Open Problem:
Does an Efficient Calibrated Forecasting Strategy Exist?

Jacob Abernethy and Shie Mannor
UC Berkeley and The Technion

Sequential Prediction Problem:

Prediction:

Outcome:

Sequential Prediction Problem:

Prediction:	0.25	0.91	0.43	0.07	....
Outcome:	**0**	**1**	**1**	**0**	

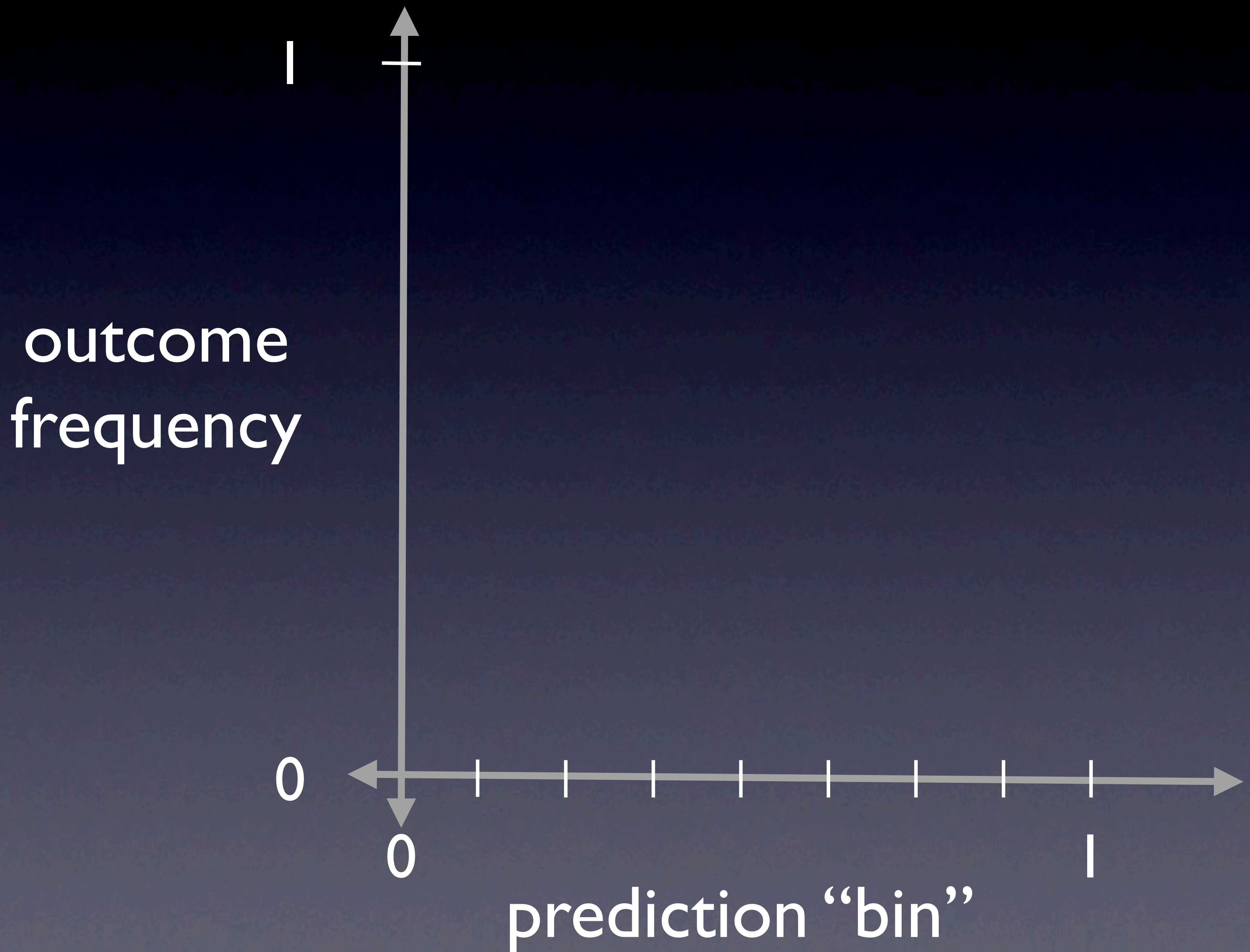
Calibrated Forecasting

- Input: a 0/1 bit sequence z_1, z_2, z_3... (i.e. rain, shine, shine, rain)
- Want: "good" probability predictions p1, p2, p3, ...
- "Good" means "for all the times I said 40%, it rained roughly 2/5 of the time", and the same for 10%, 20%, etc.

