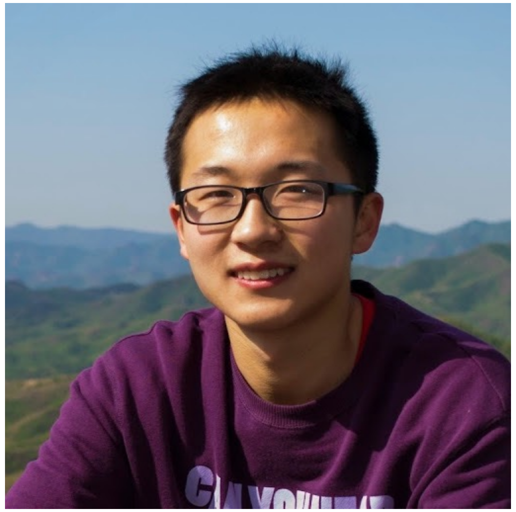
Modular Conformal Calibration

Charlie Marx*

Joint work with:

*Equal contribution



Shengjia Zhao*

Willie Neiswanger

Stefano Ermon

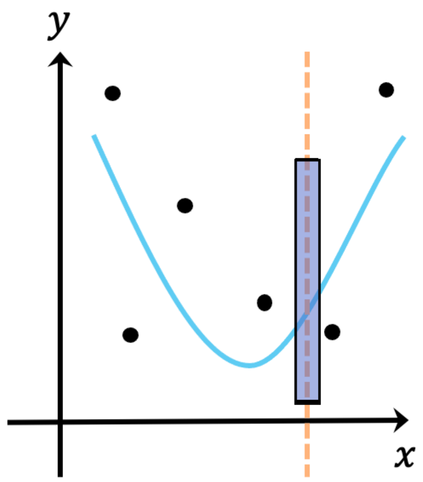
Why Quantify Uncertainty?

- Calibrate trust
- Enable efficient human oversight
- Improve performance in downstream tasks

Uncertainty Quantification: Two Perspectives

Conformal Prediction: turn any predictor into a set predictor with guaranteed *coverage*

$$P\big(Y \in S(X)\big) = 1 - \alpha$$



Uncertainty Quantification: Two Perspectives

Conformal Prediction: turn any predictor into a set predictor with guaranteed *coverage*

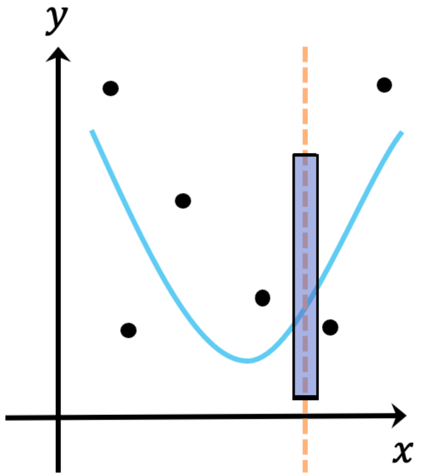
$$P\big(Y \in S(X)\big) = 1 - \alpha$$

Pros:

- weak assumptions
- applies to arbitrary predictors

Cons:

- Set predictions contain less information than distribution predictions



Uncertainty Quantification: Two Perspectives

Conformal Prediction: turn any predictor into a set predictor with guaranteed *coverage*

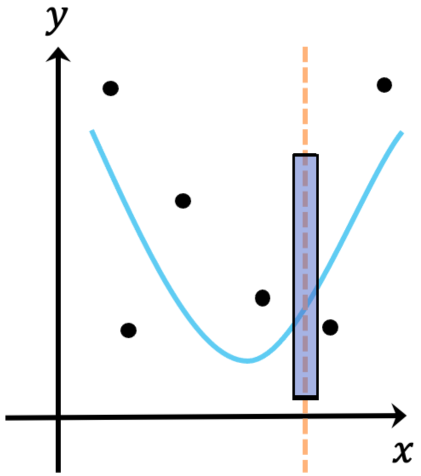
$$P\big(Y \in S(X)\big) = 1 - \alpha$$

Pros:

- weak assumptions
- applies to arbitrary predictors

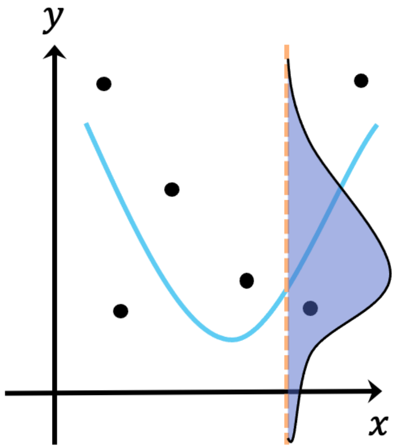
Cons:

- Set predictions contain less information than distribution predictions



Recalibration: adjust any distribution predictor to satisfy *calibration*

$$P(Y \leq Q_\alpha(X)) = \alpha, \qquad \forall \alpha \in [0, 1]$$



Uncertainty Quantification: Two Perspectives

Conformal Prediction: turn any predictor into a set predictor with guaranteed *coverage*

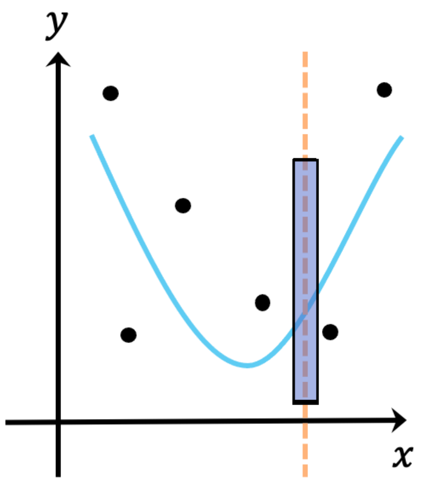
$$P\big(Y \in S(X)\big) = 1 - \alpha$$

Pros:

- weak assumptions
- applies to arbitrary predictors

Cons:

- Set predictions contain less information than distribution predictions



Recalibration: adjust any distribution predictor to satisfy *calibration*

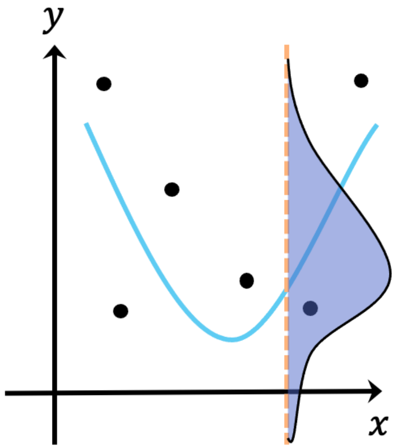
$$P(Y \leq Q_\alpha(X)) = \alpha, \qquad \forall \alpha \in [0, 1]$$

Pros:

- weak assumptions
- outputs flexible distributions

Cons:

- only applies to distribution predictors



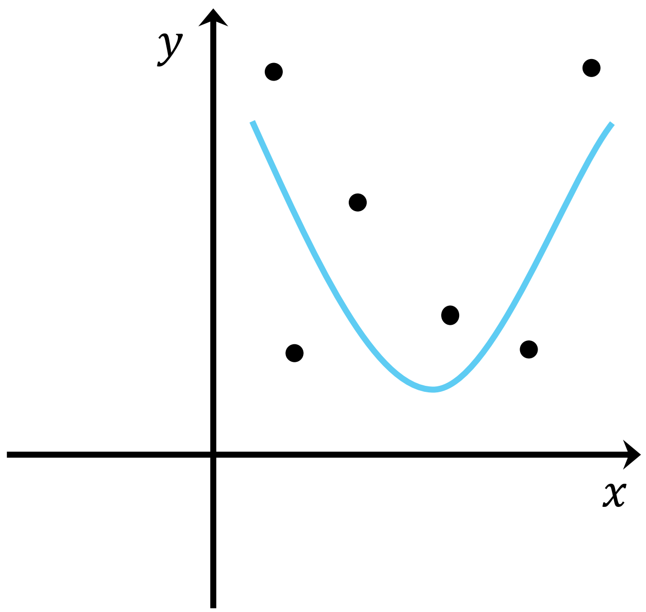
Our Work

- Transform an arbitrary regressor into a calibrated distribution predictor
- Achieve ϵ calibration error with $O(1/\epsilon)$ samples
- Unify conformal prediction and recalibration

