Blackwell Approachability and Calibrated Forecasting

Yen-Huan Li
yen-huan.li@epfl.ch

Laboratory for Information and Inference Systems (LIONS)
École Polytechnique Fédérale de Lausanne (EPFL)
Recommended reading materials


Table of Contents

1. Calibrated Probability Forecasting
2. Blackwell Approachability
3. Existence of a Calibrated Forecasting Strategy
Problem of Binary Probability Forecasting
Can One Predict the Future?

Given an arbitrary sequence of binary digits

0 1 1 0 1 0 0 0 1 0 1,

can one predict the next digit, *without any probabilistic assumption*?

---

**Weather Forecasting [3]**

Think of ‘1’ as “a rainy day”, and ‘0’ otherwise.

Given the weather history (say, of the last three years), *without any more information*, can one tell whether tomorrow will be rainy or not?
Probability Forecasting

What we typically see in a weather forecast is a probability forecast, e.g.,

The chance of rain is 80% tomorrow.

Binary Probability Forecasting Protocol

For $t = 1, 2, 3, \ldots$,
- Forecaster announces $p_t \in [0, 1]$.
- Reality announces $y_t \in \{0, 1\}$, without knowing $p_t$.

Definition (Strategy)

A strategy of Forecaster is a possibly random mapping:

$$p_t = f(p_1, y_1, p_2, y_2, \ldots, p_{t-1}, y_{t-1}, \xi_t),$$

where $\xi_t$ is a random variable (r.v.). Similarly, a strategy of Reality is a mapping

$$y_t = g(p_1, y_2, p_2, y_2, \ldots, p_{t-1}, y_{t-1}, \eta_t),$$

where $\eta_t$ is another r.v. independent of $\xi_t$. 
Calibrated Forecasting: Intuition

**Intuitive Definition**

A strategy of **FORECASTER** is *calibrated*, if for any strategy of **REALITY** and any $\rho \in [0, 1]$,

$$\frac{\sum_{t=1}^{T} \mathbb{I}\{y_t=1, p_t \approx \rho\}}{\sum_{t=1}^{T} \mathbb{I}\{p_t \approx \rho\}} \approx \rho.$$ 

**Example**

Suppose the weather forecast is calibrated. Then if the chance of rain is said to be 80%, a reasonable person should take an umbrella.

Obviously, this “definition” is not precise enough to work with. Some things to clarify:

1. What is the value of $T$?
2. What is the precise definition of “$p_t \approx \rho$”?
3. What is the precise definition of “the ratio $\approx \rho$”? 

Blackwell Approachability and Calibrated Forecasting | Prof. Volkan Cevher, volkan.cevher@epfl.ch
Consider an \( \varepsilon \)-covering \( Q_\varepsilon := \{ q_1, q_2, \ldots, q_{M_\varepsilon} \} \) of the interval \([0, 1]\). Suppose \( p_t \) only takes value in \( Q_\varepsilon \). Define the empirical frequencies

\[
\begin{align*}
    f_m(T) &:= \frac{\sum_{t=1}^{T} I\{p_t=q_m, y_t=1\}}{\sum_{t=1}^{T} I\{p_t=q_m\}}, \quad m = 1, 2, \ldots, M_\varepsilon, \\
    &\text{whenever the denominator is non-zero, and } f_m(T) := q_m \text{ otherwise.}
\end{align*}
\]

We want

\[
\limsup_{T \to \infty} |f_m(T) - q_m| \leq \varepsilon \quad \text{a.s.}
\]

We would like to ignore those \( m \)'s such that \( \sum_{t=1}^{T} I\{p_t=q_m\} = o(T) \). A compact formulation is to require

\[
\limsup_{T \to \infty} \sum_{m=1}^{M_\varepsilon} |f_m(T) - q_m| \left( \frac{\sum_{t=1}^{T} I\{p_t=q_m\}}{T} \right) \leq \varepsilon \quad \text{a.s.}
\]

Some calculation leads to the equivalent definition of \( \varepsilon \)-calibration in the next slide.
Calibrated Forecasting: Formalism (2/2)

Definition ($\varepsilon$-calibration [3, 9])

