Online Asymptotic Calibration

Juan Elenter
elenter@seas.upenn.edu

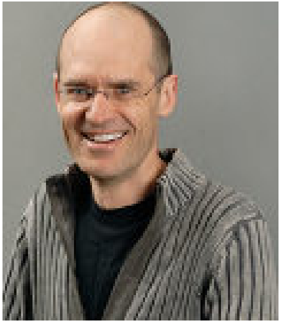
March 23, 2022

Outline

1. Context

Authors

Asymptotic Calibration, Biometrika, Vol. 85, Oxford University Press, 1991



Dean Foster

Rakesh Vohra

Online Calibration Basics

- $X^T = \overbrace{[1, \quad 0, \quad 1, \quad 1, \cdots, 0]}^{\text{length}\, T}$ $\qquad e.g:\ \begin{cases} 1 & \longrightarrow \textit{Rain} \\ 0 & \longrightarrow \text{No Rain} \end{cases}$
- Forecasting method $F : \mathcal{X}^T \to A = [0, 1]$
- Forecast at time T: $f_T = F(X^{T-1})$.
- Number of p forecasts up to time t: $n_t(p; F, X) = \sum_{t=1}^{T} 1\{f_t = p\}$
- Fraction of those forecasts where it actually rained: $\rho_t(p; X, F) = \frac{\sum_{t=1}^{T} 1\{f_t = p\} X_t}{n_t(p; X, F)}$

Traditional Calibration Definitions

- Intuitively, F is well-calibrated w.r.t X iff $\rho_t \approx p$ for all $p \in A$
- Definition of Online Asymptotic Calibration:

> F is well-calibrated w.r.t X iff $\lim_{t\to\infty} \underbrace{\sum_{p\in A} (\rho_t(p) - p)^2 \frac{n_t(p)}{t}}_{C_t(F,X)} = 0$

Context

- Asymptotic calibration is not a sufficient condition for a forecast to be good. Let $X = [0, 1, 0, 1, \cdots]$:

$$f_t = \frac{1}{2} \quad VS \quad f_t = X_t \quad \text{for all } t$$

- Oakes (1985): No **deterministic** forectasing scheme can be calibrated **for all** possible sequences.
- Schervish (1985): Why should we care about long term, worst-case criteria? "In the short run, when we are alive, a forecaster might do quite well".

The mission

- To rescue the notion of online asymptotic calibration.
- How? Broaden the definition of calibration by allowing randomized forecasts:
- Forecasting method $\cancel{F : \mathcal{X}^T \to [0, 1]}$,

$$F : \mathcal{X}^T \to L \quad \text{where L is a suitable distribution space.}$$

2. Problem Formulation

Game Theoretic Formulation

- Two Players: $\begin{cases} \text{Forecaster} & \longrightarrow & \text{Picks F} \\ \text{Nature} & \longrightarrow & \text{Picks X} \end{cases}$
- Rules of the game:
  - 1. Forecaster chooses F and reveals only the forecast distribution to Nature.
  - 2. At each $t \geq 1, f_t(X^{t-1})$ is drawn and simultaneously, Nature selects X_t.
  - 3. At the end of the game, statistician pays $C_t(F, X)$ to Nature.
- The forecaster wins if $C_t(F, X)$ is vanishingly small in some probabilistic sense.

Probabilistic Formulation

A randomized forecast F is ϵ-calibrated iff:

$$\lim_{t \to \infty} P\{C_t(F, X) < \epsilon\} > 1 - \epsilon$$

for all X.

Connection between formulations

In the context of a game: $\exists$ F ϵ-calibrated iff

$$\min_F \max_X \mathbb{E}\{C_t(F, X)\} \leq \epsilon^2$$

for t sufficiently large.

3. Foster & Vohra's approach

Preliminaries

- At round t, forecaster draws from the set $A = \{0, \frac{1}{k}, \frac{2}{k}, \cdots, 1\}$ with probability distribution μ_t:
$$\mathbb{P}(f_t = \frac{i}{k}) = \mu_t^i$$
- $EBS_T(F, X) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=0}^{k}\mu_t^i(X_t - \frac{i}{k})^2$
- $F^{i\to j}(X_{t-1}) = \begin{cases} \frac{j}{k} & \text{if } F(X_{t-1}) = \frac{i}{k} \\ F(X_{t-1}) & \text{else} \end{cases}$
- Notion of Regret:
$$R_T^{i\to j} = T[EBS_T(F, X) - EBS_T(F^{i\to j}, X)]_+$$

Main Theorem

If μ_t satisifies the following condition for each t:

$$\sum_{j\neq i} \mu_t^j R_{t-1}^{j\to i} = \mu_t^i \sum_{j\neq i} R_{t-1}^{i\to j} \quad \text{for all } i = 0, \cdots, k$$

the forecaster F with $k > \frac{1}{\epsilon^2}$ is ϵ-calibrated.

