Class conditional conformal prediction for multiple inputs by p-value aggregation

Jean-Baptiste Fermanian[1], Mohamed Hebiri[2], Joseph Salmon[1]

[1] IMAG, IROKO, Université de Montpellier, Inria, CNRS, France [2] LAMA, Université Gustave Eiffel, France

1. Setting: multi-class classification

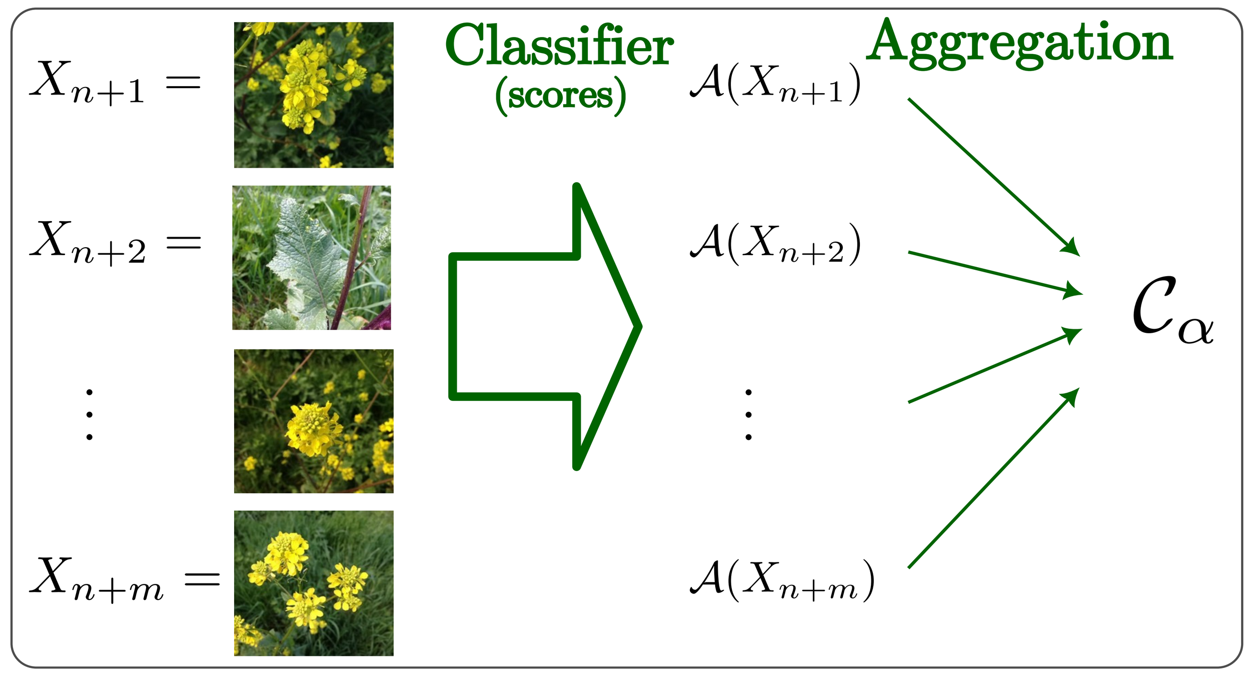
Inputs:

- $\alpha \in [0,1]$ a fixed coverage error.
- $(X_i, Y_i)_{i=1}^n$ a *calibration* set, $n_y = |\{i : Y_i = y\}|$.
- $(X_{n+j})_{j=1}^m$ *multiple* inputs of the **same unknown** class Y_{n+1}.
- $s : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ a score function *e.g.* (one minus) the softmax output.

Conformal scores:

- $S_i = s(X_i, Y_i)$, S_i^y are scores of class y, $S_{n+j}(y) = s(X_{n+j}, y)$.

