Bayesian Logistic Betting Strategy against Probability Forecasting

Masayuki KUMON, Jing LI, Akimichi TAKEMURA
and Kei TAKEUCHI

METR 2012–08
April 2012

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html
The METR technical reports are published as a means to ensure timely dissemination of scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the authors or by other copyright holders, notwithstanding that they have offered their works here electronically. It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author’s copyright. These works may not be reposted without the explicit permission of the copyright holder.
Bayesian logistic betting strategy against probability forecasting

Masayuki Kumon, Jing Li, Akimichi Takemura and Kei Takeuchi

April 2012

Abstract

We propose a betting strategy based on Bayesian logistic regression modeling for the probability forecasting game in the framework of game-theoretic probability by Shafer and Vovk [16]. We prove some results concerning the strong law of large numbers in the probability forecasting game with side information based on our strategy. We also apply our strategy for assessing the quality of probability forecasting by the Japan Meteorological Agency. We find that our strategy beats the agency by exploiting its tendency of avoiding clear-cut forecasts.

Keywords and phrases: exponential family, game-theoretic probability, Japan Meteorological Agency, probability of precipitation, strong law of large numbers.

1 Introduction

In this paper we consider assessing quality of probability forecasting for binary outcomes. A primary example of probability forecasting is the probability of precipitation announced by weather forecasting agencies. The binary outcomes are either “rain” (more precisely, precipitation above certain amount during a specified period at a particular location) or “no rain”. In the United States the National Weather Service started to announce probability of precipitation in 1965 (cf. [6]), whereas the Japan Meteorological Agency started probability forecasting in 1980 for Tokyo area and extended it to the whole Japan in 1986\(^1\). How can we assess the quality of probability forecasting? We propose to assess probability forecasting by setting up a hypothetical betting game against forecasting agencies in the framework of game-theoretic probability by Shafer and Vovk [16].

We can regard the capital process of a betting strategy as a test statistic of a statistical hypothesis ([15], [17]). Our null hypothesis is that given the probability \( p_n \) announced by the

---

\(^1\)http://www.jma.go.jp/jma/kishou/intro/gyomu/index2.html (in Japanese)
agency, the outcome is indistinguishable from the Bernoulli trial with success probability \( p_n \). If this hypothesis is true, then the capital process becomes a non-negative martingale and the capital process converges to a finite value almost surely. However if the announced probability \( p_n \) is not good, then a clever strategy may be able to beat the forecasting agency in the betting game. In our game we construct a betting strategy based on Bayesian logistic regression modeling, which is a very standard statistical model for analyzing binary responses. We will prove some results on the strong law of large numbers in probability forecasting game with side information based on our betting strategy. We also see that our strategy works well against probability of precipitation announced by the Japan Meteorological Agency.

Organization of this paper is as follows. In Section 2 we formulate the probability forecasting game with side information and derive some basic properties of betting strategies. It also serves as a brief introduction to game-theoretic probability theory. In Section 3 we introduce our betting strategy based on logistic regression model. In Section 4 we prove some properties of our logistic betting strategy in the framework of game-theoretic probability. In Section 5 we give numerical studies of our strategy. In particular we apply our strategy to the data on probability of precipitation announced by the Japan Meteorological Agency. We end the paper with some discussions in Section 6.

2 Formulation of the probability forecasting game and summary of preliminary results

In this section we formulate the probability forecasting game and extend it to include side information. We mostly follow the results in [10].

At the beginning of day \( n \) (or at the end of day \( n - 1 \)) an agency (we call it “Forecaster”) announces a probability \( p_n \) of certain event in day \( n \), such as precipitation in day \( n \). Let \( x_n = 0, 1 \) be the indicator variable for the event, i.e., \( x_n = 1 \) if the event occurs and \( x_n = 0 \) otherwise. We suppose that a player “Reality” decides the binary outcome \( x_n \). When Forecaster announces \( p_n \), it also sells a ticket with the price of \( p_n \) per ticket. The ticket pays one monetary unit when the event occurs in day \( n \), i.e., the value of the ticket at the end of day \( n \) is \( x_n \). A bettor or gambler, called “Skeptic”, buys \( M_n \) tickets with the price of \( p_n \) per ticket. Then the payoff to Skeptic in day \( n \) is \( M_n(x_n - p_n) \). We allow \( M_n \) to be negative, so that Skeptic can bet also on the non-occurrence of the event. If the probability announced by the agency is appropriate, it is hard for Skeptic to make money in this game. On the other hand, if the probability is biased in some way, Skeptic may be able to increase his capital denoted by \( K_n \). Hence we can measure the quality of probability forecasting in terms of \( K_n \).

We now give a protocol of the game, following the notational convention of [16].

**Binary Probability Forecasting (BPF)**

**Protocol:**
Skeptic announces his initial capital \( K_0 = 1 \).
FOR \( n = 1, 2, \ldots \):
Forecaster announces \( p_n \in (0, 1) \).
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $x_n \in \{0, 1\}$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n)$.

**Collateral Duty:** Skeptic must keep $\mathcal{K}_n$ non-negative.

Forecaster is supposed to decide its forecast $p_n$ based on all relevant side information available at the time of announcement. We modify the above protocol so that Forecaster also discloses the relevant side information $c_n$, which is a $d$-dimensional column vector, together with the probability $p_n$. Furthermore we define auxiliary capital processes $S_n$ and $\mathcal{V}_n$.

**Binary Probability Forecasting With Side Information (BPFSI)**

**Protocol:**

FOR $n = 1, 2, \ldots$:

- Forecaster announces $p_n \in (0, 1)$ and $c_n \in \mathbb{R}^d$.
- Skeptic announces $M_n$.
- Reality announces $x_n \in \{0, 1\}$.