4. A calibrated forecaster

How to find an ϵ-Calibrated Forecast

- Let $A = \begin{pmatrix} -\sum_{j\neq i} R^{1\to j} & R^{2\to 1} & \cdots & R^{k\to 1} \\ R^{1\to 2} & -\sum_{j\neq 2} R^{2\to j} & \cdots & R^{k\to 2} \\ \vdots & \vdots & \vdots & \vdots \\ R^{1\to k} & R^{2\to k} & \cdots & -\sum_{j\neq k} R^{k\to j} \end{pmatrix} \longrightarrow \sum_i A_{ij} = 0 \quad \text{for all } j$
- $\exists$ ϵ-calibrated F $\iff$ $\exists\, x$ probability vector s.t: $Ax = 0$.

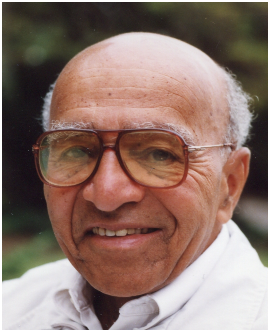
How to find an ϵ-Calibrated Forecast

- Let A_2 such that $A_{2ij} = \frac{A_{ij}}{\max_{i,j}|A_{ij}|}$
- $P = A_2 + \mathbb{I}_D$
- P positive and $\sum_i P_{ij} = 1$
- $\exists$ state x_s (stationary distribution) such that $Px_s = x_s$.
- $Px_s = x_s \quad \rightarrow \quad A_2 x_s + \mathbb{I}_D x_s = x_s \quad \rightarrow \quad A_2 x_s = 0 \quad \rightarrow \quad A x_s = 0$

μ_t can be found through Gaussian elimination of the Regret matrix A.

Sketch of Original Proof

- 1. Approximate $C_t(F,X)$ by $\tilde{C}_t(F,X) = \sum_{j=0}^{k} \frac{\sum_{t=1}^{T} \mu_t^j}{T} (\sum_{t=1}^{T} \frac{\mu_t^j X_t}{\sum_{t=1}^{T} \mu_t^j} - \frac{j}{k})^2$
- 2. Note that $C_t - \tilde{C}_t \xrightarrow{p} 0$
- 2. Bound $\tilde{C}_t(F,X) \leq k\rho + \frac{k}{2\rho t} + \frac{1}{4k^2}$
- 3. For all $\epsilon > 0$, if $k > \frac{1}{\epsilon^2}$, we have $\tilde{C}_t(F,X) \leq \epsilon$ for t big enough.
- 4. Combining 1. and 3. , for all $\epsilon > 0$, $\exists\, t_0$ such that $\mathbb{P}(C_t(F,X) < \epsilon) \geq 1 - \epsilon$ for all $t > t_0$.



David Blackwell: "Basically, I'm not interested in doing research and I never have been. I'm interested in understanding, which is quite a different thing. "

- "Over the past few years many proofs of the existence of calibration have been discovered. Does the literature really need one more? Probably not. " Foster.

Proof through Blackwell's Approachability

- Recall Strategy: $f_t \in \{0, \frac{1}{k}, \frac{2}{k}, \cdots, 1\}$ according probability distribution $\mu_t = \{\mu_t(0), \mu_t(1), \cdots, \mu_t(k)\}$.
- More general setting with $X_t \in [0, 1]$.
- Consider the vector payoff:

$$c(\mu_t, X_t) = [\, \mu_t(0)(X_t - \frac{0}{k}), \mu_t(1)(X_t - \frac{1}{k}), \cdots, \mu_t(k)(X_t - 1) \,]$$

Proof through Blackwell's Approachability

- In the context of game theory , a set S is approachable if there exists a forecasting scheme such that

$$\lim_{t\to\infty} d(\bar{c}_t(\mu, X), S) = 0 \quad a.s$$

where $d(c, S) = \inf_{s\in S} \|c - S\|$ and $\bar{c}_t$ is the average pay-off vector function until time t.

Blackwell's Approachability Theorem: Any closed convex set S is approachable if and only if it is response-satisfiable. (i.e: for all X_t, $\exists\, \mu_t$ such that $c(\mu_t, X_t) \in S$).

Proof through Blackwell's Approachability

- Let $\epsilon > 0$.
- We can select k such that the set $S = \{x : \sum |x| \leq \epsilon\}$ (ℓ_1-ball of radius ϵ) is approachable:
  - If $k > \frac{1}{2\epsilon}$ then $\exists\ i$ such that: $|\frac{i}{k} - X_t| < \epsilon \quad$ for all $X_t \in [0, 1]$.
  - By setting $\mu_t = \delta(i)$ we get: $c(\mu_t, X_t) = [0, \cdots, 0, \frac{i}{k} - y, 0, \cdots, 0]$.
  - $c(\mu_t, X_t) \ \in \ S$
- By Blackwell's Theorem: $d(\frac{1}{T}\sum_{t=1}^{T} c(\mu_t, X_t), S) \ \longrightarrow \ 0$ a.s.
- The almost sure convergence of average pay-off implies :

$$\lim_{t\to\infty} P\{C_t(\mu, X) < \epsilon\} > 1 - \epsilon$$

Key Takeaways

- "Our goal in this paper has been to rescue the notion of calibration..."
- Generalization of the original notion of calibration to allow for randomised forecasts.
- The definition is not vacuous (existence and algorithms).
- Still, Schervish's concerns remain unanswered.