Consider an $\varepsilon$-covering $Q_\varepsilon := \{q_1, q_2, \ldots, q_{M_\varepsilon}\}$ of the interval $[0, 1]$. A strategy of FORECASTER taking values in $Q_\varepsilon$ is $\varepsilon$-calibrated, if for any strategy of REALITY,

$$\limsup_{T \to \infty} \frac{1}{T} \sum_{m=1}^{M_\varepsilon} \sum_{t=1}^{T} \mathbb{I}\{p_t = q_m\} \left| q_m - y_t \right| \leq \varepsilon, \text{ a.s.}$$

Definition (Calibration [3, 13])

A strategy of FORECASTER is calibrated, if for any strategy of REALITY, $\varepsilon > 0$, and $\rho \in [0, 1],$

$$\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}\{|p_t - \rho| \leq \varepsilon\} \left| p_t - y_t \right| = 0, \text{ a.s.}$$
Other Applications and Implications

**Computing a correlated equilibrium [8]**
For almost every two-player repeated game, the set of correlated equilibria equals the set of limit empirical distribution of actions by calibrated learning.

**Testing [17]**
If the truth passes a \textit{finite test} with probability $\rho$, then there exists an ignorant algorithm that passes the test with the same probability.

**Contextual online learning [19]**
If the spaces of side information and outcomes are locally compact, and the loss function is continuous and compact type, there exists a \textit{universally consistent} prediction algorithm.

**Approximating a Nash equilibrium [7]**
There exists an uncoupled strategy, such that in every repeated game, the empirical distribution of all players’ actions converges to an $\varepsilon$-Nash equilibrium almost surely.
Some Impossibility Facts

Binary Probability Forecasting Protocol

For $t = 1, 2, 3, \ldots$,

- **Forecaster** announces $p_t = f(p_1, y_1, \ldots, p_{t-1}, y_{t-1}, \xi_t) \in [0, 1]$.
- **Reality** announces $y_t = g(p_1, y_1, \ldots, p_{t-1}, y_{t-1}, \eta_t) \in \{0, 1\}$, without knowing $p_t$.

Fact 1: Addressing **Reality** that knows $p_t$ Is Impossible

If $p_t$ is known, **Reality** can choose $y_t \sim \text{Bernoulli}(\varphi(p_t))$ independently for all $t$, using some function $\varphi : [0, 1] \to [0, 1]$ very different from the identity mapping.

Fact 2: Deterministic Forecasting Is Impossible [4, 9, 16]

If **Forecaster** adopts a deterministic strategy, i.e., $\xi_t$ is absent, then **Reality** can get the exact value of $p_t$ by computing $f(p_1, y_1, \ldots, p_{t-1}, y_{t-1})$. 
A Weaker Notion of Calibration

Definition (Weak calibration)
A strategy of FORECASTER is weakly calibrated with respect to a set $\mathcal{F}$ of functions $f : [0, 1] \rightarrow \mathbb{R}$, if for every $f \in \mathcal{F}$,

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} f(p_t) |y_t - p_t| = 0, \quad \text{a.s.}$$

Facts
▶ There exists deterministic weakly calibrated strategies for FORECASTER, even against REALITY that knows $p_t$ [12, 20].
▶ The weakly calibrated strategy in [12] followed by random rounding leads to an $\varepsilon$-calibrated strategy.
▶ Weak calibration is computationally hard in general [10].
### Goal of This Lecture

<table>
<thead>
<tr>
<th>Theorem ([6, 9, 13])</th>
</tr>
</thead>
<tbody>
<tr>
<td>There exists an ( \varepsilon )-calibrated forecasting strategy for Forecaster.</td>
</tr>
</tbody>
</table>

**Proof.**
Cast the problem of calibrated forecasting as an approachability problem. Prove by Blackwell’s approachability theorem.