$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - p_n)$.

$S_n := S_{n-1} + c_n(x_n - p_n)$.

$\mathcal{V}_n := \mathcal{V}_{n-1} + c_n c_n^T p_n(1 - p_n)$.

**Collateral Duty:** Skeptic must keep $\mathcal{K}_n$ non-negative.

In the protocol, $c_n^T$ denotes the transpose of $c_n$, $\mathcal{K}_n$ is a scalar, $S_n$ is a $d$-dimensional column vector and $\mathcal{V}_n$ is a $d \times d$ symmetric matrix.

If $d = 1$ and $c_n \equiv 1$, then $S_n = \sum_{i=1}^n (x_i - p_i)$. When we study the usual strong law of large numbers in game-theoretic probability, we are interested in the convergence $S_n/n \to 0$ as $n \to \infty$. Generalizing this case, in the presence of side information, we are interested in the convergence $\mathcal{V}_n^{-1} S_n \to 0$, although the order of $\mathcal{V}_n$ may be different from $O(n)$. We call this convergence the usual form of the strong law of large numbers in BPFSI. See Theorem 4.1 in Section 4.1. However, as we prove in Theorem 4.2 of Section 4.2, under mild regularity conditions, we can prove a stronger result

$$\lim_{n \to \infty} g(\mathcal{V}_n)^{-1} S_n = 0,$$

where $g(\mathcal{V})$ is close to $\mathcal{V}^{1/2}$ such as $g(\mathcal{V}) = \mathcal{V}^{1/2+\epsilon}$, $\epsilon > 0$.

Let

$$\nu_n = \frac{M_n}{\mathcal{K}_{n-1}}$$

denote the fraction of the capital Skeptic bets on day $n$. Then the capital process $\mathcal{K}_n$ is written as

$$\mathcal{K}_n = \prod_{i=1}^n (1 + \nu_i(x_i - p_i)). \tag{1}$$
Now suppose that Skeptic himself models Reality’s move as a Bernoulli variable with the success probability \( \hat{p}_n \in (0, 1) \). If Skeptic totally trusts Forecaster, then he sets \( \hat{p}_n = p_n \). However if Skeptic does not totally trust Forecaster he may formulate \( \hat{p}_n \) differently from \( p_n \). Furthermore suppose that Skeptic uses the “Kelly criterion” ([12], [9]) to determine \( \nu_n \) so as to maximize the expected value of the logarithm of the capital growth under \( \hat{p}_n \):

\[
\nu_n : E_{\hat{p}_n}[\log(1 + \nu(x_n - p_n))] \to \max.
\]

Writing

\[
E_{\hat{p}_n}[\log(1 + \nu(x_n - p_n))] = \hat{p}_n \log(1 + \nu(1 - p_n)) + (1 - \hat{p}_n) \log(1 - \nu p_n)
\]

and differentiating this with respect to \( \nu \), the unique maximizer \( \nu_n \) is obtained as

\[
\nu_n = \frac{\hat{p}_n - p_n}{p_n(1 - p_n)} = \frac{\hat{p}_n}{p_n} - \frac{1 - \hat{p}_n}{1 - p_n}.
\]  

(2)

With this choice of \( \nu_n \) we have

\[
1 + \nu_n(x_n - p_n) = \begin{cases} 
\hat{p}_n/p_n & \text{if } x_n = 1 \\
(1 - \hat{p}_n)/(1 - p_n) & \text{if } x_n = 0 
\end{cases}
\]

\[
= \frac{\hat{p}_n^x(1 - \hat{p}_n)^{1-x_n}}{p_n^x(1 - p_n)^{1-x_n}}.
\]

Hence (1) is written as

\[
K_n = \frac{\prod_{i=1}^{n} \hat{p}_i^x(1 - \hat{p}_i)^{1-x_i}}{\prod_{i=1}^{n} p_i^x(1 - p_i)^{1-x_i}}.
\]

In the case that Skeptic models the joint probability \( \hat{p}(x_1, \ldots, x_n) \) of Reality’s moves, \( \hat{p}_n \) is given as the conditional probability

\[
\hat{p}_n = \frac{\hat{p}(x_1, \ldots, x_{n-1}, 1)}{\hat{p}(x_1, \ldots, x_{n-1})}.
\]

In this case

\[
\hat{p}_n^x(1 - \hat{p}_n)^{1-x_n} = \frac{\hat{p}(x_1, \ldots, x_{n-1}, x_n)}{\hat{p}(x_1, \ldots, x_{n-1})}, \quad x_n = 0, 1,
\]

and \( K_n \) is written as

\[
K_n = \frac{\hat{p}(x_1, \ldots, x_n)}{\prod_{i=1}^{n} p_i^x(1 - p_i)^{1-x_i}}.
\]  

(3)

For the rest of this section we introduce some terminology of game-theoretic probability. An infinite sequence of Forecaster’s moves and Reality’s moves

\[
\xi = p_1 c_1 x_1 p_2 c_2 x_2 \ldots
\]
is called a path. The set $\Omega$ of all paths is called the sample space. A subset $E \subset \Omega$ is an event. A strategy $\mathcal{P}$ of Skeptic determines $\hat{p}_n$ based on a partial path $p_1 c_1 x_1 \ldots p_{n-1} c_{n-1} x_{n-1} p_n c_n$:

$$\mathcal{P} : p_1 c_1 x_1 \ldots p_{n-1} c_{n-1} x_{n-1} p_n c_n \mapsto \hat{p}_n, \quad n = 1, 2, \ldots .$$

