“Calibeating”: Beating Forecasters at Their Own Game

Sergiu Hart

May 2026

“Calibeating”: Beating Forecasters at Their Own Game

Sergiu Hart

Center for the Study of Rationality
Dept of Mathematics   Dept of Economics
The Hebrew University of Jerusalem

hart@huji.ac.il

http://www.ma.huji.ac.il/hart

Joint work with

Dean P. Foster

University of Pennsylvania &
Amazon Research NY

Papers

Papers

- Sergiu Hart
  "Calibration: The Minimax Proof", 1995
  Matching, Dynamics and Games for the Allocation of Resources 2025

  www.ma.huji.ac.il/hart/publ.html#calib-minmax

Papers

- Sergiu Hart
  "Calibration: The Minimax Proof", 1995
  Matching, Dynamics and Games for the Allocation of Resources 2025

  www.ma.huji.ac.il/hart/publ.html#calib-minmax

- Dean P. Foster and Sergiu Hart
  "Smooth Calibration, Leaky Forecasts, Finite Recall, and Nash Dynamics"
  Games and Economic Behavior 2018

  www.ma.huji.ac.il/hart/publ.html#calib-eq

Papers

Papers

- Dean P. Foster and Sergiu Hart
  "Forecast Hedging and Calibration"
  Journal of Political Economy 2021
  www.ma.huji.ac.il/hart/publ.html#calib-int

Papers

- Dean P. Foster and Sergiu Hart
  "Forecast Hedging and Calibration"
  Journal of Political Economy 2021

  www.ma.huji.ac.il/hart/publ.html#calib-int

- Dean P. Foster and Sergiu Hart
  " 'Calibeating': Beating Forecasters at Their Own Game"
  Theoretical Economics 2023
  - Addendum 2024; Errata 2026 (arXiv)

  www.ma.huji.ac.il/hart/publ.html#calib-beat

Papers

Papers

- Dean P. Foster and Sergiu Hart
  "Proper Calibeating"
  2025
  www.ma.huji.ac.il/hart/publ.html#calib-proper

Calibration

Calibration

- Forecaster says: "***The probability of rain tomorrow is*** p"

Calibration

- Forecaster says: "***The probability of rain tomorrow is*** p"
- Forecaster is **CALIBRATED** if

Calibration

- Forecaster says: "***The probability of rain tomorrow is*** p"
- Forecaster is **CALIBRATED** if
  - for every forecast p:
    in the days when the forecast was p, the proportion of rainy days equals p

Calibration

- Forecaster says: "***The probability of rain tomorrow is*** p"
- Forecaster is **CALIBRATED** if
  - for every forecast p:
    in the days when the forecast was p, the proportion of rainy days equals p
    (or: is close to p in the long run)

Calibration

Calibration

CALIBRATION *can be guaranteed*
(no matter what the weather will be)

Calibration

CALIBRATION ***can be guaranteed***
(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]

Calibration

CALIBRATION *can be guaranteed*
(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem

Calibration

CALIBRATION ***can be guaranteed***

(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem
- . . .

Calibration

CALIBRATION ***can be guaranteed***

(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem
- . . .
- Hart and Mas-Colell 1996 [publ 2000]: procedure by Blackwell's Approachability

Calibration

CALIBRATION *can be guaranteed*
(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem
- . . .
- Hart and Mas-Colell 1996 [publ 2000]: procedure by Blackwell's Approachability
- Foster 1999: simple procedure

Calibration

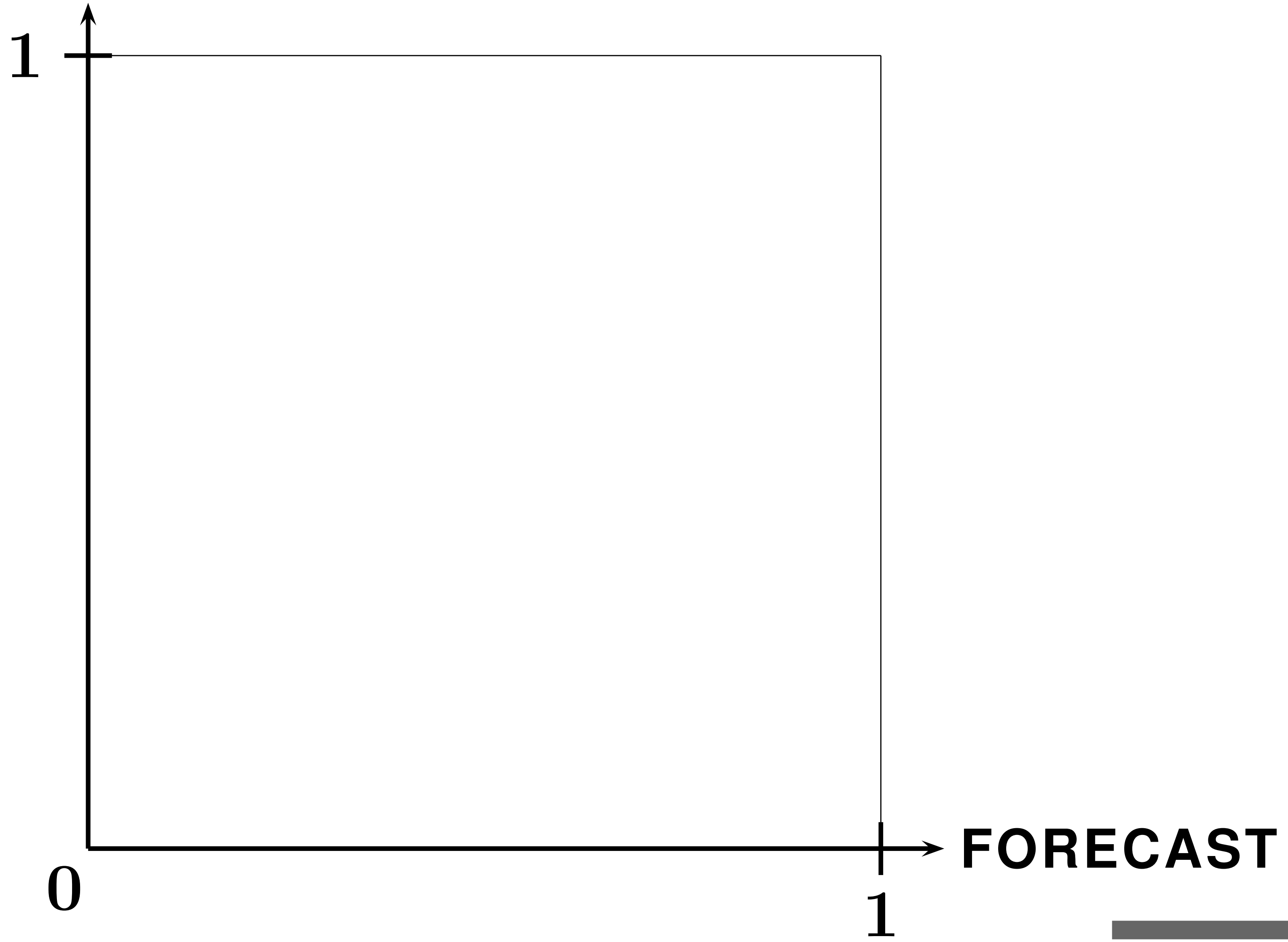
CALIBRATION ***can be guaranteed***
(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem
- . . .
- Hart and Mas-Colell 1996 [publ 2000]: procedure by Blackwell's Approachability
- Foster 1999: simple procedure
- Foster and Hart 2016 [publ 2021]: simplest procedure, by "Forecast Hedging"

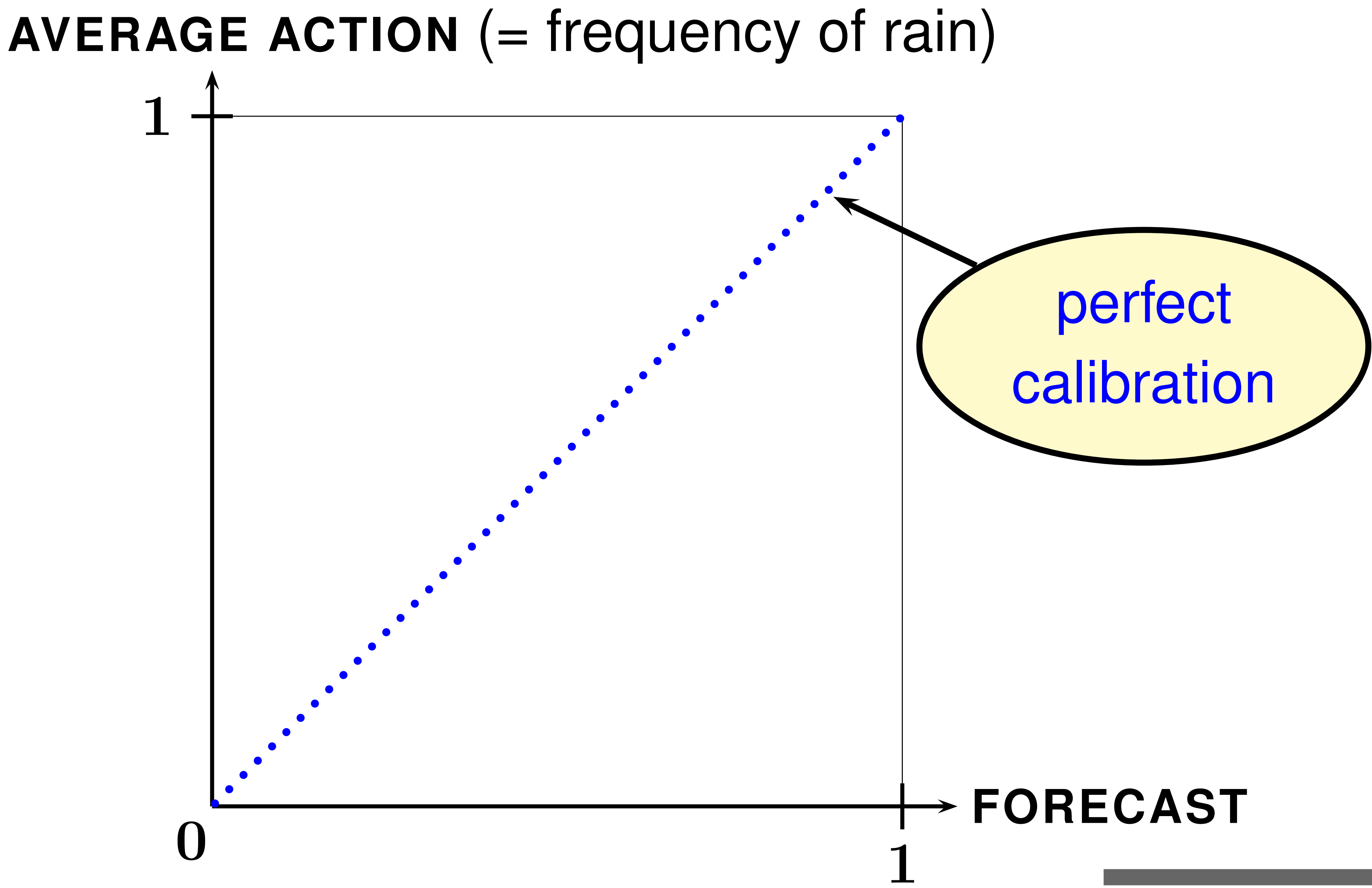
Forecast-Hedging (FH)