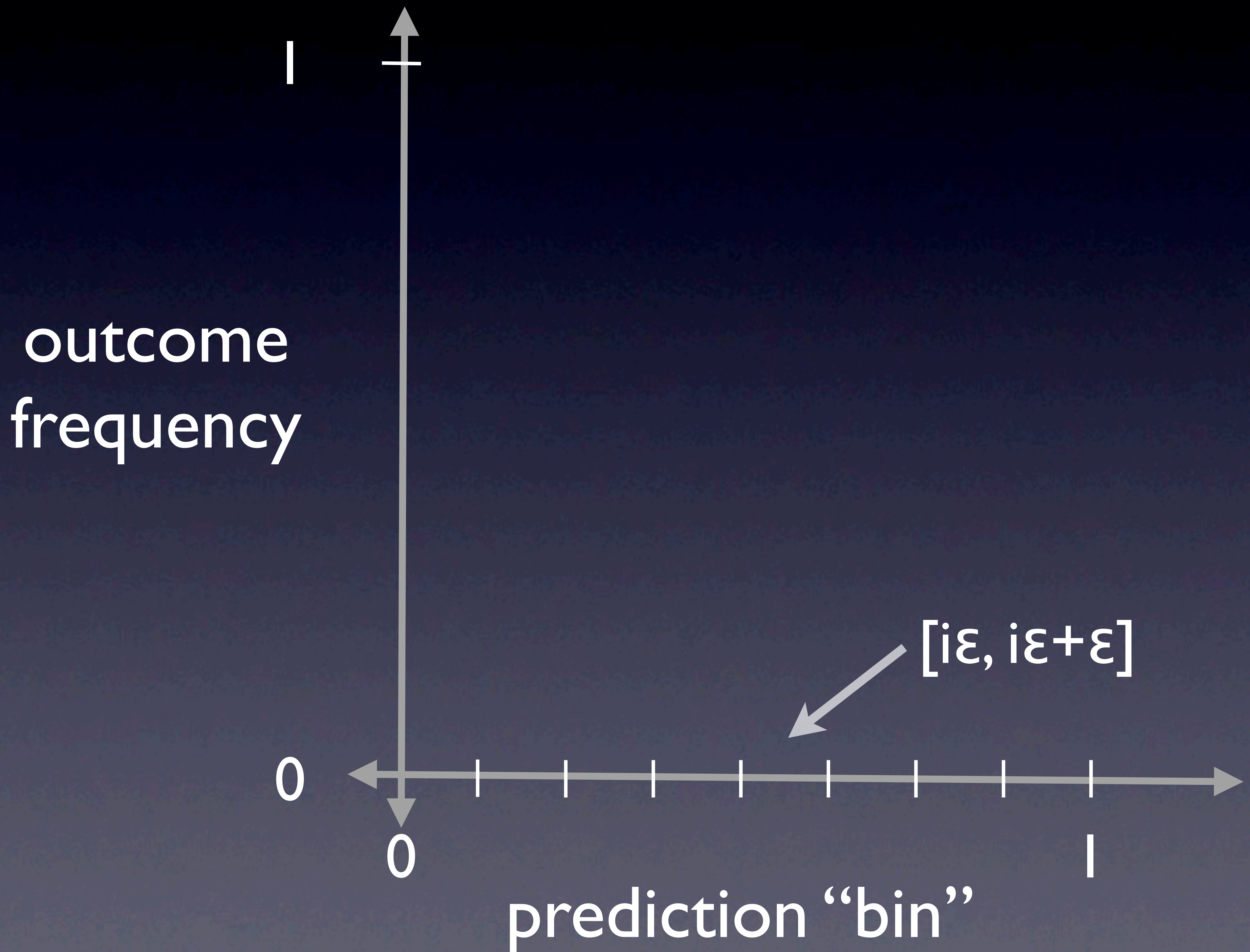
Are weather forecasts calibrated?