$K_n^\mathcal{P} = K_n^\mathcal{P}(\xi)$ denotes the capital process when Skeptic adopts the strategy $\mathcal{P}$. We say that Skeptic can weakly force an event $E$ by a strategy $\mathcal{P}$ if $K_n^\mathcal{P}$ is never negative and

$$\limsup_n K_n^\mathcal{P}(\xi) = \infty \quad \forall \xi \notin E.$$ 

For two events $E_1, E_2 \subset \Omega$, $E_1^C \cup E_2$ is denoted as $E_1 \Rightarrow E_2$, where $E_1^C$ is the complement of $E_1$. We say that by a strategy $\mathcal{P}$ Skeptic can weakly force a conditional event $E_1 \Rightarrow E_2$ if $K_n^\mathcal{P}$ is never negative and

$$\limsup_n K_n^\mathcal{P}(\xi) = \infty \quad \forall \xi \in E_1 \cap E_2^C.$$ 

$E_1$ is interpreted as a set of regularity conditions for the event $E_2$ to hold.

Let $\lambda_{\max,n}$ and $\lambda_{\min,n}$ denote the maximum and the minimum eigenvalues of $V_n$. In this paper we consider the following regularity conditions:

i) $\lim_n \lambda_{\min,n} = \infty$.

ii) $\limsup_n \lambda_{\max,n}/\lambda_{\min,n} < \infty$.

iii) $\{c_1, c_2, \ldots \}$ is a bounded set.

Namely we take $E_1$ as

$$E_1 = \{\xi \mid \lim_n \lambda_{\min,n} = \infty, \limsup_n \lambda_{\max,n}/\lambda_{\min,n} < \infty \text{ and } c_1, c_2, \ldots \text{ are bounded}\}.$$ 

The condition i) makes the meaning of $"V_n \to \infty"$ precise. The condition ii) means that $V_n$ stays away from being singular. For $d = 1$ ii) is trivial and not needed.

### 3 Logistic betting strategy

In this section we introduce a betting strategy based on logistic modeling of Reality’s moves.

As in the previous section Skeptic models $x_n$ as a Bernoulli variable with the success probability $\hat{p}_n$. Furthermore we specify that Skeptic uses the following logistic regression model for the logarithm of the odds ratio:

$$\log \frac{\hat{p}_n}{1 - \hat{p}_n} = \log \frac{p_n}{1 - p_n} + \theta' c_n, \quad (5)$$

where $\theta \in \mathbb{R}^d$ is a parameter vector.

In previous studies in game-theoretic probability, many strategies of Skeptic depend only on $x_i - p_i$, $i \leq n - 1$, and do not depend on $p_n$. However obviously it is more reasonable to consider Skeptic’s strategies which depend on $p_n$. Strategies explicitly depending on $p_n$ are also
important from the viewpoint of defensive forecasting ([20], [18]). We again discuss this point in Section 4.3.

We now consider the capital process $K^\theta_n$ of (5) for a fixed $\theta \in \mathbb{R}^d$. Solving for $\hat{p}_n$ we have

$$\hat{p}_n = \frac{p_ne^{\theta c_n}}{1 + p_n(e^{\theta c_n} - 1)}, \quad 1 - \hat{p}_n = \frac{1 - p_n}{1 + p_n(e^{\theta c_n} - 1)}.\quad (6)$$

Then

$$\hat{p}_n^{x_i}(1 - \hat{p}_n)^{1-x_i} = p_n^{x_i}(1 - p_n)^{1-x_i} \frac{e^{\theta c_{i,n}}}{1 + p_n(e^{\theta c_n} - 1)}$$

and the capital process is written as

$$K^\theta_n = \prod_{i=1}^n \frac{\hat{p}_i^{x_i}(1 - \hat{p}_i)^{1-x_i}}{p_i^{x_i}(1 - p_i)^{1-x_i}} = \frac{e^{\theta \sum_{i=1}^n c_{i,x_i}}}{\prod_{i=1}^n(1 + p_i(e^{\theta c_i} - 1))}.\quad (7)$$

Naturally it is better for Skeptic to choose the value of $\theta$ depending on the moves of other players. In this paper we consider a Bayesian strategy, which specifies a prior distribution $\pi(\theta)$ for $\theta$. Bayesian strategies for Binary Probability Forecasting with constant $p_n \equiv p$ was considered in [10]. Bayesian strategy is basically the same as the universal portfolio by Cover and his coworkers ([3], [4], [5]). In the universal portfolio, a prior is put on the betting ratio $\nu$ itself, where as we put a prior on the parameter of Skeptic’s model. Furthermore differently from [4] we allow continuous side information.

In the Bayesian logistic strategy with the prior density function $\pi(\theta)$ of $\theta$, the capital process $K^\pi_n$ is written as

$$K^\pi_n = \int_{\mathbb{R}^d} K^\theta_n \pi(\theta) d\theta = \int_{\mathbb{R}^d} \frac{e^{\theta \sum_{i=1}^n c_{i,x_i}}}{\prod_{i=1}^n(1 + p_i(e^{\theta c_i} - 1))} \pi(\theta) d\theta.$$

$K_n^\pi$ is of the form (3) where

$$\hat{p}(x_1, \ldots, x_n) = \prod_{i=1}^n p_i^{x_i}(1 - p_i)^{1-x_i} \int_{\mathbb{R}^d} \frac{e^{\theta \sum_{i=1}^n c_{i,x_i}}}{\prod_{i=1}^n(1 + p_i(e^{\theta c_i} - 1))} \pi(\theta) d\theta.$$

In this paper we consider a prior density which is positive in a neighborhood of the origin. We call such $\pi$ “a prior supporting a neighborhood of the origin”.

