Classify-and-Count with Prediction Sets: A Conformal Approach to Label Distribution Estimation

Evgueni Smirnov and Filip Schlembach

Department of Advanced Computing Sciences,
Maastricht University,
Maastricht, The Netherlands
smirnov@maastrichtuniversity.nl, filip.schlembach@maastrichtuniversity.nl

Abstract. We introduce *Mondrian Inductive Conformal Quantification* (MICQ), a set-based approach to learning to quantify that combines Mondrian conformal prediction with *probability class counting* (PCC). For each unlabeled instance X from test bag B, a Mondrian conformal predictor first constructs a prediction set $\Gamma^{\alpha}(X)$ at miscoverage level α, and then PCC distributes the instance's weights to prevalences of class labels in $\Gamma^{\alpha}(X)$ proportionally to their posteriors. The final class prevalence estimates are obtained by averaging these weights across all instances in test bag B. Due to the per-class validity provided by the Mondrian conformal predictor MICQ remains reliable under prior shift. Experiments on standard benchmarks with controlled prior shifts demonstrate that MICQ indeed improves over PCC.

Keywords: Conformal prediction, learning to quantify, prior shift, Mondrian inductive conformal classification, probability class counting.

1 Introduction

The problem of *learning to quantify* (L2Q) is to estimate the distribution of class labels in a given (test) bag B of unlabeled instances [1]. Standard quantification methods, such as the *probability class counting* (PCC) approach, first use a trained classifier to estimate the posterior probabilities of class labels for the instances from B, and then compute their average across all instances. While simple, PCC can be sensitive to *prior shift* when the class distribution changes between training and testing while the class-conditional feature distributions remain stable.

To address this limitation, we propose a quantification method based on *conformal prediction*, a framework for set-valued classification that offers rigorous finite-sample guarantees [6]. Our approach, called *Mondrian Inductive Conformal Quantification* (MICQ), employs the class-conditional guarantees of Mondrian conformal predictors to provide prevalence estimates that remain reliable under prior shift.

MICQ can be viewed as a *conformal extension of PCC*. For each instance $X \in B$ and a probabilistic classifier used in PCC, the approach first constructs a conformal prediction set $\Gamma^\alpha(X)$ of plausible labels at a given miscoverage level α. The contribution (weights) of X to class prevalences is then distributed among the class labels in $\Gamma^\alpha(X)$ in proportion to their posterior probabilities, modulated by a *temperature* parameter τ that controls the weight's sharpness. The final class prevalence estimates are obtained by averaging these weights across all instances in the test bag B. By adjusting both the size of the prediction sets $\Gamma^\alpha(X)$ and the concentration of weights, MICQ allows regularization which has a potential to improve robustness and stability of quantification under prior shift.

We evaluate MICQ against the standard PCC baseline on three datasets under controlled prior shift using Dirichlet sampling. Our experiments show that MICQ consistently achieves lower absolute error than PCC across a range of conditions.

The rest of the paper is organized as follows. The problem of L2Q and (Mondrian) conformal classification are discussed in Section 2. Section 3 presents MICQ in detail. The experiments are provided and discussed in Section 4. Section 5 concludes the paper.

2 Background

This section provides the necessary background to introduce MICQ. It first considers formally the problem of learning to quantify and then (Mondrian) conformal classification.

2.1 Problem of Learning to Quantify

We formalize the problem of learning to quantify (L2Q) as follows. Let $\mathcal{X}$ be the input space and $\mathcal{Y}$ a finite class-label set. At training time the phenomenon under study is given by an unknown distribution P on $\mathcal{X} \times \mathcal{Y}$ with $(X, Y) \sim P$. We observe M i.i.d. realizations (x_m, y_m) of (X, Y) in $\mathcal{X} \times \mathcal{Y}$ that form the training data $D := \{(x_m, y_m)\}_{m=1}^M$. At test time, we face another unknown distribution P^* on $\mathcal{X} \times \mathcal{Y}$ and we observe N i.i.d. realizations x_n of X in $\mathcal{X}$ that form a test bag $B := \{x_n\}_{n=1}^N$. For each class label $y \in \mathcal{Y}$ we denote the class priors by $\pi_y := P(Y = y)$ and $\pi_y^* := P^*(Y = y)$ and combine them in vectors $\boldsymbol{\pi} = (\pi_y)_{y \in \mathcal{Y}}$ and $\boldsymbol{\pi}^* = (\pi_y^*)_{y \in \mathcal{Y}}$, respectively. In this context, the problem of L2Q is to estimate the class priors in $\boldsymbol{\pi}^*$, given the training set D and the unlabeled test bag B [1]. Thus, a L2Q estimator is supposed to output vector $\hat{\boldsymbol{\pi}}_B$ for bag B as close as possible to $\boldsymbol{\pi}_B$.

