.

## 2 Uniform law of iterated logarithm for univariate $M$-estimators

In this section, we focus on the case of univariate $M$-estimators. This is a vast family that contains the sample mean, the sample median and many other estimators. The relevance of $M$-estimators in contaminated models has been highlighted by several studies (see Huber (1964); Maronna (1976) as well as the recent work Loh (2017) and references therein).

### 2.1 Assumptions and main result

The precise setting considered in this section is the following. Random variables $Y, Y_1, Y_2, Y_3, \ldots$ are independently drawn from a probability distribution $\mathbb{P}_Y$ on some space $\mathcal{Y}$. Let $\phi : \mathcal{Y} \times \Theta \to \mathbb{R}$ be a given loss function, where $\Theta$ is an open interval in $\mathbb{R}$. Throughout this work, we make the tacit assumption that the random variable $\phi(Y, \theta)$ has a finite expectation for all $\theta \in \Theta$. The population and the empirical risks are then defined, respectively, by the formulas

$$\Phi(\theta) = \mathbb{E}\left[\,\phi(Y,\theta)\,\right], \quad \widehat{\Phi}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \phi(Y_i,\theta),$$

where $n \geq 1$ is an integer. We denote by $\theta^*$ a minimizer of $\Phi$ on $\Theta$, and by $\widehat{\theta}_n$ a minimizer of $\widehat{\Phi}_n$ on $\Theta$.

**Assumption 2.1.** *The function $\phi(Y,\cdot)$ is convex $\mathbb{P}_Y$-almost surely and $\phi(Y,\theta) \to \infty$ as $\theta$ approaches the boundary of $\Theta$, $\mathbb{P}_Y$-almost surely (we say that the function $\phi(Y,\cdot)$ is convex and coercive).*

Assumption 2.1 requires from the loss $\phi$ to be approximately U-shaped and guarantees that $\theta^*$ and $\widehat{\theta}_n$ are

well defined. To show that $\widehat{\theta}_n$ converges fast enough (with high probability) to $\theta^*$, we will impose a local positive-curvature assumption on the population risk.

**Assumption 2.2.** *There exist two positive constants $r$ and $\alpha$ such that for all $\theta \in \Theta$ with $|\theta - \theta^*| \leq r$, $\Phi(\theta) \geq \Phi(\theta^*) + (\alpha/2)(\theta - \theta^*)^2$.*

It is worth emphasizing here that this "local positive-curvature" assumption needs to hold for the population risk only. Clearly, a sufficient condition for Assumption 2.2 to hold is that $\Phi$ is strongly convex in a neighborhood of $\theta^*$. Finally, to be able to obtain non-asymptotic guarantees that take the form of anytime Gaussian concentration, we require from the process $\theta \mapsto \phi(Y - \theta)$ to be smooth and to have sub-Gaussian tails.

**Assumption 2.3.** *There exists a positive constant $\sigma$ such that the random variables $\phi(Y, \theta) - \phi(Y, \theta^*)$ are $\sigma^2(\theta - \theta^*)^2$-sub-Gaussian[2] for all $\theta \in \Theta$.*

One checks that Assumption 2.3 is fulfilled if $\phi(Y, \cdot)$ is $\eta$-Lipschitz with a sub-Gaussian variable $\eta$. We stress that the function $\phi$ is not assumed differentiable and, more importantly, that $Y$ is not necessarily sub-Gaussian. We are now ready to state our first theorem on the uniform concentration of $M$-estimators.

**Theorem 1.** *Let Assumptions 2.1 to 2.3 hold. For any $\delta \in (0, 1)$, set*

$$t^{\mathrm{LIL}}_{n,\delta} := \frac{3.3\sigma}{\alpha}\sqrt{\frac{1.1 \ln\ln n + \ln(15/\delta) + 2.6}{n}}.$$

*Let $n_0 = n_0(\alpha, r, \delta)$ be the smallest integer $n \geq 12$ for which $t^{\mathrm{LIL}}_{n,\delta} \leq r$. Then,*

$$\mathbb{P}\left(\forall n \geq n_0, \quad |\widehat{\theta}_n - \theta^*| \leq t^{\mathrm{LIL}}_{n,\delta}\right) \geq 1 - \delta. \qquad (1)$$

While the complete proof of Theorem 1 is postponed to the supplementary material, let us make a quick comment. In our proof, we show that it is enough to establish any-time concentration inequalities for sums of sub-Gaussian random variables. For partial sums of a sequence of sub-Gaussian random variables, sharp any-time concentration inequalities were recently proved in Howard et al. (2018); Jamieson et al. (2014); Maillard (2019). However, these bounds do not apply in our case, since the terms in the sums arising in our proof change with the size of the sum. In other words, our sums are not partial sums of a given sequence of sub-Gaussian random variables.

[2]See, e.g., Koltchinskii (2011, Section 3.1) for a definition of centered sub-Gaussian random variables and their properties. We recall that a random variable is sub-Gaussian if its centered version is sub-Gaussian.

The setting described in the beginning of this section might seem disconnected from any application, since it builds on an infinite set of independent random variables. However, the validity of the bound for an infinity of values of the sample size $n$ makes it suitable for using in situations where the sample size is random and data-dependent. More precisely, the last theorem implies that for any $\delta \in (0, 1)$ and for any random variable $N$ taking values in the set of natural numbers $\mathbb{N}$, we have, with probability larger than $1 - \delta$,

$$|\widehat{\theta}_N - \theta^*| \leq \frac{3.3\sigma}{\alpha}\sqrt{\frac{1.1 \ln\ln N + \ln(15/\delta) + 2.6}{N}}$$

For instance, if we assume that the acquisition of each data point $y$ has a cost $\psi(y)$, the number $N$ might be given by $N = \max\{n : \psi(Y_1) + \ldots + \psi(Y_n) \leq B\}$, where $B$ is a given available budget.