Motivation: Pl@ntNet receives multiple images of the same plant.
↪ **Output a set containing the class with 'high' probability.**



Targeted guarantee: class conditional validity

$$\mathbb{P}\Big[Y_{n+1} \in \mathcal{C}_\alpha | Y_{n+1} = y\Big] \geq 1 - \alpha, \quad \forall y \in \mathcal{Y}.$$

↪ to avoid coverage disparities between classes

2. Background: conformal prediction

Marginal conformal prediction.
The set $\mathcal{C}_\alpha^{mrg}(x) = \big\{y : s(x,y) \leq S_{\lceil (1-\alpha)(n+1)\rceil}\big\}$ is **marginally valid**:

$$\mathbb{P}\Big[Y_{n+1} \in \mathcal{C}_\alpha^{mrg}(X_{n+1})\Big] \geq 1 - \alpha.$$

Class conditional conformal prediction.
The set $\mathcal{C}_\alpha^{cd}(x) = \big\{y : s(x,y) \leq S^y_{\lceil (1-\alpha)(n_y+1)\rceil}\big\}$ is **conditionally valid**.

P-value equivalence.

$$y \in \mathcal{C}_\alpha^{cd}(x) \iff p(y) = \frac{1}{n_y+1}\sum_{i=1}^{n_y} \mathbf{1}\{s(x,y) \leq S_i^y\} \geq \alpha.$$

3. Baseline: majority voting

- **Majority set** contains all classes present in more than half of the sets:

$$\mathcal{C}_{\alpha,m}^{MAJ} = \left\{y : \sum_{j=1}^m \mathbf{1}\{y \in \mathcal{C}_{\alpha/2}^{cd}(X_{n+j})\} \geq m/2\right\}.$$

- **Exchangeable majority set** intersects the sequential majority voting sets:

$$\mathcal{C}_\alpha^{Exch.\ MAJ} = \bigcap_{k=1}^m \mathcal{C}_{\alpha,k}^{MAJ}.$$

- Using Gasparin and Ramdas (2024), these sets are conditionally valid.

4. Key results

Exchangeability assumption

For all $y \in \mathcal{Y}$ and conditionally to $Y_{n+1} = y$, the vector of scores $\big(S_1^y, \dots, S_{n_y}^y, S_{n+1}(y), \dots, S_{n+m}(y)\big)$ is exchangeable.

Distribution of randomized conformal p-values

Theorem. For $y \in \mathcal{Y}$ and $j \in [\![m]\!]$, let

$$p_j^{\mathrm{rd}}(y) = \frac{1}{n_y}\sum_{i=1}^{n_y}\Big[\mathbf{1}\{S_{n+j}(y) < S_i^y\} + \mathbf{1}\{S_{n+j}(y) = S_i^y\}\mathbf{1}\{U_j \leq U_i^y\}\Big],$$

where $(U_j)_{j=1}^m$ and $(U_i^y)_{i=1}^{n_y}$ for $y \in \mathcal{Y}$ are i.i.d. uniform on $[0,1]$. Then

$$\mathbf{p}_\uparrow^{\mathrm{rd}}(Y_{n+1}) = \Big(p_{(1)}^{\mathrm{rd}}(Y_{n+1}), \dots, p_{(m)}^{\mathrm{rd}}(Y_{n+1})\Big) \,\Big|\, (Y_{n+1} = y) \sim \mathcal{U}(A_{n_y,m}),$$

where $A_{n,m} = \big\{a \in \{\frac{i}{n} : 0 \leq i \leq n\}^m : a_1 \leq \ldots \leq a_m\big\}$.

- Generalize results of Gazin et al. (2023) to the case of *ties*.

Distribution of conditional empirical coverage

Corollary. For $y \in \mathcal{Y}$ and $\ell \in [\![m]\!]$

$$\mathbb{P}\left[\sum_{j=1}^m \mathbf{1}\Big\{y \in \mathcal{C}_\alpha^{cd}(X_{n+j})\Big\} = \ell \,|Y_{n+1} = y\right] \geq \mathbb{P}[\beta = \ell],$$

where β follows the distribution $\mathtt{BetaBin}(m, (n_y+1) - k_y, k_y)$ with $k_y = \lfloor (n_y+1)\alpha \rfloor$.

- It is an equality in case of *no ties*.

5. Aggregation of conformal sets

Refinement of majority voting.

Let $q_\alpha^{BB}(y)$ the α-quantile of the distribution $\mathtt{BetaBin}\big(m, n_y + 1 - k_y, k_y\big)$. Then,

$$\mathcal{C}_\alpha^{Beta-Binomial} = \left\{y : \sum_{j=1}^m \mathbf{1}\{y \in \mathcal{C}_\alpha^{cd}(X_{n+j})\} \geq q_\alpha^{BB}(y)\right\},$$

is **conditionally valid**.