Prediction	Frequency
0%	7.9%
10%	5.3%
20%	10.8%
30%	19.2%
40%	26.5%
50%	27.8%
60%	46.2%
70%	58.0%
80%	58.1%
90%	63.6%
100%	66.7%

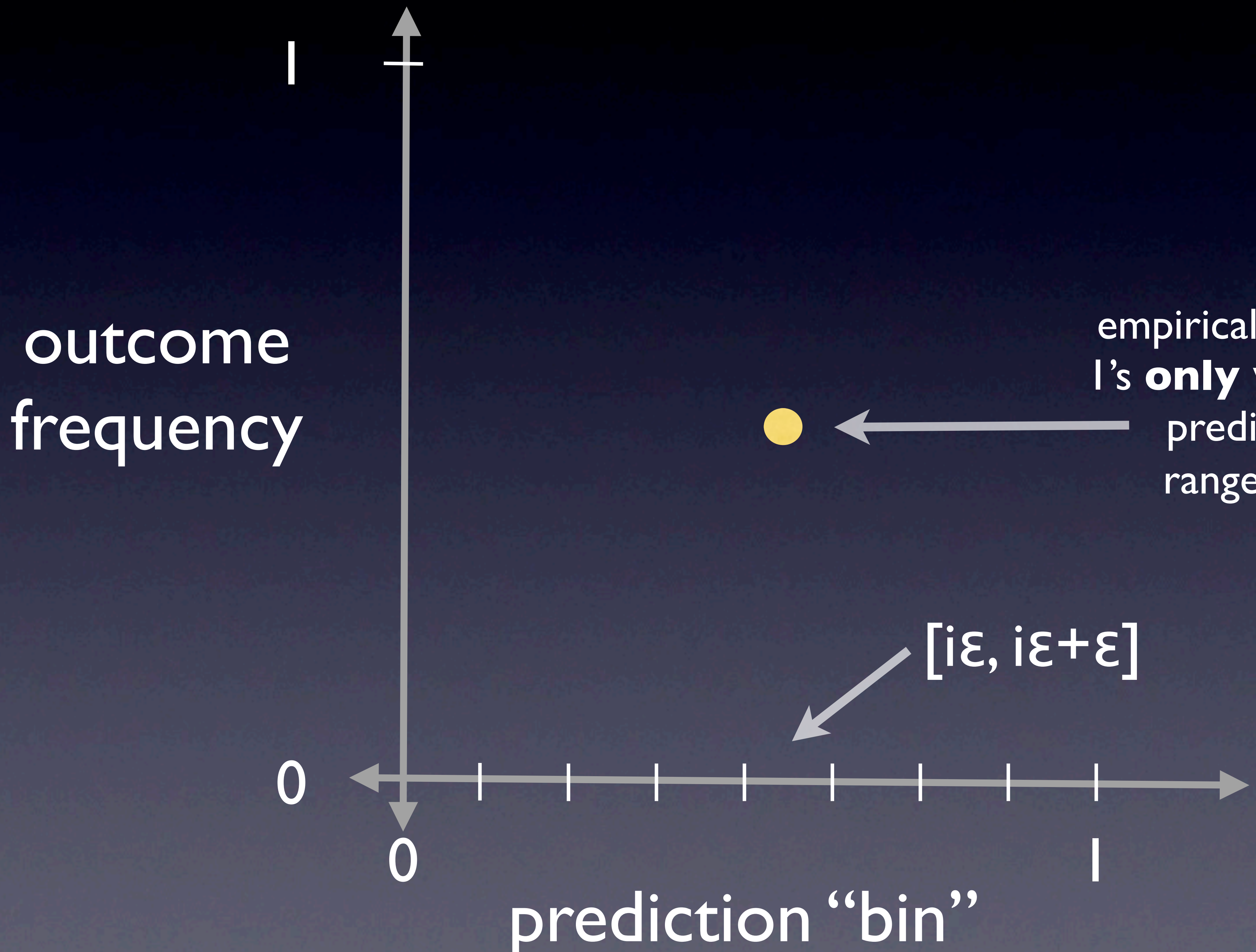
(Source: freakonomics blog)