## 2.2 Examples

We now present three common examples for which all the assumptions presented above are satisfied. In all these examples, $\mathcal{Y} = \Theta = \mathbb{R}$.

**Mean estimation** Let $\phi(x, \theta) = (x - \theta)^2$. Assume that $Y$ is $s^2$-sub-Gaussian. Then, one can check that Assumptions 2.1 to 2.3 are all satisfied with $r = +\infty$, $\alpha = 2$ and $\sigma = 2s$. For an in-depth analysis of this particular case we refer to (Howard et al., 2018).

**Median and quantile estimation** Let $\phi(x, \theta) = |x - \theta| - |x|$. Assume that $Y$ has a unique median $\theta^*$ and that its cumulative distribution function $F$ satisfies $|F(\theta) - 1/2| \geq (\alpha/2)|\theta - \theta^*|$, for all $\theta \in [\theta^* - r, \theta^* + r]$, where $\alpha, r > 0$ are fixed numbers. Then, $\theta^*$ is the unique minimizer of $\Phi$ and for all $\theta \in [\theta^* - r, \theta^* + r]$, the increment $\Phi(\theta) - \Phi(\theta^*)$ is equal to

$$2\int_{\theta^*}^{\theta} x \,\mathrm{d}F(x) - (\theta - \theta^*) + 2\big(\theta F(\theta) - \theta^* F(\theta^*)\big)$$
$$\overset{(a)}{=} 2\int_{\theta^*}^{\theta} F(x)\,\mathrm{d}x - (\theta - \theta^*),$$

where $(a)$ is obtained by integration by parts. Hence, $\Phi(\theta) - \Phi(\theta^*) = \int_{\theta^*}^{\theta}(2F(x) - 1)\,dx \geq (\alpha/2)(\theta - \theta^*)^2$, yielding Assumption 2.2. Moreover, since $\phi(Y, \theta)$ is bounded almost surely and 1-Lipschitz, for all $\theta \in \mathbb{R}$, Assumption 2.3 is automatically true (with $\sigma = 1$).

The same arguments hold true if $\phi(x, \theta) = \tau_\beta(x - \theta) - \tau_\beta(x)$, where $\tau_\beta(x) = \beta x - x_-$ with $x_- = \min(x, 0)$ the negative part. For this function $\phi$, $\theta^*$ is the $\beta$-quantile of $\mathbb{P}_Y$, for $\beta \in (0, 1)$.

**Huber's $M$-estimators** For $c > 0$, we define by $g_c(x) = x^2$ if $|x| \leq c$ and $g_c(x) = c(2|x| - c)$ if $|x| > c$.

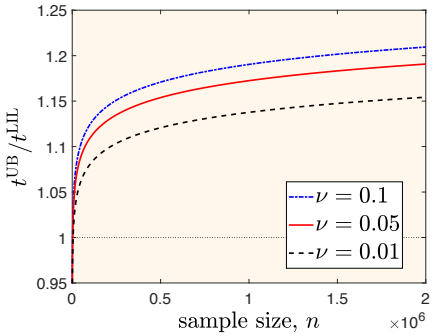


Figure 1: Ratio $t^{\mathrm{UB}}_{n,\delta'}/t^{\mathrm{LIL}}_{n,\delta}$ for different sample sizes $n$ and confidence levels $\nu$.

Let $\phi(x,\theta) = g_c(x-\theta) - g_c(x)$. This function $g_c$ being $2c$-Lipschitz, Assumption 2.3 is satisfied with $\sigma = 2c$. Assume that $Y$ has a positive density $f$ on $\mathbb{R}$. Then, it is easy to check that $\Phi$ is twice differentiable, with $\Phi''(\theta) = 2\,(\,F(\theta+c) - F(\theta-c)\,) > 0$, for all $\theta \in \mathbb{R}$, where $F$ is the cumulative distribution function of $Y$. Hence, $\theta^*$ is well-defined and unique, and if there exists $m > 0$ such that $f(x) \geq m$ for $x \in [\theta^*-2c, \theta^*+2c]$, then Assumption 2.2 is satisfied with $r = 2c$ and $\alpha = 4cm$.

### 2.3 Comparison with union bound

Let $Y_1, \ldots, Y_n$ be i.i.d. random variables and let $\phi : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be a loss such that assumptions of Theorem 1 are satisfied. Let $\widehat{\theta}_n$ be the $M$-estimator associated with the samples $Y_1, \ldots, Y_n$ and the loss $\phi$. Using the same trick we developed for the proof of Theorem 1 (see supplementary material) we obtain the following tail bound : $\forall n \geq 1$, $\mathbb{P}\big(|\widehat{\theta}_n - \theta^*| > \frac{2\sigma}{\alpha}\sqrt{2\ln(2/\delta)/n}\big) \leq \delta$. Setting

$$t^{\mathrm{UB}}_{n,\delta} := \frac{2\sigma}{\alpha}\sqrt{\frac{2\ln(2n^{1+\varepsilon}/\delta)}{n}},$$

the union bound leads to

$$\mathbb{P}\left(\forall\, n \geq 12 \quad |\widehat{\theta}_n - \theta^*| \leq t^{\mathrm{UB}}_{n,\delta}\right) \geq 1 - \sum_{n=12}^{\infty} \frac{\delta}{n^{1+\varepsilon}}. \quad (2)$$

Figure 1 shows the ratio of the sub-Gaussian upper bound $t^{\mathrm{UB}}_{n,\delta'}$ over the LIL upper bound $t^{\mathrm{LIL}}_{n,\delta}$ provided by Theorem 1 for different levels of global confidence. The parameters $\delta$ and $\delta'$ are chosen to guarantee that the right hand sides in both (1) and (2) are equal to the prescribed confidence level $1-\nu$. For $t^{\mathrm{UB}}_{n,\delta'}$, we chose $\varepsilon = 0.1$, the results for other values of $\varepsilon$ being very similar. We observe that for most sample sizes n, the LIL bound is tighter than the one obtained by the union bound. In addition, the gap between the bounds widens as the sample size grows.