### 4 Properties of logistic betting strategy from the viewpoint of game-theoretic probability

In this section we prove game-theoretic properties of our Bayesian logistic strategy.
4.1 Weak forcing of the usual form of the strong law of large numbers

The first theoretical result on our logistic betting strategy is the following theorem.

**Theorem 4.1.** In BPFSI, by a Bayesian logistic strategy with a prior supporting a neighborhood of the origin, Skeptic can weakly force

\[ E_1 \Rightarrow \lim_{n} V_n^{-1} S_n = 0, \]

where \( E_1 \) is given in (4).

The rest of this subsection is devoted to a proof of this theorem. The basic logic of our proof is the same as in Section 3.2 of [16].

We first consider the logarithm of \( K_n^r \) in (7) for a fixed \( r \):

\[ \log K_n^r = \theta' \sum_{i=1}^{n} c_i x_i - \sum_{i=1}^{n} \log(1 + p_i(e^{c_i} - 1)). \]

For notational simplicity we write

\[ u(\theta) = \log K_n^r. \]

We investigate the behavior of \( u(\theta) \) for \( \theta \) close the origin. Fix \( \theta \in \mathbb{R}^d \) with unit length (i.e. \( ||\theta|| = 1 \)) and consider \( u(s\theta), 0 \leq s \leq \epsilon \). Note that \( u(0) = 0 \). We will choose \( \epsilon \) appropriately later in (11).

The derivative of \( u(s\theta) \) with respect to \( s \) is written as follows.

\[
\frac{\partial}{\partial s} u(s\theta) = \theta' \sum_{i=1}^{n} c_i x_i - \sum_{i=1}^{n} \frac{\theta' c_i p_i e^{s\theta c_i}}{1 + p_i(e^{s\theta c_i} - 1)} \\
= \theta' \sum_{i=1}^{n} c_i (x_i - p_i) - \sum_{i=1}^{n} \frac{\theta' c_i p_i e^{s\theta c_i} - \theta' c_i p_i (1 + p_i(e^{s\theta c_i} - 1))}{1 + p_i(e^{s\theta c_i} - 1)} \\
= \theta' S_n - \sum_{i=1}^{n} \theta' c_i p_i (1 - p_i) e^{s\theta c_i} - 1 \\
1 + p_i(e^{s\theta c_i} - 1) \). \tag{8}
\]

Note that \( \theta' c_i \) and \( e^{s\theta c_i} - 1 \) have the same sign and hence each summand in the second term on the right-hand side of (8) is non-negative.

Let

\[ \gamma_p(y) = \frac{e^y - 1}{1 + p(e^y - 1)} \]

be a function of \( y \in \mathbb{R} \) depending on the parameter \( p \in [0, 1] \). Note \( \gamma_p(0) = 0 \). Its derivative is computed as

\[ \gamma'_p(y) = \frac{e^y}{(1 + p(e^y - 1))^2} > 0. \tag{9} \]

Hence

\[ \gamma_p(y) = \int_0^y \gamma'_p(z)dz = \int_0^y \frac{e^z}{(1 + p(e^z - 1))^2}dz, \]
where for negative $y < 0$ we interpret $\int_y^0 (\cdots) dz$ as $-\int_y^0 (\cdots) dz$. Now $\gamma'_p(z)$ in (9) is monotone in $p$ with $\gamma'_0(z) = e^z$ and $\gamma'_1(z) = e^{-z}$. Hence

$$e^{-|z|} = \min(e^{-z}, e^z) \leq \gamma'_p(z) \leq \max(e^{-z}, e^z) = e^{|z|}.$$

Then for $z$ between 0 and $y$ we have

$$e^{-|z|} \leq \gamma'_p(z) \leq e^{|z|}.$$  \hspace{1cm} (10)

Using the upper bound $e^{|z|}$ and integrating $\gamma'_p(z)$ we obtain

$$|\gamma_p(y)| = \frac{|e^y - 1|}{1 + p(e^y - 1)} \leq |y|e^{|y|} \quad \text{and} \quad 0 \leq y\gamma_p(y) = y \frac{e^y - 1}{1 + p(e^y - 1)} \leq y^2 e^{|y|}.$$

Let $L_{c,n} = \max_{1 \leq i \leq n} ||c_i||$. Then

$$\frac{\partial}{\partial s} u(s\theta) \geq \theta^r S_n - s \sum_{i=1}^n (\theta^r c_i)^2 p_i (1 - p_i) e^{L_{c,n}} = \theta^r S_n - s \theta^r V_n \theta e^{L_{c,n}}$$

and integrating this for $0 \leq s \leq \epsilon$ we have (for any $\theta$ and $\epsilon > 0$)

$$u(\epsilon \theta) \geq \theta^r S_n - \frac{\epsilon^2}{2} \theta^r V_n \theta e^{L_{c,n}}.$$