<table>
<thead>
<tr>
<th>Theorem ([13])</th>
</tr>
</thead>
<tbody>
<tr>
<td>There exists a calibrated forecasting strategy for Forecaster.</td>
</tr>
</tbody>
</table>

**Proof.**
Apply the Borel-Cantelli lemma.
Blackwell Approachability
Repeated Games with Vector Payoffs

Protocol of a Repeated Game with Vector Payoffs

For $t = 1, 2, \ldots$,

1. **Alice** announces $a_t \in A$.
2. **Bob** announces $b_t \in B$, without knowing $a_t$.
3. **Alice** observes the payoff vector $f(a_t, b_t) \in \mathbb{R}^p$.

Notes

1. We assume both $A$ and $B$ are of finite cardinalities.
2. A **strategy** of **Alice** is a mapping $h_A : (A \times B)^* \rightarrow \Delta(A)$; Alice chooses $a_t$ randomly according to $p_t = h_A(a_1, b_1, \ldots, a_{t-1}, b_{t-1}) \in \Delta(A)$.
3. A strategy $h_B : (A \times B)^* \rightarrow \Delta(B)$ of **Bob** is defined similarly.
4. For any $p_t \in \Delta(A)$, $q_t \in \Delta(B)$, define $f(p_t, q_t) := \sum_{a \in A, b \in B} f(a, b)p_t(a)q_t(b)$. 
(Blackwell) Approachability

Protocol of a Repeated Game with Vector Payoffs

For \( t = 1, 2, \ldots \),

1. **Alice** announces \( a_t \in A \).
2. **Bob** announces \( b_t \in B \), without knowing \( a_t \).
3. **Alice** observes the payoff vector \( f(a_t, b_t) \in \mathbb{R}^p \).

Definition (Blackwell Approachability [1])

A set \( \mathcal{X} \subseteq \mathbb{R}^p \) is **approachable**, if **Alice** has a strategy such that, for any strategy of **Bob**,

\[
\lim_{T \to \infty} \text{dist} \left( \bar{f}_T, \mathcal{X} \right) = 0, \quad \text{a.s.,}
\]

where \( \bar{f}_T := T^{-1} \sum_{t=1}^{T} f(a_t, b_t) \), and \( \text{dist}(v, \mathcal{X}) := \|v - \text{proj}_{\mathcal{X}}(v)\|_2 \).
Example: Learning with Expert Advice (1/2)

Protocol of Learning with Expert Advice

For \( t = 1, 2, \ldots \),

1. **EXPERT-\( i \)** announces \( \gamma_t(i) \in \Gamma, \ i = 1, 2, \ldots, n \).
2. **LEARNER** announces \( \gamma_t \in \Gamma \).
3. **REALITY** announces \( \omega_t \in \Omega \), **without knowing** \( \gamma_t \).
4. **LEARNER** observes the losses \( \lambda(\gamma_t, \omega_t) \) and \( \lambda_t(\gamma_t(i), \omega_t) \) for all \( i \).

Definition (No Regret)

We say that a (possibly random) strategy of **LEARNER** feels no regret, if

\[
\lim_{T \to \infty} \frac{1}{T} \left( \sum_{t=1}^{T} \lambda(\gamma_t, \omega_t) - \sum_{t=1}^{T} \lambda(\gamma_t(i), \omega_t) \right) \leq 0, \quad \text{for all } i \text{ a.s.}
\]
Example: Learning with Expert Advice (2/2)

Protocol of Learning with Expert Advice

For \( t = 1, 2, \ldots \),

1. **Expert-i** announces \( \gamma_t(i) \in \Gamma, \ i = 1, 2, \ldots, n \).
2. **Learner** announces \( \gamma_t \in \Gamma \).
3. **Reality** announces \( \omega_t \in \Omega \), without knowing \( \gamma_t \).
4. **Learner** observes the vector payoff \( f_t := (\lambda(\gamma_t, \omega_t) - \lambda(\gamma_t(i), \omega_t))_{1 \leq i \leq n} \in \mathbb{R}^n \).

Definition (No Regret)

Let \( X \) be the negative orthant, i.e., \( X := (-\infty, 0]^n \).

