Poster #54
TODAY 10:45am

Verified Uncertainty Calibration

Ananya Kumar, Percy Liang, Tengyu Ma

Stanford University




Uncertainties - Beyond Model Accuracy

Tumor size
Teratoma
...

20% chance of cancer

Reality: 40% such people have cancer (!)
Implication: Wrong Treatment

- Testicular cancer (Calster & Vickers), Bipolar disorder (Lindhiem et al), Criminal recidivism (Fazel et al)

Miscalibration of Neural Networks

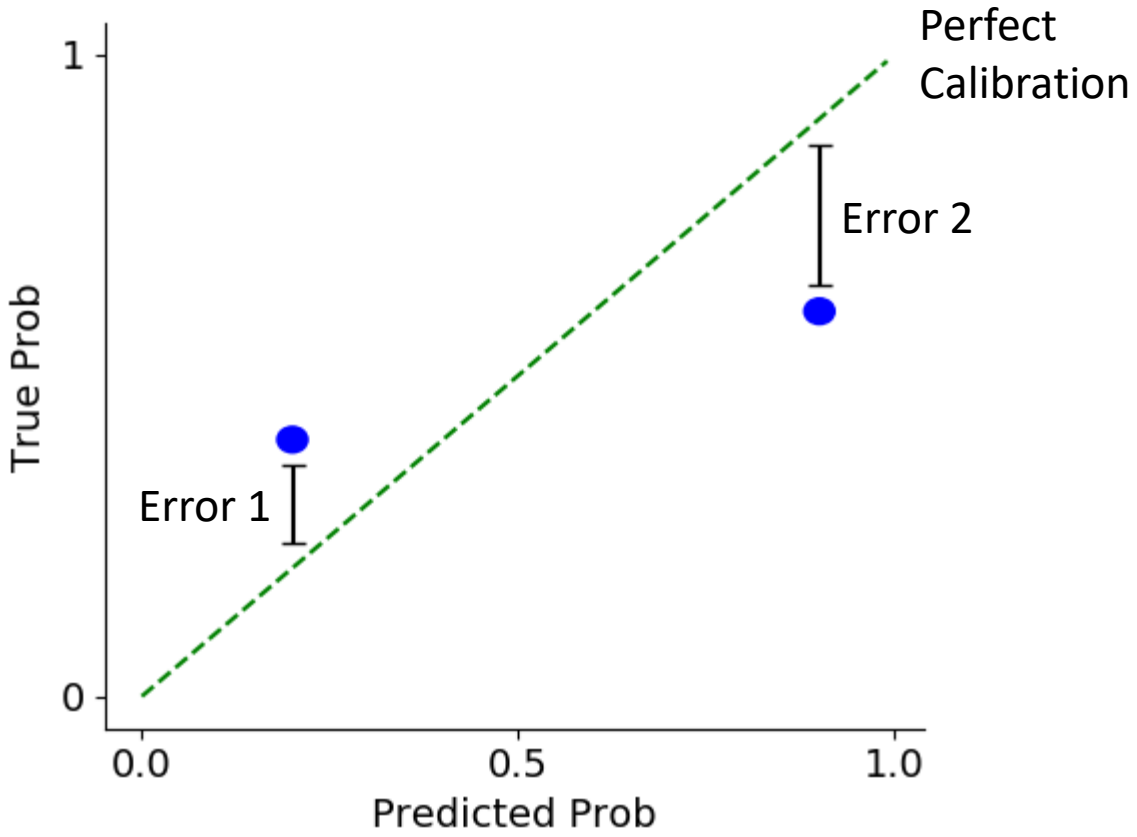
Resnet on CIFAR-100

Model's perceived accuracy	90%
Actual accuracy	70%

Cite: Guo et al 2017

Calibration Error (CE)