For the rest of our proof we arbitrary choose and fix a path $\xi \in E_1$, where $E_1$ is given in (4). Various constants ($\epsilon$'s, $L$’s etc.) below may depend on $\xi$. By iii) there exists $L_c$ such that $L_{c,n} < L_c$ for all $n$. Also there exist $n_0$ and $L_1$ such that $\lambda_{\max,n}/\lambda_{\min,n} < L_1$ for all $n \geq n_0$. Now suppose that $V_n^{-1} S_n \to 0$ for this $\xi$. Then for some $\epsilon_1 > 0$ and for infinitely many $n$ we have $||V_n^{-1} S_n|| \geq \epsilon_1$. Let $N_1 = \{n_1, n_2, \ldots\}$ be a subsequence such that $||V_n^{-1} S_n|| \geq \epsilon_1$ for $n \in N_1$. The normalized vectors

$$\eta_n = \frac{V_n^{-1} S_n}{||V_n^{-1} S_n||}, \quad n \in N_1,$$

have an accumulation point $\eta$, $||\eta|| = 1$, and hence along a further subsequence $N_2 \subset N_1$ we have

$$\lim_{n \to \infty, \ n \in N_2} \eta_n = \eta.$$

By Cauchy-Schwarz, for three vectors $a, b, c \in \mathbb{R}^d$, we have

$$\frac{|b^r V_n c|}{a^r V_n a} \leq \frac{\lambda_{\max,n} ||b|| ||c||}{\lambda_{\min,n} ||a||^2} < L_1 \frac{||b|| ||c||}{||a||^2}, \quad \forall n \geq n_0.$$

Then we can choose $0 < \epsilon_2 < 1/4$ such that for all sufficiently large $n \in N_2$ and for all $\tilde{\eta}$, $||\tilde{\eta}|| = 1$, sufficiently close to $\eta$, we have

$$u(\epsilon \tilde{\eta}) \geq \epsilon \tilde{\eta}^r S_n - \frac{\epsilon^2}{2} \tilde{\eta}^r V_n \tilde{\eta} e^{L_{c,n}}$$

$$= \epsilon ||V_n^{-1} S_n|| \tilde{\eta}^r V_n \tilde{\eta} - \frac{\epsilon^2}{2} \tilde{\eta}^r V_n \tilde{\eta} e^{L_{c,n}}$$

$$\geq \epsilon \epsilon_1 \tilde{\eta}^r V_n \eta (1 - \epsilon_2) - \frac{\epsilon^2}{2} \eta^r V_n \eta (1 + \epsilon_2) e^{L_{c,n}}$$

$$= \epsilon \eta^r V_n \eta (\epsilon_1 (1 - \epsilon_2) - \frac{\epsilon}{2} (1 + \epsilon_2) e^{L_{c,n}}).$$
We now choose small enough \( \epsilon > 0 \) such that
\[
\epsilon_1 (1 - \epsilon_2) - \frac{\epsilon}{2} (1 + \epsilon_2) e^{e \epsilon} > \frac{\epsilon_1}{2}.
\] (11)
Then
\[ u(\epsilon \bar{\eta}) \geq \frac{\epsilon \epsilon_1}{2} \lambda_{\min,n} \to \infty \quad (n \to \infty, n \in \mathbb{N}_2). \]
Note that the convergence is uniform for \( \bar{\eta} \) in some neighborhood \( N(\eta) \) of \( \eta \). Since our prior \( \pi \) puts a positive weight to \( N(\epsilon \eta) \), \( \mathcal{K}_n^\theta \to \infty \) along \( n \in \mathbb{N}_2 \). This completes our proof of Theorem 4.1.

4.2 Weak forcing of a more precise form of the strong law of large numbers

As discussed in Section 2, we can establish a much more precise rate of convergence of the strong law of large numbers based on our Bayesian logistic strategy. Our main theorem of this paper is stated as follows.

**Theorem 4.2.** In BPFSI, by a Bayesian logistic strategy with a prior distribution supporting a neighborhood of the origin, Skeptic can weakly force
\[
E_1 \Rightarrow \limsup_n \frac{S_n' V_n^{-1} S_n}{\log \det V_n} \leq 1,
\]
where \( E_1 \) is given in (4).

We give a proof of this theorem in the following three subsections.

4.2.1 Bounding the maximum likelihood estimate

We now consider the behavior of \( \mathcal{K}_n^\theta \) in (7), when \( \mathcal{K}_n^\theta \) is maximized with respect to \( \theta \). Let
\[ \hat{\theta}_n^* = \arg\max \mathcal{K}_n^\theta. \]
We call \( \hat{\theta}_n^* \) the maximum likelihood estimate, since \( \mathcal{K}_n^\theta \) is of the form of the likelihood function of the logistic regression model. It is easily seen that the maximizer \( \hat{\theta}_n^* \) is finite except for a special case that the vectors in \( \{c_i \mid x_i = 1\} \cup \{-c_i \mid x_i = 0\} \) lie on a half-space defined by a hyperplane containing the origin. More specifically in Lemma 4.3 we prove that \( \|\hat{\theta}_n^*\| \) is small when \( \|V_n^{-1} S_n\| \) is small.

The maximizing \( \hat{\theta}_n^* \) can only be computed at the end of day \( n \) after seeing all the data \( p_1, c_1, x_1, \ldots, p_n, c_n, x_n \). Hence we call a strategy using \( \hat{\theta}_n^* \) a “hindsight strategy”, which is the same as the best constant rebalanced portfolio (BCRP) in the terminology of the universal portfolio.

We prove the following lemma.
Lemma 4.3. Let $L_{c,n} = \max_{1 \leq i \leq n} \|c_i\|$ and $L_{\lambda,n} = \lambda_{\text{max},n} / \lambda_{\text{min},n}$ where we assume $\lambda_{\text{min},n} > 0$. Then

$$\|V_n^{-1}S_n\| \leq \frac{1}{3L_{c,n}L_{\lambda,n}} \Rightarrow \|\hat{\theta}_n\| \leq 3L_{\lambda,n}\|V_n^{-1}S_n\|.$$

For any fixed $\xi \in E_1$, there exist $L_c, L_{\lambda}$, such that $L_{c,n} < L_c$ and $L_{\lambda,n} < L_{\lambda}$ for all sufficiently large $n$. Also in Theorem 4.1 we proved that Skeptic can weakly force $E_1 \Rightarrow \lim_n V_n^{-1}S_n = 0$. From these results we have the following proposition.