Modular Conformal Calibration: Example

Base Predictor

$f : \mathcal{X} \to \mathcal{R}$



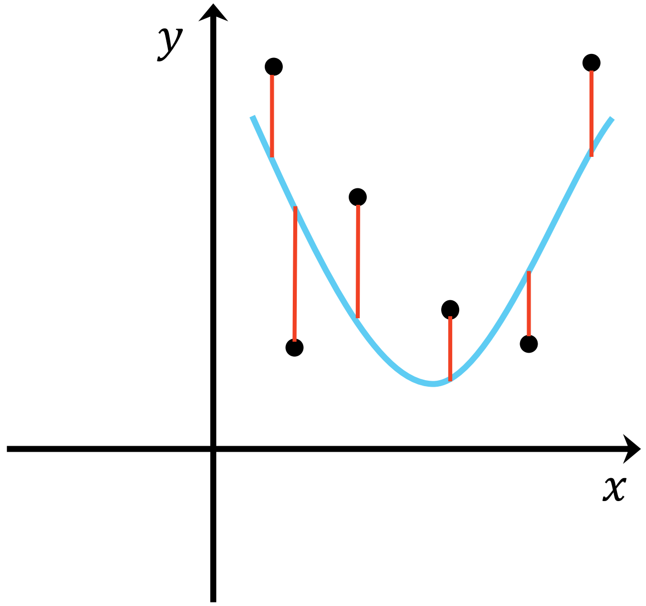
Modular Conformal Calibration: Example

Base Predictor

$f: \mathcal{X} \to \mathcal{R}$

Calibration Score

$s: \mathcal{R} \times \mathcal{Y} \to \mathbb{R}$



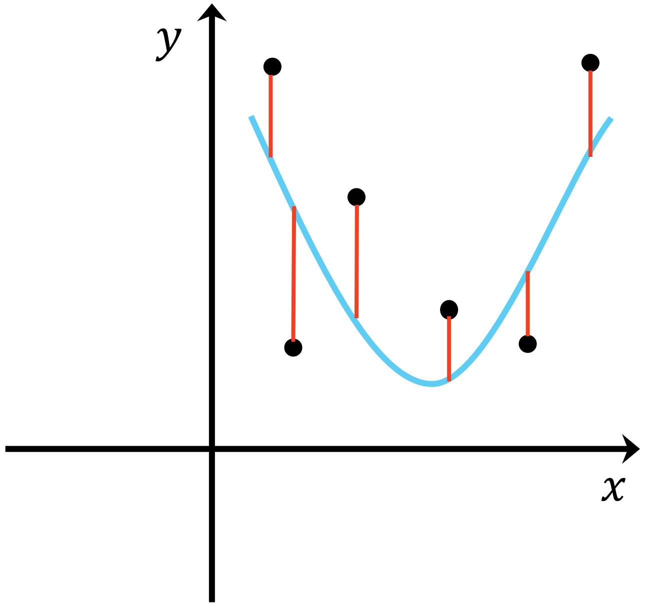
Modular Conformal Calibration: Example

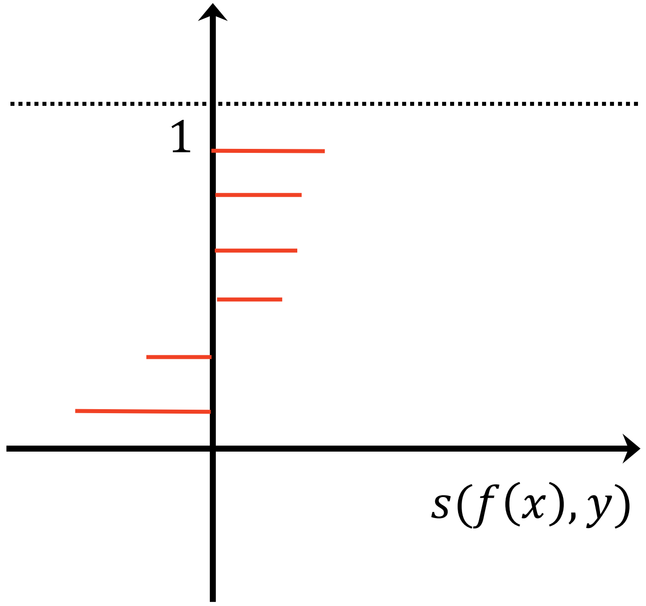
Base Predictor

$f: \mathcal{X} \rightarrow \mathcal{R}$

Calibration Score

$s: \mathcal{R} \times \mathcal{Y} \rightarrow \mathbb{R}$





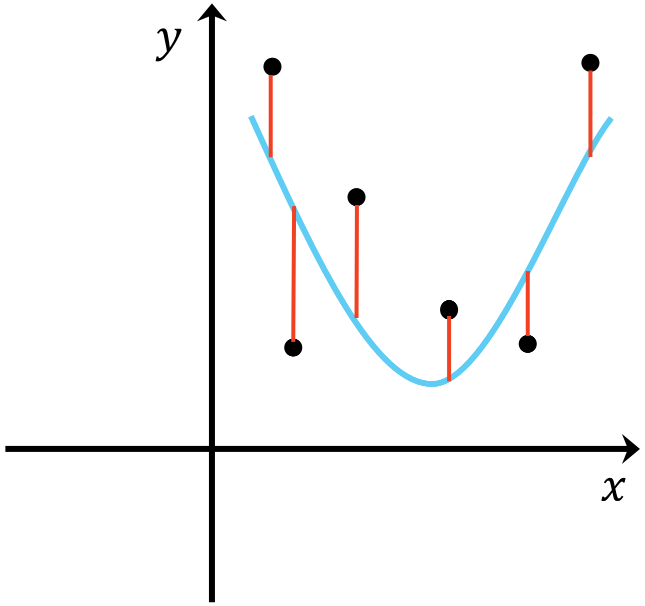
Modular Conformal Calibration: Example

Base Predictor

$f : \mathcal{X} \to \mathcal{R}$

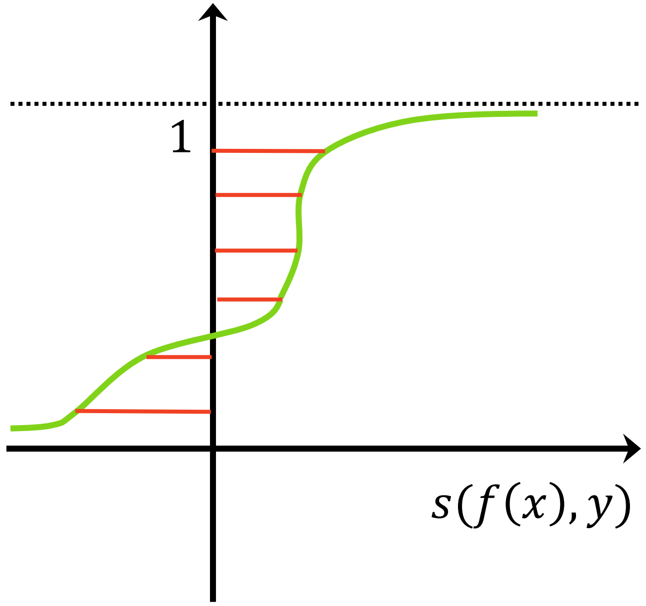
Calibration Score

$s : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}$