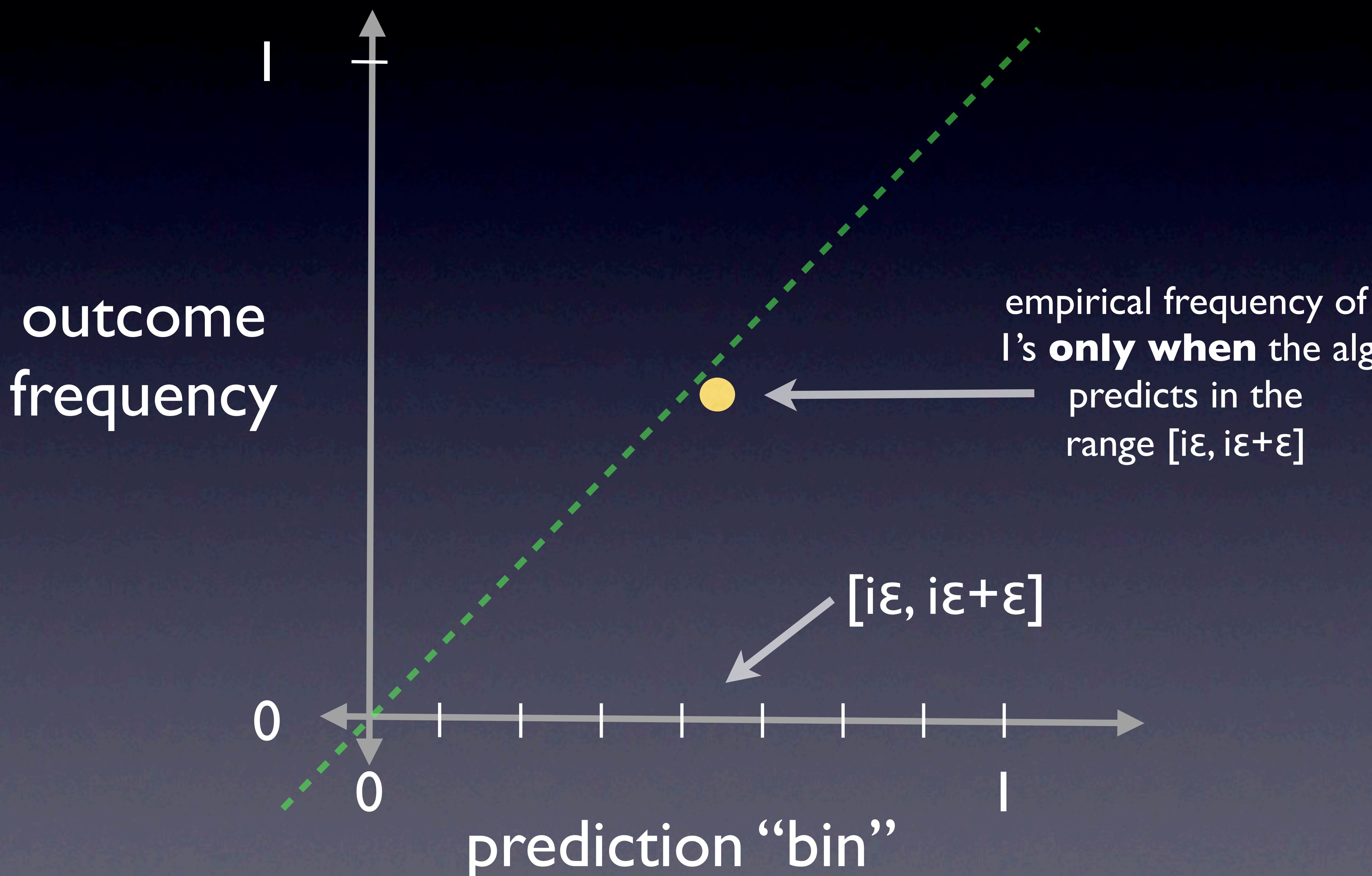
1
outcome
frequency
0
0
1
prediction “bin”