- Average difference between model's predicted prob and true prob

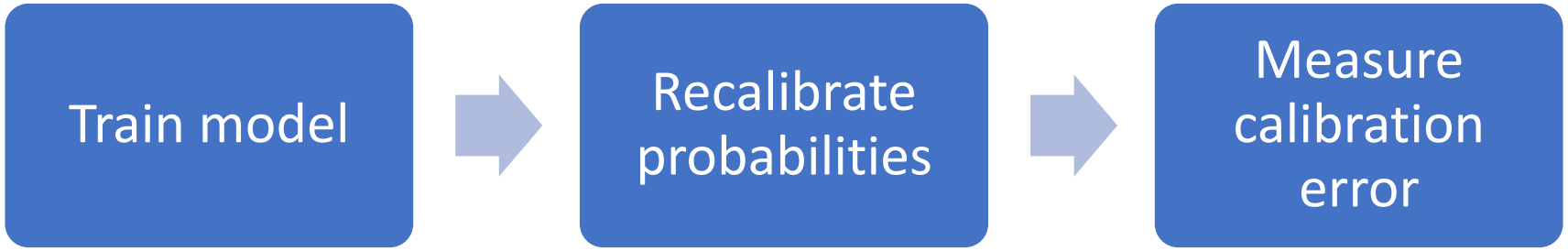


Details: $CE = \sqrt{E[(m-p)^2]}$

m is predicted prob

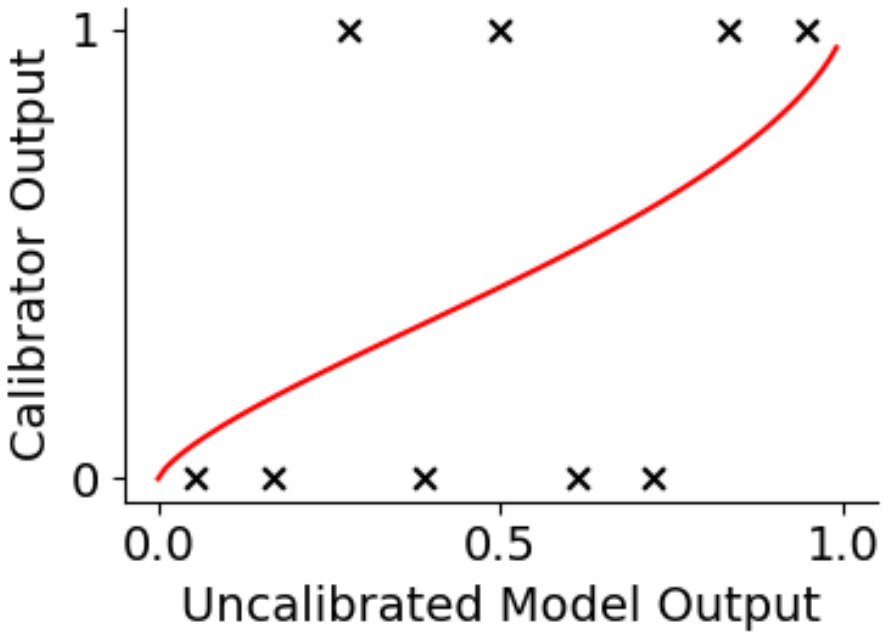
p is true prob

Recalibration Pipeline



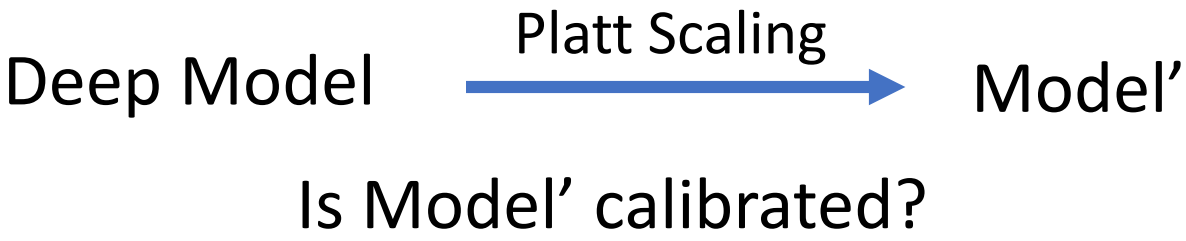
Platt Scaling

- Platt scaling, temperature scaling scale the model probabilities to improve them

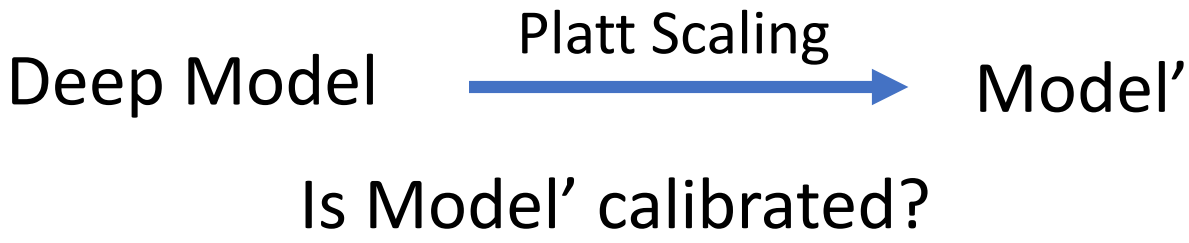


X = labels

1. Is Platt Scaling calibrated?



1. Is Platt Scaling calibrated?



- Prior work reports calibration error = 2%
- We show that calibration error greater than 4%

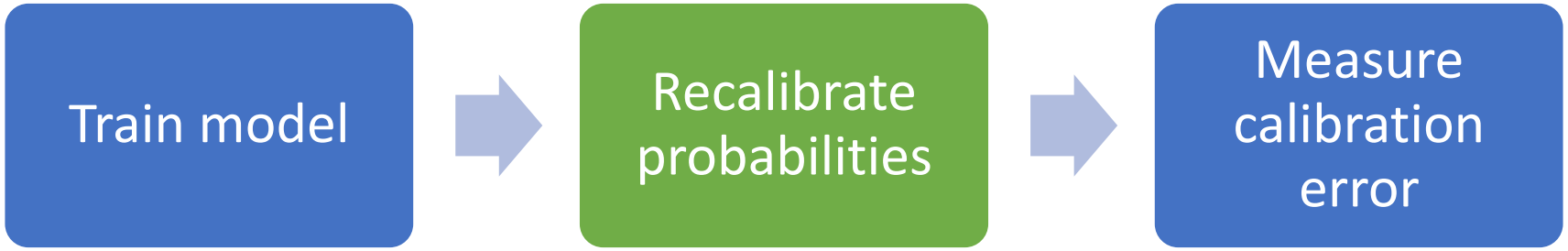
1. Is Platt Scaling calibrated?