Interpolation Algorithm

$g : \mathbb{R}^n \to (\mathbb{R} \to [0, 1])$



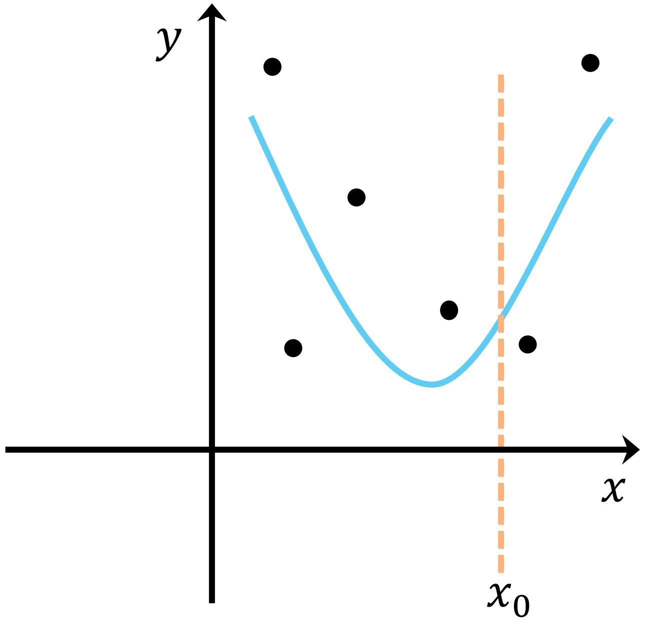
Modular Conformal Calibration: Example

Base Predictor

$f: \mathcal{X} \to \mathcal{R}$

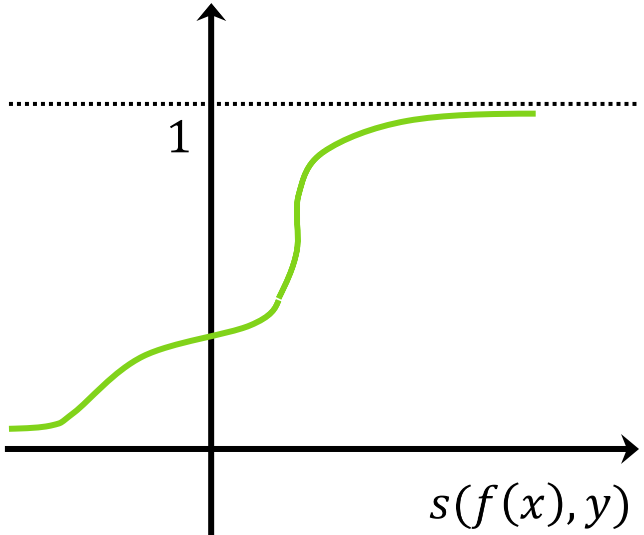
Calibration Score

$s: \mathcal{R} \times \mathcal{Y} \to \mathbb{R}$

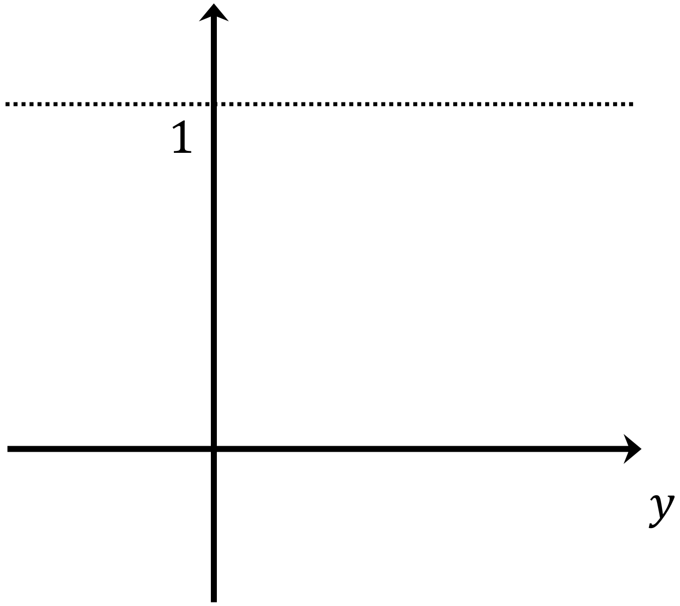


Interpolation Algorithm

$g: \mathbb{R}^n \to (\mathbb{R} \to [0,1])$



$P(Y \leq y \mid X = x_0)$



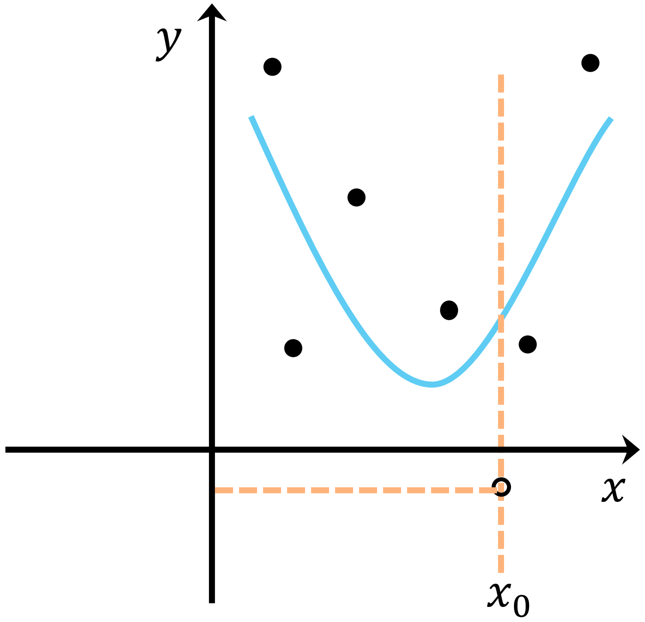
Modular Conformal Calibration: Example

Base Predictor

$f : \mathcal{X} \to \mathcal{R}$

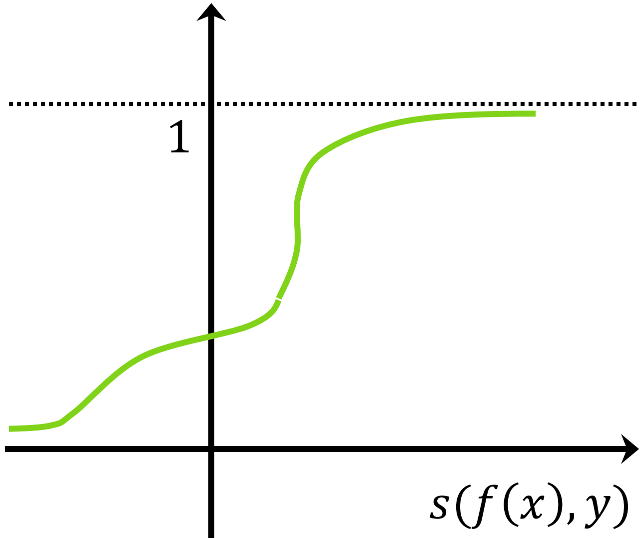
Calibration Score

$s : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}$

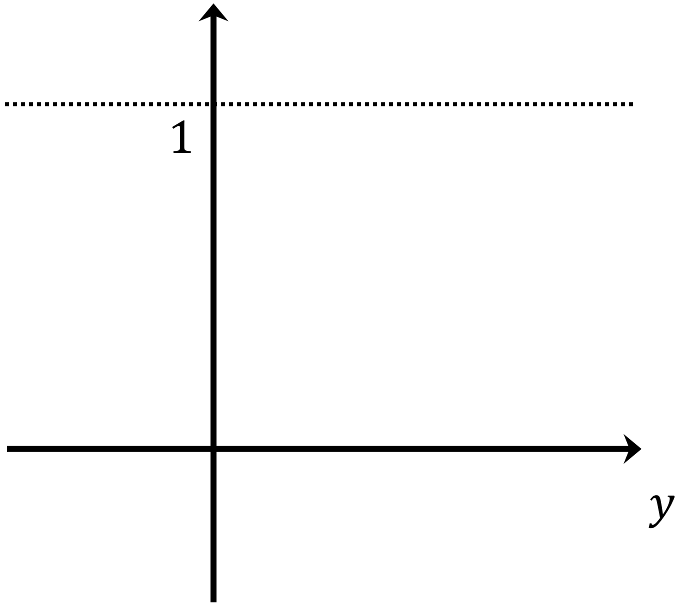