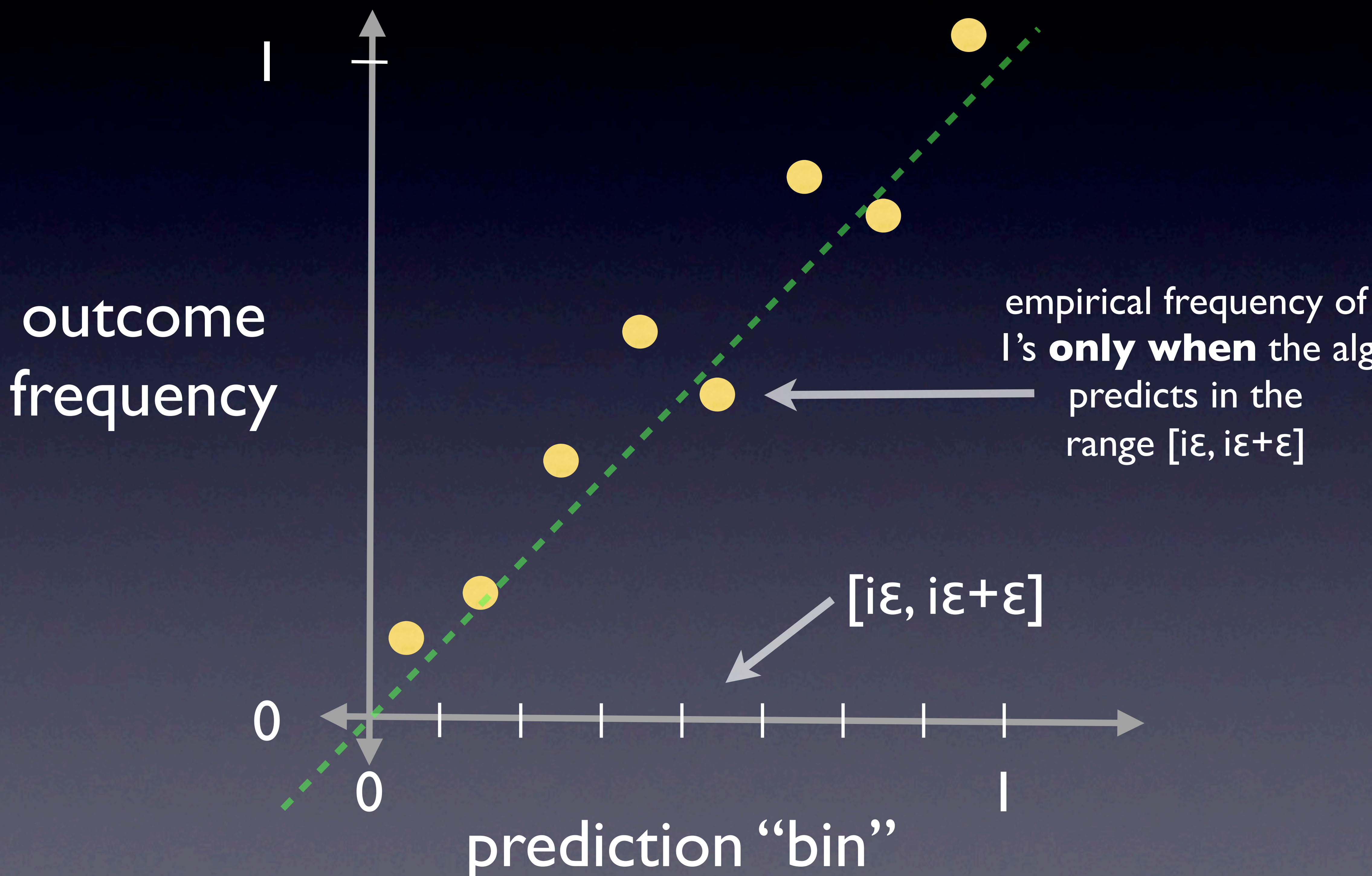
1
outcome
frequency
0
[iε, iε+ε]
0
1
prediction “bin”