Is Model' calibrated?

- Prior work reports calibration error = 2%
- We show that calibration error greater than 4%

Impossible to measure
calibration error of scaling

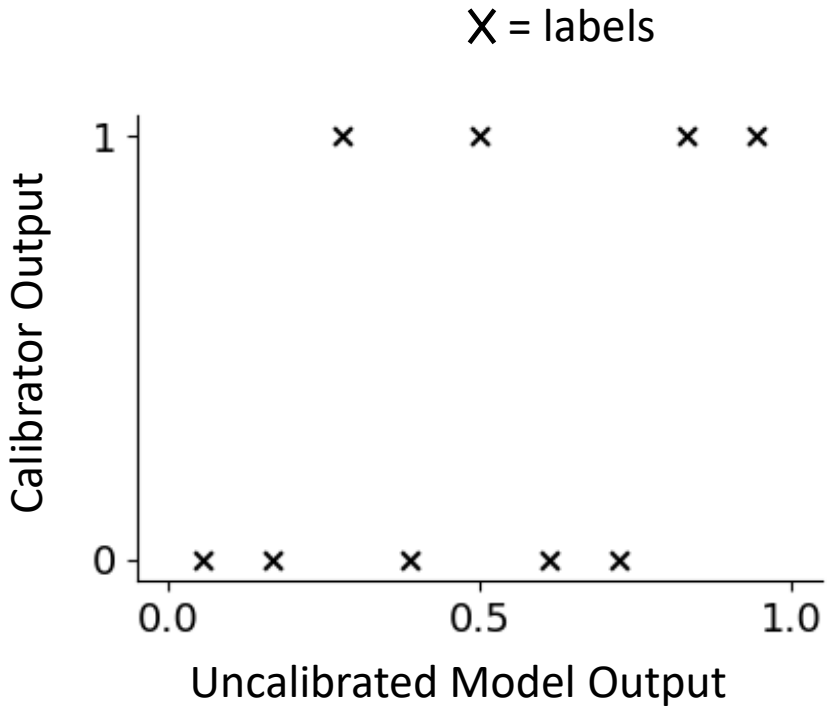
Recalibration Pipeline



Should be able to tell how calibrated we are!

2. Scaling-Binning Calibrator

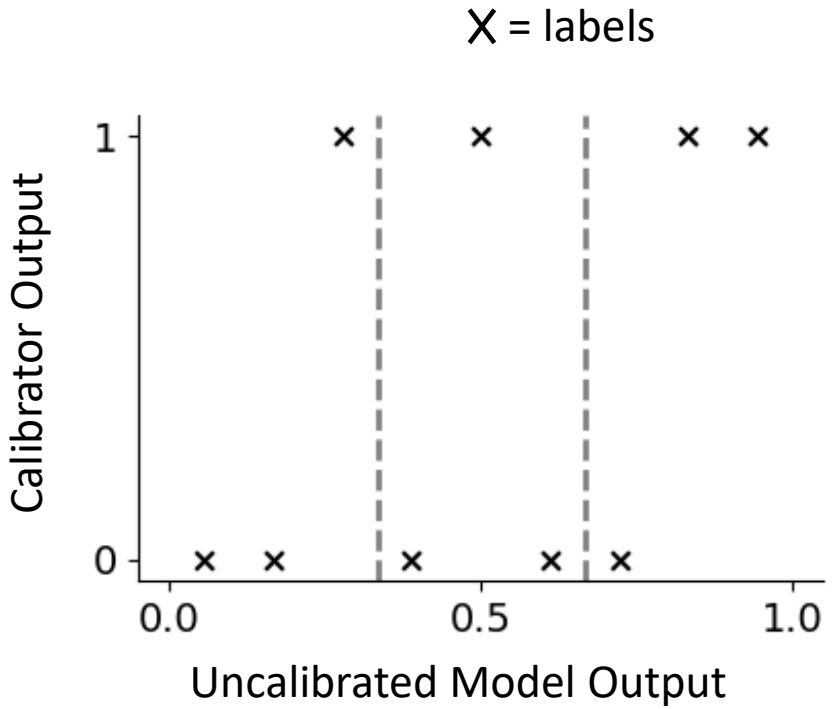
- Histogram binning outputs average label value in each bin



