Merging uncertainty sets via majority vote

Matteo Gasparin [1] Aaditya Ramdas[2]

[1]University of Padova [2]Carnegie Mellon University

Introduction

- In statistics, uncertainty is commonly captured through uncertainty sets (i.e., confidence intervals or prediction sets).
- In certain scenarios, different (dependent) uncertainty sets are generated by different agents.
- Some examples are conformal prediction intervals based on different algorithms or confidence intervals for a parameter of interest based on different methods.
- How should we combine K **arbitrarily dependent uncertainty sets**?

Problem statement

- **Input:** $\mathcal{C}_1, \dots, \mathcal{C}_K$ are $K \geq 2$ arbitrarily dependent uncertainty sets satisfying $\mathbb{P}(c \in \mathcal{C}_k) \geq 1-\alpha$, for all $k = 1, \dots, K$.
- **Output:** a single set that combines them in a black-box manner.

Two important quantities to consider: **coverage** and **size**.

Two naive solutions:

- $\cup_{k=1}^K \mathcal{C}_k$ has coverage $1-\alpha$, but it is too conservative.
- $\cap_{k=1}^K \mathcal{C}_k$ has coverage $1-K\alpha$, but it is too anti-conservative.

Majority vote

Include all the points that are contained in at least half of the sets.

$$\mathcal{C}^M := \left\{ s \in \mathcal{S} : \frac{1}{K}\sum_{k=1}^K 1\{s \in \mathcal{C}_k\} > \frac{1}{2} \right\}.$$

Using Markov's inequality: $\mathbb{P}(c \in \mathcal{C}^M) \geq 1-2\alpha$.
In addition,

$$m(\mathcal{C}^M) \leq \frac{2}{K}\sum_{k=1}^K m(\mathcal{C}_k),$$

where $m(\cdot)$ denotes the Lebesgue measure of a set.

Summary of the main results

- Majority vote is a good way to merge uncertainty sets.
- Improvements achieved through **randomization** and **exchangeability**.
- Drawback: In some cases (rarely in sims), the output is a union of intervals.
- The method can be used to derandomize statistical procedures based on **data splitting**.

Adding prior information

If there is a belief that certain agents are more accurate $\rightarrow$ incorporate prior information through a prior distribution $w = (w_1, .., w_k)$ over the agents.
Weighted majority vote:

$$\mathcal{C}^W := \left\{ s \in \mathcal{S} : \sum_{k=1}^K w_k 1\{s \in \mathcal{C}_k\} > \frac{1}{2} \right\}.$$

In this case: $\mathbb{P}(c \in \mathcal{C}^W) \geq 1-2\alpha$ and $m(\mathcal{C}^W) \leq 2\sum_{k=1}^K w_k m(\mathcal{C}_k)$.

Improving majority vote with randomization

Let $u \sim \text{Unif}(0,1)$, independent of all the data. Define

$$\mathcal{C}^R := \left\{ s \in \mathcal{S} : \sum_{k=1}^K w_k 1\{s \in \mathcal{C}_k\} > \frac{1}{2} + u/2 \right\}.$$

We obtain that $\mathcal{C}^R \subseteq \mathcal{C}^W$ and $\mathbb{P}(c \in \mathcal{C}^R) \geq 1-2\alpha$.
The proof is based on the uniformly-randomized Markov inequality.
Another possibility is to define the set $\mathcal{C}^U$ with a completely random threshold u, in this case $\mathbb{P}(c \in \mathcal{C}^U) \geq 1-\alpha$.

Merging exchangeable sets

- When $\mathcal{C}_1, \dots, \mathcal{C}_K$ are exchangeable, it is possible to obtain something better than a naive majority vote.
- We denote $\mathcal{C}^M(1:K) = \mathcal{C}^M$ to highlight that it is based on the majority vote of sets $\mathcal{C}_1, \dots, \mathcal{C}_K$.