Forecast-Hedging (FH)

AVERAGE ACTION (= frequency of rain)

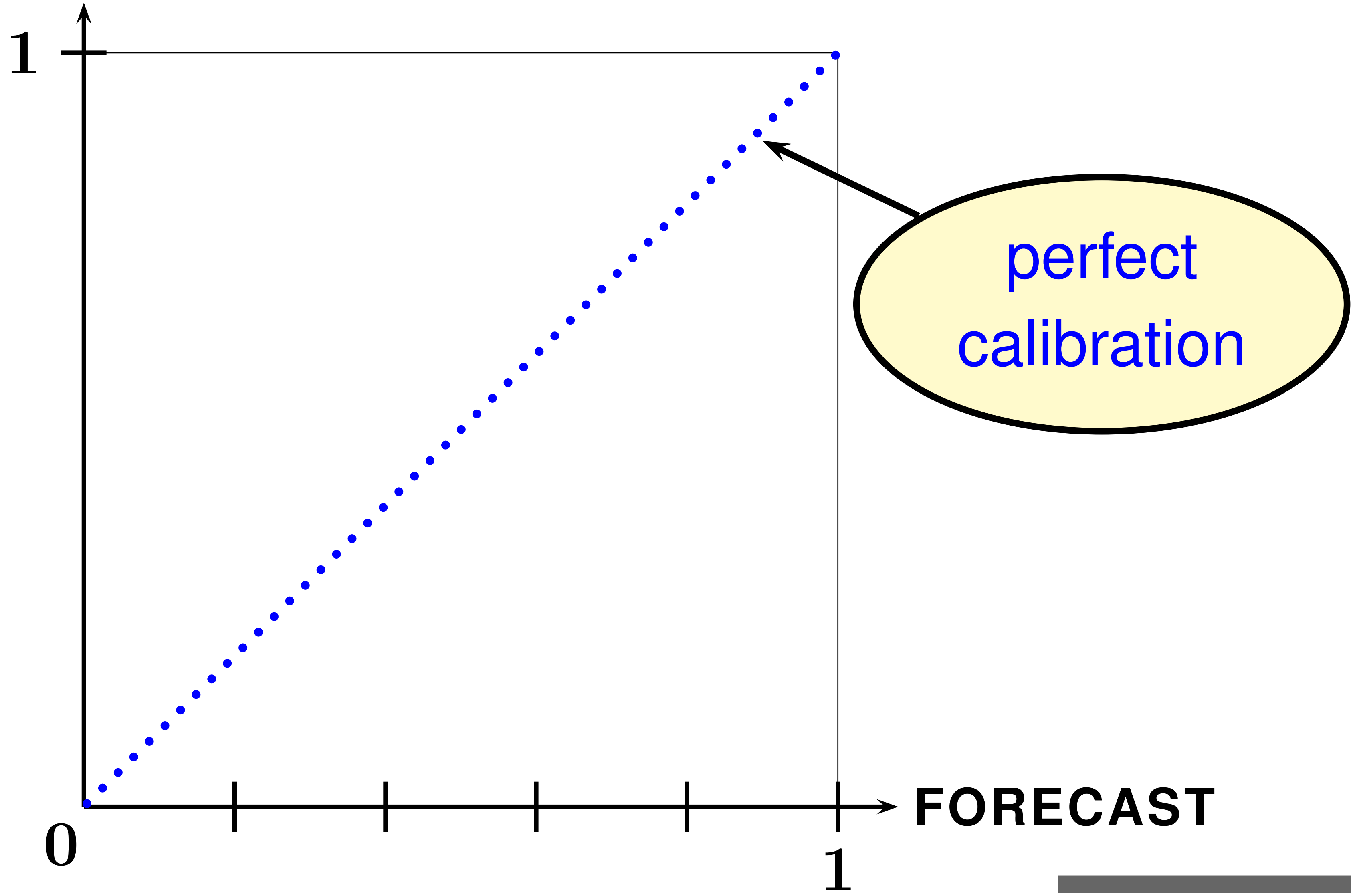


Forecast-Hedging (FH)

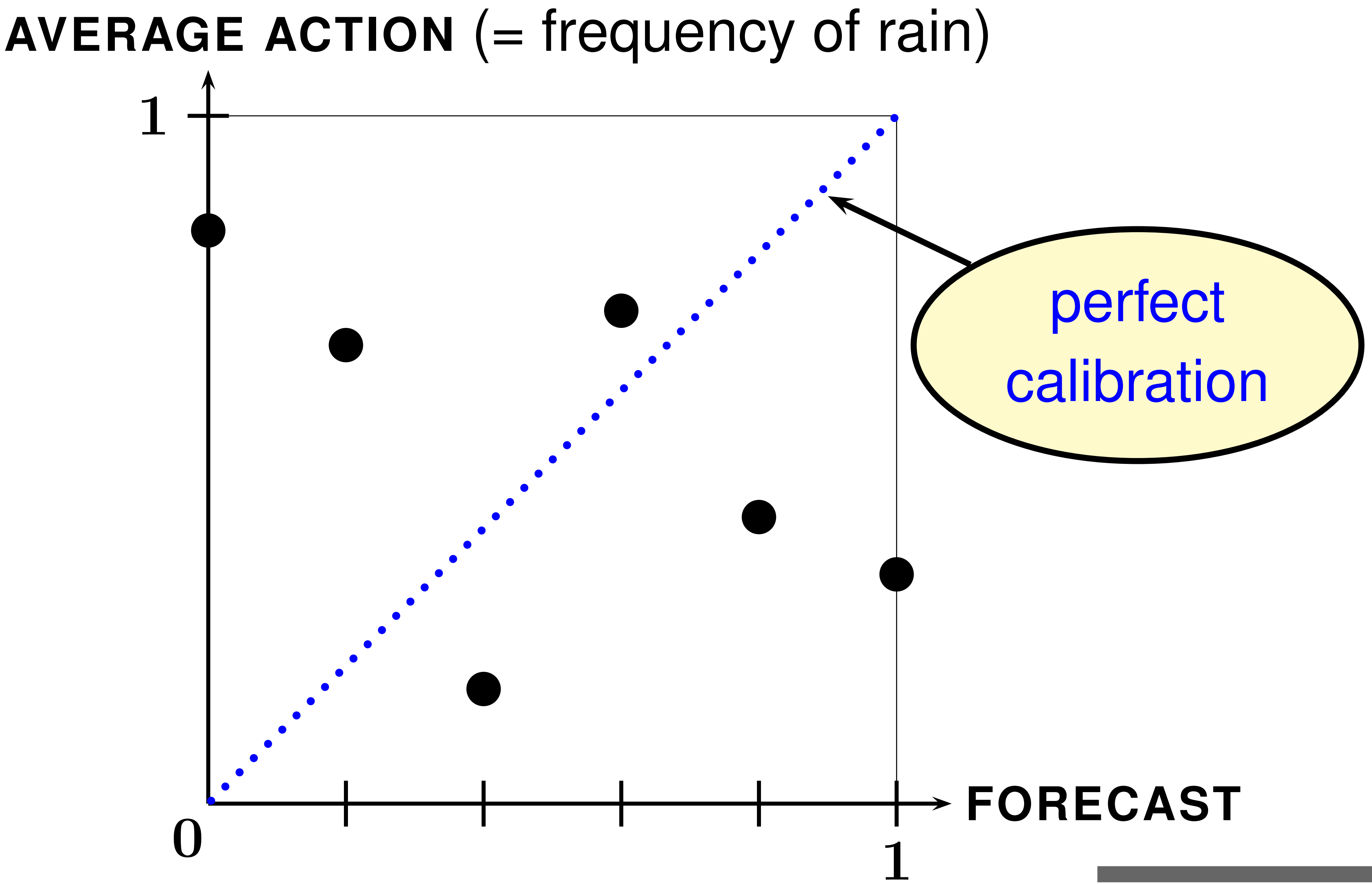


Forecast-Hedging (FH)

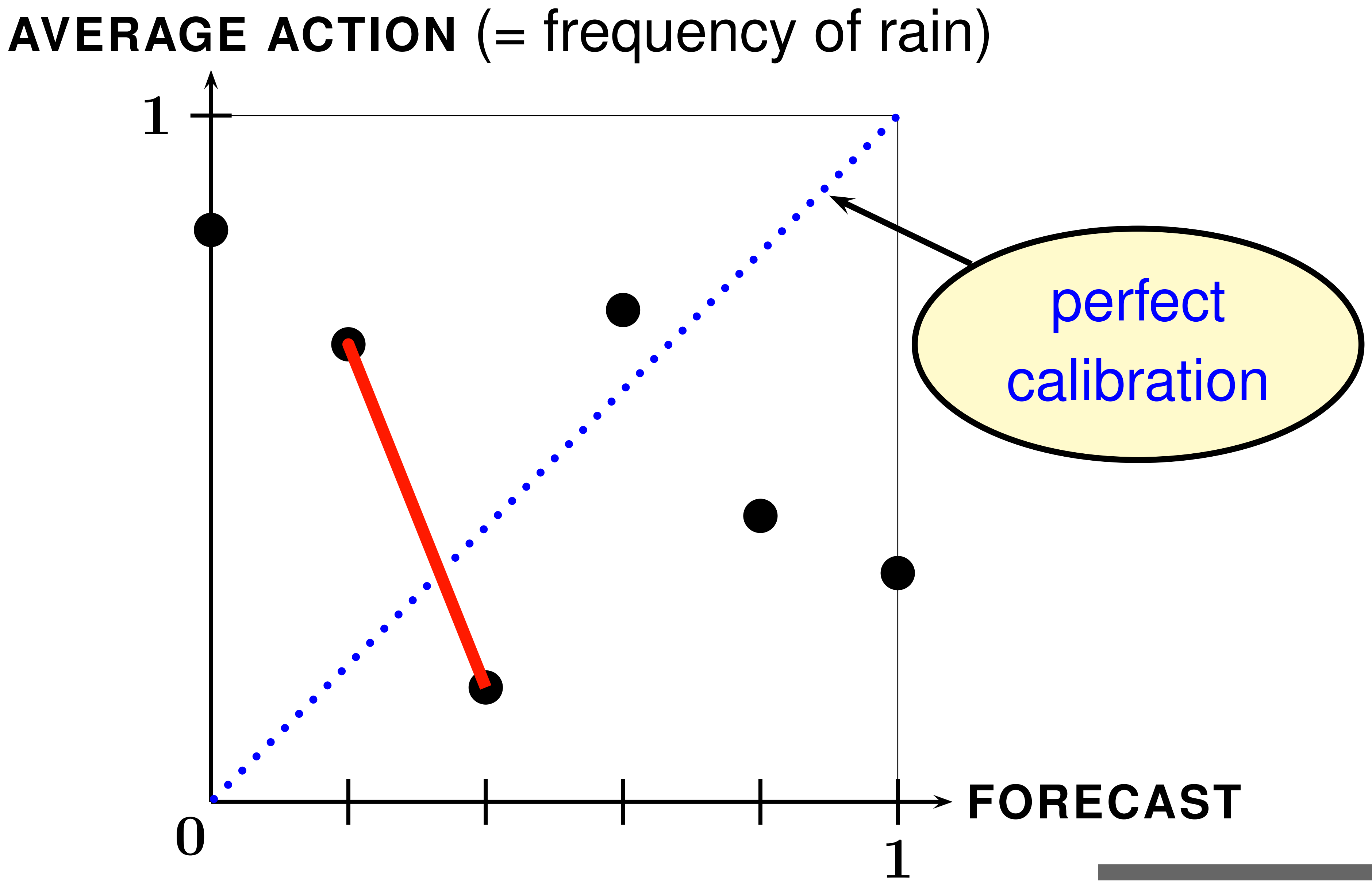
AVERAGE ACTION (= frequency of rain)