## 3 Uniform LIL for $M$-estimators of a multivariate parameter

We consider here the multivariate analog of the previous problem. The goal is to predict a real-valued label using a $d$-dimensional feature.

### 3.1 Assumptions and main result

We are given $n$ independent label-feature pairs $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$, with labels $Y_i \in \mathbb{R}$ and features $\boldsymbol{X}_i \in \mathbb{R}^d$, drawn from a common probability distribution $\mathbb{P}$. Let $\phi_n : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be a given loss function and $\rho_n : \mathbb{R}^d \to \mathbb{R}$ be a given penalty. We assume throughout that the random variable $\phi_n(Y_1, \boldsymbol{\theta}^\top \boldsymbol{X}_1)$ has a finite expectation, for every $\boldsymbol{\theta}$, with respect to the probability distribution $\mathbb{P}$.

For a sample $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$, we define the penalized empirical and population risks

$$\widehat{\Phi}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \phi_n(Y_i, \boldsymbol{\theta}^\top \boldsymbol{X}_i) + \rho_n(\boldsymbol{\theta}),$$
$$\Phi_n(\boldsymbol{\theta}) = \mathbb{E}\big[\phi_n(Y_1, \boldsymbol{\theta}^\top \boldsymbol{X}_1)\big] + \rho_n(\boldsymbol{\theta}).$$

Note that both the loss function $\phi_n$ and the penalty $\rho_n$ are allowed to depend on the sample size $n$. Since our results are non-asymptotic, this dependence will be reflected in the constants appearing in the law of iterated logarithm stated below. We also define the penalized $M$-estimator $\widehat{\boldsymbol{\theta}}_n$ and its population counterpart $\boldsymbol{\theta}^*_n$ by

$$\widehat{\boldsymbol{\theta}}_n \in \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^d} \widehat{\Phi}_n(\boldsymbol{\theta}) \quad \text{and} \quad \boldsymbol{\theta}^*_n \in \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^d} \Phi_n(\boldsymbol{\theta}). \quad (3)$$

Typical examples where such a formalism is applicable are the maximum a posteriori approach and penalized empirical risk minimization. Our goal is to establish a tight non-asymptotic bound on the error of $\widehat{\boldsymbol{\theta}}_n$, that is, with high probability, valid for every $n \in \mathbb{N}$.

The main result of this section is valid under the assumptions listed below. We will present later on some common examples in which all these assumptions are satisfied.

**Assumption 3.1.** (Lipschitz loss) *The function $\theta \mapsto \phi_n(y,\theta)$ is $L_n$-Lipschitz, for every fixed $y \in \mathbb{R}$.*

**Assumption 3.2.** (Convex penalty) *The function $\theta \mapsto \widehat{\Phi}_n(\theta)$ is convex almost surely.*

**Assumption 3.3.** (Curvature of the population risk) *There exists a positive non-increasing sequence $(\alpha_n)$ such that, for any $n \in \mathbb{N}^*$, for any $\boldsymbol{w} \in \mathbb{R}^d$, $\Phi_n(\boldsymbol{\theta}^*_n + \boldsymbol{w}) - \Phi_n(\boldsymbol{\theta}^*_n) \geq (\alpha_n/2)\|\boldsymbol{w}\|_2^2$.*

**Assumption 3.4.** (Boundedness of features) *There exists a positive constant $B$ such that $\|\boldsymbol{X}_1\|_2 \leq B$ almost surely.*

We will use the notation $\kappa_n = L_n/\alpha_n$ and refer to this quantity as the condition number. Note that all the foregoing assumptions are common in statistical learning, see for instance Rakhlin et al. (2012); Sridharan et al. (2009). They are helpful not only for proving statistical guarantees but also for designing efficient computational methods for approximating $\widehat{\boldsymbol{\theta}}_n$.

**Theorem 2.** *Let Assumptions 3.1 to 3.4 be satisfied for every $n \in \mathbb{N}$. Assume, in addition, that starting from some integer $n_0 \geq 6$, the sequence $\kappa_n^2 \ln\ln n/n$ is decreasing. Define for any $\delta \in (0,1)$, $n \geq n_0$,*

$$t_{n,\delta}^{\mathrm{MVLIL}} = 3.6\kappa_n B \frac{\sqrt{\ln\ln n + \ln(50/\delta)} + 1}{\sqrt{n}}.$$

*Then, for any $q \geq 2$ and $\delta \in (0,1)$, it holds that*

$$\mathbb{P}\left(\forall n \geq n_0, \quad \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^*\|_q \leq t_{n,\delta}^{\mathrm{MVLIL}}\right) \geq 1-\delta.$$