- BetaBinomial's quantiles can be approximated by Binomial's ones when $|\mathcal{Y}|m \ll n$.

P-value aggregation method

Let $v : [0,1]^m \times \mathcal{Y} \to \mathbb{R}$ and $T \in \mathbb{N}$, let $P_t^y \overset{i.i.d.}{\sim} \mathcal{U}(A_{n_y,m})$ and $V_t^y = v(P_t^y, y)$ for $t \in [\![T]\!]$ and $y \in \mathcal{Y}$. Then, the set

$$\mathcal{C}_\alpha^v = \left\{y : v(\mathbf{p}_\uparrow^{\mathrm{rd}}(y), y) \geq V^y_{(\lfloor (T+1)\alpha \rfloor)}\right\},$$

is **conditionally valid**.

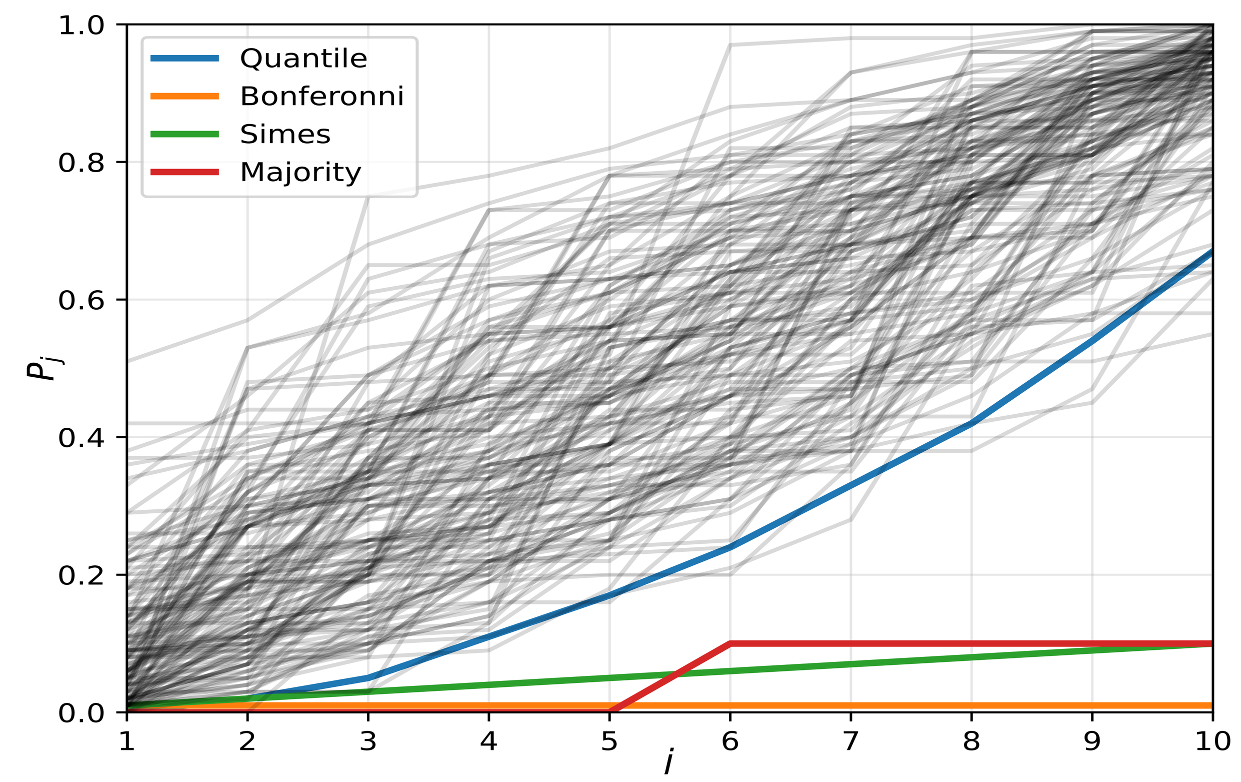
- Encompass Bonferonni correction : $v = \min p_j$; or Simes : $v = \min p_j / j$.