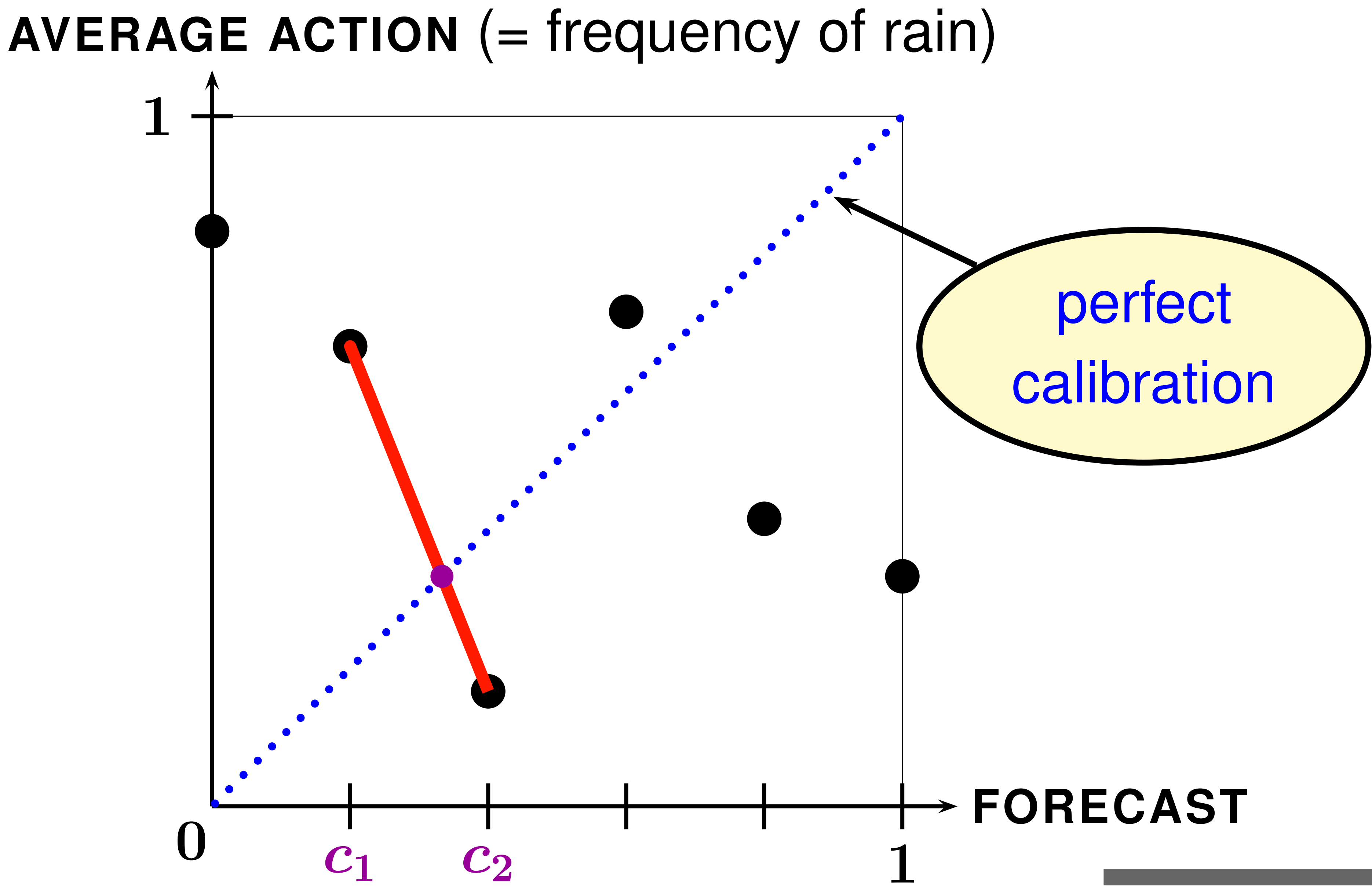
Forecast-Hedging (FH)



Forecast-Hedging (FH)



Forecast-Hedging (FH)



Forecasting

Forecasting ?

Forecasting ?

BACK-casting !

Forecasting ?

BACK-casting !

(not fore-casting)

Forecasting ?

BACK-casting !

(not fore-casting)

... prophet looking backwards ...

Forecasting ?

BACK-casting !

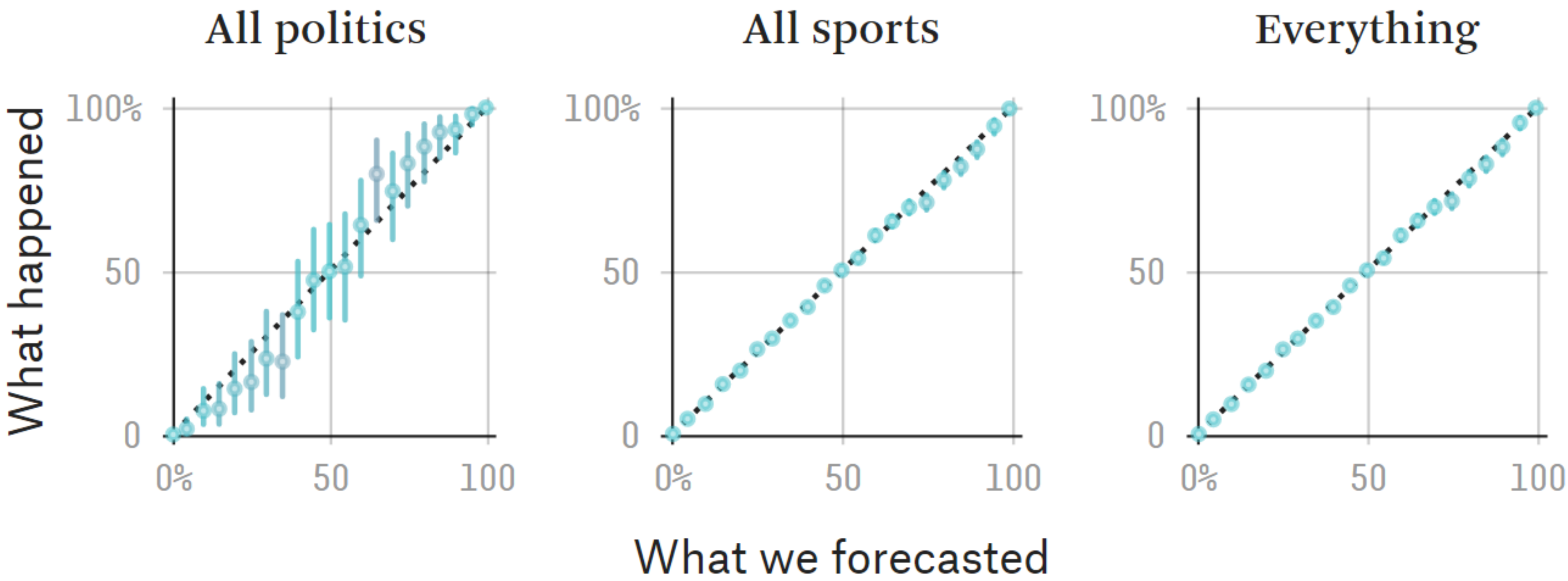
(not fore-casting)

... prophet looking backwards ...