We say that a (possibly random) strategy of **Learner** feels no regret, if \( X \) is approachable by **Learner**, i.e.,

\[
\lim_{T \to \infty} \text{dist} \left( \frac{1}{T} \sum_{t=1}^{T} f_t, X \right) = 0, \quad \text{a.s.}
\]
Theorem (Blackwell’s Approachability Theorem [1])

A closed convex set $\mathcal{X}$ is approachable, if and only if for any $q_B \in \Delta(\mathcal{B})$, there exists some $p_A \in \Delta(\mathcal{A})$, such that $f(p_A, q_B) \in \mathcal{X}$.

The only if part is obvious.

Proof of the Only If Part.

Suppose there exists some $q_B^* \in \Delta(\mathcal{B})$, such that $f(p_A, q_B^*) \notin \mathcal{X}$ for all $q_A \in \Delta(\mathcal{A})$. Then Bob can simply choose $q_t = q_B^*$ for all $t$. □
Tools (1/2): von Neumann’s Minimax Theorem

**Definition (Convex-Concave Function)**

A function $g : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ is convex-concave, if $g(\cdot, y)$ is convex for any $y \in \mathcal{Y}$, and $g(x, \cdot)$ is concave for any $x \in \mathcal{X}$.

**Theorem (von Neumann’s Minimax Theorem [18])**

Let $\mathcal{X} \subseteq \mathbb{R}^{d_1}$ and $\mathcal{Y} \subseteq \mathbb{R}^{d_2}$ be convex compact sets. Let $g : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ be convex-concave. Then it holds that

$$
\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} g(x, y) = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} g(x, y).
$$
Proof of the If Part (1/5)

Lemma (Blackwell’s Condition)

If for any \( q_B \in \Delta(B) \), there exists some \( p_A \in \Delta(A) \), such that \( f(p_A, q_B) \in \mathcal{X} \), then for any \( v \in \mathbb{R}^p \), there exists some \( p_A \in \Delta(A) \), such that

\[
\langle v - \text{proj}_{\mathcal{X}}(v), f(p_A, p_B) - \text{proj}_{\mathcal{X}}(v) \rangle \leq 0, \quad \text{for all } p_B \in \Delta(B).
\]

Proof.

By the optimality condition, one has

\[
\langle v - \text{proj}_{\mathcal{X}}(v), x - \text{proj}_{\mathcal{X}}(v) \rangle \leq 0, \quad \text{for all } x \in \mathcal{X}.
\]

Apply von Neumann’s minimax theorem, this implies

\[
\min_{p_A} \max_{p_B} \langle v - \text{proj}_{\mathcal{X}}(v), f(p_A, p_B) - \text{proj}_{\mathcal{X}}(v) \rangle = \max_{p_B} \min_{p_A} \langle v - \text{proj}_{\mathcal{X}}(v), f(p_A, p_B) - \text{proj}_{\mathcal{X}}(v) \rangle \leq 0.
\]
Proof of the If Part (2/5)

Define the average payoff:

\[
\bar{f}_T := T^{-1} \sum_{t=1}^{T} f(a_t, b_t).
\]

**A Strategy of Alice [1]**

For every \( t \in \mathbb{N} \), find some \( p_t \in \Delta(A) \) such that

\[
\langle \bar{f}_{t-1} - \text{proj}_X(\bar{f}_{t-1}), f(p_t, q) - \text{proj}_X(\bar{f}_{t-1}) \rangle \leq 0, \quad \text{for all } q \in \Delta(B).
\]

Announce \( a_t \in A \) randomly according to \( p_t \).

**Note**

- Existence of \( p_t \) is guaranteed by Blackwell’s condition.
- \( p_t \) can be found by solving the bilinear saddle point problem [11, 15]

\[
p_t \in \arg \min_{p \in \Delta(A)} \max_{q \in \Delta(B)} \langle \bar{f}_{t-1} - \text{proj}_X(\bar{f}_{t-1}), f(p, q) - \text{proj}_X(\bar{f}_{t-1}) \rangle.
\]
Proof of the If Part (3/5)

Observation

It suffices to consider the approachability of

\[ \tilde{X} := X \cap \text{conv} \left( \{ f(a, b) \mid a \in A, b \in B \} \right). \]

Lemma

Define

\[ M := \max_{a \in A, b \in B} \| f(a, b) \|_2, \quad R := \max_{v \in \tilde{X}} \| v \|_2. \]

The strategy in the previous slide ensures:

1. \( \mathbb{E} \left[ \text{dist}(\bar{f}_t, X) \right] \leq t^{-1/2}(M + R) \) for all \( t \).
2. \( \text{dist}(\bar{f}_t, X) \) converges to zero almost surely, as \( t \to \infty \).
Proof of the If Part (4/5)

Define \( f_t := f(a_t, b_t) \), \( \pi_t := \text{proj}_\mathcal{X}(\bar{f}_t) \), and \( \delta_t := \text{dist}(\bar{f}_t, \mathcal{X}) \).