(a) Histogram

2. Scaling-Binning Calibrator

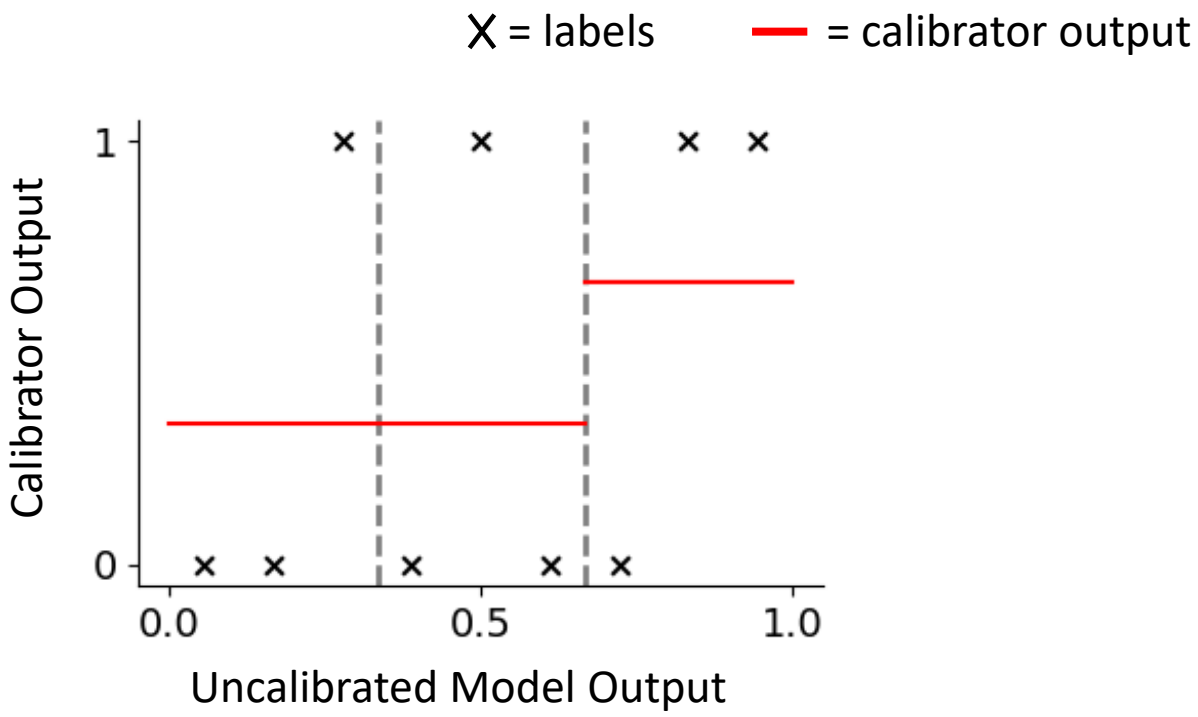
- Histogram binning outputs average label value in each bin



(a) Histogram

2. Scaling-Binning Calibrator

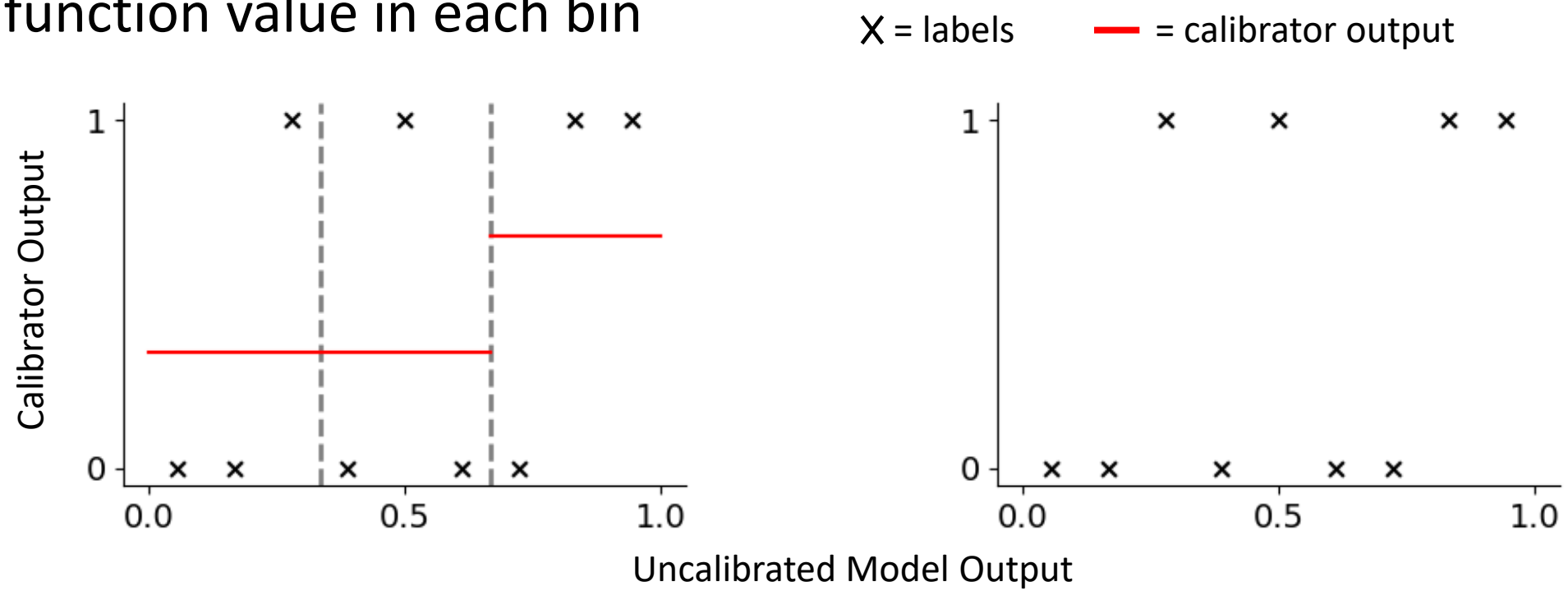
- Histogram binning outputs average label value in each bin