Friedrich von Schlegel (1797)
Walter Benjamin

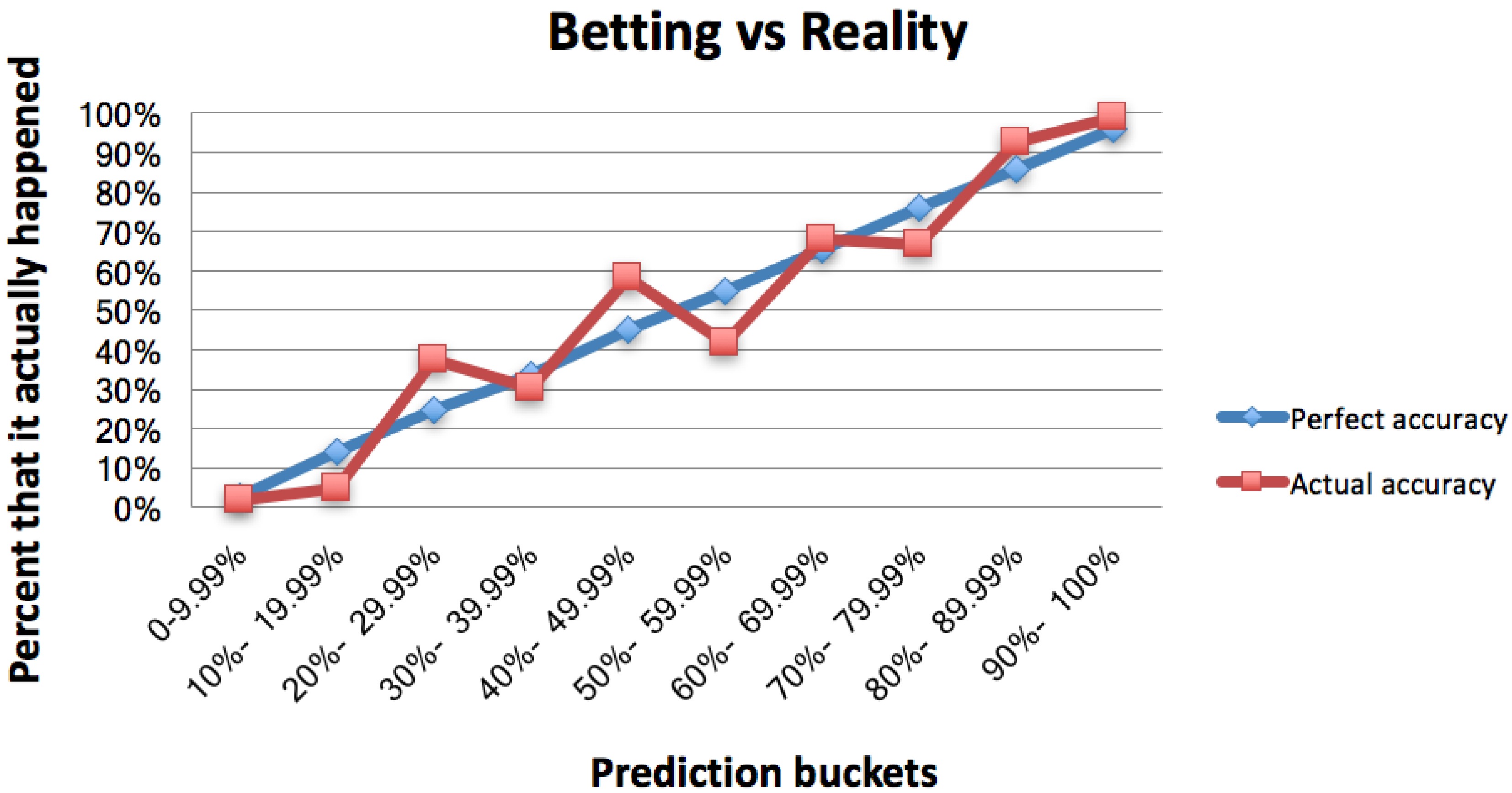
Calibration in Practice

Calibration in Practice



Calibration plots of FiveThirtyEight.com
(as of June 2019)

Calibration in Practice



Calibration plot of ElectionBettingOdds.com
(2016 – 2018)

Example

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: **CALIBRATION** $= \mathbf{0}$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: **CALIBRATION** $= \mathbf{0}$

F2: **CALIBRATION** $= \mathbf{0}$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: **CALIBRATION** $= 0$ **IN-BIN VARIANCE** $= 0$

F2: **CALIBRATION** $= 0$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: **CALIBRATION** $= \mathbf{0}$ **IN-BIN VARIANCE** $= \mathbf{0}$

F2: **CALIBRATION** $= \mathbf{0}$ **IN-BIN VARIANCE** $= \mathbf{\frac{1}{4}}$

Notations

Notations

- a_t = action at time t

Notations

- a_t = action at time t
- c_t = forecast at time t

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x

$$\bar{a}(x) = \frac{\sum_{t=1}^{T} \mathbf{1}_x(c_t)\, a_t}{\sum_{t=1}^{T} \mathbf{1}_x(c_t)}$$

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$

$$\mathcal{K} = \frac{1}{T} \sum_{t=1}^{T} \|c_t - \bar{a}(c_t)\|^2$$

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$
- $\mathcal{R} \equiv \mathcal{R}_T$ = **REFINEMENT** score = average variance inside the bins

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$
- $\mathcal{R} \equiv \mathcal{R}_T$ = **REFINEMENT** score = average variance inside the bins

$$\mathcal{R} = \frac{1}{T} \sum_{t=1}^{T} \| a_t - \bar{a}(c_t) \|^2$$

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$
- $\mathcal{R} \equiv \mathcal{R}_T$ = **REFINEMENT** score = average variance inside the bins

Brier score

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ **BRIER** (1950) score $=$ average distance between a_t and c_t

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ BRIER (1950) score = average distance between a_t and c_t

$$\mathcal{B} = \frac{1}{T} \sum_{t=1}^{T} \|a_t - c_t\|^2$$

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ **BRIER** (1950) score = average distance between a_t and c_t

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ **Brier** (1950) score $=$ average distance between a_t and c_t
- $$\mathcal{B} = \mathcal{R} + \mathcal{K}$$

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ BRIER (1950) score $=$ average distance between a_t and c_t
- $$\mathcal{B} = \mathcal{R} + \mathcal{K}$$

  BRIER $=$ REFINEMENT $+$ CALIBRATION

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ **BRIER** (1950) score $=$ average distance between a_t and c_t
-

$$\mathcal{B} = \mathcal{R} + \mathcal{K}$$

BRIER $=$ **REFINEMENT** $+$ **CALIBRATION**

Proof.

$$\mathbb{E}[(X - c)^2] = \mathbb{V}ar(X) + (\bar{X} - c)^2$$

where c is a constant and X is a random variable with $\bar{X} = \mathbb{E}[X]$

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ Brier (1950) score $=$ average distance between a_t and c_t
- $$\mathcal{B} = \mathcal{R} + \mathcal{K}$$

  BRIER $=$ REFINEMENT $+$ CALIBRATION

Example

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: **CALIBRATION** $= 0$ **IN-BIN VARIANCE** $= 0$

F2: **CALIBRATION** $= 0$ **IN-BIN VARIANCE** $= \frac{1}{4}$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: $\mathcal{K} = 0 \quad \mathcal{R} = 0$

F2: $\mathcal{K} = 0 \quad \mathcal{R} = \frac{1}{4}$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: $\mathcal{K} = \mathbf{0}$ $\mathcal{R} = \mathbf{0}$ $\mathcal{B} = \mathbf{0}$

F2: $\mathcal{K} = \mathbf{0}$ $\mathcal{R} = \frac{1}{4}$ $\mathcal{B} = \frac{1}{4}$

“Experts”

“Experts”

Testing experts:

“Experts”

Testing experts:

✓ BRIER score

“Experts”

Testing experts:

✓ BRIER score

✗ CALIBRATION score

“Expertise”

“Expertise”

- Recognize ***patterns*** and ***regularities*** in the data

“Expertise”

- Recognize ***patterns*** and ***regularities*** in the data
- ***Sort*** the days into bins that consist of similar days

“Expertise”

- Recognize ***patterns*** and ***regularities*** in the data
- ***Sort*** the days into bins that consist of similar days
- Make the binning as ***refined*** as possible

“Expertise”

- Recognize ***patterns*** and ***regularities*** in the data
- ***Sort*** the days into bins that consist of similar days
- Make the binning as ***refined*** as possible

$\longleftrightarrow$ ***LOW*** **REFINEMENT SCORE**

“Expertise” and Calibration

"Expertise" and Calibration

- **CALIBRATION** ($\mathcal{K} \approx 0$) can always be ***guaranteed*** in the long run

"Expertise" and Calibration

- **CALIBRATION** ($\mathcal{K} \approx 0$) can always be ***guaranteed*** in the long run
- But: **CALIBRATION** procedures ignore whatever "**EXPERTISE**" one has

"Expertise" and Calibration

- **CALIBRATION** ($\mathcal{K} \approx 0$) can always be ***guaranteed*** in the long run
- But: **CALIBRATION** procedures ignore whatever **"EXPERTISE"** one has

Question:

Can one **GAIN CALIBRATION**

without **LOSING "EXPERTISE"** ?

“Expertise” and Calibration

- **CALIBRATION** ($\mathcal{K} \approx 0$) can always be ***guaranteed*** in the long run
- But: **CALIBRATION** procedures ignore whatever “**EXPERTISE**” one has

Question:

Can one **GAIN CALIBRATION**

without **LOSING “EXPERTISE”** ?

- Can one get $\mathcal{K}$ to 0 without increasing $\mathcal{R}$?

"Expertise" and Calibration

- **CALIBRATION** ($\mathcal{K} \approx 0$) can always be ***guaranteed*** in the long run
- But: **CALIBRATION** procedures ignore whatever "**EXPERTISE**" one has

Question:

Can one **GAIN CALIBRATION**

without **LOSING "EXPERTISE"** ?

- Can one get $\mathcal{K}$ to 0 without increasing $\mathcal{R}$?
- Can one decrease $\mathcal{B} = \mathcal{R} + \mathcal{K}$ by $\mathcal{K}$?

“Expertise” and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?

“Expertise” and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

“Expertise” and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

  $\Rightarrow \quad \mathcal{K}' = 0$

“Expertise” and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

  $\Rightarrow \quad \mathcal{K}' = 0 \qquad \mathcal{R}' = \mathcal{R}$

“Expertise” and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

$$\Rightarrow \quad \mathcal{K}' = 0 \qquad \mathcal{R}' = \mathcal{R} \qquad \mathcal{B}' = \mathcal{B} - \mathcal{K}$$

"Expertise" and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

  $\Rightarrow \quad \mathcal{K}' = 0 \qquad \mathcal{R}' = \mathcal{R} \qquad \mathcal{B}' = \mathcal{B} - \mathcal{K}$

  - **IN RETROSPECT** / **OFFLINE** (when the $\bar{a}(c)$ are known)

"Expertise" and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

$$\Rightarrow \quad \mathcal{K}' = 0 \qquad \mathcal{R}' = \mathcal{R} \qquad \mathcal{B}' = \mathcal{B} - \mathcal{K}$$

  - **IN RETROSPECT** / **OFFLINE** (when the $\bar{a}(c)$ are known)

Question:

Can one do this **ONLINE** ?

