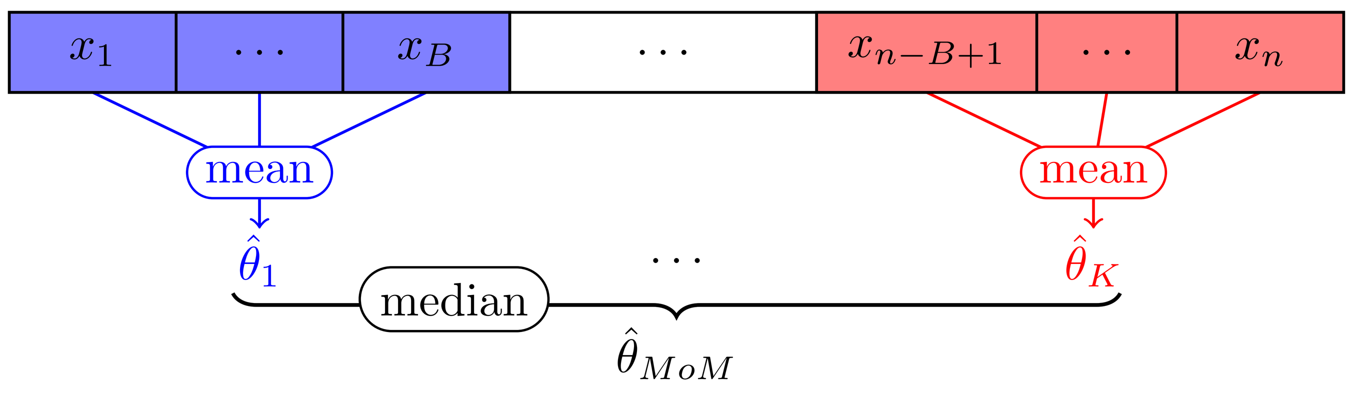
On Medians of (Randomized) Pairwise Means

P. Laforgue⋆, S. Clémençon⋆, P. Bertail‡

⋆ LTCI, Télécom Paris, Institut Polytechnique de Paris, 75013, Paris, France

‡ Modal'X, Université Paris-Nanterre, 92001, Nanterre Cedex, France

Median of Means (MoM)

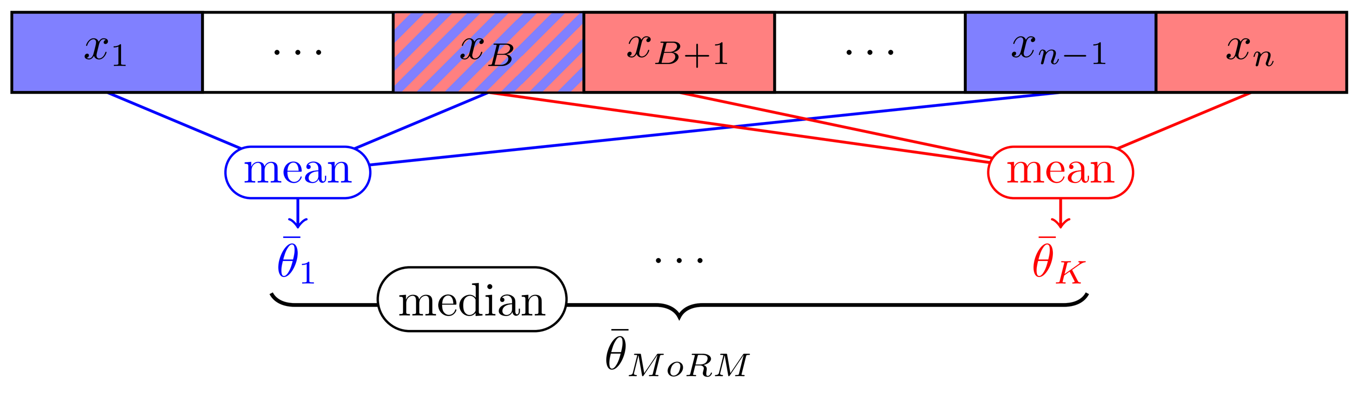


If $x_1, \ldots, x_n$ are n independent realizations of a r.v. X such that $\mathbb{E}[X] = \theta$, and $\mathrm{Var}(X) = \sigma^2$, for any $\delta \in [e^{1-n/2}, 1[$, choosing $K = \lceil \log(1/\delta) \rceil$ it holds:

$$\mathbb{P}\left\{ \left| \hat{\theta}_{\mathrm{MoM}} - \theta \right| > 2\sqrt{2}e\sigma\sqrt{\frac{1 + \log(1/\delta)}{n}} \right\} \leq \delta.$$