We write
\[
\delta^2_{t+1} \leq \left\| \bar{f}_{t+1} - \pi_t \right\|^2_2 \\
= \left( \frac{t}{t+1} \right)^2 \delta^2_t + \left( \frac{1}{t+1} \right)^2 \|f_{t+1} - \pi_t\|^2_2 + \frac{2t}{(t+1)^2} \langle \bar{f}_t - \pi_t, f_{t+1} - \pi_t \rangle.
\]

Denote by \( E_t \) the conditional expectation w.r.t. \((a_1, b_1, \ldots, a_t, b_t)\). By the Blackwell condition, we write
\[
E_t \left[ \delta^2_{t+1} \right] \leq \left( \frac{t}{t+1} \right)^2 \delta^2_t + \left( \frac{1}{t+1} \right)^2 \|f_{t+1} - \pi_t\|^2_2 \\
\leq \left( \frac{t}{t+1} \right)^2 \delta^2_t + \left( \frac{1}{t+1} \right)^2 (M + R)^2.
\]

By induction, we obtain
\[
E \left[ \delta^2_t \right] \leq \frac{(M + R)^2}{t}.
\]
**Tools (2/2): Doob’s Martingale Convergence Theorem**

**Definition (Supermartingale)**

A sequence of random variables $X_1, X_2, \ldots$ forms a supermartingale, if $E[|X_t|] < +\infty$ for all $t$, and

$$E[X_{t+1}|X_1, X_2, \ldots, X_t] \leq X_t.$$

**Theorem (Doob’s Martingale Convergence Theorem [5])**

A non-negative supermartingale converges to an integrable random variable almost surely.
Proof of the If Part (5/5)

Define the random variables

\[ \xi_t := \delta_t^2 + (M + R)^2 \sum_{\tau > t} \frac{1}{(\tau + 1)^2}. \]

Then

\[ \mathbb{E}_t [\xi_{t+1}] \leq \left( \frac{t}{t + 1} \right)^2 \delta_t^2 + \sum_{\tau > t} \frac{1}{(\tau + 1)^2} \leq \xi_t, \]

implying that \((\xi_t)_{t \in \mathbb{N}}\) is a non-negative supermartingale.

Hence \((\delta_t)_{t \in \mathbb{N}}\) converges to an integrable random variable \(\delta_\infty\) almost surely. Since \(\mathbb{E} \left[ \delta_t^2 \right] \to 0\), we have \(\delta_\infty = 0\) almost surely.
Blackwell’s Approachability Theorem and an Achieving Strategy

**Theorem (Blackwell’s Approachability Theorem [1])**

A closed convex set $\mathcal{X}$ is approachable, if and only if for any $q_B \in \Delta(B)$, there exists some $p_A \in \Delta(A)$, such that $f(p_A, q_B) \in \mathcal{X}$.

**A Strategy of Alice [1]**

For every $t \in \mathbb{N}$, find some $p_t \in \Delta(A)$ such that

$$
\left\langle \bar{f}_{t-1} - \text{proj}_{\mathcal{X}}(\bar{f}_{t-1}), f(p_t, q) - \text{proj}_{\mathcal{X}}(\bar{f}_{t-1}) \right\rangle \leq 0, \quad \text{for all } q \in \Delta(B).
$$

Announce $a_t \in A$ randomly according to $p_t$.

**Remark**

- Existence of $p_t$ is guaranteed by Blackwell’s condition.
- Computing $p_t$ can be done by solving a bilinear saddle point problem.
Non-Asymptotic Result

We have proved the almost sure convergence of $\text{dist}(\bar{f}_t, \mathcal{X})$ to zero. How fast can the convergence rate be?