We define

$$\mathcal{C}^E := \bigcap_{k=1}^K \mathcal{C}^M(1:k).$$

By definition $\mathcal{C}^E \subseteq \mathcal{C}^M$, in addition $\mathbb{P}(c \in \mathcal{C}^E) \geq 1-2\alpha$.

A simple way to improve the majority vote for arbitrarily dependent sets: process them in a random order ($\mathcal{C}^\pi$).

Derandomizing statistical procedures

It can be used also for **point estimators**.

Theorem: Suppose $\hat\theta_1, \dots, \hat\theta_K$ are K univariate point estimators of θ that are based using n data points and satisfy a high probability concentration bound

$$\mathbb{P}(|\hat\theta_k - \theta| \leq w(n,\alpha)) \geq 1-\alpha,$$

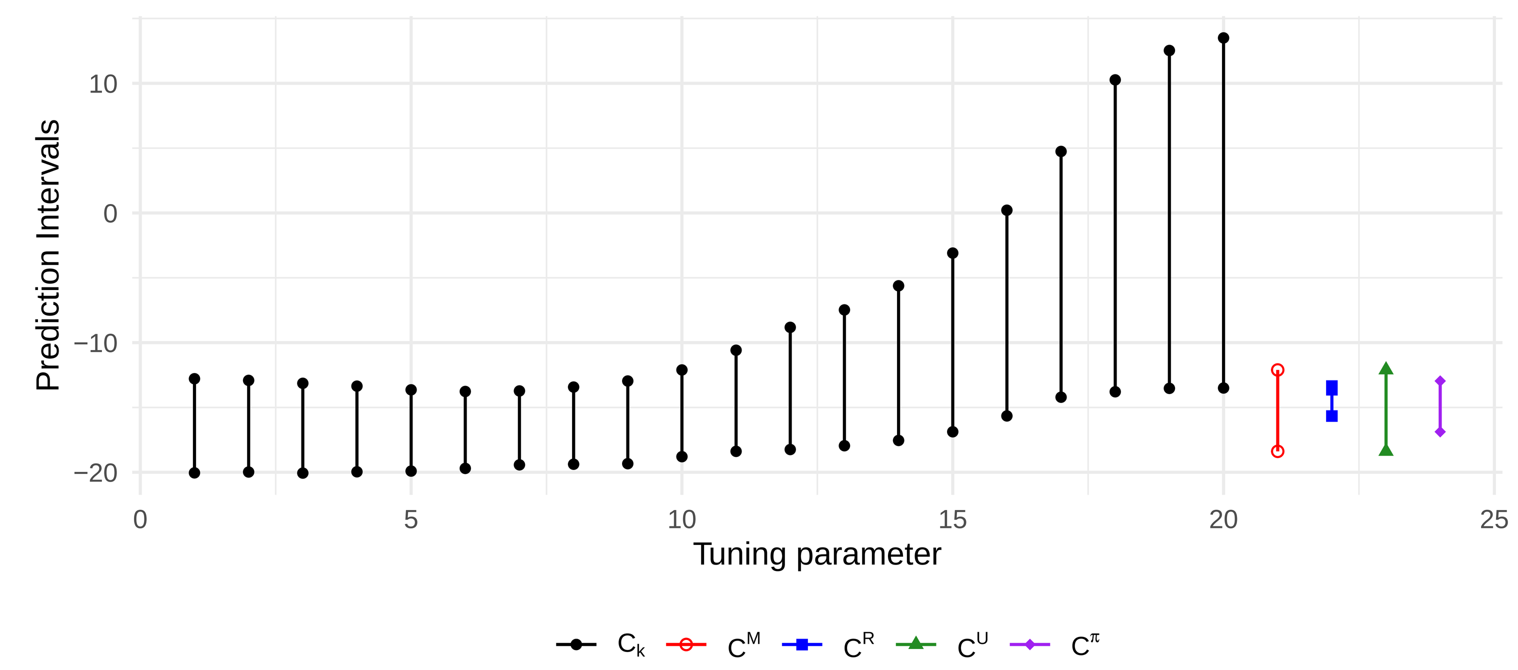
for some function w. Then, their median $\theta_{(\lceil K/2 \rceil)}$ satisfies

$$\mathbb{P}(|\hat\theta_{(\lceil K/2 \rceil)} - \theta| \leq w(n,\alpha)) \geq 1-2\alpha. \quad (1)$$

Further, if $\hat\theta_1, \dots, \hat\theta_K, \dots$ are exchangeable, then (1) is uniformly valid.