Proof: Let $I_k^\varepsilon := \mathbb{1}_{|\hat{\theta}_k - \theta| > \varepsilon}$, then $\mathbb{P}\{|\hat{\theta}_{\mathrm{MoM}} - \theta| > \varepsilon\} \leq \mathbb{P}\{\sum_{k=1}^K I_k^\varepsilon \geq \frac{K}{2}\}$. Bound using Hoeffding (or binomial law), with $\mathbb{E}[I_k^\varepsilon] \leq \sigma^2/(B\varepsilon^2)$.

Median of Randomized Means



If blocks are formed by SWoR, for any $\tau \in]0, 1/2[$, for any $\delta \in [2e^{-8\tau^2 n/9}, 1[$, choosing $K = \lceil \log(2/\delta)/(2(1/2 - \tau)^2) \rceil$, $B = \lfloor 8\tau^2 n/(9\log(2/\delta)) \rfloor$, it holds:

$$\mathbb{P}\left\{ \left| \bar{\theta}_{\mathrm{MoRM}} - \theta \right| > \frac{3\sqrt{3}\,\sigma}{2\,\tau^{3/2}}\sqrt{\frac{\log(2/\delta)}{n}} \right\} \leq \delta.$$

Proof: Let $\mathcal{I}_k^\varepsilon := \mathbb{1}_{|\bar{\theta}_k - \theta| > \varepsilon}$, $U_n^\varepsilon := \mathbb{E}_\epsilon[\frac{1}{K}\sum_{k=1}^K \mathcal{I}_k^\varepsilon \mid \mathcal{S}_n]$, and $p^\varepsilon := \mathbb{E}_{\mathcal{S}_n}[U_n^\varepsilon]$. Then $\mathbb{E}_{\mathcal{S}_n}[\mathbb{P}_\epsilon\{\frac{1}{K}\sum_{k=1}^K \mathcal{I}_k^\varepsilon - U_n^\varepsilon \geq \frac{1}{2} - \tau | \mathcal{S}_n\}]$, $\mathbb{P}_{\mathcal{S}_n}\{U_n^\varepsilon - p^\varepsilon \geq \tau - p^\varepsilon\}$.

Motivations and Remarks

Randomization motivations

- Classic alternative to segmentation
- Natural in *MoM Gradient Descent*
- Extension to incomplete U-stats

Remarks on bound

- K is arbitrary (may exceed n)
- B is arbitrary (always ≥ 1)
- Additional τ: tradeoff K/B

Possible extensions

- Other sampling schemes (Poisson, Monte Carlo) are more challenging as they do not benefit from the B-degree U-statistic's concentration.
- Extension to multivariate random variables (see geometric MoMs in [2]) is direct theoretically and cheap computationally.

U-statistics & Pairwise Learning

A natural estimate of $\mathbb{E}[h(X_1, X_2)]$, with X_1 and X_2 i.i.d. random vectors and h symmetric, from an i.i.d. sample $x_1, \ldots, x_n$ is the U-statistic

$$U_n(h) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(x_i, x_j).$$

Encountered *e.g.* in *pairwise ranking* or in *metric learning*:

$$\widehat{\mathcal{R}}_n(r) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbb{1}\left\{ r(x_i, x_j) \cdot (y_i - y_j) \leq 0 \right\},$$

$$\widehat{\mathcal{R}}_n(d) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbb{1}\left\{ y_{ij} \cdot (d^2(x_i, x_j) - \epsilon) \geq 0 \right\}.$$