In this paper we consider the problem of L2Q for two cases:

- No distribution shift: $P = P^*$, and

[1] We note that the individual test class labels are never observed and the goal is *not to classify*, but to estimate label proportions.

– Prior distribution shift: $P(Y) \neq P^*(Y)$ (i.e. $\pi_y \neq \pi_y^*$ for any $y \in \mathcal{Y}$) and $P(X \mid Y = y) = P^*(X \mid Y = y)$ for any class label $y \in \mathcal{Y}$.

2.2 Conformal Classification

Conformal prediction is a framework for producing set-valued predictions with rigorous, statistical finite-sample guarantees [6]. Conformal predictors construct valid prediction sets $\Gamma^\alpha(X)$ for any new test point X, meaning that the true class label Y lies inside the prediction set with probability

$$\Pr_{(X,Y)\sim P}(Y \in \Gamma^\alpha(X)) \geq 1 - \alpha \tag{1}$$

for any user-specified miscoverage rate $\alpha \in (0, 1)$. This validity guarantee holds when there is no distribution shift ($P = P^*$) which implies that the training and test instances are exchangeable [6]. We later show how to maintain validity in the case of prior distribution shift.

The validity guarantee above is achieved by comparing the nonconformity score of a test example to those of the instances from the training set D. A nonconformity score function $s : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ measures how unusual a labeled instance $x \in \mathcal{X}$ appears w.r.t to other instances in D. Function s has to be defined *label-symmetric* s.t. for any fixed $x \in \mathcal{X}$ and for any permutation $\tau : \mathcal{Y} \to \mathcal{Y}$, the set of values $\{s(x, y) : y \in \mathcal{Y}\}$ equals the set $\{s(x, \tau(y)) : y \in \mathcal{Y}\}$; i.e., permuting the class labels does not change the distribution of non-conformity scores across the class-label space. One common choice for the function s is $s(x, y) = 1 - \widehat{p}(y \mid x)$, where $\widehat{p}(\cdot \mid x)$ is the posterior probability output by a probabilistic classifier h.

The most computationally efficient setting for conformal prediction is the *inductive setting* [2,3]. In this setting, the training set D is partitioned into two disjoint subsets, a *proper training set* $D_t = \{(X_m, Y_m)\}_{m=1}^{M_t}$ and a *calibration set* $D_c = \{(X_m, Y_m)\}_{m=M_t+1}^{M}$ with sizes M_t and $M_c = M - M_t$, respectively. A probabilistic classifier h is first trained on the proper training set D_t as a predictive function that can provide posterior probability estimates $\widehat{p}(y \mid X)$ for any class label $y \in \mathcal{Y}$ and input instance X; i.e. $h(X) = \{\widehat{p}(y \mid X)\}_{y\in\mathcal{Y}}$. The classifier h is first applied to all the calibration instances $(X_m, Y_m) \in D_c$ to obtain set S of their nonconformity scores $s_m := s(X_m, Y_m) = 1 - \widehat{p}(Y_m \mid X_m)$. Then, whenever we have a new test instance $X \sim P^*(X)$ the probabilistic classifier h provides a posterior class-label distribution $\{\widehat{p}(y \mid X)\}_{y\in\mathcal{Y}}$. These class-label probability estimates allow computing their associated nonconformity scores $s(X, y) = 1 - \widehat{p}(y \mid X)$ for X and every possible class label $y \in \mathcal{Y}$.