Interpolation Algorithm

$g : \mathbb{R}^n \to (\mathbb{R} \to [0, 1])$



$P(Y \leq y \mid X = x_0)$



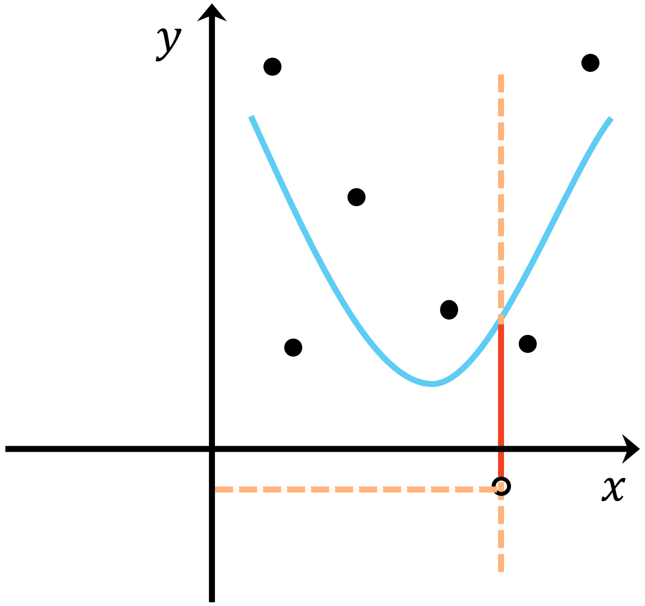
Modular Conformal Calibration: Example

Base Predictor

$f : \mathcal{X} \to \mathcal{R}$

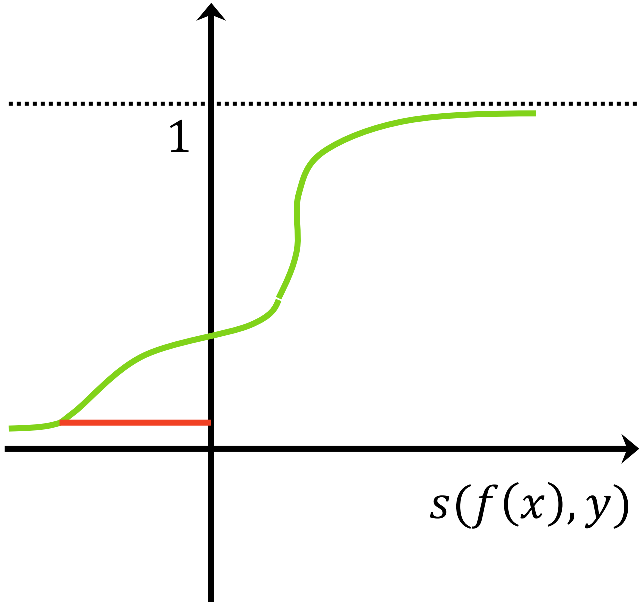
Calibration Score

$s : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}$

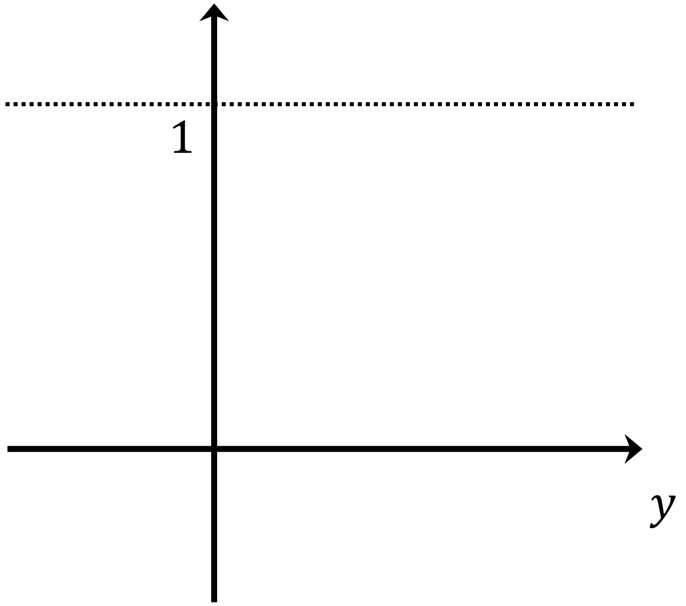


Interpolation Algorithm

$g : \mathbb{R}^n \to (\mathbb{R} \to [0,1])$

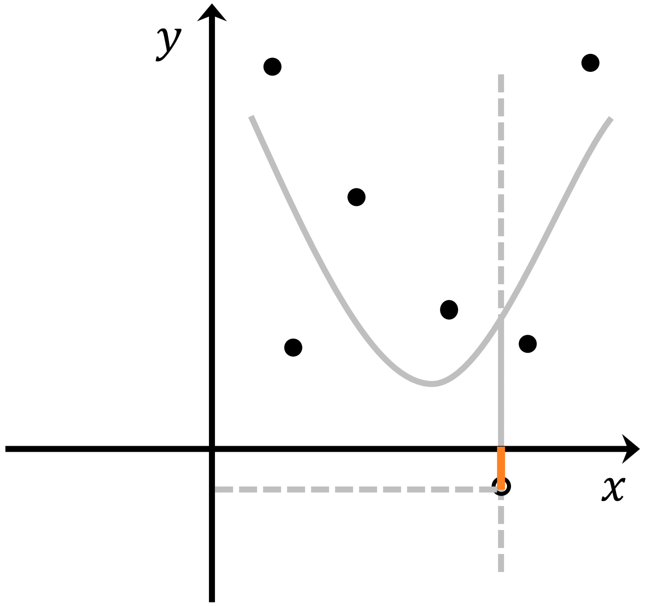


$P(Y \leq y \mid X = x_0)$

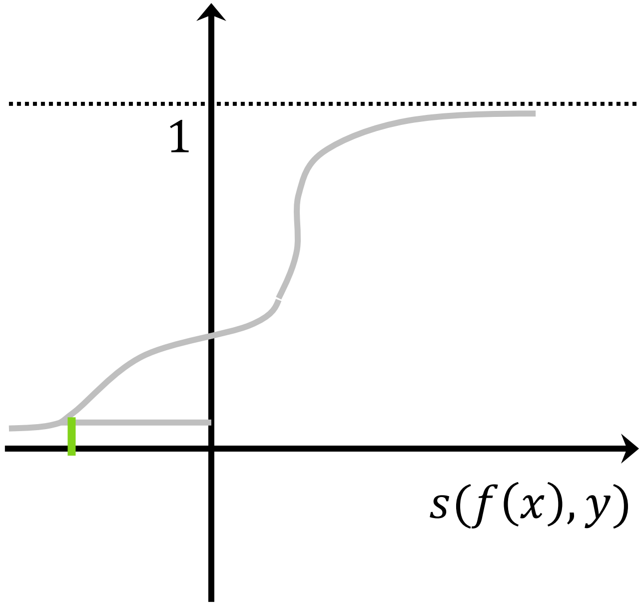


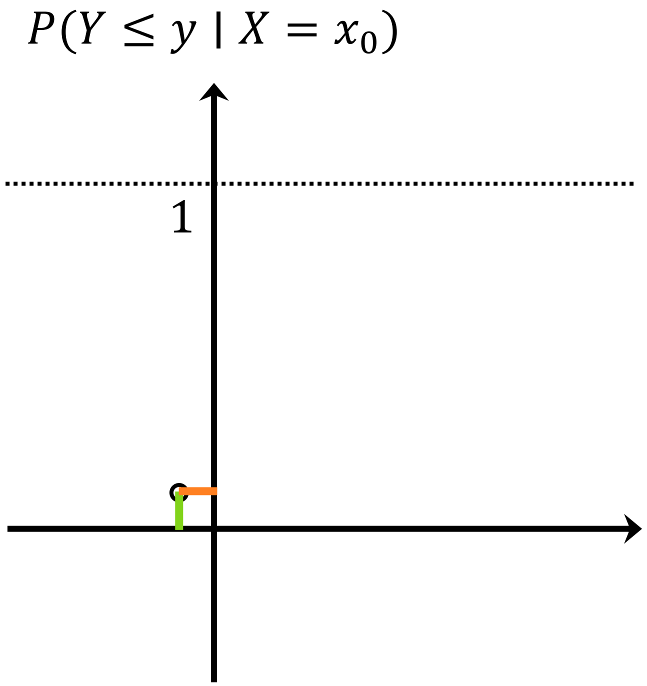
Modular Conformal Calibration: Example

Base Predictor
$f: \mathcal{X} \to \mathcal{R}$
Calibration Score
$s: \mathcal{R} \times \mathcal{Y} \to \mathbb{R}$