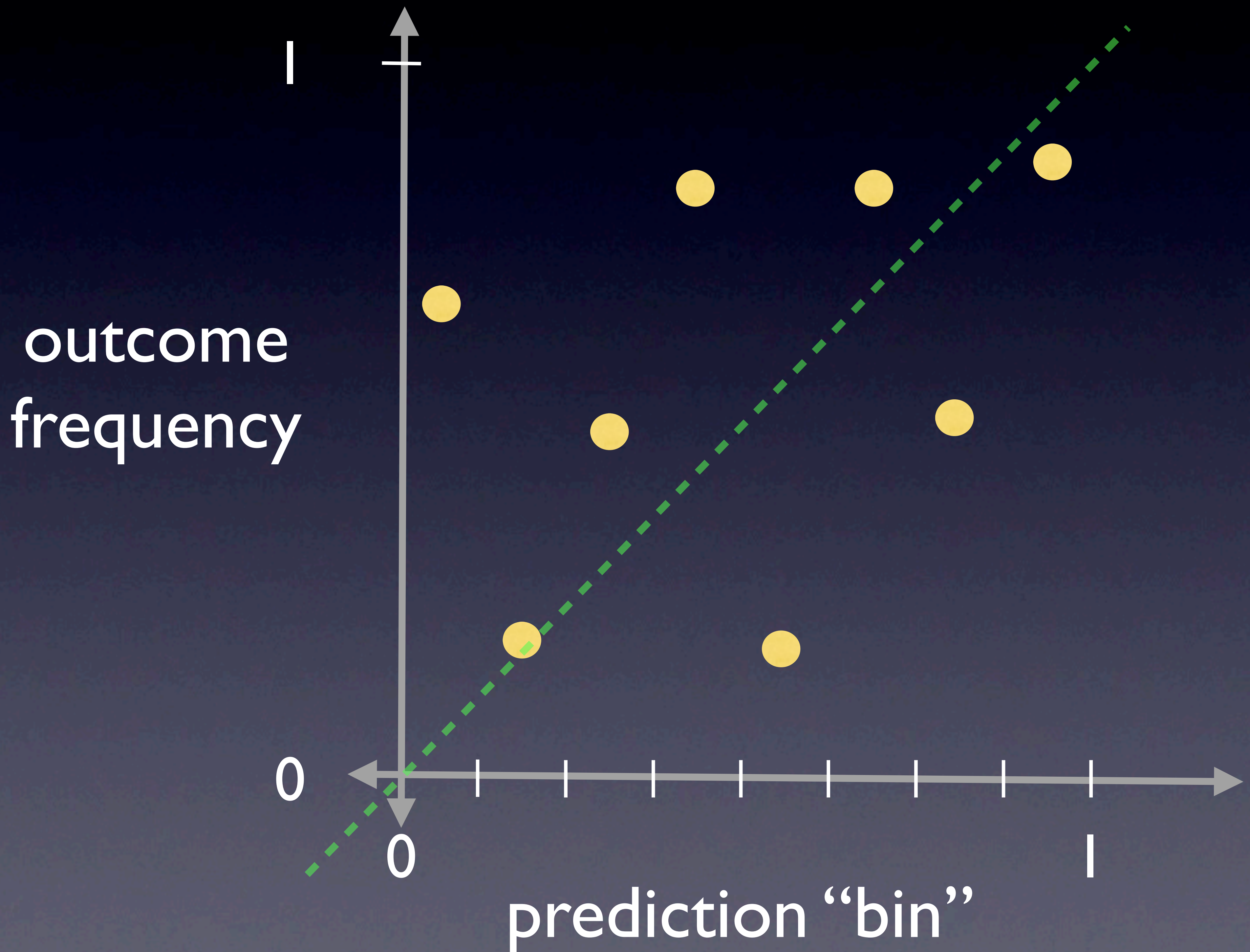
1
outcome
frequency
empirical frequency of
1's **only when** the alg
predicts in the
range [iε, iε+ε]
[iε, iε+ε]
0
0
1
prediction "bin"


1
outcome
frequency
empirical frequency of
1's **only when** the alg
predicts in the
range [iε, iε+ε]
[iε, iε+ε]
0
0
1
prediction "bin"

Well-calibrated Forecaster



Poorly-calibrated Forecaster



Calibration + Adversarial Data

- If data is IID, calibration is **easy!**
  (strategy: predict the empirical frequency)
- If data is not IID we can **still calibrate**!
- True even against an adaptive adversary!
- But requires randomization.

Calibration Scores: Strong + Weak

A **strong** calibration score:

$$\forall p \in [0,1] : \left| \frac{1}{T} \sum_{t=1}^{T} \mathbf{I}_{p_t \in (p-\epsilon, p+\epsilon)} (p_t - z_t) \right| \to 0$$