(a) Histogram

2. Scaling-Binning Calibrator

- Scaling-binning calibrator fits a function to data, and outputs average function value in each bin

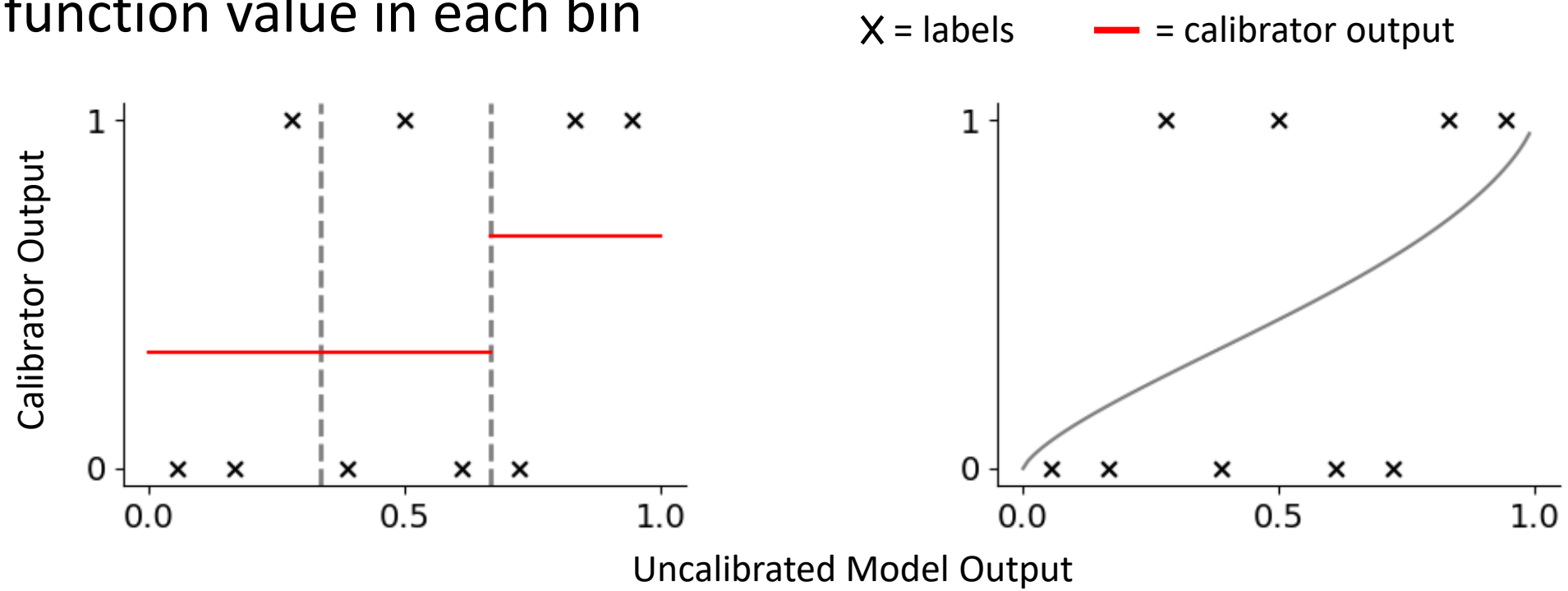


(a) Histogram

(b) Scaling-binning (ours)

2. Scaling-Binning Calibrator

- Scaling-binning calibrator fits a function to data, and outputs average function value in each bin