- Consider a forecasting sequence b_t (in a [finite] set B)

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - **ONLINE**: use only b_t and history

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - ONLINE: use only b_t and history
  - such that

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - ONLINE: use only b_t and history
  - such that

$$\mathcal{B}^{\mathrm{c}}_T \leq \mathcal{B}^{\mathrm{b}}_T - \mathcal{K}^{\mathrm{b}}_T + \mathrm{o}(1) \quad \text{as } T \to \infty$$

for ALL sequences a_t and b_t (uniformly)

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - **ONLINE**: use only b_t and history
  - such that

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - **ONLINE**: use only b_t and history
  - such that

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}} = \mathcal{R}^{\mathrm{b}}$$

"Calibeating"

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - ONLINE: use only b_t and history
  - such that

$$\boxed{\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}} = \mathcal{R}^{\mathrm{b}}$$

c "BEATS" b by b's CALIBRATION score

"Calibeating"

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - ONLINE: use only b_t and history
  - such that

$$\boxed{\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}} = \mathcal{R}^{\mathrm{b}}$$

c "BEATS" b by b's CALIBRATION score

- GUARANTEED for ALL sequences of actions and forecasts

Example

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
b	80%	40%	80%	40%	80%	40%	

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
$\boldsymbol{b}$	80%	40%	80%	40%	80%	40%	

$\boldsymbol{b}$: $\mathcal{K}^{\mathrm{b}} = \mathbf{0.1}$ $\mathcal{R}^{\mathrm{b}} = \mathbf{0}$ $\mathcal{B}^{\mathrm{b}} = \mathbf{0.1}$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
$\boldsymbol{b}$	80%	40%	80%	40%	80%	40%	
$\boldsymbol{c}$	100%	0%	100%	0%	100%	0%	

$\boldsymbol{b}$: $\mathcal{K}^{\mathrm{b}} = 0.1$ $\quad \mathcal{R}^{\mathrm{b}} = 0$ $\quad \mathcal{B}^{\mathrm{b}} = 0.1$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
b	80%	40%	80%	40%	80%	40%	
c	100%	0%	100%	0%	100%	0%	

b: $\mathcal{K}^{\mathrm{b}} = 0.1$ $\quad \mathcal{R}^{\mathrm{b}} = 0$ $\quad \mathcal{B}^{\mathrm{b}} = 0.1$

c: $\mathcal{K}^{\mathrm{c}} = 0$ $\quad \mathcal{R}^{\mathrm{c}} = 0$ $\quad \mathcal{B}^{\mathrm{c}} = 0$

Calibeating

time	1	2	3	4	5	6	...
rain	1	0	1	0	1	0	
b	80%	40%	80%	40%	80%	40%	
c	100%	0%	100%	0%	100%	0%	

b: $\mathcal{K}^{\text{b}} = 0.1$ $\mathcal{R}^{\text{b}} = 0$ $\mathcal{B}^{\text{b}} = 0.1$

c: $\mathcal{K}^{\text{c}} = 0$ $\mathcal{R}^{\text{c}} = 0$ $\mathcal{B}^{\text{c}} = 0$

c ***calibeats*** b: $\mathcal{B}^{\text{c}} \leq \mathcal{B}^{\text{b}} - \mathcal{K}^{\text{b}}$

Calibeating

time	1	2	3	4	5	6	...
rain	1	0	1	0	1	0	
b	80%	40%	80%	40%	80%	40%	
c	100%	0%	100%	0%	100%	0%	

b: $\mathcal{K}^{\mathrm{b}} = 0.1$ $\mathcal{R}^{\mathrm{b}} = 0$ $\mathcal{B}^{\mathrm{b}} = 0.1$

c: $\mathcal{K}^{\mathrm{c}} = 0$ $\mathcal{R}^{\mathrm{c}} = 0$ $\mathcal{B}^{\mathrm{c}} = 0$

c ***calibeats*** b: $\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}} = \mathcal{R}^{\mathrm{b}}$

Calibeating

Calibeating

(that was easy ...)

Calibeating

(that was easy ...)

Can one **CALIBEAT** ***in general, non-stationary, situations ?***

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary
- **Forecasts of** b are arbitrary

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary
- **Forecasts of** b are arbitrary
- **Binning of** b is not perfect ($\mathcal{R}^{\mathrm{b}} > 0$)

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary
- **Forecasts of** b are arbitrary
- **Binning of** b is not perfect ($\mathcal{R}^{\mathrm{b}} > 0$)
- **Bin averages** do not converge

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary
- **Forecasts of** b are arbitrary
- **Binning of** b is not perfect ($\mathcal{R}^{\mathrm{b}} > 0$)
- **Bin averages** do not converge
- **ONLINE**

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary
- **Forecasts of** b are arbitrary
- **Binning of** b is not perfect ($\mathcal{R}^{\mathrm{b}} > 0$)
- **Bin averages** do not converge
- ONLINE
- GUARANTEED (even against adversary)

Calibeating

Calibeating

Theorem

There exists a **CALIBEATING** procedure

A Way to Calibeat

A Way to Calibeat

Theorem

The procedure

$$c_t = \bar{a}^{\mathrm{b}}_{t-1}(b_t)$$

GUARANTEES b-CALIBEATING

A Simple Way to Calibeat

Theorem

The procedure

$$c_t = \bar{a}^{\mathrm{b}}_{t-1}(b_t)$$

GUARANTEES b-**CALIBEATING**

Forecast the average action of the current b-forecast

Proof

Proof

$$\mathbb{V}\mathbf{ar} \;=\; \frac{1}{T}\sum_{t=1}^{T}\left\|x_t - \bar{x}_T\right\|^2$$

Proof

$$\mathbb{V}\text{ar} = \frac{1}{T}\sum_{t=1}^{T}\|x_t - \bar{x}_T\|^2$$

$$= \frac{1}{T}\sum_{t=1}^{T}\left(1-\frac{1}{t}\right)\|x_t - \bar{x}_{t-1}\|^2$$

Proof

$$\mathbb{V}\text{ar} = \frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_{\color{red}T}\|^2$$

$$= \frac{1}{T}\sum_{t=1}^{T} \left(1 - \frac{1}{t}\right) \|x_t - \bar{x}_{\color{red}t-1}\|^2$$

Proof

$$\mathbb{V}\text{ar} = \frac{1}{T}\sum_{t=1}^{T}\left\|x_t - \bar{x}_T\right\|^2$$

$$= \frac{1}{T}\sum_{t=1}^{T}\left(1-\frac{1}{t}\right)\left\|x_t - \bar{x}_{t-1}\right\|^2$$

Proof

$$
\begin{aligned}
\mathbb{V}\mathrm{ar} &= \frac{1}{T}\sum_{t=1}^{T} \left\| x_t - \bar{x}_T \right\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(1 - \frac{1}{t}\right) \left\| x_t - \bar{x}_{t-1} \right\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T} \left\| x_t - \bar{x}_{t-1} \right\|^2 - \mathrm{o}(1)
\end{aligned}
$$

Proof

$$\mathbb{V}\mathrm{ar} = \frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_T\|^2$$

$$= \frac{1}{T}\sum_{t=1}^{T} \left(1 - \frac{1}{t}\right) \|x_t - \bar{x}_{t-1}\|^2$$

$$= \frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_{t-1}\|^2 - \mathrm{o}(1)$$

(*) $\mathrm{o}(1) = \mathrm{O}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\right) = \mathrm{O}\left(\frac{\log T}{T}\right)$

Proof

$$
\begin{aligned}
\mathbb{V}\mathrm{ar} &= \frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_T\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(1 - \frac{1}{t}\right) \|x_t - \bar{x}_{t-1}\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_{t-1}\|^2 - \mathrm{o}(1)
\end{aligned}
$$

Proof: "Online Variance"

$$\begin{aligned}
\mathbb{V}\text{ar} &= \frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_T\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(1 - \frac{1}{t}\right) \|x_t - \bar{x}_{t-1}\|^2 \\
&= \underbrace{\frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_{t-1}\|^2}_{\widetilde{\mathbb{V}\text{ar}}} - \text{o}(1) \\
&= \widetilde{\mathbb{V}\text{ar}} - \text{o}(1)
\end{aligned}$$

Proof: “Online Variance”

$$\mathbb{V}\mathrm{ar} \;=\; \widetilde{\mathbb{V}\mathrm{ar}} - o(1)$$

Proof: “Online Refinement”

$$\mathbb{V}\mathbf{ar} = \widetilde{\mathbb{V}\mathbf{ar}} - \mathbf{o}(1)$$

$$\mathcal{R}^{\mathbf{b}} = \widetilde{\mathcal{R}}^{\mathbf{b}} - \mathbf{o}(1)$$

Proof: “Online Refinement”

$$\mathbb{V}\mathrm{ar} = \widetilde{\mathbb{V}\mathrm{ar}} - \mathrm{o}(1)$$

$$\mathcal{R}^{\mathrm{b}} = \tilde{\mathcal{R}}^{\mathrm{b}} - \mathrm{o}(1)$$

$$= \frac{1}{T}\sum_{t=1}^{T} \|a_t - \bar{a}_{t-1}(b_t)\|^2 - \mathrm{o}(1)$$

Proof: “Online Refinement”

$$\begin{aligned}
\mathbb{V}\mathrm{ar} &= \widetilde{\mathbb{V}\mathrm{ar}} - \mathrm{o}(1) \\
\mathcal{R}^{\mathrm{b}} &= \widetilde{\mathcal{R}}^{\mathrm{b}} - \mathrm{o}(1) \\
&= \underbrace{\frac{1}{T}\sum_{t=1}^{T} \|a_t - \bar{a}_{t-1}(b_t)\|^2}_{} - \mathrm{o}(1) \\
&= \qquad\qquad \mathcal{B}^{\mathrm{c}} \qquad\qquad - \mathrm{o}(1)
\end{aligned}$$

$c_t = \bar{a}_{t-1}(b_t)$

Calibeating

Calibeating

Theorem

$$c_t = \bar{a}^{\mathrm{b}}_{t-1}(b_t)$$

GUARANTEES b-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

Self-Calibeating

Theorem

$$c_t = \bar{a}^{\mathrm{b}}_{t-1}(b_t)$$

GUARANTEES b-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

Theorem

$$c_t = \bar{a}^{\mathrm{c}}_{t-1}(c_t)$$

GUARANTEES c-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{c}} - \mathcal{K}^{\mathrm{c}}$$

Self-Calibeating

Theorem

$$c_t = \bar{a}^{\mathrm{b}}_{t-1}(b_t)$$

GUARANTEES b-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

Theorem

$$c_t = \bar{a}^{\mathrm{c}}_{t-1}(c_t)$$

GUARANTEES c-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{c}} - \mathcal{K}^{\mathrm{c}}$$

$$\Leftrightarrow \quad \mathcal{K}^{\mathrm{c}} = 0$$

Self-Calibeating = Calibrating

Theorem

$$c_t = \bar{a}^{\mathrm{b}}_{t-1}(b_t)$$

GUARANTEES b-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

Theorem

$$c_t = \bar{a}^{\mathrm{c}}_{t-1}(c_t)$$

GUARANTEES CALIBRATION:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{c}} - \mathcal{K}^{\mathrm{c}}$$
$$\Leftrightarrow \quad \mathcal{K}^{\mathrm{c}} = 0$$

“Fixed Point”

How do we get c_t “close to” $\bar{a}_{t-1}(c_t)$?

“Fixed Point”

How do we get c_t “close to” $\bar{a}_{t-1}(c_t)$?

- $C \subset \mathbb{R}^m$ compact convex set
- $D \subset C$ finite δ-grid of C (for $\delta > 0$)
- $f : D \longrightarrow \mathbb{R}^m$ **arbitrary** function

“Fixed Point”

How do we get c_t “close to” $\bar{a}_{t-1}(c_t)$?