Median of (Randomized) U-statistics

Blocks are formed either by partitioning or by SWoR. **Complete** U-statistics are computed on each block. The medians of the (randomized) U-statistics verify with probability at least $1 - \delta$:

$$\left| \hat{\theta}_{\mathrm{MoU}} - \theta(h) \right| \leq \sqrt{\frac{C_1 \log\frac{1}{\delta}}{n} + \frac{C_2 \log^2(\frac{1}{\delta})}{n\left(2n - 9\log\frac{1}{\delta}\right)}},$$

$$\left| \bar{\theta}_{\mathrm{MoRU}} - \theta(h) \right| \leq \sqrt{\frac{C_1(\tau) \log\frac{2}{\delta}}{n} + \frac{C_2(\tau) \log^2(\frac{2}{\delta})}{n\left(8n - 9\log\frac{2}{\delta}\right)}},$$

with C_1 and C_2 only depending on h, $C_1(\tau) = C_1/(2\tau)^3$, $C_2(\tau) = C_2/(2\tau)^3$. Extension to incomplete U-statistics made hard by replications in a block.

Pairwise Tournament Procedure

Adapted from [1], we want to find $f^* \in \operatorname{argmin}_{f \in \mathcal{F}} \mathcal{R}(f) = \mathbb{E}[\ell(f, (X, X'))]$.

For $f \in \mathcal{F}$, let $H_f := \sqrt{\ell(f, X, X')}$. For every candidates pair $(f, g) \in \mathcal{F}^2$:

1) On a 1st part of the sample, compute the MoU estimate of $\|H_f - H_g\|_{L_1}$

$$\Phi_{\mathcal{S}}(f, g) = \operatorname{median}\left(\hat{U}_1|H_f - H_g|, \ldots, \hat{U}_K|H_f - H_g| \right).$$

2) If it is *large enough*, on a 2nd part of the sample, compute the *match*

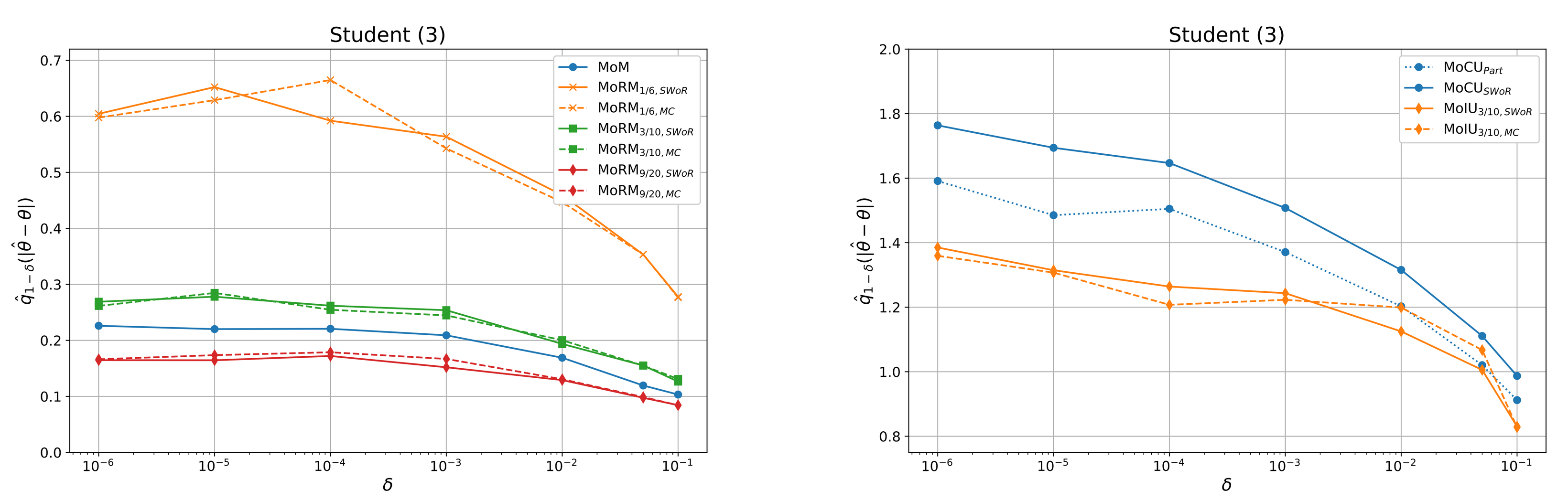
$$\Psi_{\mathcal{S}'}(f, g) = \operatorname{median}\left(\hat{U}_1(H_f^2 - H_g^2), \ldots, \hat{U}_{K'}(H_f^2 - H_g^2) \right).$$