**Proposition 4.4.** In the same setting as in Theorem 4.1 Skeptic can weakly force $E_1 \Rightarrow \lim_n \hat{\theta}_n = 0$.

The rest of this subsection is devoted to a proof of Lemma 4.3. Consider the inner product $\theta'\nabla u(\theta) = \theta'\ \text{grad} \ u(\theta)$ of $\theta$ and the gradient of $u(\theta)$. If $\theta'\nabla u(\theta) \leq 0$, then the gradient points toward the interior of the ball with radius $r = \|\theta\|$ as shown in Figure 1. If $\theta'\nabla u(\theta) \leq 0$ for all $\theta$ with $\|\theta\| = r$, then $\|\hat{\theta}_n\| \leq r$. This can be seen as follows. Suppose $\|\hat{\theta}_n\| > r$. Let $\bar{\theta}$ be the maximizer of $u(\theta)$ on the sphere (the boundary of the ball). Then at $\bar{\theta}$ the gradient of $\nabla u(\bar{\theta})$ is a positive multiple of $\theta$ and this contradicts $\theta'\nabla u(\bar{\theta}) \leq 0$.

As in the previous subsection, using this time the lower bound in (10), we have

$$\theta'\nabla u(\theta) \leq \theta'\ S_n - \theta'\ V_n\theta e^{-L_{c,n}\|\theta\|}.$$ 

Now

$$|\theta'\ S_n| = |\theta'\ V_n V_n^{-1}S_n| \leq \|\theta'\ V_n\| : \|V_n^{-1}S_n\|$$

and

$$\|\theta'\ V_n\|^2 = \theta'\ V_n^2\theta \leq \|\theta\|^2\lambda_{\text{max},n}^2.$$ 

Hence

$$|\theta'\ S_n| \leq \|\theta\|\lambda_{\text{max},n}\|V_n^{-1}S_n\|.$$ 

Furthermore

$$\theta'\ V_n\theta e^{-L_{c,n}\|\theta\|} \geq \lambda_{\text{min},n}\|\theta\|^2 e^{-L_{c,n}\|\theta\|}.$$ 

Figure 1: Gradient of $u(\theta)$
Therefore
\[ \theta' \nabla u(\theta) \leq \lambda_{\min,n} \| \theta \| (L_{L,n} \| V_n^{-1} S_n \| - \| \theta \| e^{-L_{L,n} \| \theta \|}). \]

For \( \| \theta \| = 3L_{L,n} \| V_n^{-1} S_n \| \)
\[ L_{L,n} \| V_n^{-1} S_n \| - \| \theta \| e^{-L_{L,n} \| \theta \|} = L_{L,n} \| V_n^{-1} S_n \| (1 - 3e^{-3L_{L,n} \| V_n^{-1} S_n \|}) \]
Then for \( \| V_n^{-1} S_n \| \leq 1/(3L_{c,n} L_{L,n}) \)
\[ 3e^{-3L_{L,n} \| V_n^{-1} S_n \|} \geq 3e^{-1} > 1. \]

Hence, if \( \| V_n^{-1} S_n \| \leq 1/(3L_{c,n} L_{L,n}) \), we have \( \theta' \nabla u(\theta) < 0 \) for all \( \theta \) with \( \| \theta \| = 3L_{L,n} \| V_n^{-1} S_n \| \). By the remark just after Proposition 4.4, this completes the proof of Lemma 4.3.

### 4.2.2 Behavior of the hindsight strategy

We summarize properties of \( \log \mathcal{K}_n^{\hat{\theta}} \) in view of the standard theory of exponential families ([2]) in statistical inference. Define
\[ \psi_i(\theta) = \log(1 + p_i(e^{\theta c_i} - 1)), \quad \psi(\theta) = \sum_{i=1}^{n} \psi_i(\theta). \]

Note that \( \psi_i(\theta) \) is the cumulant generating function (potential function) for the logistic regression model, which is an exponential family model with the natural parameter \( \theta \). Hence each \( \psi_i(\theta) \) and \( \psi(\theta) \) are convex in \( \theta \). Indeed by (9), the Hessian matrix \( H_{\psi_i}(\theta) \) of \( \psi_i \) is given as
\[ H_{\psi_i}(\theta) = c_i c_i' \frac{p_i(1 - p_i)e^{\theta c_i}}{(1 + p_i(e^{\theta c_i} - 1))^2}, \]
which is non-negative definite. The Hessian matrix
\[ H_{\psi}(\theta) = \sum_{i=1}^{n} H_{\psi_i}(\theta) \]
of \( \psi \) is positive definite if \( V_n \) is positive definite, which is the Fisher information matrix in terms of the natural parameter \( \theta \).

Convexity of \( \psi_i \) implies concavity of \( \log \mathcal{K}_n^{\hat{\theta}} = \theta' T_n - \psi(\theta) \), where
\[ T_n = \sum_{i=1}^{n} c_i x_i = S_n + \sum_{i=1}^{n} c_i p_i. \]
Hence if the maximum of \( \log \mathcal{K}_n^{\hat{\theta}} \) is attained at a finite value \( \hat{\theta}_n^* \), then the “maximum likelihood estimate” \( \hat{\theta}_n \) satisfies “the likelihood equation”
\[ \frac{\partial}{\partial \theta} \log \mathcal{K}_n^{\hat{\theta}} = 0. \]
or equivalently
\[ T_n = \nabla \psi(\hat{\theta}_n^*) . \] (12)

The likelihood equation can also be written as
\[ 0 = \sum_{i=1}^{n} (x_i - \hat{\mu}_i^*) c_i, \quad \hat{\mu}_i^* = \hat{\mu}_{c_i} = \frac{p_i(\hat{\theta}_i^*)}{1 + p_i(\hat{\theta}_i^*)}. \]

From this it follows that \( \hat{\theta}_n^* = 0 \) if and only if \( T_n = \sum_{i=1}^{n} c_i p_i \).

