On Medians of (Randomized) Pairwise Means

In Proceedings of ICML 2019

Pierre Laforgue[1], Stephan Clémençon[1], Patrice Bertail[2]

[1] LTCI, Télécom Paris, Institut Polytechnique de Paris, France
[2] Modal'X, UPL, Université Paris-Nanterre, France

Outline

Reminders on the Simple Mean and the Simple Median

The Median of (Randomized) Means Estimator

The Median of (Randomized) U-Statistics Estimator

Learning from MoRM

Conclusion

The Simple Mean and Simple Median

$$\text{Sample } \mathcal{S}_n = \{Z_1, \ldots, Z_n\} \sim Z \text{ i.i.d. such that } \mathbb{E}[Z] = \theta$$

Simple Mean

- $\hat{\theta}_{\text{avg}} = \frac{1}{n} \sum_{i=1}^{n} Z_i$
- $\hat{\theta}_{\text{avg}} = \underset{\mu}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} |Z_i - \mu|^2$
- Estimator of $\theta = \mathbb{E}[Z]$

The Simple Mean and Simple Median

$$\text{Sample } \mathcal{S}_n = \{Z_1, \ldots, Z_n\} \sim Z \ \text{ i.i.d. such that } \mathbb{E}[Z] = \theta$$

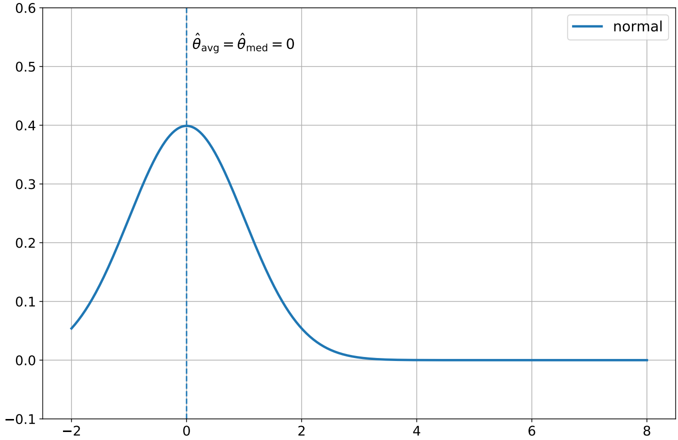
Simple Mean

- $\hat{\theta}_{\text{avg}} = \frac{1}{n} \sum_{i=1}^{n} Z_i$
- $\hat{\theta}_{\text{avg}} = \underset{\mu}{\operatorname{argmin}} \ \frac{1}{n} \sum_{i=1}^{n} |Z_i - \mu|^2$
- Estimator of $\theta = \mathbb{E}[Z]$

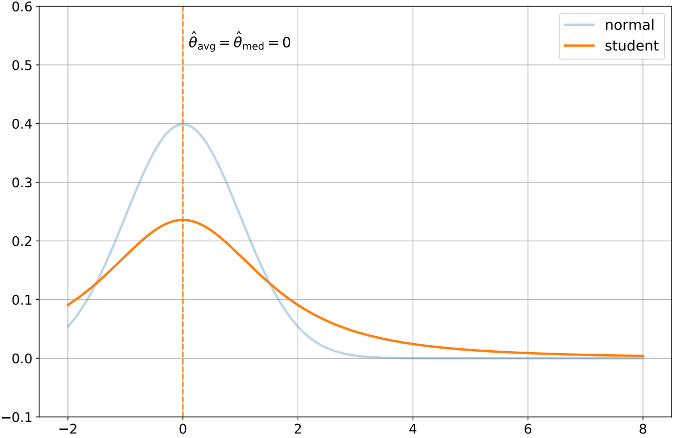
Simple Median

- $\hat{\theta}_{\text{med}} = Z_{\sigma(\frac{n+1}{2})}$, with $Z_{\sigma(1)} \leq \ldots \leq Z_{\sigma(n)}$
- $\hat{\theta}_{\text{med}} = \underset{\mu}{\operatorname{argmin}} \ \frac{1}{n} \sum_{i=1}^{n} |Z_i - \mu|$
- Estimator of $q_{\frac{1}{2}}$ such that
  $\mathbb{P}\left\{Z \geq q_{\frac{1}{2}}\right\} = \mathbb{P}\left\{Z \leq q_{\frac{1}{2}}\right\} = \frac{1}{2}$