Example: conformal prediction with lasso

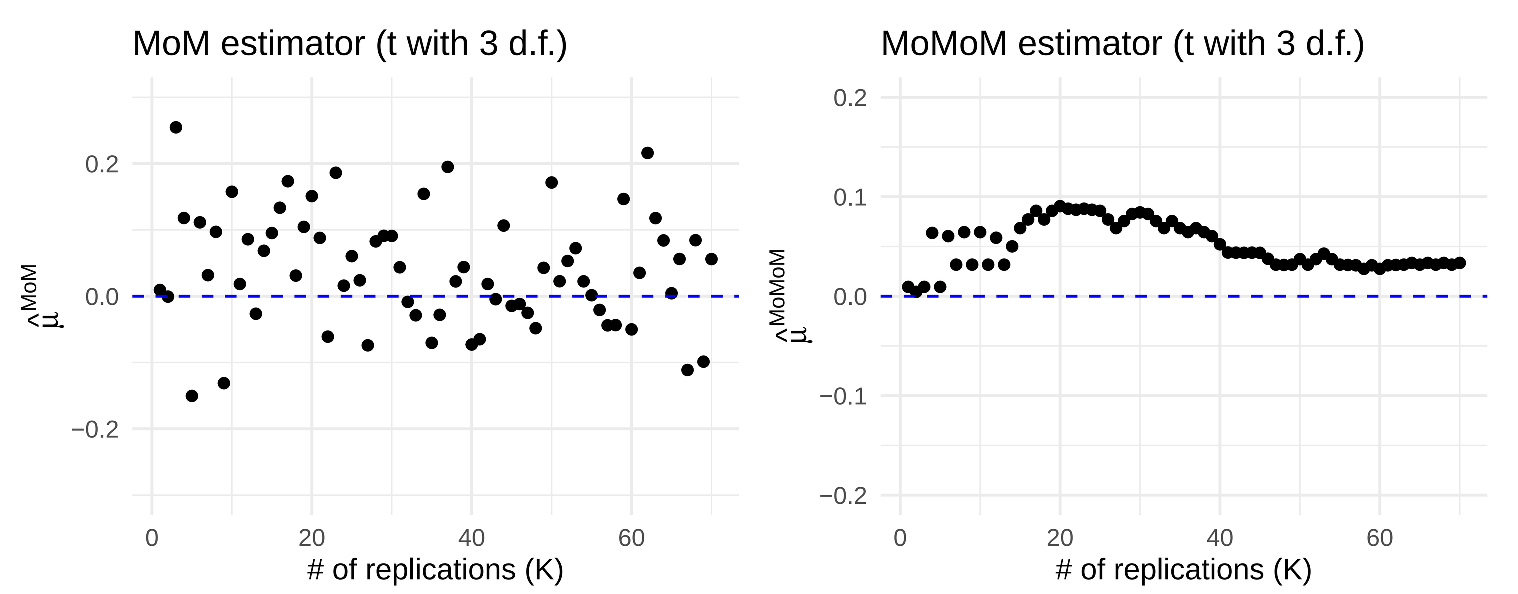
Fit lasso regression to data, with different penalty parameters λ and $\alpha = 0.05$.



Randomized sets used $u = 1/2$ for visualization.
Coverage: $\mathcal{C}^M = 0.97$, $\mathcal{C}^R = 0.92$, $\mathcal{C}^U = 0.96$, $\mathcal{C}^\pi = 0.93$.