Calibration Scores: Strong + Weak

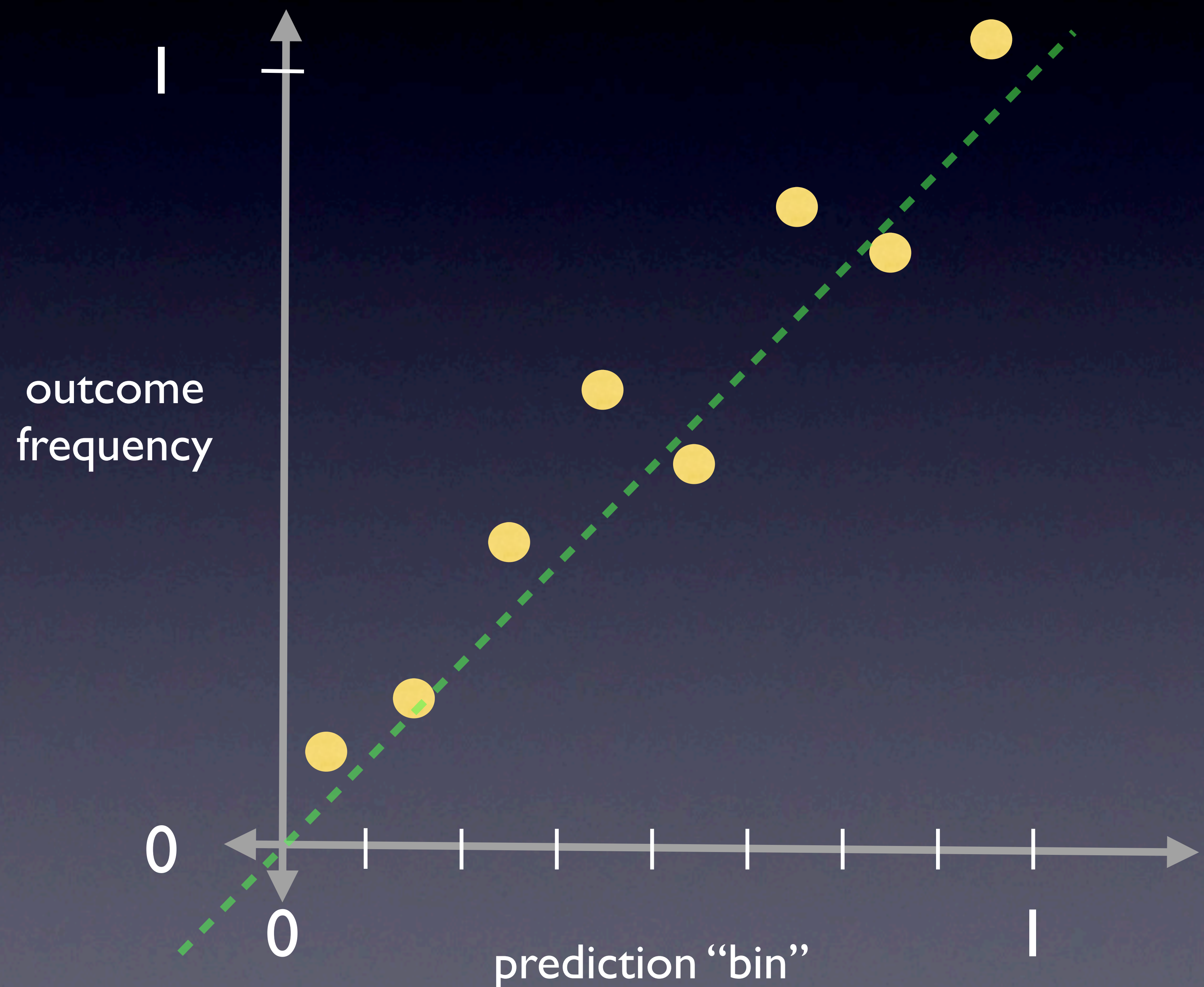
A **strong** calibration score:

$$\forall p \in [0,1] : \left| \frac{1}{T} \sum_{t=1}^{T} \mathbf{I}_{p_t \in (p-\epsilon, p+\epsilon)} (p_t - z_t) \right| \to 0$$