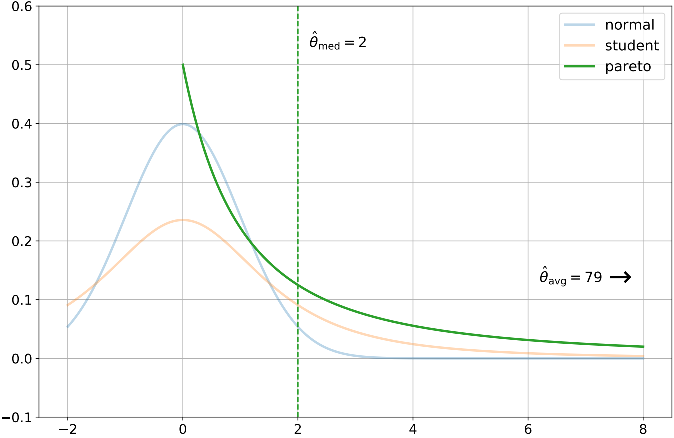
Graphically



Graphically



Graphically



Deviation Probabilities

Ways to asses an estimator:

- Quadratic Risk:

$$R(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right] = \underbrace{\left(\mathbb{E}[\hat{\theta}] - \theta\right)^2}_{\mathsf{Bias}^2(\hat{\theta})} + \mathsf{Variance}(\hat{\theta}).$$

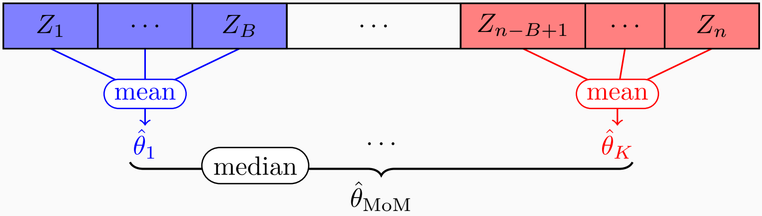
- Deviation Probabilities [Catoni, 2012]: $\mathbb{P}\left\{|\hat{\theta} - \theta| > t\right\}$.

- If X is bounded (see Hoeffding's Inequality) or sub-Gaussian:

$$\mathbb{P}\left\{\left|\hat{\theta}_{\mathsf{avg}} - \theta\right| > \sigma\sqrt{\frac{2\ln(2/\delta)}{n}}\right\} \leq \delta.$$

Outline

The Median of Means

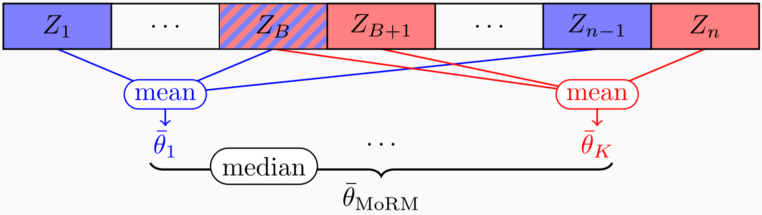


$Z_1, \ldots, Z_n$ i.i.d. realizations of r.v. Z s.t. $\mathbb{E}[Z] = \theta$, $\mathrm{Var}(Z) = \sigma^2$.

$\forall \delta \in [e^{1-\frac{n}{2}}, 1[$, set $K := \lceil \ln(1/\delta) \rceil$, it holds [Devroye et al., 2016]:

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

The Median of Randomized Means



With blocks formed by SWoR, $\forall\ \tau \in]0, 1/2[,\ \forall\ \delta \in [2e^{-\frac{8\tau^2 n}{9}}, 1[$, set $K := \left\lceil \frac{\ln(2/\delta)}{2(1/2-\tau)^2} \right\rceil$, and $B := \left\lfloor \frac{8\tau^2 n}{9\ln(2/\delta)} \right\rfloor$, it holds:

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

Remarks

Randomization motivations

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

Remarks

Randomization motivations

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

Remarks on bound

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

Remarks

Randomization motivations

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

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)
- Multivariate MoMs [Minsker et al., 2015]

Outline

U-statistics & Pairwise Learning

Estimate $\mathbb{E}[h(Z_1, Z_2)]$ from an i.i.d. sample $Z_1, \ldots, Z_n$:

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

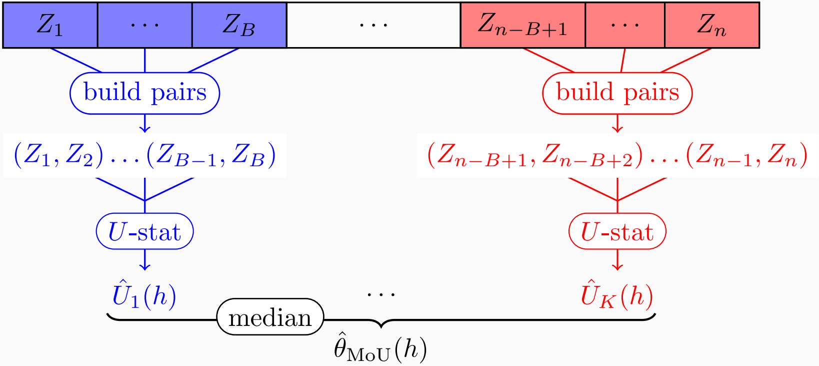
Ex: the empirical variance when $h(z, z') = \frac{(z-z')^2}{2}$.

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

$$\widehat{\mathcal{R}}_n(r) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbb{1}\left\{ r(X_i, X_j) \cdot (Y_i - Y_j) \leq 0 \right\}.$$

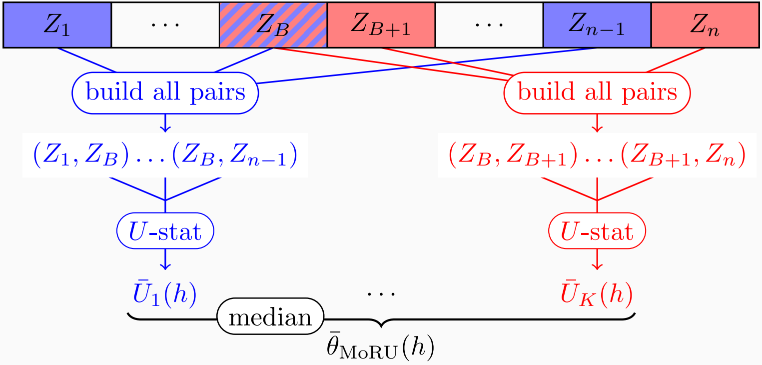
$$\widehat{\mathcal{R}}_n(d) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbb{1}\left\{ Y_{ij} \cdot (d(X_i, X_j) - \epsilon) > 0 \right\}.$$

The Median of U-statistics



$$\text{w.p.a.l. } 1-\delta, \quad \left|\hat{\theta}_{\mathrm{MoU}} - \theta(h)\right| \leq C_1(h)\sqrt{\frac{1+\ln(1/\delta)}{n}} + C_2(h)\,\frac{1+\ln(1/\delta)}{n}.$$

The Median of Randomized U-statistics



$$\text{w.p.a.l. } 1-\delta, \quad \left|\bar{\theta}_{\mathrm{MoRU}} - \theta(h)\right| \leq C_1(h,\tau)\sqrt{\frac{\ln(2/\delta)}{n}} + C_2(h,\tau)\,\frac{\ln(2/\delta)}{n}.$$

Outline

MoM Gradient Descent (GD)

- **ML:** $\min_{f \in \mathcal{F}} \quad \mathbb{E}_Z[\ell(f, Z)]$
- **ERM:** $\min_{f \in \mathcal{F}} \quad \frac{1}{n} \sum_{i=1}^{n} \ell(f, Z_i)$
- **GD:** $f \leftrightarrow f_\theta, \quad \theta_{t+1} = \theta_t - \gamma_t \nabla_\theta \ell(f_\theta, Z_i)$

What to do with $MoM(\ell(f_\theta, Z))$?