### 3.2 Discussion

As an immediate consequence of Theorem 2 we get the following result. Let $N$ be a randomly chosen integer that can depend on the infinite sequence $\{(\boldsymbol{X}_i, Y_i), i \in \mathbb{N}^*\}$ of random feature-label pairs drawn from $\mathbb{P}$. We observe only the first $N$ elements of this sequence and wish to make a prediction of the label $Y$ at a point $\boldsymbol{x} \in \mathbb{R}^d$. Assume that the best linear prediction is of the form $g(\boldsymbol{x}^\top \boldsymbol{\theta}^*)$, where $\boldsymbol{\theta}^*$ is the minimizer of the expected loss and $g$ is a known, $L$-Lipschitz, link function. Then, we can predict the label at $\boldsymbol{x}$ by $g(\boldsymbol{x}^\top \widehat{\boldsymbol{\theta}}_N)$, where $\widehat{\boldsymbol{\theta}}_N$ is the empirical risk minimizer. According to the last theorem, this predicted value satisfies

$$|g(\boldsymbol{x}^\top \widehat{\boldsymbol{\theta}}_N) - g(\boldsymbol{x}^\top \boldsymbol{\theta})| \leq L\|\boldsymbol{x}\|_2 \, t_{N,\delta}^{\mathrm{MVLIL}},$$

with probability at least $1-\delta$.

A bound for the case $q \in [1,2)$ can be obtained by using the fact that the $\ell_q$ norm is upper bounded by $d^{(2-q)/(2q)}$ times the $\ell_2$ norm (Hölder's inequality). The resulting bound corresponds to the $\ell_2$ bound multiplied by this factor. Note this dependence on $d$ is optimal, even in batch setting.

As noted above, all the foregoing assumptions are common in statistical learning. For instance, if $\rho_n(\boldsymbol{\theta}) = \lambda_n\|\boldsymbol{\theta}\|_2^2$ is the ridge penalty (Hoerl and Kennard, 2000) and $\phi_n$ is either the absolute deviation ($\phi_{abs}(y,y') = |y-y'|$, see for instance Wang et al. (2014)), the hinge ($\phi_{abs}(y,y') = (1-yy')_+$ with $y \in [-1,1]$) or the logistic ($\phi_{log}(y,y') = \ln(1+e^{-yy'})$ with $y \in [-1,1]$) loss, the aforementioned assumptions are satisfied with $L_n = 1$ and $\alpha_n = \lambda_n$. One can also consider the usual squared loss $\phi(y,y') = (y-y')^2$ under the additional assumption that $Y$ is bounded by a known constant $B_y$. Under this condition, if the minimization problems in (3) are constrained to the ball of radius $R$, Assumptions 3.1 and 3.3 are satisfied with $\alpha_n = 1$ and $L_n = 2B_y + BR$. It should be noted that Assumption 3.3 is satisfied, for instance, when $\Phi_n$ is strongly convex. Remarkably, as opposed to some other papers (Hsu and Sabato, 2016; Shalev-Shwartz et al., 2010), Theorem 2 requires this assumption for the population risk only.

Assumption 3.4 can be replaced, with some extra work, by sub-Gaussianity of $\|\boldsymbol{X}\|_2$. The statement of this extension and its proof can be found in Section 7.3.1.

### 3.3 Possible extensions

The conditions under which Theorem 2 holds can be further relaxed. We have in mind the following two extensions. First, the curvature condition can be restricted to a neighborhood of $\boldsymbol{\theta}_n^*$ only, by letting $\Phi_n$ grow linearly outside the neighborhood. Second, the Lipschitz assumption on $\phi_n$ can be replaced by the following one: for a constant $\beta$ and a sub-Gaussian random variable $\eta$, the function $u \mapsto \phi_n(Y,u) - \beta u^2$ is $\eta$-Lipschitz. This last extension will allow us to cover the case of squared loss without restriction to a bounded domain. All these extensions are fairly easy to implement, but they significantly increase the complexity of the statement of the theorem. In this work, we opted for sacrificing generality in order to get better readability of the result.

Another interesting avenue for future research is the extension of the presented results to the high-dimensional online setting, *i.e.*, when the dimension might be larger than the sample size. In the batch setting, an in-depth analysis of $M$-estimators can be found in Negahban et al. (2012). It is also important in such a high-dimensional setting to avoid the factor $B$ in the expression of $t_{n,\delta}^{\mathrm{MVLIL}}$, since it might scale as $\sqrt{d}$.

Finally, we can consider a more general setting in which the terms $\phi(Y_i, \boldsymbol{\theta}^\top \boldsymbol{X}_i)$ are replaced by $\psi(Z_i, \theta)$, where $Z_i$ are i.i.d. random variables. The only change to be made is in replacing Assumptions 3.1 and 3.4 by a new assumption, that requires the function $[\psi(Z_i, \boldsymbol{\theta}) - \psi(Z_i, \boldsymbol{\theta}')]$ to be bounded by $|\boldsymbol{V}_i^\top(\boldsymbol{\theta} - \boldsymbol{\theta}')|$, for all $\boldsymbol{\theta}$, $\boldsymbol{\theta}'$, with a random vector $\boldsymbol{V}_i$ which has a bounded (or sub-Gaussian) norm. This setting has the advantage of being more general than the one adopted in Section 3. However, the relevant examples we have in mind at correspond all to partial linear models.