The p-values

$$p_y(X) = \frac{\#\{s_m \in S : s_m \geq s(X, y)\} + 1}{M_c + 1},$$

of test instance X for each class label $y \in \mathcal{Y}$ determine if the class label y is included in prediction set. If $p_y(X)$ is greater than a miscoverage rate α, y is

included in prediction set $\Gamma^\alpha(X)$. Thus, the final prediction set for X at α equals

$$\Gamma^\alpha(X) = \{\, y \in \mathcal{Y} : p_y(X) > \alpha\}.$$

As stated above, when there is no distribution shift, $P = P^*$, and exchangeability holds, so does the validity guarantee expressed in Equation (1). However, when there exists prior distribution shift, $P(Y) \neq P^*(Y)$, and $P(X \mid Y) = P^*(X \mid Y)$, validity is no longer guaranteed. To address this problem we describe below the Mondrian conformal predictor (MCP) in the inductive setting.

Mondrian Inductive Conformal Classification Mondrian inductive conformal predictor (MICP) is a class-conditional conformal predictor [6]. To handle distribution prior shift, it divides the nonconformity scores of the calibration instances by class labels y [6]. For each class label y we receive set $\mathcal{S}_y = \{s(X_m, Y_m) : (X_m, Y_m) \in D_c, Y_m = y\}$ with size M_y. This allows us to compute the Mondrian class-conditional p-value $p_y(X)$ of test instance X for candidate class label y

$$p_y(X) = \frac{\#\{s_m \in S_y : s_m \geq s(X, y)\} + 1}{M_y + 1}, \tag{2}$$

and, subsequently, the class-conditional prediction set

$$\Gamma^\alpha(X) = \{y : p_y(X) > \alpha\}. \tag{3}$$

We note that under the prior-shift assumption, test instance $X \mid (Y = y)$ is drawn from the same distribution as each calibration instance for class label y, so the set $S_y \cup \{s(X, y)\}$ is exchangeable, i.e. every permutation of these $M_y + 1$ scores has the same joint distribution. Exchangeability implies that the rank R_y of $s(X, y)$ among the $M_y + 1$ scores is uniformly distributed:

$$\Pr(R_y = k \mid Y = y) = \frac{1}{M_y + 1}, \qquad k = 1, \ldots, M_y + 1.$$

In this context, following Equation 2, the Mondrian class-conditional p-value can be redefined as

$$p_y(X) = \frac{R_y}{M_y + 1}.$$

Therefore,

$$\Pr\left(p_y(X) \leq \alpha \mid Y = y\right) = \frac{\lfloor \alpha(M_y + 1) \rfloor}{M_y + 1} \leq \alpha.$$

Because the event $Y \notin \Gamma^\alpha(X)$ is equivalent to $p_y(X) \leq \alpha$,

$$\Pr\left(Y \notin \Gamma^\alpha(X) \mid Y = y\right) = \Pr\left(p_y(X) \leq \alpha \mid Y = y\right) \leq \alpha. \tag{4}$$

Hence,

$$\Pr(Y \in \Gamma^\alpha(X) \mid Y = y) \geq 1 - \alpha$$

for every class y, showing that Mondrian inductive conformal predictors are valid for every class y.

3 Mondrian Inductive Conformal Quantification (MICQ)

Mondrian inductive conformal quantification (MICQ) is an approach to estimating class proportions in an unlabeled dataset. It offers robust estimations in cases of no distribution shift and distribution prior shift.

3.1 MICQ Description

The pseudocode of MICQ for the training and calibration phase is given in Algorithm 1. The input required for this phase consists of: a proper training set (D_t) and a calibration set (D_c). MICQ starts by training a probabilistic classifier h on the training set D_t. Then it iterates over the calibration instances in D_c to compute for each instance $(X_m, Y_m) \in D_c$ its probability estimate $\hat{p}(Y_m \mid X_m) = h(X_m)(Y_m)$ and its nonconformity score $s(X_m, Y_m) = 1 - \hat{p}(Y_m \mid X_m)$. Once the scores are known, MICQ forms the sets $\mathcal{S}_y$ of nonconformity scores for each class. At the end of the training and calibration phase MICQ outputs the probabilistic classifier h and the set of the calibration nonconformity score sets $\{\mathcal{S}_y\}_{y \in \mathcal{Y}}$.