MoM Gradient Descent (GD)

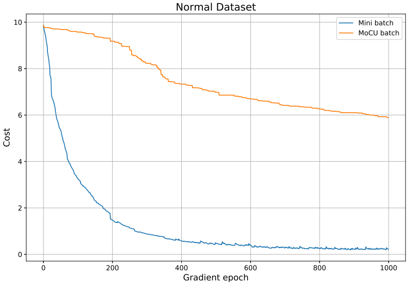
- **ML:** $\min_{f \in \mathcal{F}} \quad \mathbb{E}_Z[\ell(f, Z)]$
- **ERM:** $\min_{f \in \mathcal{F}} \quad \frac{1}{n} \sum_{i=1}^{n} \ell(f, Z_i)$
- **GD:** $f \leftrightarrow f_\theta, \quad \theta_{t+1} = \theta_t - \gamma_t \nabla_\theta \ell(f_\theta, Z_i)$

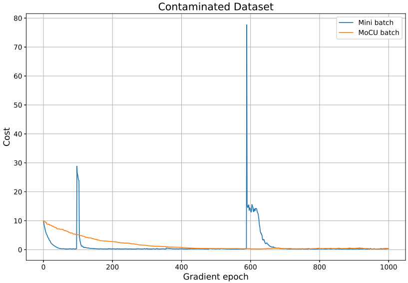
What to do with $MoM(\ell(f_\theta, Z))$?

$\rightarrow$ Mini-batch GD with specific criterion to select the mini-batch.

$\rightarrow$ Necessitates randomizing the partition.

Graphically





The Tournament Procedure

Adapted from [Lugosi and Mendelson, 2016]. We want to find $f^* \in \underset{f \in \mathcal{F}}{\operatorname{argmin}}\ \mathcal{R}(f) = \mathbb{E}[\ell(f, Z)]$. For any pair $(f, g) \in \mathcal{F}^2$:

1) Compute the MoM estimate of $\|f - g\|_{L_1}$

$$\Phi_{\mathcal{S}}(f, g) = \text{median}\left(\hat{\mathbb{E}}_1|f - g|, \ldots, \hat{\mathbb{E}}_K|f - g|\right).$$

2) If it is *large enough*, compute the *match*

$$\Psi_{\mathcal{S}'}(f, g) = \text{median}\left(\hat{\mathbb{E}}_1(f^2 - g^2), \ldots, \hat{\mathbb{E}}_K(f^2 - g^2)\right).$$

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

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

Outline

Conclusion

- MoM exhibits good guarantees with few assumptions

Conclusion

- MoM exhibits good guarantees with few assumptions
- Guarantees preserved through randomization

Conclusion

- MoM exhibits good guarantees with few assumptions
- Guarantees preserved through randomization
- Nicely extends to (randomized) U-statistics of arbitrary degree

Conclusion

- MoM exhibits good guarantees with few assumptions
- Guarantees preserved through randomization
- Nicely extends to (randomized) U-statistics of arbitrary degree
- Extension of MoM GD and tournament procedures

Conclusion

- MoM exhibits good guarantees with few assumptions
- Guarantees preserved through randomization
- Nicely extends to (randomized) U-statistics of arbitrary degree
- Extension of MoM GD and tournament procedures
- Remaining questions on sampling schemes, incomplete U-stats

Conclusion

- MoM exhibits good guarantees with few assumptions
- Guarantees preserved through randomization
- Nicely extends to (randomized) U-statistics of arbitrary degree
- Extension of MoM GD and tournament procedures
- Remaining questions on sampling schemes, incomplete U-stats
- Paper available in the proceedings of ICML 2019

References I

Catoni, O. (2012).
Challenging the empirical mean and empirical variance: a deviation study.
In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 48, pages 1148–1185. Institut Henri Poincaré.

Devroye, L., Lerasle, M., Lugosi, G., Oliveira, R. I., et al. (2016).
Sub-gaussian mean estimators.
The Annals of Statistics, 44(6):2695–2725.

References II

Lugosi, G. and Mendelson, S. (2016).
Risk minimization by median-of-means tournaments.
arXiv preprint arXiv:1608.00757.

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