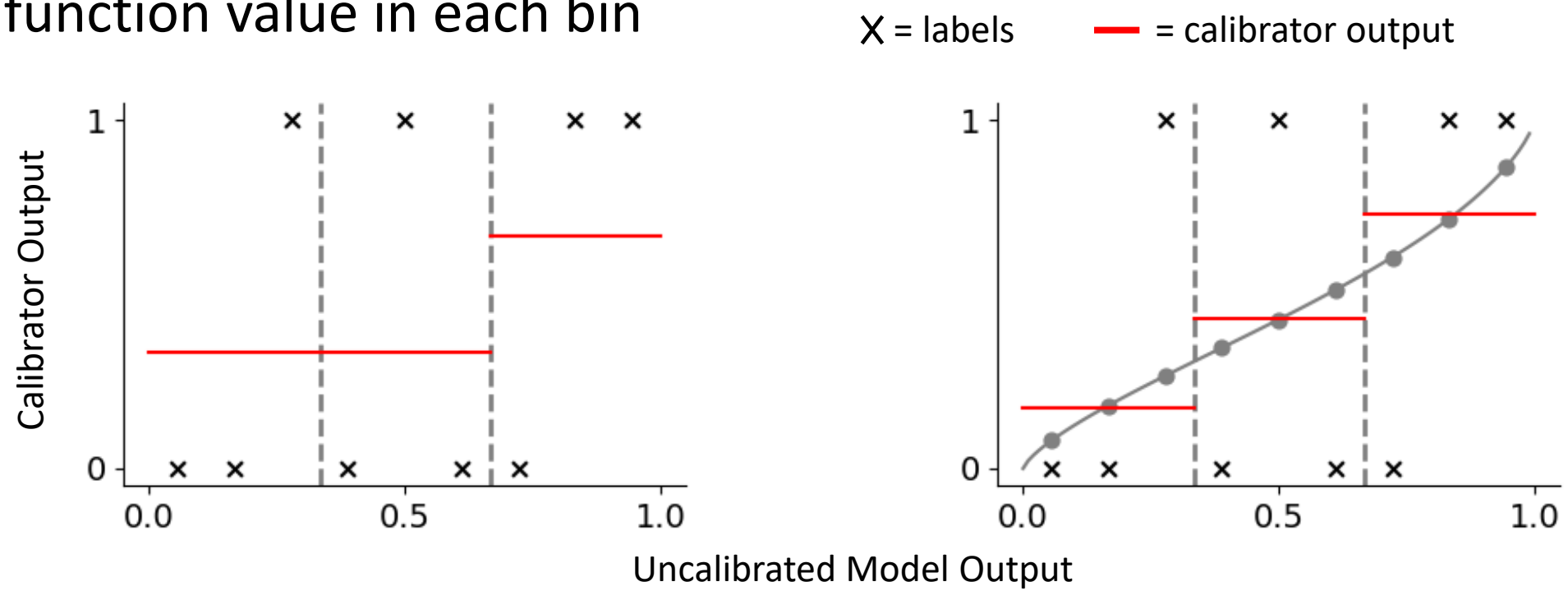


(a) Histogram

(b) Scaling-binning (ours)

2. Scaling-Binning Calibrator

- Scaling-binning calibrator fits a function to data, and outputs average function value in each bin

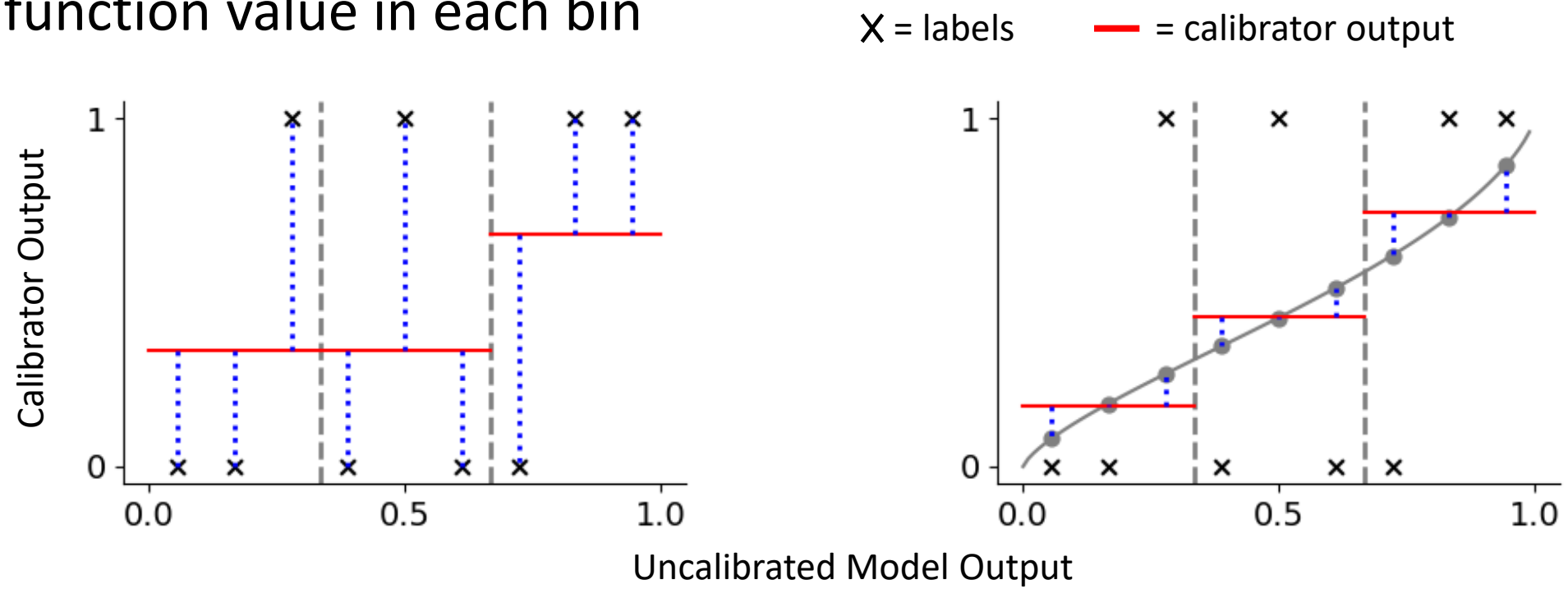


(a) Histogram

(b) Scaling-binning (ours)