A candidate $\hat{f}$ winning all its matches verify w.p.a.l. $1 - \exp(c_0 n \min\{1, r^2\})$

$$\mathcal{R}(\hat{f}) - \mathcal{R}(f^*) \leq cr.$$

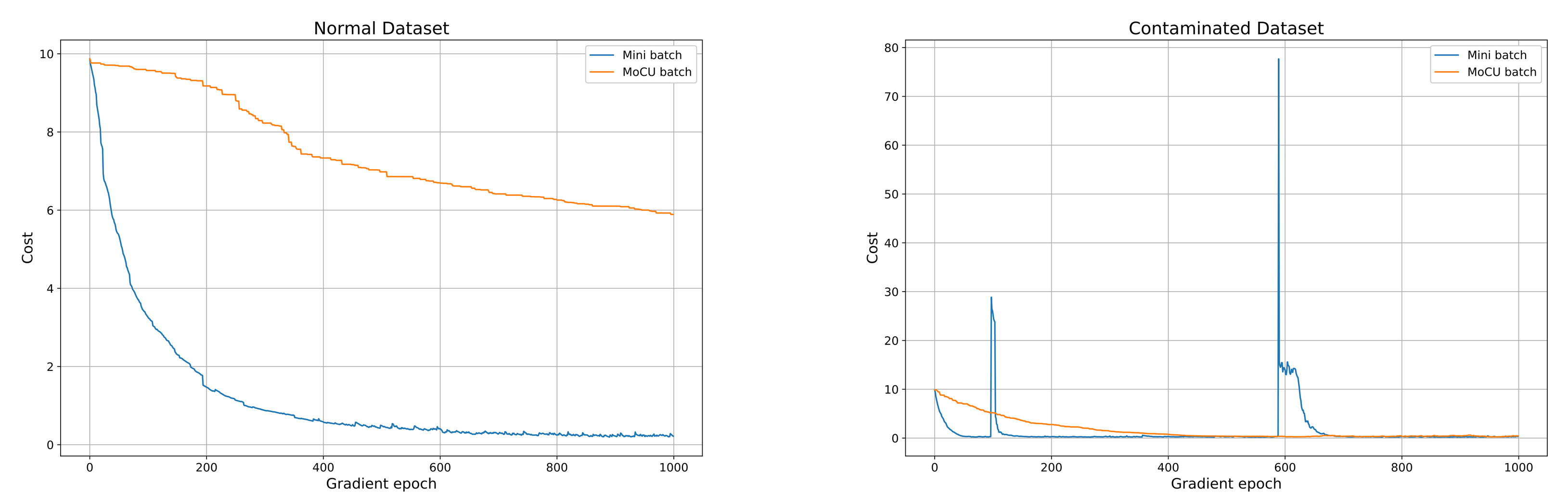
Estimation Experiments

Empirical deviation quantiles for estimations with MoRM (left, several τ settings) and MoU (right, with incomplete versions also).



Metric Learning Experiments

Standard (blue) and MoU (orange) gradient descents on a metric learning problem for a sane (left) and a contaminated (right) dataset.



References

[1] G. Lugosi and S. Mendelson. Risk minimization by median-of-means tournaments. *arXiv preprint arXiv:1608.00757*, 2016.

[2] S. Minsker et al. Geometric Median and Robust Estimation in Banach Spaces. *Bernoulli*, 21(4):2308–2335, 2015.