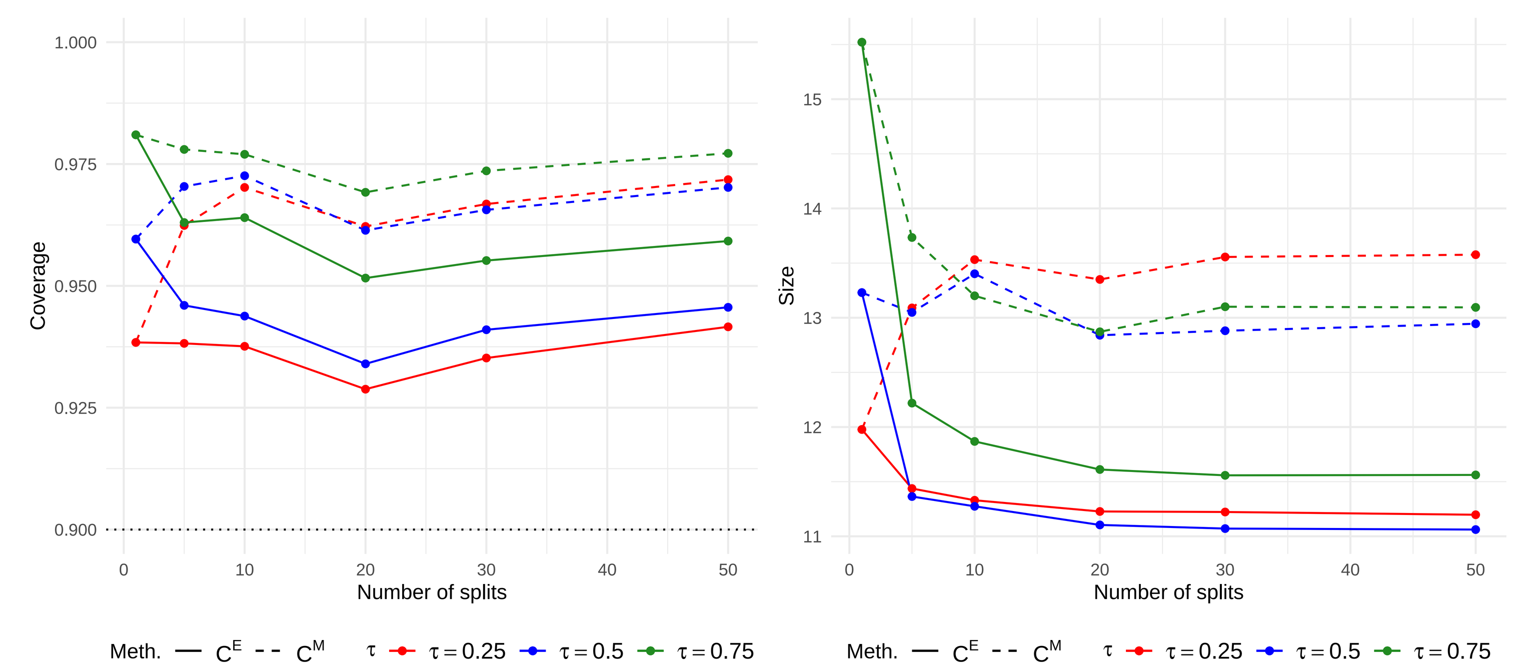
Derandomizing MoM (Median-of-Means)

$\hat\mu^{\text{MoM}}$: Estimator of the mean for $X_1, ..., X_n \overset{iid}{\sim} P$ based on data-splitting.
$\hat\mu^{\text{MoMoM}} :=$ median$(\hat\mu_1^{\text{MoM}}, \dots, \hat\mu_K^{\text{MoM}})$



Multi-split conformal inference

Construct K *split* conformal prediction intervals + (*exchangeable*) majority vote.



$\mathcal{C}^E$: smaller sets and coverage closer to the level $1-\alpha = 0.9$.

arXiv/2401.09
matteo.gasparin.1@phd.unipd.it