- $C \subset \mathbb{R}^m$ compact convex set
- $D \subset C$ finite δ-grid of C (for $\delta > 0$)
- $f : D \longrightarrow \mathbb{R}^m$ **arbitrary** function

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

“Fixed Point”

How do we get c_t “close to” $\bar{a}_{t-1}(c_t)$?

- $C \subset \mathbb{R}^m$ compact convex set
- $D \subset C$ finite δ-grid of C (for $\delta > 0$)
- $f : D \to \mathbb{R}^m$ **arbitrary** function

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2\right] \le \mathbb{E}_{x \sim P}\left[\|v - f(x)\|^2\right] + \delta^2$$

$$\forall v \in C$$

Stochastic "Fixed Point"

How do we get c_t "close to" $\bar{a}_{t-1}(c_t)$?

- $C \subset \mathbb{R}^m$ compact convex set
- $D \subset C$ finite δ-grid of C (for $\delta > 0$)
- $f : D \to \mathbb{R}^m$ **arbitrary** function

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2\right] \leq \mathbb{E}_{x \sim P}\left[\|v - f(x)\|^2\right] + \delta^2$$

$$\forall v \in C$$

Stochastic "Fixed Point"

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2\right] \le \mathbb{E}_{x \sim P}\left[\|v - f(x)\|^2\right] + \delta^2$$

$$\forall v \in C$$

Stochastic "Fixed Point"

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - f(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

Stochastic “Fixed Point”

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - f(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

Stochastic "Fixed Point"

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - f(x)\|^2\right] \le \delta^2 \quad \forall v \in C$$

Proof.

Stochastic "Fixed Point"

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - f(x)\|^2\right] \le \delta^2 \quad \forall v \in C$$

Proof.

- wlog (project $f(x)$ to C): $\quad f : D \to C$

Stochastic "Fixed Point"

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x\sim P}\left[\|v-x\|^2 - \|v-f(x)\|^2\right] \le \delta^2 \quad \forall v \in C$$

Proof.

- wlog (project $f(x)$ to C): $f : D \to C$
- interpolate: *continuous* $f : C \to C$

Stochastic "Fixed Point"

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x\sim P}\left[\|v-x\|^2-\|v-f(x)\|^2\right]\le\delta^2 \quad \forall v\in C$$

Proof.

- wlog (project $f(x)$ to C): $f : D \longrightarrow C$
- interpolate: *continuous* $f : C \longrightarrow C$
- apply Brouwer **FIXED POINT** theorem

Stochastic “Fixed Point”

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x\sim P}\left[\|v-x\|^2-\|v-f(x)\|^2\right]\leq\delta^2 \quad \forall v\in C$$

Proof.

- wlog (project $f(x)$ to C): $f:D\rightarrow C$
- interpolate: *continuous* $f:C\rightarrow C$
- apply Brouwer **FIXED POINT** theorem

“Wrong” Proof !

Stochastic “Fixed Point”

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x\sim P}\left[\|v - x\|^2 - \|v - f(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

Stochastic “Fixed Point”

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x\sim P}\left[\|v - x\|^2 - \|v - f(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

“Right” Proof.

Stochastic "Fixed Point"

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - f(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

"Right" Proof. Use the **MINIMAX** theorem

Stochastic "Fixed Point"

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - f(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

"Right" Proof. Use the MINIMAX theorem

- $h(x, v) := \|v - x\|^2 - \|v - f(x)\|^2$

Stochastic "Fixed Point"

- $h(x, v) := \|v - x\|^2 - \|v - f(x)\|^2$

Stochastic "Fixed Point"

- $h(x, v) := \|v - x\|^2 - \|v - f(x)\|^2$
- For every distribution Q on $v \in C$:

Stochastic "Fixed Point"

- $h(x, v) := \|v - x\|^2 - \|v - f(x)\|^2$
- For every distribution Q on $v \in C$:

  $$h(x, Q) = \mathbb{E}_{v \sim Q}\left[\|v - x\|^2 - \|v - f(x)\|^2\right]$$

Stochastic "Fixed Point"

- $h(x, v) := \|v - x\|^2 - \|v - f(x)\|^2$
- For every distribution Q on $v \in C$:

  $$h(x, Q) = \mathbb{E}_{v \sim Q}\left[\|v - x\|^2 - \|v - f(x)\|^2\right]$$

  - For $x = \mathbb{E}_Q[v]$ we get $h(x, Q) \leq 0$

Stochastic “Fixed Point”

- $h(x, v) := \|v - x\|^2 - \|v - f(x)\|^2$
- For every distribution Q on $v \in C$:

$$h(x, Q) = \mathbb{E}_{v \sim Q}\left[\|v - x\|^2 - \|v - f(x)\|^2\right]$$

  - For $x = \mathbb{E}_Q[v]$ we get $h(x, Q) \le 0$
  - For $x \approx \mathbb{E}_Q[v]$ in D we get $h(x, Q) \le \delta^2$

Stochastic "Fixed Point"

- $h(x, v) := \|v - x\|^2 - \|v - f(x)\|^2$
- For every distribution Q on $v \in C$:
  - For $x \approx \mathbb{E}_Q[v]$ in D we get $h(x, Q) \leq \delta^2$

Stochastic "Fixed Point"

- $h(x, v) := \|v - x\|^2 - \|v - f(x)\|^2$
- For every distribution Q on $v \in C$:
  - For $x \approx \mathbb{E}_Q[v]$ in D we get $h(x, Q) \leq \delta^2$
- $\Rightarrow$ [**MINIMAX**]
  There exists a distribution P on $x \in D$ s.t.

  $$h(P, v) \leq \delta^2 \quad \text{for every } v \in C$$

Stochastic “Fixed Point”

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - f(x)\|^2\right] \le \delta^2 \quad \forall v \in C$$

Stochastic "Fixed Point"

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x\sim P}\left[\|v-x\|^2-\|v-f(x)\|^2\right]\leq\delta^2\quad\forall v\in C$$

- **Obtained by solving a MINIMAX problem (LP)**

Outgoing Minimax (FH)

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x\sim P}\left[\|v-x\|^2-\|v-f(x)\|^2\right]\leq\delta^2\quad\forall v\in C$$

- **Obtained by solving a MINIMAX problem (LP)**

Stochastic "Fixed Point"

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - f(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

- **Obtained by solving a MINIMAX problem (LP)**

Stochastic "Fixed Point"

Theorem. There exists a probability distribution P on the δ-grid D of C such that:

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - f(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

- **Obtained by solving a MINIMAX problem (LP)**
- Moreover: solving a FIXED POINT problem yields a probability distribution P that is ALMOST DETERMINISTIC: its support is included in a ball of size δ

Calibrating

Calibrating

Theorem

There is a stochastic procedure that **GUARANTEES CALIBRATION**

Calibrating

Theorem

There is a stochastic procedure that **GUARANTEES CALIBRATION**

Proof. Self-calibeating + Stochastic Fixed Point

Calibrating

Theorem

There is a stochastic procedure that **GUARANTEES** **CALIBRATION**

Proof. Self-calibeating + Stochastic Fixed Point

Note. δ-**CALIBRATION**

Calibrated Calibeating

Calibrated Calibeating

Theorem

There is a stochastic procedure that **GUARANTEES CALIBEATING**

Calibrated Calibeating

Theorem

There is a stochastic procedure that **GUARANTEES CALIBEATING** and **CALIBRATION**

Calibrated Calibeating

Theorem

There is a stochastic procedure that **GUARANTEES** **CALIBEATING** and **CALIBRATION**

Proof. Calibeat the **joint** binning of $\boldsymbol{b}$ and $\boldsymbol{c}$, by applying Stochastic Fixed Point

Calibrated Calibeating

Theorem

There is a stochastic procedure that **GUARANTEES CALIBEATING** and **CALIBRATION**

Proof. Calibeat the **joint** binning of b and c, by applying Stochastic Fixed Point

---

STRONG CALIBEATING:

$$\mathcal{R}(\mathbf{c}) \leq \mathcal{R}(\mathbf{b}) \quad \text{and} \quad \mathcal{K}(\mathbf{c}) = \mathbf{0}$$

Continuous Calibration

Continuous Calibration

- **CONTINUOUS CALIBRATION**: combine the days when the forecast was ***close to*** p

Continuous Calibration

- **CONTINUOUS CALIBRATION**: combine the days when the forecast was ***close to*** p (smooth out the calibration score)

Continuous Calibration

- **CONTINUOUS CALIBRATION**: combine the days when the forecast was ***close to*** p (smooth out the calibration score)
- **Theorem**:

> There exists a ***deterministic*** procedure that is **CONTINUOUSLY CALIBRATED**.

Continuous Calibration

- **CONTINUOUS CALIBRATION**: combine the days when the forecast was ***close to*** p (smooth out the calibration score)
- **Theorem**:

> There exists a ***deterministic*** procedure that is **CONTINUOUSLY CALIBRATED**.

Deterministic $\Rightarrow$ holds also when the forecasts are ***"leaked"***

Continuous Calibration

- **CONTINUOUS CALIBRATION**: combine the days when the forecast was ***close to*** p (smooth out the calibration score)
- **Theorem**:

> There exists a ***deterministic*** procedure that is **CONTINUOUSLY CALIBRATED**.

Deterministic $\Rightarrow$ holds also when the forecasts are ***"leaked"***

Foster and Kakade (2004, 2006)
Foster and Hart (2018, **2021**)

Continuous-Calibrated Calibeating

Continuous-Calibrated Calibeating

Theorem

There is a ***deterministic*** procedure that **GUARANTEES CALIBEATING**

Continuous-Calibrated Calibeating

Theorem

There is a ***deterministic*** procedure that **GUARANTEES** **CALIBEATING** and **CONTINUOUS CALIBRATION**

Continuous-Calibrated Calibeating

Theorem

There is a ***deterministic*** procedure that **GUARANTEES CALIBEATING** and **CONTINUOUS CALIBRATION**

Proof. Calibeat the **joint** binning of $\boldsymbol{b}$ and $\boldsymbol{c}$, by a Fixed Point result (Brouwer)

Multi-Calibeating

Multi-Calibeating

Theorem

There is a ***deterministic*** procedure
that **GUARANTEES**

simultaneous CALIBEATING
of several forecasters

Multi-Calibeating

Theorem

There is a ***stochastic*** procedure
that **GUARANTEES**

simultaneous CALIBEATING
of several forecasters

and **CALIBRATION**

Multi-Calibeating

Theorem

There is a ***stochastic*** procedure
that **GUARANTEES**

simultaneous CALIBEATING
of several forecasters

and **CALIBRATION**

Proof. Calibeat the **joint** binning

In all the results above:

In all the results above:

	CALIBRATION	
Obtained by	***Minimax***	
Procedure	***stochastic***	

... and Continuous Calibration

In all the results above:

	CALIBRATION	CONTINUOUS CALIBRATION
Obtained by	***Minimax***	***Fixed Point***
Procedure	***stochastic***	***deterministic***