## 4 Application to Bandits

In this section, we apply the univariate uniform law of iterated logarithm established in Section 2 to the multi-armed bandit problem. More precisely, we study

the Best Arm Identification (BAI) problem in the fixed confidence setting. It consists in identifying, for a given confidence level and as fast as possible, which arm produces the highest expected outcome, (see Audibert et al. (2010); Gabillon et al. (2012); Kaufmann et al. (2016)). This means that we are able to collect data by sampling from $K$ unknown distributions $\mathbb{P}_1, \dots, \mathbb{P}_K$ and our goal is to identify the distribution having the largest expectation. Naturally, the same problem can be formulated for finding the distribution with the largest median, or the largest quantile of a given order. In particular, such a formulation of the problem might be of interest in cases where the expectations of the outcomes of each arm may not be defined (rewards are heavy-tailed) or are not meaningful (rewards are subject to some arbitrary contamination), see Altschuler et al. (2018). We show in this section that theoretical results of previous sections provide an extension of the lil'UCB algorithm of Jamieson et al. (2014) to this framework.

### 4.1 Robust Best Arm Identification (RBAI)

We consider a robust version of BAI, which we call Robust BAI (RBAI). Suppose there are $K$ arms, each arm $k \in [K]$ producing i.i.d. rewards

$$Y_{1,k}, Y_{2,k}, Y_{3,k}, \dots \overset{\text{iid}}{\sim} \mathbb{P}_k.$$

At each round $n = 1, 2, \dots$, the player chooses an arm $I_n \in [K]$ and receives the corresponding reward $Y_{T_{I_n}(n-1),I_n}$, where $T_k(n-1) = \mathbb{1}(I_1 = k) + \dots + \mathbb{1}(I_{n-1} = k)$ is the number of times the arm $k$ was pulled during the rounds $1, \dots, n-1$.

For a given loss function $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, convex with respect to its second argument, we define

$$\theta_k \in \arg\min_{\theta \in \mathbb{R}} \mathbb{E}_{\mathbb{P}_k}[\phi(Y, \theta)].$$

From a statistical perspective, the problem under consideration encompasses that of finding the maximum point (by active learning) in a quantile regression problem (Chernozhukov, 2005). For instance, consider the case of median regression. The aim is to maximize a function $f : [0,1] \to \mathbb{R}$ over a grid of points $x_1, \dots, x_K \in [0,1]$, using noisy evaluations of $f$. At each round $n$, we can choose one $x_k$ and observe the value

$$Y_n = f(x_k) + \xi_n,$$

where $\{\xi_n\}$ is a sequence of i.i.d. random variables with median equal to zero. Clearly, this enters into the framework described in the previous paragraph with $\theta_k = f(x_k)$ and each $\mathbb{P}_k$ is just a shifted-by-$\theta_k$ version of the distribution of $\xi_n$.

We use the rewards of the $k$-th arm for estimating $\theta_k$ by empirical risk minimisation: for every arm $k \in [K]$ and every sample size $n \geq 1$, we let

$$\widehat{\theta}_{k,n} \in \arg\min_{\theta \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} \phi(Y_{i,k}, \theta).$$

With this notation, after $n$ rounds, we are able to compute the quantities $\widehat{\theta}_{k,T_k(n)}$ for $k \in [K]$. These quantities, combined with the confidence bounds furnished by the LIL of Theorem 1, lead to $M$-estimator lil'UCB algorithm described in Algorithm 1[3].

**Algorithm 1** M-estimator lil'UCB.

**Input** $\nu$, $\lambda$, $\gamma > 0$ and $n_0 \in \mathbb{N}$

Sample each arm $n_0$ times
Set $\delta = ((\sqrt{16\nu + 9} - 3)/16)^2$
**for** $k$ in $1 : K$ **do**
    $T_k \leftarrow n_0$
    Sample $k$th arm $n_0$ times
    Compute $\widehat{\theta}_{k,T_k}$
    Set $s(k) \leftarrow \widehat{\theta}_{k,T_k} + \gamma\sqrt{\frac{\ln \ln T_k + \ln(1/\delta)}{T_k}}$
**end for**
$n \leftarrow Kn_0$
**while** $(1+\lambda)\max_{k\in[K]} T_k < 1 + \lambda n$ **do**
    $I \leftarrow \arg\max_{k\in[K]} s(k)$
    Sample arm $I$
    Update $T_I \leftarrow T_I + 1$, $n \leftarrow n + 1$
    Compute $\widehat{\theta}_{I,T_I}$
    Set $s(I) \leftarrow \widehat{\theta}_{I,T_I} + \gamma\sqrt{\frac{\ln \ln T_I + \ln(1/\delta)}{T_I}}$
**end while**

**Output** $\arg\max_{k\in[K]} T_k$

### 4.2 Main results

To state the theoretical results, let $k^* = \operatorname{argmax}_{k\in[K]} \theta_k$ be the subscript corresponding to the best arm. We assume $k^*$ to be unique, and we define, for $k \neq k^*$, the sub-optimality gaps $\Delta_k = \theta_{k^*} - \theta_k$. We introduce the quantities

$$\mathbf{H}_1 = \sum_{k \neq k^*} \frac{1}{\Delta_k^2}, \quad \mathbf{H}_2 = \sum_{k \neq k^*} \frac{\ln\left(2 + \ln_+(1/\Delta_k^2)\right)}{\Delta_k^2}.$$

Those quantities play a key role in characterizing the complexity of the BAI problem.

**Theorem 3.** *Let $\theta \mapsto \phi(y, \theta)$ be a convex function for every $y \in \mathbb{R}$ and let the distributions $\mathbb{P}_k$ satisfy Assumptions 2.2 and 2.3 with parameters $\alpha, \sigma > 0$. For any $\nu \in (0, 0.2)$ and $\beta \in (0, 4.8)$, there exist positive constants[4] $\lambda$ and $C$ such that with probability at least*

[3] $\lambda$, $\gamma$ and $n_0$ should be seen as tuning parameters for which our theoretical results give some guidance.