**Theorem (Essentially Exercise 7.23 in [2])**

Assume that $\mathcal{X}$ satisfies the Blackwell condition. Define $\tilde{\mathcal{X}}$ as the intersection of $\mathcal{X}$ and the convex hull of $\{f(a, b) | a \in A, b \in B\}$.

There exists a strategy for Alice, such that for any strategy of Bob and any $T \in \mathbb{N}$, it holds with probability at least $1 - \delta$ that

$$\text{dist} \left( \bar{f}_T, \mathcal{X} \right) \leq \frac{M + R}{\sqrt{T}} + 4 \sqrt{\frac{M^2}{cT} \left[ \log \left( \frac{1}{\delta} \right) + 2.5 \right]} = O(T^{-1/2}),$$

where $M := \max_{a, b} \|f(a, b)\|_2$, $R := \max_{v \in \tilde{\mathcal{X}}} \|v\|_2$, and $c$ is a positive constant.
**Proof (1/3)**

Define the average expected payoff:

\[ \tilde{f}_T := T^{-1} \sum_{t=1}^{T} f(p_t, b_t). \]

**An Improved Strategy of ALICE [2]**

For every \( t \in \mathbb{N} \), find some \( p_t \in \Delta(A) \) such that

\[ \langle \tilde{f}_{t-1} - \text{proj}_X(\tilde{f}_{t-1}), f(p_t, q) - \text{proj}_X(\tilde{f}_{t-1}) \rangle \leq 0, \quad \text{for all } q \in \Delta(B). \]

Announce \( a_t \in A \) randomly according to \( p_t \).

**Note**

- This is slightly different from the original strategy Blackwell proposed.
- Existence of \( p_t \) is guaranteed by Blackwell’s condition.
- Computing \( p_t \) can be done by solving a bilinear saddle point problem.
Proof (2/3)

Define $\tilde{\pi}_t := \text{proj}_{\mathcal{X}}(\tilde{f}_t)$ and $\tilde{\delta}_t := \text{dist}(\tilde{f}_t, \mathcal{X})$.

Similarly as in the previous proof, we write

$$\tilde{\delta}_{t+1}^2 \leq \left\| \tilde{f}_{t+1} - \tilde{\pi}_t \right\|_2^2$$

$$= \left( \frac{t}{t+1} \right)^2 \tilde{\delta}_t^2 + \left( \frac{1}{t+1} \right)^2 \left\| f(p_{t+1}, b_t) - \tilde{\pi}_t \right\|_2^2 + \frac{2t}{(t+1)^2} \left\langle \tilde{f}_t - \tilde{\pi}_t, f(p_t, b_t) - \tilde{\pi}_t \right\rangle$$

$$\leq \left( \frac{t}{t+1} \right)^2 \tilde{\delta}_t^2 + \left( \frac{1}{t+1} \right)^2 (M + R)^2.$$

By induction, we get

$$\tilde{\delta}_t \leq \frac{M + R}{\sqrt{t}}.$$
**Tool: Vector-Valued Martingale**

**Definition (Vector-Valued Martingale)**

Consider a sequence of random vectors \((X_t)_{t \in \mathbb{N}}\). We say that \((X_t)_{t \in \mathbb{N}}\) is a (vector-valued) martingale, if for all \(t\),

\[
\begin{align*}
\mathbb{E}[\|X_t\|_2] &< +\infty, \\
\mathbb{E}[X_{t+1}|X_1, X_2, \ldots, X_t] &= X_t.
\end{align*}
\]

**Theorem (Azuma Inequality for Vector-Valued Martingales [14])**

There exists a positive constant \(c\), such that for any vector-valued martingale \((X_t)_{t \in \mathbb{N}}\) satisfying \(\|X_{t+1} - X_t\|_2 \leq a_t\) for all \(t\), it holds that

\[
P\left(\|X_{t+1} - X_1\|_2 \geq u\right) \leq e^{2.5} \exp \left(-\frac{cu^2}{\sum_{\tau=1}^{t} a_{\tau}^2}\right),
\]

for any \(t \in \mathbb{N}\) and \(u > 0\).
Proof (3/3)

Recall that $\bar{f}_T := T^{-1} \sum_{t=1}^T f(a_t, b_t)$.