2. Scaling-Binning Calibrator

- Scaling-binning calibrator fits a function to data, and outputs average function value in each bin



(a) Histogram

(b) Scaling-binning (ours)

2. Scaling-Binning Calibrator

Recalibration Method	*Samples Needed*	*Can Estimate Calibration?*
Platt Scaling	**Few**: $O\left(\frac{1}{\varepsilon^2}\right)$	✗

B = # Bins
ε = desired CE

2. Scaling-Binning Calibrator

Recalibration Method	*Samples Needed*	*Can Estimate Calibration?*
Platt Scaling	**Few**: $O\left(\frac{1}{\varepsilon^2}\right)$	✗
Histogram Binning	More: $O\left(\frac{B}{\varepsilon^2}\right)$	✓

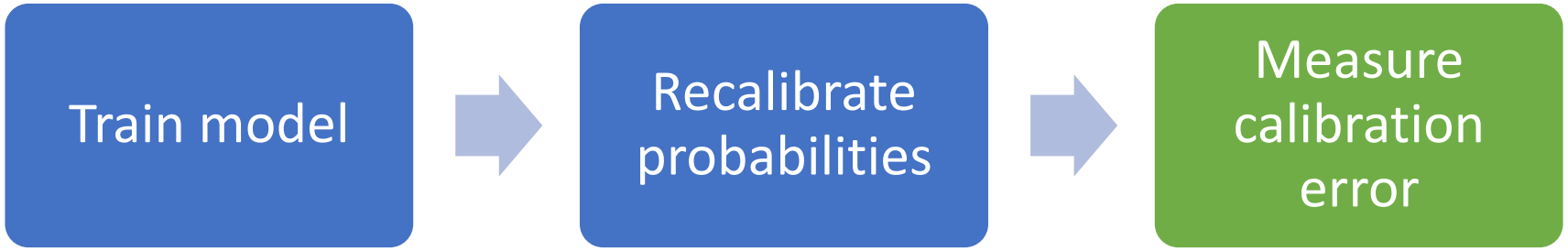
B = # Bins
ε = desired CE

2. Scaling-Binning Calibrator

Recalibration Method	*Samples Needed*	*Can Estimate Calibration?*
Platt Scaling	**Few**: $O\left(\frac{1}{\varepsilon^2}\right)$	✗
Histogram Binning	More: $O\left(\frac{B}{\varepsilon^2}\right)$	✓
Scaling-Binning (Ours)	**Few**: $O\left(\frac{1}{\varepsilon^2} + B\right)$	✓