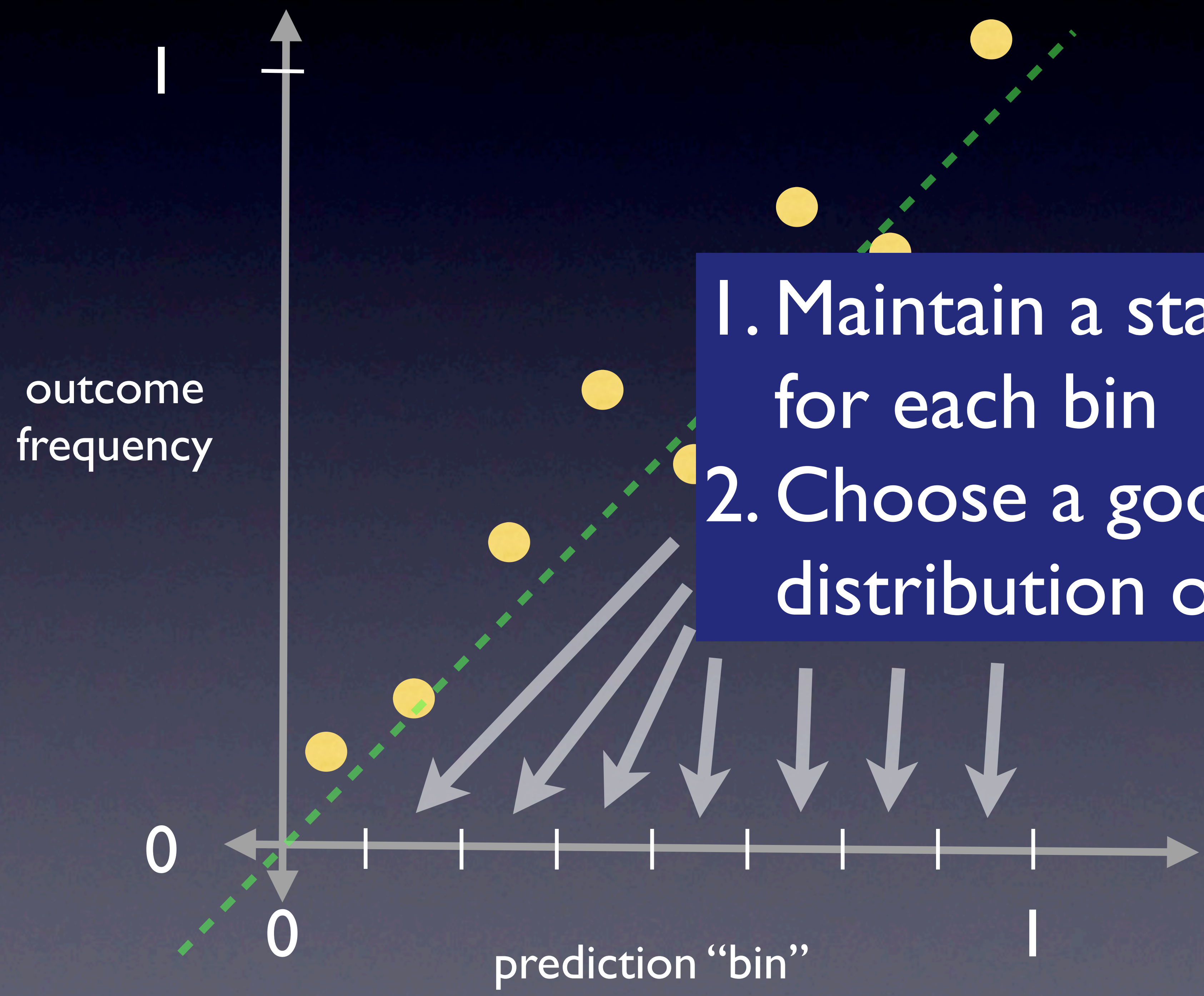
And **weak** calibration score:

$$\forall \text{ smooth fn's } g : \left| \frac{1}{T} \sum_{t=1}^{T} g(p_t)(p_t - z_t) \right| \to 0$$

Typical Calibration Alg



Typical Calibration Alg



1. Maintain a statistic for each bin
2. Choose a good distribution over bins

Challenges

- This typical approach requires $O(1/\varepsilon)$ **time** and **space**!
- This computation is required for **every prediction!**
- When #outcomes is $d > 2$, this gets *really bad*: $O(1/\varepsilon^d)$ **time** and **space**!

Good News!

	Strong Calibration	Weak Calibration
Time	$O(\log \varepsilon^{-1})$	$O(1)$
Space	$O(1/\varepsilon)$	$O(1)$
	[ABH-COLT2011]	[MSA-JASA2007]

Good News!

	Strong Calibration	Weak Calibration
Time	$O(\log \varepsilon^{-1})$	$O(1)$
Space	$O(1/\varepsilon)$	$O(1)$
	[ABH-COLT2011]	[MSA-JASA2007]

Bad News: This **only** works for **binary** case!

OPEN PROBLEM: Non-binary case, d > 2

	Strong Calibration	Weak Calibration
Time	**poly(d, log ε^{-1}) ????**	**poly(d, log ε^{-1}) ????**
Space	**poly(d, log ε^{-1}) ????**	**poly(d, log ε^{-1}) ????**

Win Cash Prizes!!!!!!!

$50 for answering each box below!

A **positive** or **negative** solution is OK!

	Strong Calibration	Weak Calibration
Time	poly(d,log ε^{-1}) ????	poly(d,log ε^{-1}) ????
Space	poly(d,log ε^{-1}) ????	poly(d,log ε^{-1}) ????

(... prizes paid
by my mother...)