Algorithm 1 MICQ – Training and Calibration Phase

Require: Proper training set $D_t = \{(X_m, Y_m)\}_{m=1}^{M_t}$, calibration set $D_c = \{(X_m, Y_m)\}_{m=M_t+1}^{M}$
Ensure: Trained classifier h, label-conditional score sets $\{\mathcal{S}_y\}$
1: Train a probabilistic classifier h on D_t
2: **for** each calibration instance $(X_m, Y_m) \in D_c$ **do**
3: $\quad \hat{p}(Y_m \mid X_m) = h(X_m)(Y_m)$
4: $\quad s(X_m, Y_m) = 1 - \hat{p}(Y_m \mid X_m)$
5: **end for**
6: **for** each class $y \in \mathcal{Y}$ **do**
7: $\quad \mathcal{S}_y = \{s(X_m, y) : (X_m, Y_m) \in D_c, Y_m = y\}$
8: **end for**
9: **return** h and $\{\mathcal{S}_y\}_{y \in \mathcal{Y}}$

The pseudocode of MICQ for the test and quantification phase is given in Algorithm 2. The input required for this phase consists of an unlabeled test bag B which class proportions need to be estimated, a user-defined miscoverage rate (α), a probabilistic classifier h, a set of the calibration nonconformity score sets $\{\mathcal{S}_y\}_{y \in \mathcal{Y}}$, a miscoverage rate α, and a temperature parameter τ. MICQ starts by initializing the prevalence value $\hat{\pi}_B^*(y)$ equal to 0 for all class labels $y \in \mathcal{Y}$. Then it iterates over instances X in the test bag B. For each test instance X and each possible class $y \in \mathcal{Y}$ MICQ computes its probability estimate $\hat{p}(y \mid X) = h(X)(y)$ using the probabilistic classifier h, and its nonconformity score $s(X, y) = 1 - \hat{p}(y \mid X)$. Then the p-value for the labeled instance (X, y) is computed following the MICP procedure (see Equation (2)). Once the p-values p_y are available for test

instance X and all class labels $y \in \mathcal{Y}$, MICQ computes the prediction set $\Gamma^\alpha(X)$ for X according to Equation (3). The resulting prediction set $\Gamma^\alpha(X)$ is composed of all class labels with p-values greater than α. Once the prediction set $\Gamma^\alpha(X)$ is available for test instance X, the weight $w_y(X)$ of X for y is computed according to:

$$w_y(X) = \begin{cases} \dfrac{\widehat{p}(y \mid X)^\tau}{\sum_{z \in \Gamma^\alpha(X)} \widehat{p}(z \mid X)^\tau}, & y \in \Gamma^\alpha(X), \\ 0, & \text{otherwise.} \end{cases}$$

where τ is a temperature parameter that controls the weight's sharpness ($\tau \geq 0$).

MICQ then updates the prevalence value $\hat{\pi}_B^*(y)$ for each class label y by adding the weight $w_y(X)$ of X to $\hat{\pi}_B^*(y)$. Then this entire procedure is repeated for all instances X in the test bag B. Upon completion, the final vector $\hat{\boldsymbol{\pi}}_B^*$ that is output by MICQ gives the estimated prevalence for each class y:

$$\hat{\pi}_B^*(y) = \frac{1}{n} \sum_{X \in B} w_y(X).$$

Algorithm 2 MICQ – Test and Quantification Phase

Require: Test bag $B = \{X_n\}_{n=1}^N$, set of the calibration nonconformity score sets $\{\mathcal{S}_y\}_{y \in \mathcal{Y}}$, miscoverage rate α, and temperature parameter τ
Ensure: Estimated prevalence vector $\hat{\boldsymbol{\pi}}_B^*$

Initialize $\hat{\pi}_B^*(y) = 0$ for all $y \in \mathcal{Y}$
for each test instance $X \in B$ **do**
  for each class $y \in \mathcal{Y}$ **do**
    $\hat{p}(y \mid X) = h(X)(y)$
    $s(X, y) = 1 - \hat{p}(y \mid X)$
    $p_y(X) = \frac{\#\{s_m \in \mathcal{S}_y : s_m \geq s(X,y)\}+1}{|\mathcal{S}_y|+1}$
  end for
  $\Gamma^\alpha(X) = \{y \in \mathcal{Y} : p_y(X) > \alpha\}$
  for each class $y \in \Gamma^\alpha(X)$ **do**
    $w_y(X) = \dfrac{\widehat{p}(y \mid X)^\tau}{\sum_{z \in \Gamma^\alpha(X)} \widehat{p}(z \mid X)^\tau}$
    $\hat{\pi}_B^*(y) \leftarrow \hat{\pi}_B^*(y) + w_y(X)$
  end for
end for
for each class $y \in \mathcal{Y}$ **do**
  $\hat{\pi}_B^*(y) \leftarrow \frac{1}{N} \cdot \hat{\pi}_B^*(y)$
end for
return $\hat{\boldsymbol{\pi}}_B^* = (\hat{\pi}_B^*(y))_{y \in \mathcal{Y}}$