[4] $\lambda$ and $C$ depend only on $\beta$ and $\sigma/\alpha$.

$1-\nu$, *Algorithm 1, used with parameters* $\nu$, $\lambda$, $\gamma = 4.4(1+\beta)\sigma/\alpha$ *and* $n_0 \geq 12$, *stops after at most*

$$Kn_0 + C\big(\mathbf{H}_1 \ln(1/\nu) + \mathbf{H}_2\big)$$

*steps and returns the best arm.*

Note that $(\ln 2)\mathbf{H}_1 \leq \mathbf{H}_2$. Therefore, for a fixed confidence level $\nu$, the number of pulls provided by Theorem 3 is $O(\mathbf{H}_2)$. The next result shows that this order of magnitude is optimal.

**Theorem 4.** *Consider the RBAI framework with fixed confidence* $\delta \in (0, 1/2)$ *described above and assume* $K = 2$. *Let* $\theta_1, \theta_2 \in \mathbb{R}$ *be such that* $\Delta = |\theta_1 - \theta_2| > 0$. *Let* $\phi(y, \theta) = \phi_0(|y - \theta|)$ *for some function* $\phi_0$ *and the arm distributions be* $\mathcal{N}(\theta_1, 1)$ *and* $\mathcal{N}(\theta_2, 1)$. *Then, any algorithm that finds after* $T$ *rounds the best arm with probability at least* $1-\delta$, *for all values of* $\Delta > 0$, *must satisfy*

$$\limsup_{\Delta \to 0} \frac{\mathbb{E}[T]}{\Delta^{-2} \ln \ln(\Delta^{-2})} \geq 2 - 4\delta.$$

Proofs of these two theorems are provided in the supplementary material.

## 5 Numerical experiments

To illustrate the results of the previous section, we conducted the following experiment. We chose the values of $\theta_k$'s according to the "$\alpha$-model" from Jamieson et al. (2014) with $\alpha = 0.3$. It imposes an exponential decay on the parameters, that is $\theta_k = 1-(k/K)^\alpha$. Along with these parameters, we consider three reward generating processes:

- *Gaussian rewards*, where $Y_{i,k} \overset{\text{iid}}{\sim} \mathcal{N}(\theta_k, \sigma^2)$,
- *Gaussian rewards subject to Cauchy contamination*, where $Y_{i,k} \overset{\text{iid}}{\sim} (1-\varepsilon)\mathcal{N}(\theta_k, \sigma^2) + \varepsilon\,\text{Cauchy}(\theta_k)$ for $\varepsilon = 5\%$,
- *Student rewards*, where $Y_{i,k} \overset{\text{iid}}{\sim} t_2(\theta_k)$ (*i.e.*, Student distribution with 2 degrees of freedom).

Note that all these processes are median centered at $\theta_k$'s. In the case of Gaussian and Student rewards, they are also mean-centered at $\theta_k$, while in the case of contaminated Gaussian rewards the mean is not defined. To test the robustness of the compared algorithms, we tuned their parameters to fit the Gaussian reward scenario.

In this set-up, we compared the original lil'UCB algorithm from Jamieson et al. (2014)—see also Jamieson and Nowak (2014) for a more comprehensive experimental evaluation—and the $M$-estimator lil'UCB described in Algorithm 1, where $\widehat{\theta}_{k,n}$ is the empirical median of rewards from arm $k$ up to time $n$. This corresponds to the $M$-estimator associated with the absolute deviation loss. This version of the $M$-estimator lil'UCB is hereafter referred to as median lil'UCB or med-lil'UCB.

In order to conduct a fair comparison, we assigned the same values to parameters shared by both procedures and set the values as in Jamieson et al. (2014): $\beta = 1$, $\lambda = (1 + 2/\beta)^2$, $\sigma = 0.5$, $\varepsilon = 0.01$ and $\nu = 0.1$. Note that, as underlined in Jamieson et al. (2014), the choice of $\lambda$ does not fit their theoretical result. This choice is justified by the fact that $\lambda$ should theoretically be proportional to $(1+2/\beta)^2$ with a constant converging to 1 when the confidence approaches 0. For our algorithm we chose $\gamma = 2$ and $n_0 = 20$.

The results, for several values of $K$ (the total number of arms), obtained by 200 independent runs of each algorithm in all the three settings, are summarized in Figure 2 and in Table 1. Numbers reported in Table 1 represent the proportion of times each algorithm succeeded to find the best arm, while Figure 2 displays the number of pulls for each algorithm. Table 1 shows that lil'UCB performed poorly on the non-Gaussian models. For contaminated Gaussian rewards, the performance of lil'UCB deteriorates as the number of arms grows, while it does not seem to be affected by the number of arms in the case of Student rewards : it identifies correctly the best arm for only around 60% of the runs in this last case. In contrast, median lil'UCB performs well in all the three scenarios, giving perfect identification over all runs.