Interpolation Algorithm
$g: \mathbb{R}^n \to (\mathbb{R} \to [0,1])$





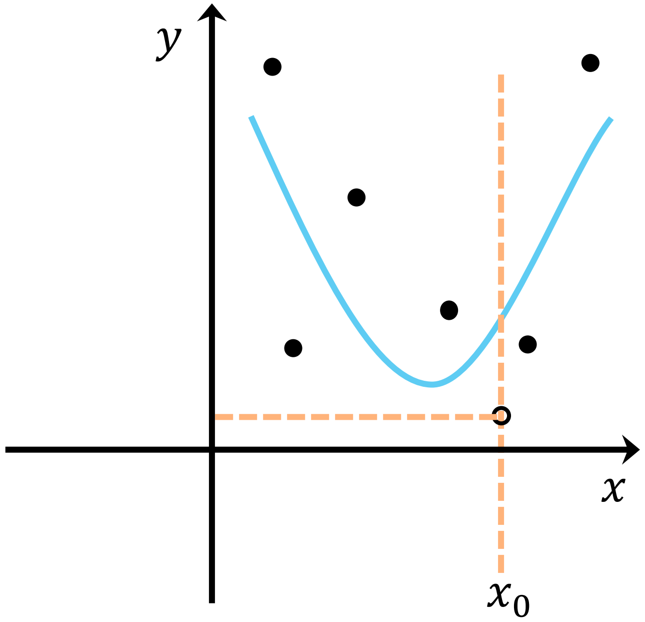
Modular Conformal Calibration: Example

Base Predictor

$f : \mathcal{X} \to \mathcal{R}$

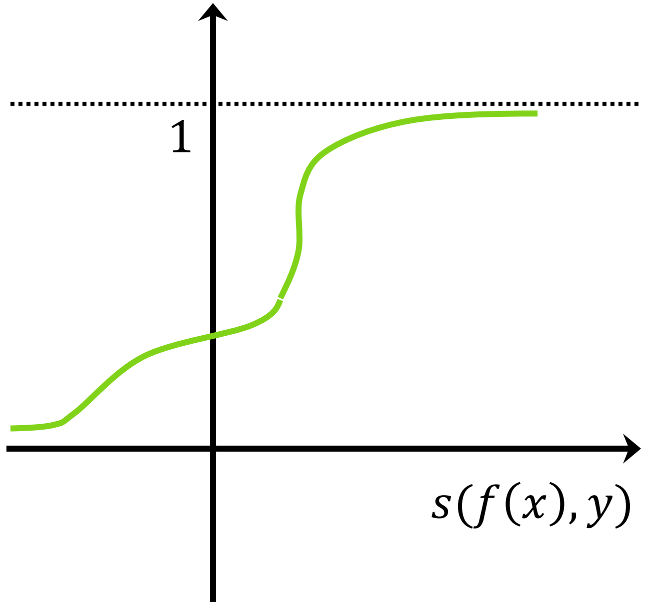
Calibration Score

$s : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}$

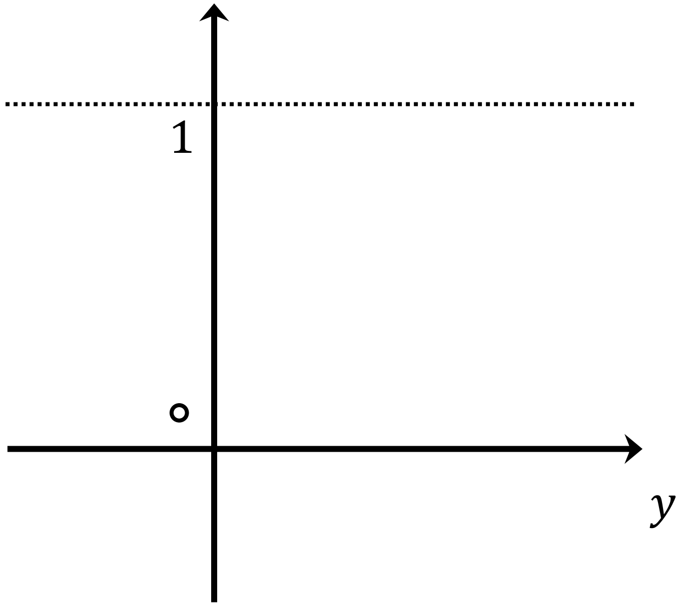


Interpolation Algorithm

$g : \mathbb{R}^n \to (\mathbb{R} \to [0,1])$



$P(Y \leq y \mid X = x_0)$



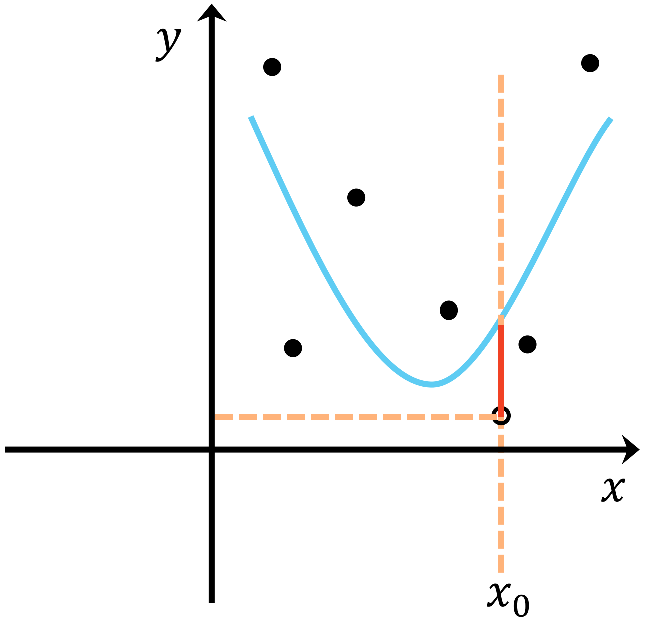
Modular Conformal Calibration: Example

Base Predictor

$f: \mathcal{X} \rightarrow \mathcal{R}$