3.2 Robustness to Prior Shift

MICQ is robust to *prior shift*; i.e. when there are changes in class priors $\pi_y \neq \pi_y^*$ while the class-conditional distribution $P(X \mid Y) = P^*(X \mid Y)$ remain fixed. This property is mainly due to class-conditional calibration of the Mondrian conformal predictor employed in MICQ. This predictor splits the calibration scores by class label ($\mathcal{S}_y$ for each y) ensuring that the feature distribution *within each class* is the same in training and test. Thus, any multiset $\mathcal{S}_y \cup \{s(X, y)\}$ stays exchangeable for every class y and the class-conditional p-value $p_y(X) = \frac{\#\{s \in \mathcal{S}_y : s \geq s(X,y)\}+1}{|\mathcal{S}_y|+1}$ satisfies the finite-sample coverage guarantee $\Pr\big(Y \in \Gamma^\alpha(X) \mid Y = y\big) \geq 1 - \alpha$ *even when class priors change.*

3.3 Regularization in MICQ

MICQ has two parameters that jointly control its bias and variance:

- **Miscoverage rate α (set size).** Smaller α yields *larger* prediction sets Γ^α which spreads each instance's contribution over more class labels and thus decreases variance and increases bias (toward more uniform class prevalences). Larger α yields *smaller* sets Γ^α, and, thus, sharpens contributions and reduces bias at the cost of higher variance.
- **Temperature parameter τ (within-set concentration).** For a fixed set Γ^α, when τ approaches 0^+, the weight w_y for any class label $y \in \Gamma^\alpha$ becomes closer to $1/|\Gamma^\alpha|$ which *increases bias* and *decreases variance*. When τ approaches $+\infty$, the weight w_y becomes closer to 1 for class label y with maximal posterior probability ($y = \arg\max_{z \in \Gamma^\alpha}$) and 0 for the remaining class labels. This *reduces bias* when posterior probabilities are informative and can *increase variance*.

The two parameters interact: the influence of τ is most pronounced when sets $|\Gamma^\alpha| > 1$ (moderate α). In this case we need to choose (α, τ) so that the sets are informative sets (through α) and appropriately concentrated within-set weights (through τ). In this context, we note that when sets Γ^α are singletons, τ has no effect.

3.4 Remark on Fisher Consistency

Fisher consistency requires that, when applied to the true test distribution, the expected estimate equals the true prevalence for all y. When $\alpha = 0$ and $\tau = 1$, the MICQ reduces to PCC, which *is* Fisher consistent.

For any $\alpha > 0$, however, Fisher consistency no longer holds. By construction, the conformal set $\Gamma^\alpha(X)$ excludes the true label with probability at most α (exactly α under continuous scores or randomized p-values), so the correct class receives weight 0 in a nonzero fraction of cases. In the remaining cases, the correct label is included but its contribution can be diminished due to the within-set normalization.

4 Experiments

This section evaluates and compares the performances of the MICQ quantifier and the standard probability class counting (PCC) quantifier. It provides experimental setup, experiment results, and discussion.

4.1 Experimental Setup

The L2Q quantifiers used in the experiments are the MICQ quantifier and PCC quantifier. To ensure model comparability both quantifiers employ the same base classifier the *Gaussian Naïve Bayes*. The Mondrian inductive conformal predictor in MICQ employs 67%/33% split, meaning that 67% of the data is used as training data D_t of the base classifier (Gaussian Naïve Bayes) and 33% of the data is used for the calibration set D_c.

4.2 Datasets

We conduct experiments on three standard benchmarking datasets from `scikit-learn` [4]. The datasets are described in Table 1.

Table 1. Datasets used in experiments

Dataset	**# Classes**	**# Features**	**# Instances**
Iris	3	4	150
Breast Cancer	2	30	569
Digits	10	64	1797

4.3 Validation Protocol

We follow a *10-fold stratified cross-validation* protocol. To simulate realistic quantification tasks under prior shift, we generate *bootstrap bags* as follows:

- For each test fold, we generate $T = 20$ test bags of the same size as the test set.
- Bags are sampled with replacement from the test fold.
- Each bag's class proportions are drawn from a Dirichlet distribution with concentration parameter $\alpha_{\text{Dir}} \in \{10, 100, 1000\}$ to simulate various prior shift conditions. We note that bigger (smaller) values for α_{Dir} imply smaller (bigger) prior shifts respectively.

