STAT 593
New trends and conclusion

Joseph Salmon
http://josephsalmon.eu

Télécom Paristech, Institut Mines-Télécom
&
University of Washington, Department of Statistics
(Visiting Assistant Professor)

Outline

Other robustness applications

Non-asymptotic results / concentration

Table of Contents

Points non covered during the course

- Robust Covariance estimation / Robust PCA
  (see Maronna *et al.* (2006), Ch. 6)
- Robust GLM (see Maronna *et al.* (2006), Ch. 7)
- Robust Time Series (see Maronna *et al.* (2006), Ch. 8)

Table of Contents

Motivation: non-asymptotic guarantees

Hoeffding Inequality[1]:

Let $X_1, \dots, X_n$ be independent random variables such that for each $i \in [n], X_i \in [a_i, b_i]$ a.s, then for $S_n = \sum_{i=1}^n (X_i - \mathbb{E}(X_i))$ and for any $t > 0$

$$\mathbb{P}(|S_n| \geq t) \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

Consequence: for *i.i.d.* $X_i \in [a, b]$ with $V = \frac{(b-a)^2}{2}$ and any $\delta \in]0, \frac{1}{2}[$

$$|\bar{X}_n - \mathbb{E}(X)| \leq \sqrt{\frac{B \log(\delta^{-1})}{n}}, \text{with probability at least } 1 - 2\delta$$

(often referred to as **Sub-Gaussian** concentration)

[1] W. Hoeffding. "Probability inequalities for sums of bounded random variables". In: *J. Amer. Statist. Assoc.* 58.301 (1963), pp. 13–30.

Proof:

Let X be a random variable such that $X \in [a, b]$ a.s. Defined $\varphi(s) = \log \mathbb{E}(\exp^{sX})$.

1. Show that $\varphi'(s)$ and $\varphi''(s)$ are the first and second moments of the random variable X under the distribution $\frac{\exp(sx)d\mathbb{P}(x)}{\int \exp(sx')d\mathbb{P}(x')}$
2. Show that for any r.v. Z such that $Z \in [a, b]$ a.s. then:
$$\mathrm{var}(Z) \leq \frac{(b-a)^2}{4}$$
3. Apply Taylor expansion with integral reminder
$$f(x) = f(a) + f'(a)(x-a) + \int_a^x f''(t)(x-t)\,\mathrm{d}t$$

Heavy Tail

Definition

A random variable X is said to be heavy-tailed at $+\infty$ (resp. $-\infty$) if for any $\lambda > 0$,

$$\lim_{t \to +\infty} \exp(\lambda t)\mathbb{P}(X > t) = +\infty \,\text{(reps. } \lim_{t \to +\infty} \exp(\lambda t)\mathbb{P}(-X > t))$$

Robustness (for some people): obtaining a **Sub-Gaussian** concentration around the expectation, for heavy tail data

Attempts: Trimming, Truncation, Flattened M-estimation

Example : consider Huber's estimator for location using M-estimation theory with the truncation function $\psi_B : x \to x\mathbb{1}_{\{|x| \leq B\}} + B\mathbb{1}_{\{|x| \geq B\}}$

Limits:

- the concentration is biased due to the shrinkage part: Using Hoeffding Inequality provides concentration around $\mathbb{E}(\psi_B(X))$ not around the true expectation of the distribution.
- difficulty of setting B in practice...

M-estimation through soft truncation[2]

Requirement: $\hat{\mu}_\alpha$ satisfies $\frac{1}{n\alpha}\sum_{i=1}^{n}\phi(\alpha[X_i]-\hat{\mu}_\alpha)=0$ where

$$-\log\left(1-x+\frac{x^2}{2}\right)\leq\phi(x)\leq\log\left(1-x+\frac{x^2}{2}\right)\,,\quad\forall x\in\mathbb{R}$$

Theorem

Let $X, X_1, \ldots, X_n$ be *i.i.d.* real-valued random variables. Let V be such that $\mathrm{Var}(X)\leq V<+\infty$. Let $\delta\in]0,1[$ and $\alpha=\sqrt{\frac{2\log(\delta^{-1})}{nV}}$. Suppose $n\geq 4\log(\delta^{-1})$. Then, with probability at least $1-2\delta$

$$|\hat{\mu}_\alpha-\mathbb{E}(X)|\leq 2\sqrt{2}\sqrt{\frac{V\log(\delta^{-1})}{n}}$$

Rem: α depends on δ

[2] O. Catoni. "Challenging the empirical mean and empirical variance: a deviation study". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 48. 4. Institut Henri Poincaré. 2012, pp. 1148–1185.

Medians of mean

Originated in the 70's[3]

Let $X, X_1, \dots, X_n$ be *i.i.d.* real-valued random variables and $\mathrm{Var}(X) < +\infty$.

Assume that $B_1, \dots, B_{n_b}$ are n_b groups forming a partition of $[n]$.

We call it a regular partition if

$$\left| |B_k| - \frac{n}{n_b} \right| \leq 1$$

And for a group B_k we write

$$\mu_{B_k} = \frac{1}{|B_k|} \sum_{i \in B_k} X_i$$

The **Medians of Means** (MOM) estimator is defined by

$$\mu_{\mathrm{MOM}, n_b} = \mathrm{Med}(\mu_{B_1}, \dots, \mu_{B_{n_b}})$$

[3] A. S. Nemirovski and D. B. Yudin. *Problem complexity and method efficiency in optimization*. A Wiley-Interscience Publication. New York: John Wiley & Sons Inc., 1983, pp. xv+388.