Table 1: Proportion of correct best arm identification (over 200 runs per scenario/algorithm).

| $K$ | Algorithm | Gauss | Contam. | Student |
|---|---|---|---|---|
| 2 | lil'UCB | 1.00 | 0.81 | 0.61 |
| | med-lil'UCB | 1.00 | 1.00 | 1.00 |
| 4 | lil'UCB | 1.00 | 0.75 | 0.61 |
| | med-lil'UCB | 1.00 | 1.00 | 1.00 |
| 8 | lil'UCB | 1.00 | 0.69 | 0.63 |
| | med-lil'UCB | 1.00 | 1.00 | 1.00 |
| 16 | lil'UCB | 1.00 | 0.66 | 0.60 |
| | med-lil'UCB | 1.00 | 1.00 | 1.00 |
| 32 | lil'UCB | 1.00 | 0.57 | 0.61 |
| | med-lil'UCB | 1.00 | 1.00 | 1.00 |
| 64 | lil'UCB | 1.00 | 0.54 | 0.62 |
| | med-lil'UCB | 1.00 | 1.00 | 1.00 |
| 128 | lil'UCB | 1.00 | 0.44 | 0.60 |
| | med-lil'UCB | 1.00 | 1.00 | 1.00 |

The curves in Figure 2 represent the median number of pulls over the 200 runs while the colored areas around the curves are delimited by the 10% and 90% quantiles

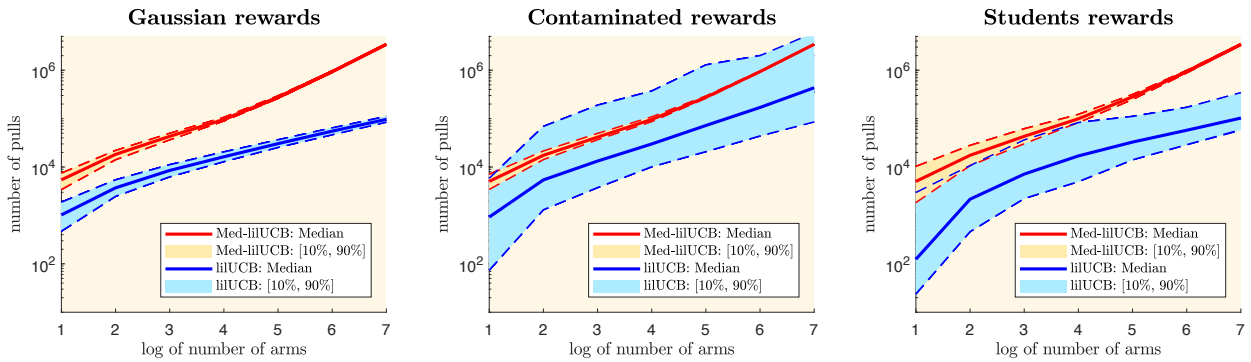


Figure 2: Total number of pulls done by the median lil'UCB and the lil'UCB algorithms for $K \in \{2, 4, 8, 16, 32, 64, 128\}$.

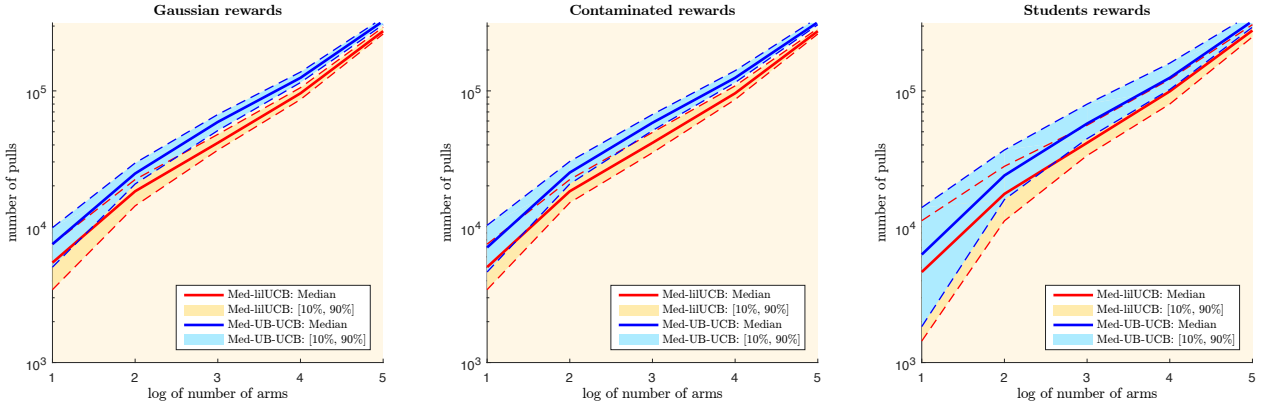


Figure 3: Total number of pulls done by the median lil'UCB and the median UCB based on the union bound.

of the number of pulls over these 200 runs. We observe that the spread of the number of pulls of lil'UCB is large for non-Gaussian models, while the curves for median lil'UCB are almost identical in the three models of rewards. The number of pulls for median lil'UCB is higher than the number of pulls for lil'UCB in the Gaussian and Student models. However, in the contaminated Gaussian model, lil'UCB might require more pulls when the number of arms is large.

Moreover, we noticed that the performance of our procedure is not sensitive to the level of contamination : we conducted the same experiment with $\varepsilon \in \{5, 10, 20, 40, 60\}$ and in all cases our procedure is 100% successful in finding the best arm. Furthermore, the number of pulls does not increase when $\varepsilon$ increases. In contrast, the performance of the original lil'UCB procedure drops down to 35% of correct identification when $\varepsilon = 60\%$ and there are 4 arms. Finally, we observed that if we replace the LIL by the naive union bound in our algorithm, the detection accuracy remains the same, but the running time increases (between 10% and 30%), see Figure 3.

These experiments illustrate the lack of robustness of lil'UCB to heavy tailed rewards and the effective robustness of median lil'UCB. Since this robustness comes with a higher number of pulls, median lil'UCB should be preferred to vanilla lil'UCB only if one suspects non-Gaussian or heavy-tailed rewards.