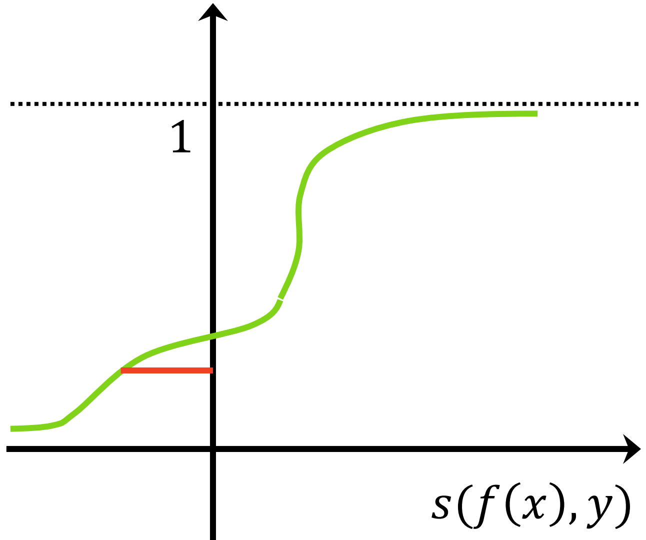
Calibration Score

$s: \mathcal{R} \times \mathcal{Y} \rightarrow \mathbb{R}$

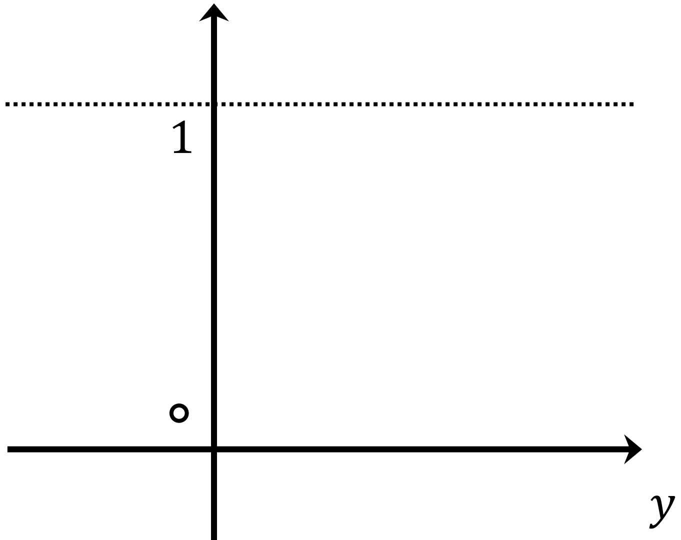


Interpolation Algorithm

$g: \mathbb{R}^n \rightarrow (\mathbb{R} \rightarrow [0,1])$



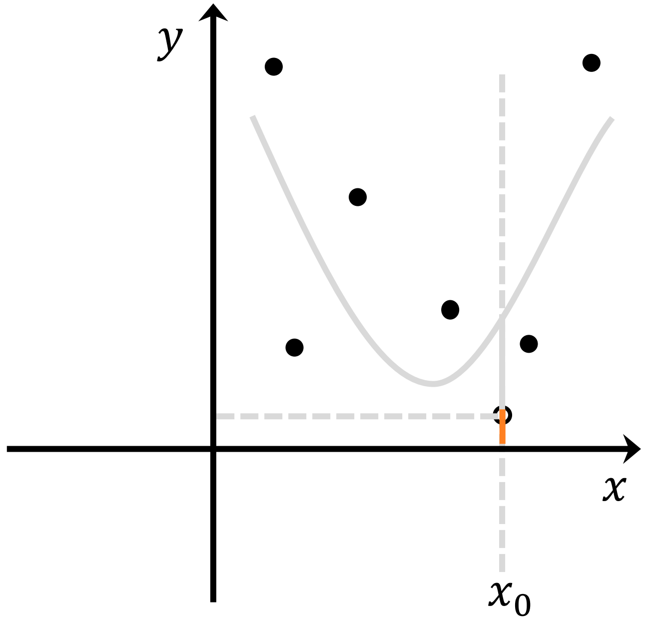
$P(Y \leq y \mid X = x_0)$

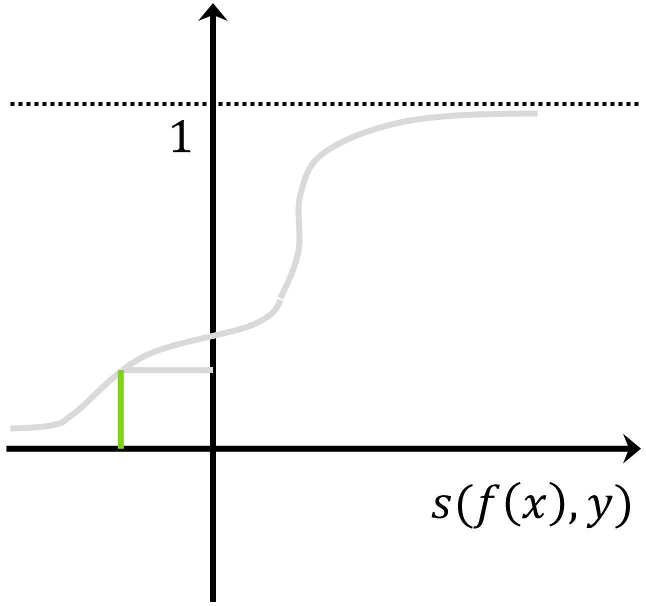


Modular Conformal Calibration: Example

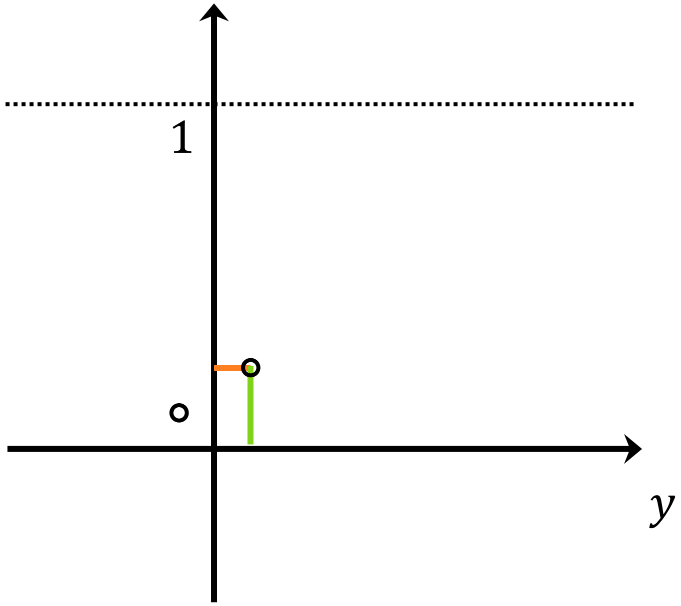
Base Predictor
$f : \mathcal{X} \to \mathcal{R}$
Calibration Score
$s : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}$