Refinement Score and Brier Score

Refinement Score and Brier Score

Claim. The REFINEMENT score is the ***minimal*** BRIER score over all ***relabelings of the bins***:

Refinement Score and Brier Score

Claim. The REFINEMENT score is the ***minimal*** BRIER score over all ***relabelings of the bins***:

$$\mathcal{R}_T(\mathrm{b}) = \min_{\phi} \mathcal{B}_T(\phi(\mathrm{b}))$$

Refinement Score and Brier Score

Claim. The REFINEMENT score is the ***minimal*** BRIER score over all ***relabelings of the bins***:

$$\mathcal{R}_T(\mathrm{b}) = \min_{\phi} \mathcal{B}_T(\phi(\mathrm{b}))$$

where the minimum is taken over all

$$\phi : B \longrightarrow \Delta(A)$$

and

$$\phi(\mathrm{b}) \equiv (\phi(b_1), ..., \phi(b_T))$$

Refinement Score and Brier Score

Claim. The REFINEMENT score is the ***minimal*** BRIER score over all ***relabelings of the bins***:

$$\mathcal{R}_T(\mathrm{b}) = \min_{\phi} \mathcal{B}_T(\phi(\mathrm{b}))$$

where the minimum is taken over all

$$\phi : B \to \Delta(A)$$

and

$$\phi(\mathrm{b}) \equiv (\phi(b_1), ..., \phi(b_T))$$

Proof. Label each bin b with $\phi(b) = \bar{a}_T(b)$

Calibeating

Calibeating

- c CALIBEATS b:

$$\mathcal{B}_T(\mathbf{c}) \leq \mathcal{R}_T(\mathbf{b}) + o(1)$$

Calibeating

- c **CALIBEATS** b:

$$\mathcal{B}(\mathrm{c}) \leq \mathcal{R}(\mathrm{b})$$

Calibeating

- c **CALIBEATS** b:

$$\mathcal{B}(\mathrm{c}) \le \mathcal{R}(\mathrm{b})$$

$$\mathcal{B}(\mathrm{c}) \le \min_{\phi} \mathcal{B}(\phi(\mathrm{b}))$$

where $\phi : B \longrightarrow \Delta(A)$

Calibeating

- c **CALIBEATS** b:

$$\mathcal{B}(\mathrm{c}) \leq \mathcal{R}(\mathrm{b})$$

$$\mathcal{B}(\mathrm{c}) \leq \min_{\phi} \mathcal{B}(\phi(\mathrm{b}))$$

where $\phi : B \rightarrow \Delta(A)$

- c **MULTI-CALIBEATS** $\mathrm{b}_1, ..., \mathrm{b}_N$:

$$\mathcal{B}(\mathrm{c}) \leq \min_{1 \leq n \leq N} \min_{\phi^n} \mathcal{B}(\phi^n(\mathrm{b}^n))$$

Calibeating

- c **CALIBEATS** b:

$$\mathcal{B}(c) \le \mathcal{R}(b)$$

$$\mathcal{B}(c) \le \min_{\phi} \mathcal{B}(\phi(b))$$

where $\phi : B \to \Delta(A)$

- c **MULTI-CALIBEATS** $b_1, ..., b_N$:

$$\mathcal{B}(c) \le \min_{1 \le n \le N} \min_{\phi^n} \mathcal{B}(\phi^n(b^n))$$

where $\phi^n : B^n \to \Delta(A)$

Calibeating

- c **MULTI-CALIBEATS** $b_1, ..., b_N$:

$$\mathcal{B}(c) \le \min_{1 \le n \le N} \min_{\phi^n} \mathcal{B}(\phi^n(b^n))$$

where $\phi^n : B^n \to \Delta(A)$

Calibeating

- c MULTI-CALIBEATS $\mathrm{b}^1, \ldots, \mathrm{b}^N$:

$$\mathcal{B}(\mathrm{c}) \le \min_{1 \le n \le N} \min_{\phi^n} \mathcal{B}(\phi^n(\mathrm{b}^n))$$

where $\phi^n : B^n \longrightarrow \Delta(A)$

Calibeating

- c **MULTI-CALIBEATS** b^1, ..., b^N:

$$\mathcal{B}(\mathrm{c}) \le \min_{1 \le n \le N} \min_{\phi^n} \mathcal{B}(\phi^n(\mathrm{b}^n))$$

where $\phi^n : B^n \longrightarrow \Delta(A)$

- c **CALIBEATS THE JOINT** of b^1, ..., b^N:

$$\mathcal{B}(\mathrm{c}) \le \min_{\phi} \mathcal{B}(\phi(\mathrm{b}^1, ..., \mathrm{b}^N))$$

Calibeating

- c **MULTI-CALIBEATS** $\mathrm{b}^1, ..., \mathrm{b}^N$:

$$\mathcal{B}(\mathrm{c}) \le \min_{1 \le n \le N} \min_{\phi^n} \mathcal{B}(\phi^n(\mathrm{b}^n))$$

where $\phi^n : B^n \longrightarrow \Delta(A)$

- c **CALIBEATS THE JOINT** of $\mathrm{b}^1, ..., \mathrm{b}^N$:

$$\mathcal{B}(\mathrm{c}) \le \min_{\phi} \mathcal{B}(\phi(\mathrm{b}^1, ..., \mathrm{b}^N))$$

where $\phi : B^1 \times ... \times B^N \longrightarrow \Delta(A)$

Stronger Expert

Stronger Expert

- c is **AS STRONG AS** $\mathrm{b}^1, \ldots, \mathrm{b}^N$:

$$\mathcal{B}(\mathrm{c}) \le \min_{1 \le n \le N} \mathcal{B}(\mathrm{b}^n)$$

Stronger Expert

- c is **AS STRONG AS** $\mathrm{b}^1, ..., \mathrm{b}^N$:

$$\mathcal{B}(\mathrm{c}) \leq \min_{1 \leq n \leq N} \mathcal{B}(\mathrm{b}^n)$$

- c is **AS STRONG AS THE LINEAR SPAN** of $\mathrm{b}^1, ..., \mathrm{b}^N$ if

$$\mathcal{B}(\mathrm{c}) \leq \min_{w^1, ..., w^N} \mathcal{B}(w^1 \mathrm{b}^1 + ... + w^N \mathrm{b}^N)$$

Calibeating vs Stronger Expert

Calibeating vs Stronger Expert

- c **MULTI-CALIBEATS** $b_1, \ldots, b_N$:

$$\mathcal{B}(c) \le \min_{1 \le n \le N} \min_{\phi^n} \mathcal{B}(\phi^n(b^n))$$

where $\phi^n : B^n \to \Delta(A)$

Calibeating vs Stronger Expert

- c **MULTI-CALIBEATS** $b_1, ..., b_N$:

$$\mathcal{B}(\mathrm{c}) \le \min_{1 \le n \le N} \min_{\phi^n} \mathcal{B}(\phi^n(\mathrm{b}^n))$$

where $\phi^n : B^n \longrightarrow \Delta(A)$

- c is **AS STRONG AS** $b^1, ..., b^N$:

$$\mathcal{B}(\mathrm{c}) \le \min_{1 \le n \le N} \mathcal{B}(\mathrm{b}^n)$$

Calibeating $\succ$ Stronger Expert

- c **MULTI-CALIBEATS** $\mathrm{b}_1, \ldots, \mathrm{b}_N$:

$$\mathcal{B}(\mathrm{c}) \le \min_{1\le n\le N} \min_{\phi^n} \mathcal{B}(\phi^n(\mathrm{b}^n))$$

where $\phi^n : B^n \longrightarrow \Delta(A)$

- c is **AS STRONG AS** $\mathrm{b}^1, \ldots, \mathrm{b}^N$:

$$\mathcal{B}(\mathrm{c}) \le \min_{1\le n\le N} \mathcal{B}(\mathrm{b}^n)$$

CALIBEATING is ***stronger*** !

Calibeating vs Stronger Expert

Calibeating vs Stronger Expert

- c **CALIBEATS THE JOINT** of $\mathrm{b}_1, ..., \mathrm{b}_N$:

$$\mathcal{B}(\mathrm{c}) \le \min_{\phi} \mathcal{B}(\phi(\mathrm{b}^1, ..., \mathrm{b}^N))$$

where $\phi : B^1 \times ... \times B^N \to \Delta(A)$

Calibeating vs Stronger Expert

- $\mathbf{c}$ **CALIBEATS THE JOINT** of $\mathbf{b}_1, ..., \mathbf{b}_N$:

$$\mathcal{B}(\mathbf{c}) \leq \min_{\phi} \mathcal{B}(\phi(\mathbf{b}^1, ..., \mathbf{b}^N))$$

where $\phi : B^1 \times ... \times B^N \rightarrow \Delta(A)$

- $\mathbf{c}$ is **AS STRONG AS THE LINEAR SPAN** of $\mathbf{b}^1, ..., \mathbf{b}^N$ if

$$\mathcal{B}(\mathbf{c}) \leq \min_{w^1,...,w^N} \mathcal{B}(w^1\mathbf{b}^1 + ... + w^N\mathbf{b}^N)$$

Calibeating $\succ$ Stronger Expert

- $\mathbf{c}$ **CALIBEATS THE JOINT** of $\mathrm{b}_1, \ldots, \mathrm{b}_N$:

$$\mathcal{B}(\mathrm{c}) \le \min_{\phi} \mathcal{B}(\phi(\mathrm{b}^1, \ldots, \mathrm{b}^N))$$

where $\phi : B^1 \times \ldots \times B^N \to \Delta(A)$

- $\mathbf{c}$ is **AS STRONG AS THE LINEAR SPAN** of $\mathrm{b}^1, \ldots, \mathrm{b}^N$ if

$$\mathcal{B}(\mathrm{c}) \le \min_{w^1, \ldots, w^N} \mathcal{B}(w^1 \mathrm{b}^1 + \ldots + w^N \mathrm{b}^N)$$

CALIBEATING is ***stronger*** !