We write

$$\text{dist}(\bar{f}_t, X) = \|\bar{f}_t - \tilde{f}_t + \tilde{f}_t - \tilde{\pi}_t + \tilde{\pi}_t - \pi_t\|_2$$

$$\leq \|\bar{f}_t - \tilde{f}_t\|_2 + \|\tilde{f}_t - \tilde{\pi}_t\|_2 + \|\tilde{\pi}_t - \pi_t\|_2$$

$$\leq \frac{M + R}{\sqrt{t}} + 2\|\bar{f}_t - \tilde{f}_t\|_2,$$

where we have used the non-expansiveness of projections.

Define $X_t := \sum_{\tau=1}^t f(a_\tau, b_\tau) - f(p_\tau, b_\tau)$. Then $(X_t)_{t \in \mathbb{N}}$ is a vector-valued martingale, and $\|X_{t+1} - X_t\|_2 \leq 2M$. By the Azuma inequality, we have

$$P\left(\|\bar{f}_t - \tilde{f}_t\|_2 \geq u\right) \leq e^{2.5} \exp\left(\frac{-c t u^2}{4M^2}\right).$$
Calibrated Forecasting
$\varepsilon$-Calibrated Binary Probability Forecasting

Consider an $\varepsilon$-covering $Q_\varepsilon := \{q_1, q_2, \ldots, q_{M_\varepsilon}\}$ of the interval $[0, 1]$.

**Binary Probability Forecasting Protocol**

For $t = 1, 2, 3, \ldots$,

- **Forecaster** announces $p_t \in Q_\varepsilon$.
- **Reality** announces $y_t \in \{0, 1\}$, without knowing $p_t$.

**Definition ($\varepsilon$-calibration [3, 9])**

A strategy of **Forecaster** is $\varepsilon$-calibrated, if for any strategy of **Reality**,\[\limsup_{T \to \infty} \frac{1}{T} \sum_{m=1}^{M_\varepsilon} \sum_{t=1}^{T} \mathbb{I}\{p_t = q_m\} |q_m - y_t| \leq \varepsilon, \quad \text{a.s.}\]
Equivalent Formulation (1/2)

Binary Probability Forecasting Protocol

For $t = 1, 2, 3, \ldots$,
- **FORECASTER** announces $p_t \in Q_{\varepsilon}$.
- **REALITY** announces $y_t \in \{0, 1\}$, without knowing $p_t$.

Formulation as a Repeated Game with Vector Payoffs

Set $A := Q_{\varepsilon}$ and $B := \{0, 1\}$. Define the payoff function

$$f(a, b) := (0, \ldots, 0, a - b, 0, \ldots, 0) \in \mathbb{R}^{M_{\varepsilon}}.$$  

For $t = 1, 2, 3, \ldots$,
- **ALICE** announces $a_t \in Q_{\varepsilon}$.
- **BOB** announces $b_t \in \{0, 1\}$, without knowing $a_t$.
- **ALICE** observes the payoff $f(a, b) \in \mathbb{R}^{M_{\varepsilon}}$.  

Equivalent Formulation (2/2)

**Definition (ε-calibration [3, 9])**
A strategy of **FORECASTER** is **ε-calibrated**, if for any strategy of **REALITY**,

\[
\limsup_{T \to \infty} \frac{1}{T} \sum_{m=1}^{M_\varepsilon} \sum_{t=1}^{T} \mathbb{I}_{\{p_t = q_m\}} |q_m - y_t| \leq \varepsilon, \text{ a.s.}
\]

**Equivalent definition**
A strategy of **ALICE** is **ε-calibrated**, if \( B_\varepsilon \), the \( \ell_1 \)-norm ball of \( \ell_1 \) radius \( \varepsilon \), is approachable.
Checking Approachability

**Proposition**

The $\ell_1$-norm ball $B_\varepsilon$ is approachable by Alice/Forecaster.