4.4 Quantification and Evaluation

For each test bag, we compute the true class prevalences $\boldsymbol{\pi}^*$, the PCC estimates, and the MICQ estimates. Performance is evaluated using *Mean Absolute Error (MAE)*:

$$\mathrm{MAE}(\hat{\boldsymbol{p}}_B, \boldsymbol{\pi}^*) = \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} \left| \hat{\pi}_B(y) - \pi_y^* \right|,$$

We study MAE as a function of miscoverage level $\alpha \in [0.005, 0.500]$ (step size $= 0.005$) to observe how the conformal miscoverage rate α acts as a *regularization parameter* in MICQ.

Table 2. Summary of experimental settings

Parameter	**Value / Description**
Base Classifier	Gaussian Naïve Bayes
Cross-Validation	10-fold Stratified CV
Calibration Ratio	0.33
Bags per Fold	$T = 20$
Bag Size	Equal to test fold size
Bag Sampling	With replacement (bootstrap)
Dirichlet Parameters	$\alpha_{\mathrm{Dir}} \in \{10, 100, 1000\}$
Miscoverage Levels	$\alpha \in [0.005, 0.500]$
Temperature	$\tau = 2$
Evaluation Metric	Mean Absolute Error (MAE)

4.5 Results and Discussion

Figure 1 presents the results for the Iris, Breast Cancer, Digits datasets under varying Dirichlet prior strengths. For the Iris and Digits datasets, MICQ consistently outperforms the PCC baseline, with its MAE curve lying below the horizontal PCC baseline for all miscoverage levels α. In contrast, on the Breast Cancer dataset the picture is different: for small miscoverage levels ($\alpha < 0.12$) the MICQ's MAE is below the PCC baseline with exception of very low α values. For $\alpha > 0.12$ MICQ is worse than PCC. The Breast-Cancer curve also exhibits a spike near $\alpha \approx 0.07$ due to discrete p-values and small per-class calibration counts. Thus, we may conclude that MICQ can outperform PCC across a spectrum of prior shift scenarios, from no shift to substantial shift.

The results presented in Figure 1 show a clear bias-variance trade-off in Mondrian-weighted MICQ. For very small miscoverage rate α, prediction sets are large, leading to low variance but high bias and thus higher MAE. For moderate α, prediction sets shrink, bias drops, and MAE reaches its minimum, often outperforming the PCC baseline. For large α, MAE rises again due to high variance from prediction sets containing mostly single class labels. The trade-off

pattern holds across varying degrees of prior shift ($\alpha_{\text{Dir}} = 10, 100, 1000$) and it confirms that Mondrian-weighted MICQ balances bias and variance and adapts well to prior shift.

5 Conclusion

We introduced *Mondrian Inductive Conformal Quantification* (MICQ), a simple and effective set-based extension of the *probability class counting* (PCC) approach to learning to quantify. By leveraging prediction sets from Mondrian conformal prediction, MICQ assigns soft fractional weights to classes, enabling accurate prevalence estimation under prior shift. We showed empirically that MICQ is capable of outperforming PCC. These results demonstrate that conformal prediction offers a way to regularize quantification and ensure robustness to distributional changes.

Future work may continue in two possible directions. The first one is to apply conformal prediction to more advanced L2Q techniques. The second direction is to apply conformal predictors specifically designed for covariate shifts [5].