Scoring Rules

Scoring Rules

- **SCORING RULE**: quadratic loss

$$\mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{d}}[||\boldsymbol{a}-\boldsymbol{c}||^{\mathbf{2}}]$$

Scoring Rules

- **SCORING RULE**: quadratic loss

$$\mathbb{E}_{a \sim d}[||a - c||^2]$$

  - minimal when $c = d$ (true distribution of a)

Scoring Rules

- **SCORING RULE**: quadratic loss

$$\mathbb{E}_{a \sim d}[||a - c||^2]$$

  - minimal when $c = d$ (true distribution of a)
- **PROPER** scoring rule $L(d, c)$:
  - expected loss is minimized when the forecast is correct, i.e., $c = d$

Scoring Rules

- **SCORING RULE**: quadratic loss

$$\mathbb{E}_{a \sim d}[||a - c||^2]$$

  - minimal when $c = d$ (true distribution of a)
- **PROPER** scoring rule $L(d, c)$:
  - expected loss is minimized when the forecast is correct, i.e., $c = d$
  - $\longrightarrow$ **DIVERGENCE** $D \equiv D^L$
    $D(d, c) \geq 0$ and $D(d, d) = 0$
    (replaces $\|\cdot\|^2$)

Brier, Calibration, Refinement

Brier, Calibration, Refinement

$$\mathcal{B}_T^L \quad := \quad \frac{1}{T}\sum_{t=1}^{T} D^L(a_t, c_t)$$

Brier, Calibration, Refinement

$$\mathcal{B}_T^L \quad := \quad \frac{1}{T}\sum_{t=1}^{T} D^L(a_t, c_t)$$

$$\mathcal{K}_T^L \quad := \quad \frac{1}{T}\sum_{t=1}^{t} D^L(\bar{a}_T(c_t), c_t)$$

Brier, Calibration, Refinement

$$\mathcal{B}_T^L := \frac{1}{T}\sum_{t=1}^{T} D^L(a_t, c_t)$$

$$\mathcal{K}_T^L := \frac{1}{T}\sum_{t=1}^{t} D^L(\bar{a}_T(c_t), c_t)$$

$$\mathcal{R}_T^L := \frac{1}{T}\sum_{t=1}^{T} D^L(a_t, \bar{a}_T(c_t))$$

Proper Calibration / Calibeating

Proper Calibration / Calibeating

- **PROPER CALIBRATION / CALIBEATING**: calibration / calibeating w.r.t. all 1-bounded proper scoring rules L

Proper Calibration / Calibeating

- **PROPER CALIBRATION / CALIBEATING**: calibration / calibeating w.r.t. all 1-bounded proper scoring rules L (uniformly in L)

Proper Calibration / Calibeating

- **PROPER CALIBRATION / CALIBEATING**: calibration / calibeating w.r.t. all 1-bounded proper scoring rules L (uniformly in L)
  - **PROPER-CALIBRATION**:

$$\sup_{L} \sup_{\mathbf{a}} \mathbb{E}\left[\mathcal{K}_T^L(\mathbf{c})\right] \leq o(1) \quad \text{(as } T \to \infty\text{)}$$

Proper Calibration / Calibeating

- **PROPER CALIBRATION / CALIBEATING**: calibration / calibeating w.r.t. all 1-bounded proper scoring rules L (uniformly in L)
  - **PROPER-CALIBRATION**:

$$\sup_{L} \sup_{\mathbf{a}} \mathbb{E}\left[\mathcal{K}_T^L(\mathbf{c})\right] \leq o(1) \quad \text{(as } T \to \infty\text{)}$$

  - **PROPER-CALIBEATING**:

$$\sup_{L} \sup_{\mathbf{a},\mathbf{b}} \mathbb{E}\left[\mathcal{B}_T^L(\mathbf{c}) - \mathcal{R}_T^L(\mathbf{b})\right] \leq o(1)$$

(as $T \to \infty$)

Proper Calibration / Calibeating

Proper Calibration / Calibeating

Results

Proper Calibration / Calibeating

Results

1. Calibration $\Longrightarrow$ **PROPER-CALIBRATION**

Proper Calibration / Calibeating

Results

1. Calibration $\Longrightarrow$ **PROPER-CALIBRATION**
2. Calibeating $\Longrightarrow$ **PROPER-CALIBEATING**

Proper Calibration / Calibeating

Results

1. Calibration $\Longrightarrow$ **PROPER-CALIBRATION**
2. Calibeating $\not\Longrightarrow$ **PROPER-CALIBEATING !!**

Proper Calibration / Calibeating

Results

1. Calibration $\Longrightarrow$ **PROPER-CALIBRATION**
2. Calibeating $\not\Longrightarrow$ **PROPER-CALIBEATING !!**
3. Calibeating the joint is **PROPER-CALIBEATING**

Proper Calibration / Calibeating

Results

1. Calibration $\Longrightarrow$ **PROPER-CALIBRATION**
2. Calibeating $\not\Longrightarrow$ **PROPER-CALIBEATING !!**
3. Calibeating the joint is **PROPER-CALIBEATING**
4. “A simple way to calibeat” ($c_t = \bar{a}^{\mathrm{b}}_{t-1}$) is **LIPSCHITZ PROPER-CALIBEATING**

Proper Calibration / Calibeating

Results

1. Calibration $\Longrightarrow$ **PROPER-CALIBRATION**
2. Calibeating $\not\Longrightarrow$ **PROPER-CALIBEATING !!**
3. Calibeating the joint is **PROPER-CALIBEATING**
4. "A simple way to calibeat" ($c_t = \bar{a}^{\mathrm{b}}_{t-1}$) is **PROPER-LI-CALIBEATING**

Proper Calibration / Calibeating

Results

1. Calibration $\Longrightarrow$ **PROPER-CALIBRATION**
2. Calibeating $\not\Longrightarrow$ **PROPER-CALIBEATING !!**
3. Calibeating the joint is **PROPER-CALIBEATING**
4. "A simple way to calibeat" ($c_t = \bar{a}^{\mathbf{b}}_{t-1}$) is **PROPER-LI-CALIBEATING**
5. **CONTINUOUS PROPER-LI-CALIBEATING** by a deterministic procedure

Proper Calibration / Calibeating

Results

1. Calibration $\Longrightarrow$ **PROPER-CALIBRATION**
2. Calibeating $\not\Longrightarrow$ **PROPER-CALIBEATING !!**
3. Calibeating the joint is **PROPER-CALIBEATING**
4. "A simple way to calibeat" ($c_t = \bar{a}^{\mathrm{b}}_{t-1}$) is **PROPER-LI-CALIBEATING**
5. **CONTINUOUS PROPER-LI-CALIBEATING** by a deterministic procedure
6. **PROPER-MULTICALIBEATING**

Decision Making Under Uncertainty

Decision Making Under Uncertainty

- $u : A \times X \to \mathbb{R}$
  - A : states of nature
  - X : decisions

Decision Making Under Uncertainty

- $\boldsymbol{u} : \boldsymbol{A} \times \boldsymbol{X} \longrightarrow \mathbb{R}$
  - $\boldsymbol{A}$: states of nature
  - $\boldsymbol{X}$: decisions
- For every distribution $\boldsymbol{d} \in \boldsymbol{\Delta}(\boldsymbol{A})$:
  - **OPTIMAL DECISION**:
    $\boldsymbol{x}^*(\boldsymbol{d}) \in \mathbf{argmax}_{\boldsymbol{x} \in \boldsymbol{X}} \mathbb{E}_{\boldsymbol{a} \sim \boldsymbol{d}} \left[\boldsymbol{u}(\boldsymbol{a}, \boldsymbol{x}) \right]$

Decision Making Under Uncertainty

- $u : A \times X \longrightarrow \mathbb{R}$
  - A : states of nature
  - X : decisions

- For every distribution $d \in \Delta(A)$:
  - **OPTIMAL DECISION**:
    $x^*(d) \in \text{argmax}_{x \in X} \mathbb{E}_{a \sim d} [u(a, x)]$
  - **LOSS**: $L^u(a, c) = -u(a, x^*(c))$

Decision Making Under Uncertainty

- $u : A \times X \longrightarrow \mathbb{R}$
  - A : states of nature
  - X : decisions
- For every distribution $d \in \Delta(A)$:
  - **OPTIMAL DECISION**:
    $x^*(d) \in \operatorname{argmax}_{x \in X} \mathbb{E}_{a \sim d}\left[u(a, x)\right]$
  - **LOSS**: $L^u(a, c) = -u(a, x^*(c))$

$\Rightarrow$ L^u is a **PROPER SCORING RULE**

Decision Making Under Uncertainty

FORECASTING SEQUENCE $\mathbf{c} = (c_t)$

Decision Making Under Uncertainty

FORECASTING SEQUENCE $\mathbf{c} = (c_t)$
generates **DECISION SEQUENCE** $x^*(c_t)$

Decision Making Under Uncertainty

FORECASTING SEQUENCE $\mathbf{c} = (c_t)$
generates **DECISION SEQUENCE** $x^*(c_t)$

- **AVERAGE UTILITY**

$$\mathcal{U}_T(\mathbf{c}) := \frac{1}{T} \sum_{t=1}^{T} u(a_t, x^*(c_t))$$

Decision Making Under Uncertainty

FORECASTING SEQUENCE $\mathbf{c} = (c_t)$
generates **DECISION SEQUENCE** $x^*(c_t)$

- **AVERAGE UTILITY**

$$\mathcal{U}_T(\mathbf{c}) := \frac{1}{T}\sum_{t=1}^{T} u(a_t, x^*(c_t))$$

- **REGRET OF (BEST REPLYING TO) FORECASTS** $\mathbf{c}$

$$\text{REG}_T^u(\mathbf{c}) := \max_{\xi:\Delta(A)\to X} \frac{1}{T}\sum_{t=1}^{T} u(a_t, \xi(c_t)) - \mathcal{U}_T(\mathbf{c})$$

Decision Making Under Uncertainty

Decision Making Under Uncertainty

REGRET $=$ **CALIBRATION**

$$\text{REG}_T^u(\mathbf{c}) = \mathcal{K}_T^{L^u}(\mathbf{c})$$

Decision Making Under Uncertainty

REGRET = **CALIBRATION**

$$\mathbf{REG}_T^u(\mathbf{c}) = \mathcal{K}_T^{L^u}(\mathbf{c})$$

a forecasting procedure is

PROPER-CALIBRATED

$\Leftrightarrow$

EVERY bounded decision maker

has **NO REGRET**

Decision Making Under Uncertainty

REGRET $=$ **CALIBRATION**

$$\mathbf{REG}_T^u(\mathbf{c}) = \mathcal{K}_T^{L^u}(\mathbf{c})$$

a forecasting procedure is

PROPER-CALIBRATED

$\Leftrightarrow$

EVERY bounded decision maker

has **NO REGRET**

PROPER-CALIBRATION = UNIVERSAL NO REGRET

Decision Making Under Uncertainty

Decision Making Under Uncertainty

- c **PROPER-CALIBEATS** b $\Rightarrow$

Decision Making Under Uncertainty

- c **PROPER-CALIBEATS** b $\Rightarrow$

$$\mathcal{U}_T(\mathrm{c}) \geq \mathcal{U}_T(\mathrm{b}) + \textbf{REG}_T^u(\mathrm{b}) - o(1)$$

for **EVERY** bounded u

What does every forecaster deserve ?

What does every forecaster deserve ?

A PROPER CALIBEATING

What does every forecaster deserve ?

A PROPER CALIBEATING