Theory for MOM[4]

Theorem

Let $X, X_1, \ldots, X_n$ be *i.i.d.* real-valued random variables with $\mathrm{Var}(X) < +\infty$. Let $\delta \in]0, 1/2[$ and let n_b be an integer such that $\log(\delta^{-1}) \leq n_b \leq n/2$ and let the $[n]$ be partitioned by n_b groups (regular). Then, with probability at least $1 - 2\delta$

$$|\hat{\mu}_{MOM,n_b} - \mathbb{E}(X)| \leq 2\sqrt{6e}\sqrt{\frac{\mathrm{Var}(X)n_b}{n}}$$

Rem: n_b depends on δ to get the best bound: $n_b \approx \log(\delta^{-1})$

[4] M. Lerasle and R. I. Oliveira. "Robust empirical mean estimators". In: *arXiv preprint arXiv:1112.3914* (2011).

Extensions towards high-dimensional spaces

For $d \geq 1$, let $X, X_1, \ldots, X_n \in \mathbb{R}^p$ be *i.i.d.* random variables.
Assume that $B_1, \ldots, B_{n_b}$ are n_b groups forming a regular partition of $[n]$ and that for a group B_k we write

$$\mu_{B_k} = \frac{1}{|B_k|} \sum_{i \in B_k} X_i$$

The Medians of Means (MOM) estimator is defined by

$$\mu_{\mathrm{MOM}, n_b} = \mathrm{Med}_{n_b}(\mu_{B_1}, \ldots, \mu_{B_{n_b}})$$

where we remind the **(Geometric) Median** definition:

$$\mathrm{Med}_{n_b}(\mu_{B_1}, \ldots, \mu_{B_{n_b}}) \in \arg\min_{\mu \in \mathbb{R}^d} \sum_{i=1}^{n_b} \|\mu - \mu_{B_i}\|$$

Results[5]

Theorem

Let $\alpha \in]0, 1/2[$. Let $X_1, \ldots, X_n \in \mathbb{R}^p$ be *i.i.d.* random variables taking values in $\mathbb{R}^p$. Let $\delta \in]0, 1[$, let n_b be an integer such that $\log(\delta^{-1}) \leq n_b \leq \frac{n}{2}$ and let $B = (B_1, \ldots, B_{n_b})$ be a regular partition. Then, with probability at least $1 - \delta$,

$$\|\mu_{\mathrm{MOM}, n_b} - \mathbb{E}(X)\| \leq K_\alpha \sqrt{\frac{\mathbb{E}[\|X - \mathbb{E}(X)\|^2] n_b}{n}}$$

where K_α is a constant depending on α

[5] S. Minsker. "Geometric median and robust estimation in Banach spaces". In: *Bernoulli* 21.4 (2015), pp. 2308–2335.

Elements of proof

A crucial Lemma states the following:
"if a given point z is "far" from the geometric median, then it is also "far" from a constant fraction of the points"

Lemma

Let $X_1, \ldots, X_n \in \mathbb{R}^p$ and let $\hat{\mu}$ there geometric median. Fix $\alpha \in]0, \frac{1}{2}[$ and assume that $z \in \mathbb{R}^p$ is such that $\|\hat{\mu} - z\| > C_\alpha r$, where $r > 0$ and

$$C_\alpha = (1 - \alpha)\sqrt{\frac{1}{1 - 2\alpha}}$$

Then there exists a subset $J \subset [n]$ of cardinality $|J| > \alpha n$ such that for all $j \in J$, $\|X_j - z\| > r$.

Additional elements on MOM

- see E. Joly's PhD thesis
- More recent machine learning point of view on MOM : Lecué and Lerasle (2017)

References I

- Catoni, O. "Challenging the empirical mean and empirical variance: a deviation study". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 48. 4. Institut Henri Poincaré. 2012, pp. 1148–1185.
- Hoeffding, W. "Probability inequalities for sums of bounded random variables". In: *J. Amer. Statist. Assoc.* 58.301 (1963), pp. 13–30.
- Joly, E. "Robust estimation of heavy-tailed distributions". PhD thesis. Université Paris-Sud, 2015.
- Lecué, G. and M. Lerasle. "Robust machine learning by median-of-means: theory and practice". In: *arXiv preprint arXiv:1711.10306* (2017).
- Lerasle, M. and R. I. Oliveira. "Robust empirical mean estimators". In: *arXiv preprint arXiv:1112.3914* (2011).
- Maronna, R. A., R. D. Martin, and V. J. Yohai. *Robust statistics: Theory and methods*. Chichester: John Wiley & Sons, 2006.

References II

- Minsker, S. “Geometric median and robust estimation in Banach spaces”. In: *Bernoulli* 21.4 (2015), pp. 2308–2335.
- Nemirovski, A. S. and D. B. Yudin. *Problem complexity and method efficiency in optimization*. A Wiley-Interscience Publication. New York: John Wiley & Sons Inc., 1983, pp. xv+388.