References

1. Esuli, A., Fabris, A., Moreo, A., Sebastiani, F.: Learning to Quantify, The Information Retrieval Series, vol. 47. Springer International Publishing, Cham (2023). https://doi.org/10.1007/978-3-031-20467-8, https://link.springer.com/10.1007/978-3-031-20467-8
2. Papadopoulos, H., Proedrou, K., Vovk, V., Gammerman, A.: Inductive Confidence Machines for Regression. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) Machine Learning: ECML 2002. pp. 345–356. Springer Berlin Heidelberg, Berlin, Heidelberg (2002)
3. Papadopoulos, H., Vovk, V., Gammerman, A.: Qualified predictions for large data sets in the case of pattern recognition. In: Wani, M., Arabnia, H., Cios, K., Hafeez, K., Kendall, G. (eds.) Proceedings of the International Conference on Machine Learning and Applications. pp. 159–163. CSREA Press (2002)
4. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
5. Tibshirani, R.J., Foygel Barber, R., Candes, E., Ramdas, A.: Conformal Prediction Under Covariate Shift. In: Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F.d., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 32. Curran Associates, Inc. (2019), https://proceedings.neurips.cc/paper/2019/file/8fb21ee7a2207526da55a679f0332de2-Paper.pdf
6. Vovk, V., Gammerman, A., Shafer, G.: Algorithmic Learning in a Random World. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-06649-8, https://link.springer.com/10.1007/978-3-031-06649-8

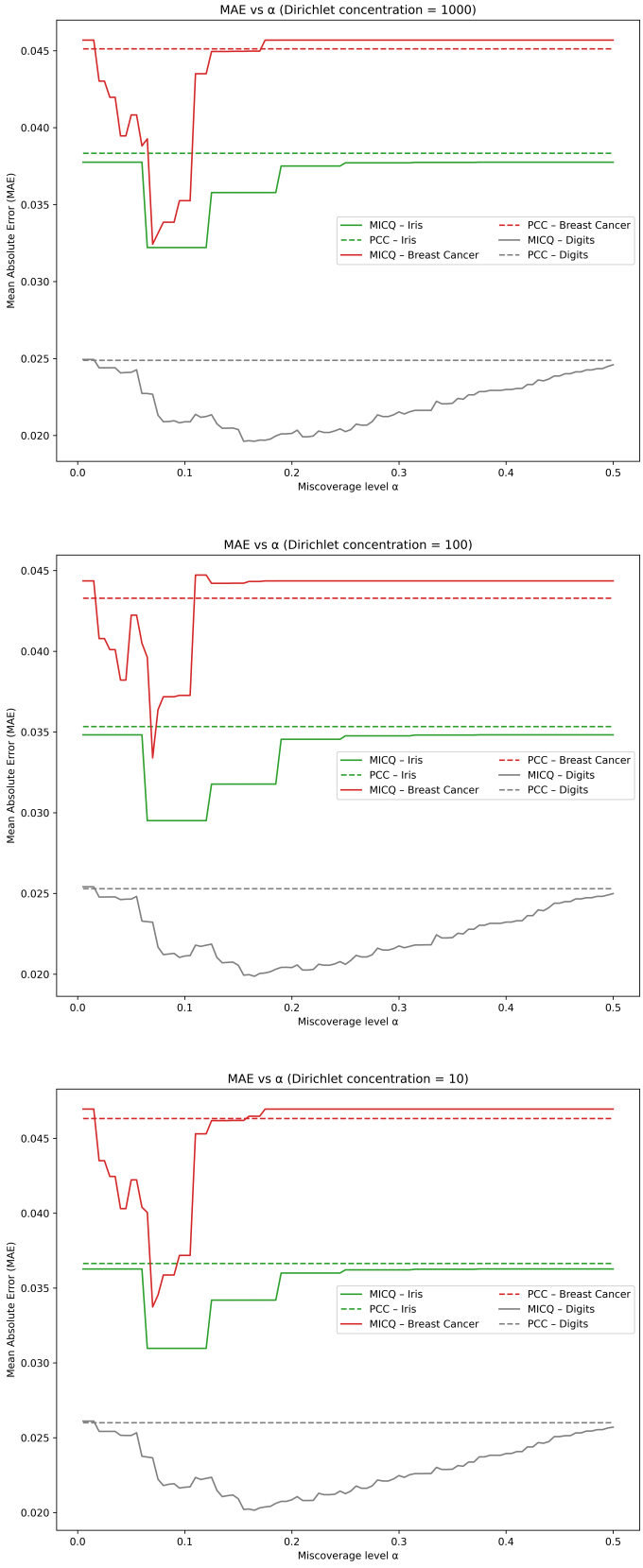

Fig. 1. Mean Absolute Error of MICQ vs. PCC across miscoverage levels for different Dirichlet strengths. From top to bottom: $\alpha_{\text{Dir}} = 1000$, $\alpha_{\text{Dir}} = 100$, $\alpha_{\text{Dir}} = 10$.