6. P-value simulation

- Simulation of 150 samples of $P \sim \mathcal{U}(A_{n,m})$ for $m = 10$ and $n = 100$.
- Comparison of the score Quantile with classical aggregation method for $\alpha = 0.1$.

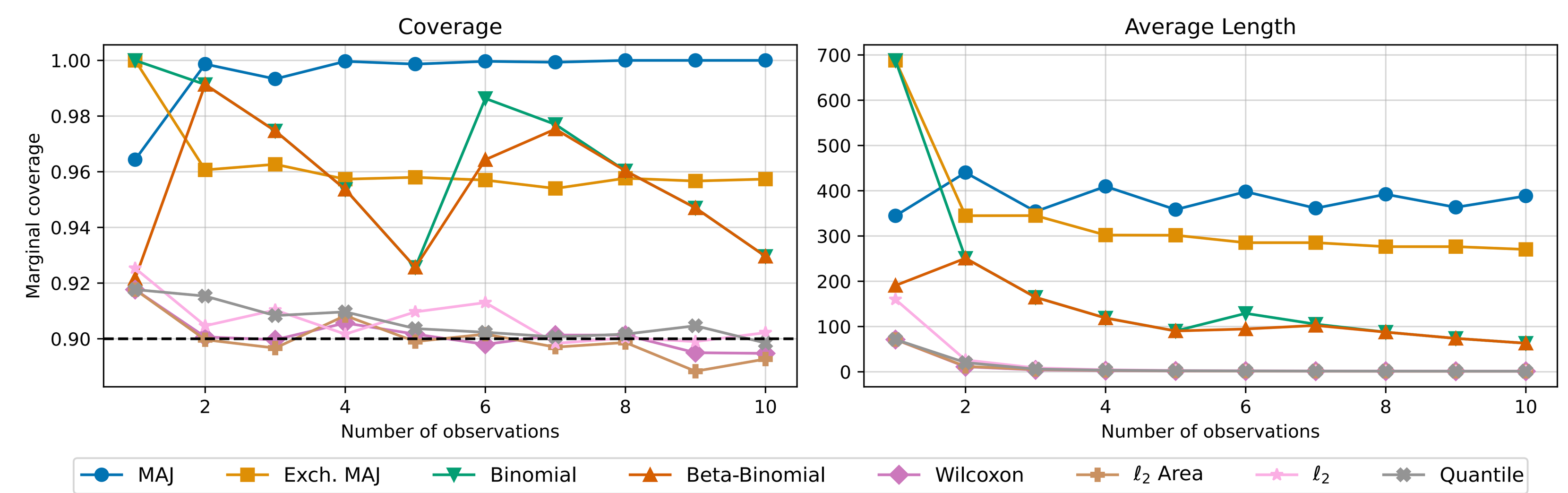
Algorithm 1 Simulation of $P \sim \mathcal{U}(A_{n,m})$

1: **Input:** n, m.
2: Draw $R_1, \dots, R_m$ w./o. replacement from $[\![n+m]\!]$.
3: $P_i \leftarrow R_{(i)}$ for $1 \leq i \leq m$.
4: $P_i \leftarrow P_i - i$ for $1 \leq i \leq m$.
5: **Output:** $(P_1/n, \dots, P_m/n)$.



7. Experiments on LifeCLEF 2015 dataset

- 688 classes, $n = 3.10^4$, classifier is a ResNet50 with top1 accuracy of 43%.
- Multi-inputs are drawn randomly from plants of the same class.
- **Used scores:**
  - ℓ_2 Area : $v = \|\mathbf{p}\|_2$,
  - Wilcoxon : $v = \|\mathbf{p}\|_1$,
  - ℓ_2 : $v = -\|\mathbf{p} - \mathtt{Id}\|_2$,
  - Quantile: $v = \min_j F_j^y(p_j)$, where F_j^y is the cdf of P_j^y.



- Majority voting $\ll$ (Beta)-Binomial approaches $\ll$ P-value aggregation methods.

8. Strengths and weaknesses

- Leverage exact distribution of the ordered randomized conformal p-values.
- Manage ties between scores by an adaptive randomization.
- Exchangeability assumption: non-robust to correlation between multi-inputs.