**Proof.**

For any $p_B \in \Delta(B)$, let $p_A$ be the point closest to $p_B$ in $Q_\varepsilon$; then we have $f(p_A, p_B) \in B_\varepsilon$, verifying the Blackwell condition. □

Recall the achieving strategy for Alice/Forecaster.

**An $\varepsilon$-Calibrated Forecasting Strategy**

Compute

$$P_t \in \arg \min_{P \in \Delta(A)} \max_{Q \in \Delta(B)} \left\langle \tilde{f}_t - \text{proj}_{B_1}(\tilde{f}_t), f(P, Q) - \text{proj}_{B_1}(\tilde{f}_t) \right\rangle,$$

where $\tilde{f}_T := T^{-1} \sum_{t=1}^{T} f(P_t, b_t)$. Choose $p_t \in Q_\varepsilon$ randomly, according to $P_t$. 
Non-Asymptotic Result

**Theorem**

For any $\varepsilon > 0$, there exists an $\varepsilon$-calibrated forecasting strategy, such that for all $T \in \mathbb{N}$, with probability at least $1 - \delta$,

$$
\frac{1}{T} \sum_{m=1}^{M_\varepsilon} \sum_{t=1}^{T} \mathbb{I}\{p_t=q_m\} \left| q_m - y_t \right|
\leq \varepsilon + \sqrt{\frac{2}{\varepsilon}} \left( \sqrt{\frac{1 + \varepsilon}{T}} + 4 \sqrt{\frac{1}{cT} \left[ \log \left( \frac{1}{\delta} \right) + 2.5 \right]} \right).
$$

**Proof.**

Choose $Q_\varepsilon$ as the uniform quantization. Then we have $M_\varepsilon \leq 2/\varepsilon$.

$$
\frac{1}{T} \sum_{m=1}^{M_\varepsilon} \sum_{t=1}^{T} \mathbb{I}\{p_t=q_m\} \left| q_m - y_t \right|
= \left\| \bar{f}_t \right\|_1 \leq \left\| \pi_t \right\|_1 + \left\| \bar{f}_t - \pi_t \right\|_1 \leq \left\| \pi_t \right\|_1 + \sqrt{M_\varepsilon} \left\| \bar{f}_t - \pi_t \right\|_2.
$$

□
Existence of a Calibrated Forecasting Strategy

Definition (Calibration [3, 13])

A strategy of Forecaster is calibrated, if for any strategy of Reality, \( \varepsilon > 0 \), and \( \rho \in [0, 1] \),

\[
\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}\{|p_t - \rho| \leq \varepsilon\} |p_t - y_t| = 0, \quad \text{a.s..}
\]

Theorem ([13])

There exists a calibrated forecasting strategy.

Proof.

Consider a multiple-round approach. For each round \( r \in \mathbb{N} \), set \( T_r = 2^r \) and \( \varepsilon_r = 2^{-0.5r} \). Check the almost sure convergence by the non-asymptotic result and Borel-Cantelli lemma.
References

   An analog of the minimax theorem for vector payoffs.

   *Prediction, Learning, and Games*.

   The well-calibrated Bayesian.

   The impossibility of inductive inference.

   *Stochastic Processes*.

   A proof of calibration via Blackwell’s approachability theorem.
References II

[7] Dean P. Foster and Sergiu Hart.
Smooth calibration, leaky forecasts, finite recall, and Nash dynamics.
2017.

Calibrated learning and correlated equilibrium.

Asymptotic calibration.

[10] Elad Hazan and Sham Kakade.
(weak) calibration is computationally hard.

Solving variational inequalities with monotone operators on domains given by linear minimization oracles.

[12] Sham M. Kakade and Dean P. Foster.
Deterministic calibration and Nash equilibrium.
References III

A geometric proof of calibration.

On the Banach-space-valued Azuma inequality and small-set isoperimetry of
Alon-Roichman graphs.

Barrier subgradient method.

[16] David Oakes.
Self-calibrating priors do not exist.

The reproducible properties of correct forecasts.

Zur theorie der gesellschaftsspiele.
Predictions as statements and decisions.
2006.

Non-asymptotic calibration and resolution.