B = # Bins
ε = desired CE

Recalibration Pipeline

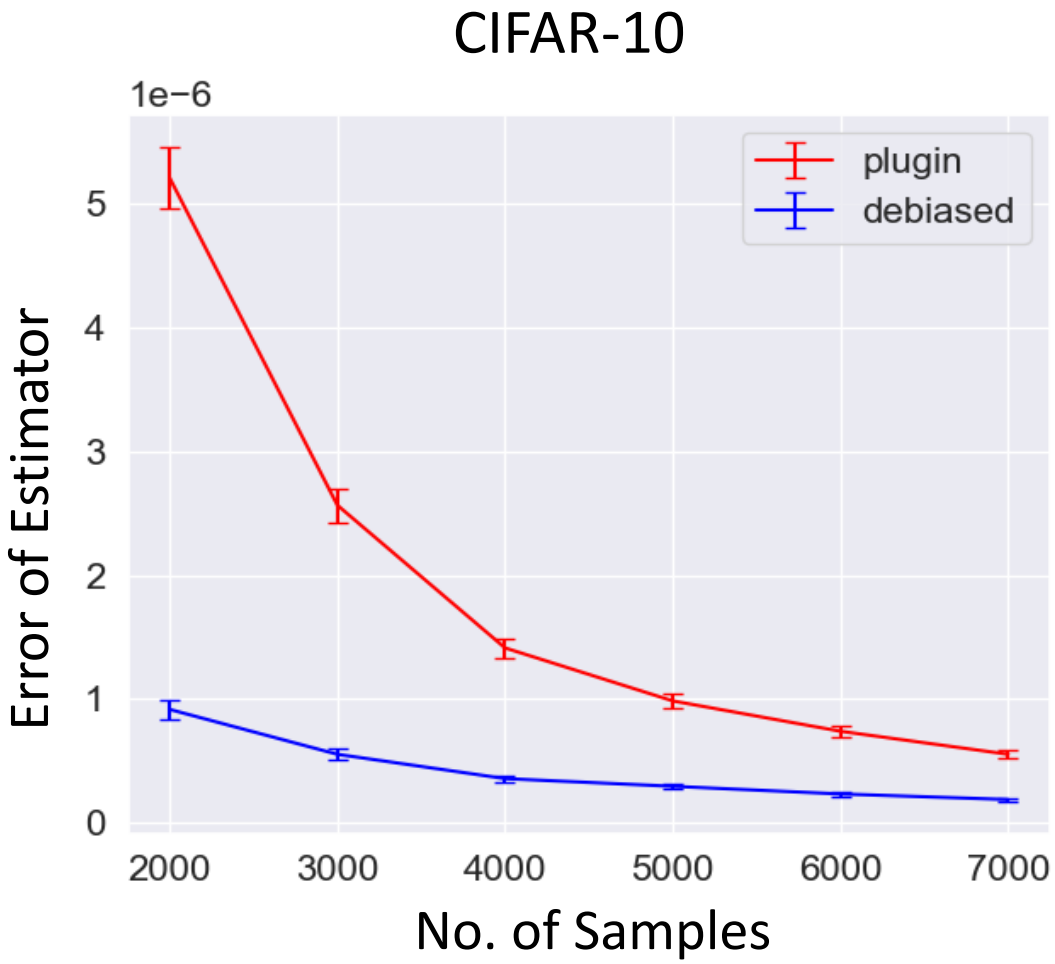


3. Verifying Calibration

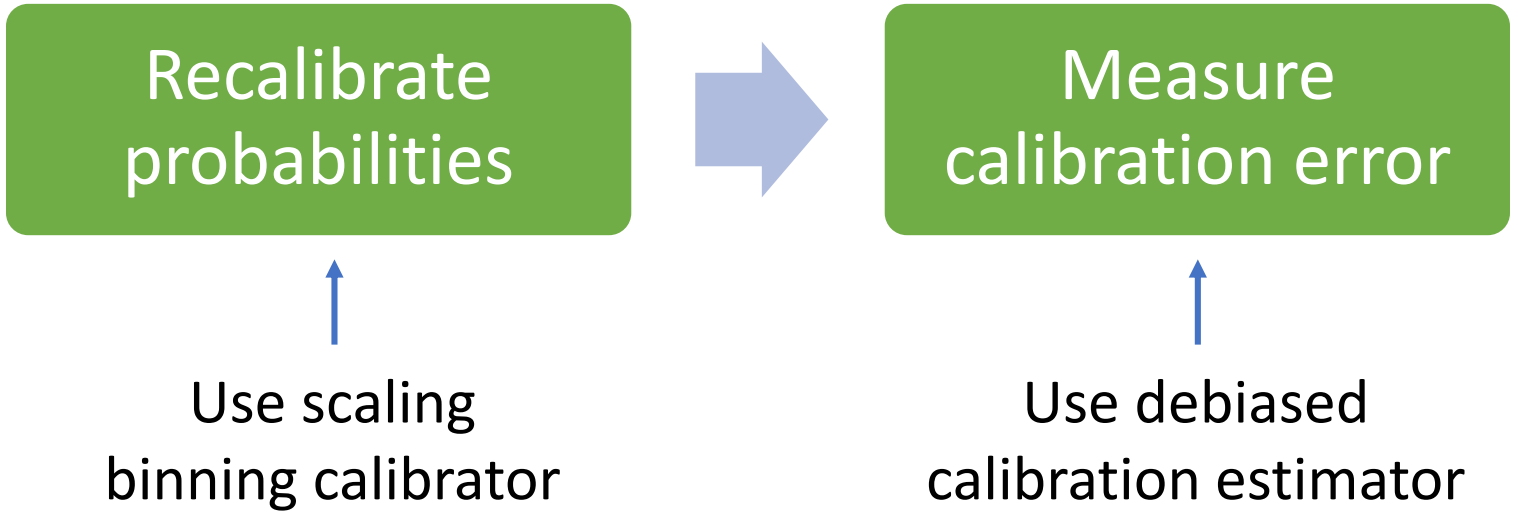
- Plugin: standard, everyone uses it (just average)
- Debiased: more sophisticated estimator from meteorology community

Estimation Method	*Samples Needed*
Plugin	More: $O\left(\frac{B}{\varepsilon^2}\right)$
Debiased	**Fewer: $\boldsymbol{O\left(\frac{\sqrt{B}}{\varepsilon^2}\right)}$**

3. Verifying Calibration



Takeaways



- For scaling methods: can only lower bound calibration error
  - Still use debiased estimator: estimates lower bound with fewer samples

Calibration Library

- Code at https://github.com/AnanyaKumar/verified_calibration
- Measure model accuracy **and calibration**

```
pip install uncertainty-calibration
```

```
import calibration as cal
ce = cal.get_calibration_error(logits, labels)
```

Poster #54

East Exhibition Hall B + C

TODAY 10:45am