THE FREQUENTIAL APPROACH TO PROBABILITY AND STATISTICS

A. P. DAWID

UNIVERSITY COLLEGE LONDON
PREQUENTIAL
PROBABILITY

Forecaster $F$ \{ successive moves
Nature $N$ \} with perfect
information

Each day: $F$ issues a probability distribution $P_{F,i}$
for r.v. $X_i$.
$N$ chooses a value $x_i$ for $X_i$.

DATA: $D = \begin{pmatrix} P_{F,1} & P_{F,2} & P_{F,3} & \cdots & P_{F,\infty} \\ x_1 & x_2 & x_3 & \cdots & x_\infty \end{pmatrix}$

PROBLEM: In the light of the outcomes $x_i$ is $P_{F,i}$ a good set of forecasts?
Additional structure

- A (non-randomised) strategy for $F$ is specified by a joint probability distribution $P_F$ for $x_i$:
  \[ P_{F,i}(x_i) = P_F(x_i | x_i^{-1} = x_i^{i-1}) \]

- A (randomised, oblivious) strategy for $N$ is specified by a joint probability distribution $P_N$ for $x_i$:
  \[ x_i \text{ simulated from } P_N(x_i | x_i^{-1} = x_i^{i-1}) \]

Initially, suppose both $F$ and $N$ operate such strategies

(—will drop later)
A. $P_N, P_F$ both public.

Match?

(i) $P_F = P_N$

-or (ii) $P_{F,i} = P_{N,i}$, all $i$.

-but rare for $P_N$ to be known.

B. $P_N$ unknown, $P_F$ public.

Develop MATCH CRITERION

$T(\bar{I}_{PF}, \bar{x})$

NULL DISTRIBUTION:

of $T(\bar{I}_{PF}, \bar{x})$

when $\bar{x} \sim P_N = P_{PF} = P_0$, say

-could depend on $P_0$. 
$T$ satisfies:

- **Weak Frequential Principle** if $T(P_F > \alpha)$ depends only on data $D = (P_F, \alpha)$.

- **Strong Frequential Principle** if relevant aspects of the null distribution of $T$ does not depend on $P_0$.

Is this possible?

**Yes**
EXAMPLE 1

\[ x_i = 0 \text{ or } 1 \]
\[ p_i = P_{F,0} (X_i = 1) \]

\[ T_n = \frac{1}{n} \sum_{i=1}^{n} (x_i - p_i) \quad (WPP) \]

Then

\[ P_o (T_n \to 0) = 1 \quad (SPP) \]

-so acceptable match if \( T_n \) is "close to 0"

How close?
EXAMPLE 2

\[ T_n = \frac{\sum_{i=1}^{n} (x_i - \phi_i)}{\left[ \sum_{i=1}^{n} \phi_i (1 - \phi_i) \right]^{1/2}} \]  

(WPP)

Then, so long as

\[ P_0 \left( \sum_{i=1}^{n} \phi_i (1 - \phi_i) \to \infty \right) = 1, \]

\[ P_0 (T_n \leq t) \to \Phi(t) \]  

(SPP)

so acceptable match if

\[ |T_n| \] not large in comparison to \( N(0, 1) \)

(significance level)
EXAMPLE 3

\[ x_i \in [a, b] \]

\[ P_{F_i} \text{ a distribution on } [a, b] \]

(continuous)

Define

\[ u_i = P_{F_i}(X_i \leq x_i) \quad \text{(WPP)} \]

Then, under \( P_0 \), the \((u_i)\) are independently uniform on \([0, 1]\) \quad \text{(SPP)}

so can assess observed \((u_i)\) for conformity with this:

[Uniform? Independent?]
Game-theoretic analysis

Third player, Statistician, plays between F and N.

Move = size h of bet against F.

E.g. binary case:
F pays $ \text{hi} \times (x_i - \pi_i)

Let $K_n = $ S's fortune at n.
$K_0 = 1.$

S WINS GAME if:

(i) $K_n$ always $\geq 0$

(ii) EITHER $K_n \to \infty$

("F discredited")

OR. A occurs

some property of data D

(WPP)
Call A **FULL** if $S$ has a winning strategy.

Then "A fails"

$\Rightarrow$ "F discredited"

Martingale theory $\Rightarrow$

If $A$ is full,

$P_0(A) = 1 \quad (SPP)$

**Example 1**: Can show

$A := \frac{1}{n} \sum (x_i - p_i) \rightarrow 0$

is full.