## 6 Conclusion and further work

We have proved a nonasymptotic law of iterated logarithm for general $M$-estimators both in univariate and in multivariate settings. These results can be seen as off-the-shelf deviation bounds that are uniform in the sample size and, therefore, suitable for online learning problems and problems in which the sample size may depend on the observations. There are several avenues for future work. For simplicity, in the multivariate case, the population risk was assumed to be above an elliptic paraboloid on the whole space. First in our agenda is to replace this condition by a local curvature one. A second interesting line of research is to establish an any-time deviation bound for sequential estimators such as the online gradient descent. It would also be of interest to obtain "in-expectation" bounds of the same type as those in (Shin et al., 2019).

## References


J. Altschuler, V.-E. Brunel, and A. Malek. Best Arm Identification for Contaminated Bandits. *arXiv e-prints*, art. arXiv:1802.09514, Feb. 2018.

M. A. Arcones. Some strong limit theorems for m-estimators. *Stochastic Processes and Their Applications*, 53(2):241–268, 1994.

J. Audibert, S. Bubeck, and R. Munos. Best arm identification in multi-armed bandits. In *COLT 2010*, pages 41–53, 2010.

S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

V. Chernozhukov. Extremal quantile regression. *Ann. Statist.*, 33(2):806–839, 2005.

J. R. Collins. Upper bounds on asymptotic variances of $M$-estimators of location. *Ann. Statist.*, 5(4):646–657, 1977.

R. H. Farrell. Asymptotic behavior of expected sample size in certain one sided tests. *The Annals of Mathematical Statistics*, pages 36–72, 1964.

V. Gabillon, M. Ghavamzadeh, and A. Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In *Advances in NeurIPS 25*, pages 3221–3229, 2012.

X. He and G. Wang. Law of the iterated logarithm and invariance principle for m-estimators. *Proceedings of the American Mathematical Society*, 123(2):563–573, 1995.

A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 42(1):80–86, 2000.

S. R. Howard, A. Ramdas, J. McAuliffe, and J. Sekhon. Uniform, nonparametric, non-asymptotic confidence sequences. *arXiv preprint arXiv:1810.08240*, 2018.

D. Hsu and S. Sabato. Loss minimization and parameter estimation with heavy tails. *Journal of Machine Learning Research*, 17(18):1–40, 2016.

P. J. Huber. Robust estimation of a location parameter. *The annals of mathematical statistics*, 35(1):73–101, 1964.

P. J. Huber and E. M. Ronchetti. *Robust statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition, 2009.

K. Jamieson, M. Malloy, R. Nowak, and S. Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439, 2014.

K. G. Jamieson and R. D. Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *48th Annual Conference on Information Sciences and Systems, CISS*, pages 1–6, 2014.

E. Kaufmann and W. M. Koolen. Mixture martingales revisited with applications to sequential tests and confidence intervals. *CoRR*, abs/1811.11419, 2018.

E. Kaufmann, O. Cappé, and A. Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17:1:1–1:42, 2016.

A. Khintchine. Über einen satz der wahrscheinlichkeitsrechnung. *Fundamenta Mathematicae*, 6(1):9–20, 1924.

A. Kolmogoroff. Über das gesetz des iterierten logarithmus. *Mathematische Annalen*, 101:126–135, 1929.

V. Koltchinskii. *Oracle inequalities in empirical risk minimization and sparse recovery problems*, volume 2033 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011.

G. Lecué and P. Rigollet. Optimal learning with $Q$-aggregation. *Ann. Statist.*, 42(1):211–224, 02 2014.

P.-L. Loh. Statistical consistency and asymptotic normality for high-dimensional robust $M$-estimators. *Ann. Statist.*, 45(2):866–896, 04 2017.

O.-A. Maillard. Sequential change-point detection: Laplace concentration of scan statistics and non-asymptotic delay bounds. In *Proceedings of ALT*, volume 98 of *PMLR*, pages 610–632, 2019.

R. A. Maronna. Robust m-estimators of multivariate location and scatter. *The Annals of Statistics*, 4(1): 51–67, 1976.

S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu. A unified framework for high-dimensional analysis of $m$-estimators with decomposable regularizers. *Statist. Sci.*, 27(4):538–557, 11 2012.

S. Portnoy. Asymptotic behavior of $M$-estimators of $p$ regression parameters when $p^2/n$ is large. I. Consistency. *Ann. Statist.*, 12(4):1298–1309, 1984.

A. Rakhlin, O. Shamir, and K. Sridharan. Making gradient descent optimal for strongly convex stochastic optimization. In *ICML 2012*. icml.cc / Omnipress, 2012.

S. Shalev-Shwartz, N. Srebro, and T. Zhang. Trading accuracy for sparsity in optimization problems with sparsity constraints. *SIAM Journal on Optimization*, 20(6):2807–2832, 2010.

J. Shin, A. Ramdas, and A. Rinaldo. On the bias, risk and consistency of sample means in multi-armed bandits. *CoRR*, abs/1902.00746, 2019. URL http://arxiv.org/abs/1902.00746.

K. Sridharan, S. Shalev-shwartz, and N. Srebro. Fast rates for regularized objectives. In *Advances in NeurIPS 21*, pages 1545–1552. 2009.

A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

J. Wang, P. Wonka, and J. Ye. Scaling SVM and least absolute deviations via exact data reduction. In *ICML 2014*, volume 32 of *JMLR W.& C.P.*, pages 523–531, 2014.