Regard (12) as determining \( \hat{\theta}_n^* \) in terms of \( t = T_n \), i.e., \( \hat{\theta}_n^* = \hat{\theta}^*(t) = T_n \). This is the inverse map of \( t = \nabla \psi(\theta) \). Differentiating \( t = \nabla \psi(\theta) \) again with respect to \( \theta \) we obtain the Jacobi matrix
\[ J = \frac{\partial t}{\partial \theta} = H(\theta) \]
as the Hessian matrix of \( \psi \). Hence the Jacobi matrix \( \frac{\partial \hat{\theta}_n^*}{\partial T_n} \) is written as
\[ \frac{\partial \hat{\theta}_n^*}{\partial T_n} = H(\hat{\theta}_n^*(T_n))^{-1}. \] (13)

Now \( \log K_n^\theta = \log K_n^\theta(\theta_n^*) \) is the Legendre transformation (cf. Chapter 3 of [1]) of \( \log K_n^\theta \):
\[ \log K_n^\theta(t) = \hat{\theta}_n^*(t)^{t} - \psi(\hat{\theta}_n^*(t)), \quad t = T_n. \]

Differentiating \( \log K_n^\theta(t) \) with respect to \( t \), by (12) we obtain
\[ \frac{\partial}{\partial t} \log K_n^\theta(t) = \hat{\theta}_n^*(t) + (t - \nabla \psi(\hat{\theta}_n^*(t))) \frac{\partial \hat{\theta}_n^*}{\partial t} = \hat{\theta}_n^*(t). \] (14)

By (13) the Hessian matrix of \( \log K_n^\theta(t) \) is given by \( H(\hat{\theta}_n^*(t))^{-1} \).

We are now ready to prove the following proposition.

**Proposition 4.5.** With the same setting as in Lemma 4.3,
\[ \| V_n^{-1} S_n \| \leq \frac{1}{3L_{c,n}L_{\lambda, n}} \Rightarrow e^{-C_n \| V_n^{-1} S_n \|} \leq \frac{\log K_n^\theta}{S^r V_n S_n/2} \leq e^{C_n \| V_n^{-1} S_n \|}, \]
where \( C_n = 3L_{c,n}L_{\lambda, n} \).

**Proof.** For given \( T_n, \bar{T}_0 = \sum_{i=1}^{n} c_i p_i \) and for \( s \in [0, 1] \), consider
\[ g(s) = \hat{\theta}_n^*(\bar{T}_0 + s S_n)(\bar{T}_0 + s S_n) - \psi(\hat{\theta}_n^*(\bar{T}_0 + s S_n)). \]

Then \( \log K_n^\theta(\bar{T}_0) = g(1) \). It is easily seen that \( g(0) = 0 \). By (14)
\[ g'(s) = \hat{\theta}_n^*(\bar{T}_0 + s S_n)' S_n. \]
Again it is easily seen that \( g'(0) = 0 \), since \( \hat{\theta}_n(\bar{T}_0) = 0 \). Then

\[
g(1) = \int_0^1 \int_0^s g''(u) duds.
\]

Now

\[
g''(u) = S_n' H_\phi(\hat{\theta}_n(\bar{T}_0 + uS_n))^{-1} S_n.
\]

By (10)

\[
e^{-\| \hat{\theta}_n(\bar{T}_0 + uS_n) \|_{L_\infty} S_n' S_n^{-1} \text{det} V_n S_n}
\leq e^{-\| \hat{\theta}_n(\bar{T}_0 + uS_n) \|_{L_\infty} S_n' S_n^{-1} \text{det} V_n S_n}
\]

Also \( \int_0^1 \int_0^s duds = 1/2 \). Furthermore by Lemma 4.3, if \( \| V_n^{-1} S_n \| \leq 1/(3L_{c,n} L_{\lambda,n}) \) then \( \| \hat{\theta}_n(\bar{T}_0 + uS_n) \| \leq 3L_{\lambda,n} \| V_n^{-1} S_n \| \) for all \( 0 \leq u \leq 1 \). Combining these results we have the proposition. \( \square \)

As in Proposition 4.5 we have the following corollary.

**Corollary 4.6.** In the same setting as in Theorem 4.1 Skeptic can weakly force

\[
E_1 \Rightarrow \lim_n \frac{\log \mathcal{K}_n^{\hat{\theta}_n}}{S_n' S_n^{-1} \text{det} V_n} = 1.
\]

### 4.2.3 Laplace method for evaluating the difference of the hindsight strategy and the logistic strategy

In the last subsection we clarified the behavior of the capital process for the hindsight strategy. Now we employ the standard Laplace method to evaluate the difference of the hindsight strategy and the logistic strategy (Section 5 of [3], Chapter 3.1 of [8]).