Interpolation Algorithm
$g : \mathbb{R}^n \to (\mathbb{R} \to [0,1])$





$P(Y \leq y \mid X = x_0)$



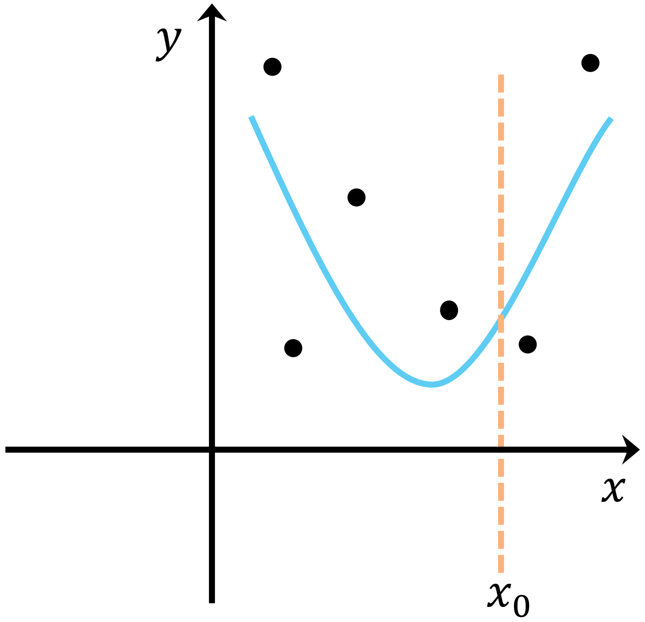
Modular Conformal Calibration: Example

Base Predictor

$f : \mathcal{X} \to \mathcal{R}$

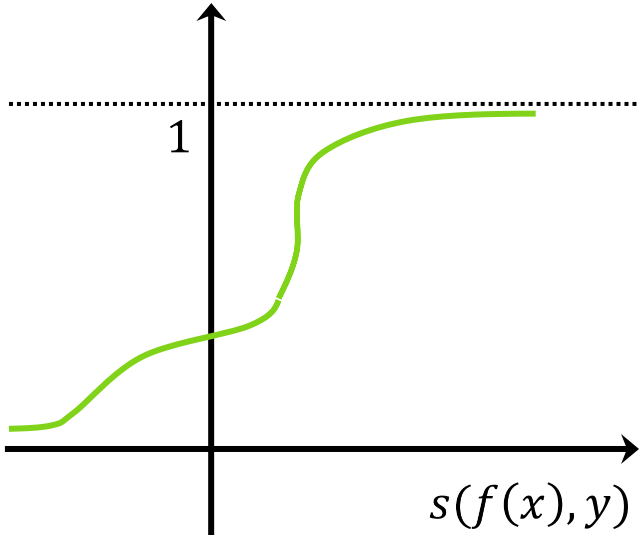
Calibration Score

$s : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}$

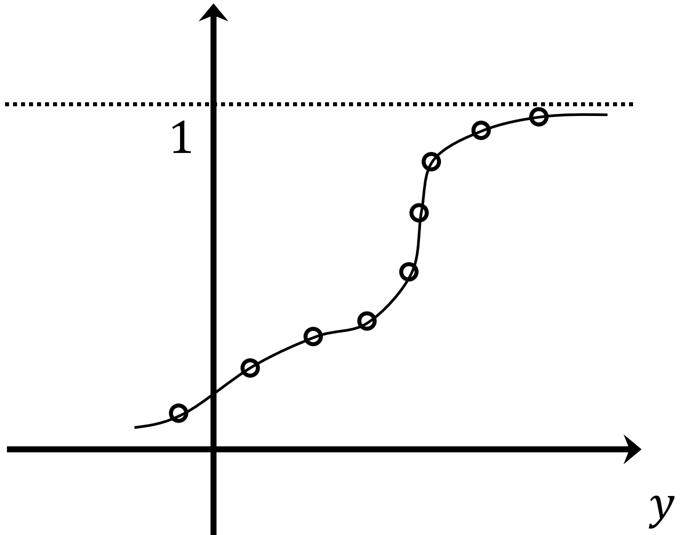


Interpolation Algorithm

$g : \mathbb{R}^n \to (\mathbb{R} \to [0, 1])$

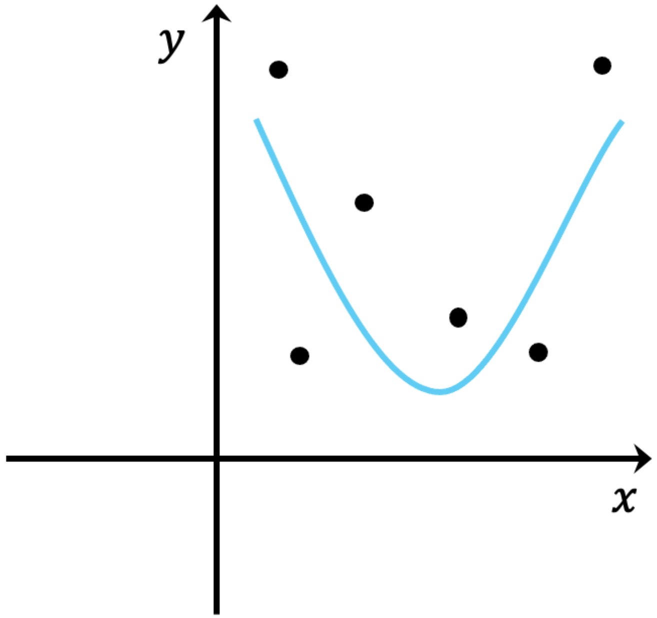


$P(Y \leq y \mid X = x_0)$

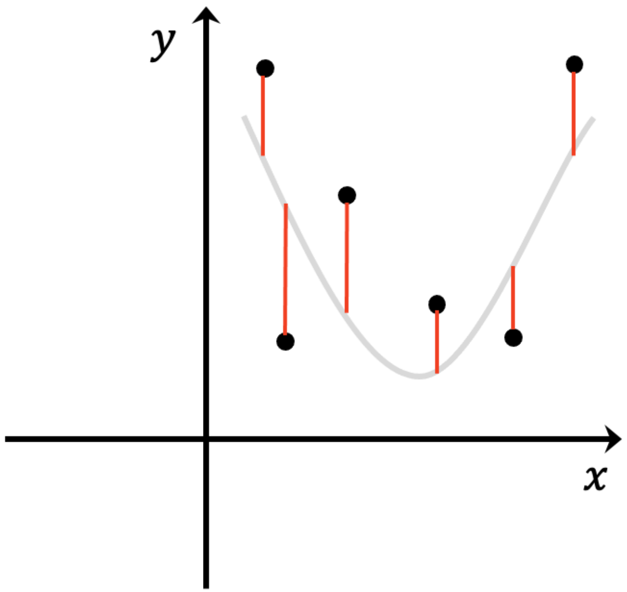


A Modular Design

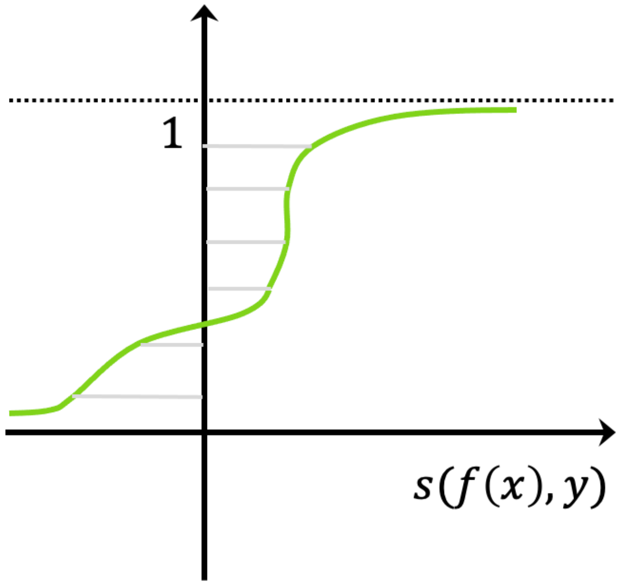
Base Predictor



Calibration Score



Interpolation Algorithm



Modular Conformal Calibration

Charles Marx [*1] Shengjia Zhao [*1] Willie Neiswanger [1] Stefano Ermon [1]

Abstract

Uncertainty estimates must be calibrated (i.e., accurate) and sharp (i.e., informative) in order to be useful. This has motivated a variety of methods for *recalibration*, which use held-out data to turn an uncalibrated model into a calibrated model. However, the applicability of existing methods is limited due to their assumption that the original model is also a probabilistic model. We introduce a versatile class of algorithms for recalibration in regression that we call *modular conformal calibration* (MCC). This framework allows one to transform any regression model into a calibrated probabilistic model. The modular design of MCC allows us to make simple adjustments to existing algorithms that enable well-behaved distribution predictions. We also provide finite-sample calibration guarantees for MCC algorithms. Our framework recovers isotonic recalibration, conformal calibration, and conformal interval prediction, implying that our theoretical results apply to those methods as well. Finally, we conduct an empirical study of MCC on 17 regression datasets. Our results show that new algorithms designed in our framework achieve near-perfect calibration and improve sharpness relative to existing methods.

1. Introduction

Uncertainty estimates can inform human decisions (Pratt et al., 1995; Berger, 2013), flag when an automated decision system requires human review (Kang et al., 2021), and serve as an internal component of automated systems. For example, uncertainty informs treatment decisions in medicine (Begoli et al., 2019) and supports safety in autonomous navigation (Michelmore et al., 2018). In such settings, the benefits of accounting for uncertainty hinge on our ability to produce *calibrated* uncertainty estimates—e.g., of those events to which one assigns a probability of 90%, the events should indeed occur 90% of the time. A model that is not calibrated can consistently make confident predictions that are incorrect.

Many models, such as neural networks (Guo et al., 2017) and Gaussian processes (Rasmussen, 2003; Tran et al., 2019), achieve high accuracy but have poorly calibrated or absent uncertainty estimates. In other cases, a pretrained model is released for wide use and it is difficult to guarantee that it will produce calibrated uncertainty estimates in new settings (Zhao et al., 2021). This leads us to the question: *how can we safely deploy models with high predictive value but poor or absent uncertainty estimates?*

These challenges have motivated work on *recalibration*, whereby a model with poor uncertainty estimates is transformed into a probabilistic model that outputs calibrated probabilities (Kuleshov et al., 2018; Vovk et al., 2020; Niculescu-Mizil & Caruana, 2005; Chung et al., 2021). Recalibration methods are attractive because they require only black-box access to a given model and can return well-calibrated probabilistic predictions.

However, calibration is not the only goal of probabilistic models. It is also important for a probabilistic model to predict sharp (i.e., low variance) distributions to convey more information. Furthermore, recalibration methods need to be data efficient to calibrate models in data poor regimes.

In this paper, we introduce *modular conformal calibration* (MCC), a class of algorithms that unifies existing recalibration methods and gives well-behaved distribution predictions from any model. Our main contributions are:

1. We introduce modular conformal calibration, a class of algorithms for recalibration in regression, which can be applied to recalibrate almost any regression model. MCC unifies isotonic calibration (Kuleshov et al., 2018), conformal calibration (Vovk et al., 2020), and conformal interval prediction (Vovk et al., 2005) under a single theoretical framework, and additionally leads to new algorithms.
2. We provide finite-sample calibration guarantees, showing that MCC can achieve ϵ calibration error with $O(1/\epsilon)$ samples. These results also apply to the afore-

[*]Equal contribution [1]Computer Science Department, Stanford University. Correspondence to: Charles Marx <ctmarx@stanford.edu>.

Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.

Thank you!

Paper

Modular Conformal Calibration

https://arxiv.org/abs/2206.11468

Twitter

@CharlieTMarx

Contact

Charlie Marx	http://charliemarx.github.io
Shengjia Zhao	https://szhao.me/
Willie Neiswanger	https://willieneis.github.io/
Stefano Ermon	https://cs.stanford.edu/~ermon/