(+ many similar results)
If we change requirements from \( K_0 = 1 \) to 
\[ K_0 = \beta \in [0,1] \]
and from \( K_n \rightarrow \infty \) to 
\( (K_n) \) reaches 1 before 0 we can define
\[
PP(A) = \text{smallest } \beta \text{ s.t. } S \text{ has a winning strategy.}
\]
Then \( P_0(A) \leq PP(A) \) (SSP)

— applies to Examples 2, 3 and many more.
PREFERENTIAL STATISTICS

More players:

Pool \( E = \{ E_\theta : \theta \in \Theta \} \)

of “experts”

- plays before \( F \), issuing a family of probability forecasts

\( P_i = \{ P_{\theta, i} : \theta \in \Theta \} \)

for \( x_i \).

Now interested in comparing \( F \) and \( E \), rather than \( F \) and \( N \).
STRAATEGIES

Again, strategy $\leftrightarrow$ joint distn.

$N: \ P_N \ (\text{unknown})$

$E: \ P_E = \{P_\theta : \theta \in \Theta \}$

$F: \ P_F$

(A) $P_E$ public.

(E) "True model": $P_N = P_{0*} \in P_E$  

"compare" $P_F$ with (unknown) $P_{0*}$

- by evaluating probabilistic behaviour under $P_\theta \in P_E$, all $\theta$.

- "Classical Statistics"

E.g.: $\{-\ln(\hat{\theta}_n)\}^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,1)$

(WPP, SPP)  

$[P_{0*}]$
Prequential consistency:

$$\| P_{F,i} - P_{\theta,i} \| \rightarrow 0 \ [P_\theta]$$

Prequential efficiency:

For any joint distribution \( R \),

$$\prod_{i=1}^{n} \frac{P_{F,i}(x_i)}{P_{R,i}(x_i)} \rightarrow 0$$

\([P_\theta, \text{almost all } \theta]\)

Both usually hold for:

- **Plug-in Strategy:**
  
  $$P_{F,i} = P_{\hat{\theta}_{i-1},i}$$

- **Bayes Strategy:**
  
  $$P_F = \int P_{\theta_i} \pi(\theta_i) d\theta$$

In both cases, \( P_{F,i} \) only depends on moves to \((i-1)\), not on strategies \( P_F \) - PREQUENTIAL
For many purposes, an efficient $P_F$ can replace the whole family $P_e$.

E.g. Example 2:

$$T_n = \frac{\sum (x_i - P_F(i))}{\sum P_F(i) (1 - P_F(i))^2}$$

Then $T_n \sim N(0,1)$ \[ P_F \]

---so we can test fit of $P_e$ by testing fit of $P_F$. 

(II) "False model": do not assume $P_N \in P_E$.

Performance measure at time $i$: $D(P_{N,i}, P_{F,i})$

E.g. $(P_{N,i} - P_{F,i})^2$

For given expert strategies $P_E$, $N$'s moves $\pi$, let $\Theta^*$ (unknown) achieve

$$\min_{\Theta} \lim_{n \to \infty} \frac{1}{n} \sum D(P_{N,i}, P_0, i)$$

(closet expert to true $P_N$ for observed data).

Then can construct estimators $(\Theta^*_n)$ such that

$$\Theta^*_n \to \Theta^* \quad [P_N]$$

PREQUENTIAL

OUT-OF-MODEL

CONSISTENCY
(III) "No model"
- do not even assume $N$ is using a strategy
→ seek $P_F$ to optimise some variant of regret:
$$\inf_{P^*} \sup_{x} \left[ \sup_{\theta} \prod_{i=1}^{n} p_{\theta, i}(x_i) / \prod_{i=1}^{n} p_{F, i}(x_i) \right]$$
- "learning with expert advice"
$n$ fixed?? On-line ??
[Shtarkov]
- Bayes strategy generally good.

How related to (I) and (II)?
Pe not known
experts just announce their
\( P_{0,i} \) at each time point.
\( \Rightarrow P_F \) must be a PREQUENTIAL strategy.

Could look for extensions
of (II): \( \theta_n^+ \rightarrow \theta^+ \) ALWAS

(III) minimax regret

What results possible?

What conditions?

"Unlink" forecasts from labels
(switching)?