**Lemma 4.7.** Let \( \pi \) be a prior density supporting a neighborhood of the origin and let \( \mathcal{K}_n^\theta \) denote its capital process. For \( \xi \in E_1 \) such that \( \lim_n \mathcal{V}_n^{-1} S_n = 0 \),

\[
\lim_n \frac{\log \mathcal{K}_n^{\hat{\theta}_n^{\xi}} - \log \mathcal{K}_n^\theta}{(1/2) \log \text{det} V_n} = 0.
\]

**Proof.** For \( \theta \) close to the origin, expanding \( \log \mathcal{K}_n^\theta \) around \( \hat{\theta}_n \) we have

\[
\log \mathcal{K}_n^\theta = \log \mathcal{K}_n^{\hat{\theta}_n} - \frac{1}{2}(\theta - \hat{\theta}_n)' H_\phi(\hat{\theta}_n)(\theta - \hat{\theta}_n),
\]

where \( \hat{\theta}_n \) is on the line segment joining \( \theta \) and \( \hat{\theta}_n \). Hence

\[
\mathcal{K}_n^\theta = \mathcal{K}_n^{\hat{\theta}_n} \times \exp\left(-\frac{1}{2}(\theta - \hat{\theta}_n)' H_\phi(\hat{\theta}_n)(\theta - \hat{\theta}_n)\right).
\]

Now by the standard Laplace method we obtain (15). \( \square \)
Finally we give a proof of Theorem 4.2.

**Proof of Theorem 4.2.** By Corollary 4.6 and Lemma 4.7

\[
\log K_n^\alpha = \frac{1}{2} \log \det V_n \left( \frac{S_n' V_n^{-1} S_n}{\log \det V_n} - 1 + o(1) \right).
\]

Hence if \( \lim \sup_n S_n' V_n^{-1} S_n / \log \det V_n > 1 \), then \( \lim \sup_n \log K_n^\alpha = \infty \). \( \square \)

### 4.3 Monotonicity with respect to the forecast probability

Here we consider the case that \( \log(p_n/(1 - p_n)) \) itself is an element of the vector of the side information \( c_n \) and hence is multiplied by a coefficient in (5). For notational convenience we here eliminate \( \log(p_n/(1 - p_n)) \) from \( c_n \) and write (5) as

\[
\log \hat{p}_n = \beta \log \frac{p_n}{1 - p_n} + \tau_n,
\]

where \( \tau_n \) denotes the effect of side information other than \( \log(p_n/(1 - p_n)) \). Intuitively \( \beta \) represents how much trust Skeptic puts in Forecaster. If \( \beta = 0 \) then Skeptic entirely ignores Forecaster’s \( p_n \) and if \( \beta = 1 \) then Skeptic takes \( p_n \) for granted. The value of \( \beta \in (0, 1) \) corresponds to partial trust in \( p_n \). It is somewhat surprising to see that \( \beta > 1 \) in the case of probability of precipitation announced by the Japan Meteorological Agency in Section 5.2.

We now investigate how \( \nu_n \) in (2) behaves with respect to \( p_n \) for given \( p_1, c_1, x_1, \ldots, p_{n-1}, c_{n-1}, x_{n-1} \). This is an important question from the viewpoint of defensive forecasting ([20], [18]), because in defensive forecasting we want to obtain \( p_n \) for which \( \nu_n = 0 \). For notational simplicity we now omit the subscript \( n \) and write (6) as

\[
\hat{p} = \frac{p \left( \frac{p}{1 - p} \right)^{\beta - 1} e^\tau}{1 + p \left( \frac{p}{1 - p} \right)^{\beta - 1} e^\tau}.
\]

Then

\[
\nu(p) = \frac{\hat{p} - p}{p(1 - p)} = \frac{p^{\beta - 1} e^\tau - (1 - p)^{\beta - 1}}{p^\beta e^\tau + (1 - p)^\beta}.
\]

Differentiating this with respect to \( p \) we obtain

\[
\frac{d\nu(p)}{dp} = \frac{-e^{2\tau} p^{2(\beta - 1)} + e^\tau p^{\beta - 2}(1 - p)^{\beta - 2}(\beta - 2 + 2p(1 - p)) - (1 - p)^{2\beta - 2}}{(p^\beta e^\tau + (1 - p)^\beta)^2}.
\]

The numerator of \( d\nu(p)/dp \) can be written as

\[-(e^\tau p^{\beta - 1} - (1 - p)^{\beta - 1})^2 + e^\tau (\beta - 1)p^{\beta - 2}(1 - p)^{\beta - 2},\]

which is non-positive for \( \beta \leq 1 \). Hence we have the following proposition.
Proposition 4.8. Under the logistic regression model (16), for \( \beta \leq 1 \) the betting ratio \( v_n(p_n) \) is monotone decreasing in \( p_n \).

It is natural that \( v_n \) is monotone decreasing in \( p_n \), because if \( p_n \) is too high and Skeptic does not believe it, then Skeptic will bet on the non-occurrence \( x_n = 0 \).

For the special case of \( \beta = 1 \),

\[
v_n(p_n) = \frac{e^{r_n} - 1}{1 + p_n(e^{r_n} - 1)},
\]

which is bounded and monotone in \( p_n \in [0, 1] \). For \( \beta < 1 \), \( v_n(p_n) \) is unbounded and it can be easily seen that

\[
\lim_{p_n \uparrow 0} \frac{v_n(p_n)}{1/p_n} = 1, \quad \lim_{p_n \uparrow 1} \frac{v_n(p_n)}{1/(1 - p_n)} = -1.
\]

We can interpret the first limit as follows. Suppose that \( p_n = 1/1000 \), i.e. the price of a ticket is \( 1/1000 \) of a dollar. In this case Skeptic can buy 1000 tickets with one dollar and has the chance of winning 1000 dollars. Hence Skeptic may want to buy 1000 tickets. Thus it is reasonable that \( v \) and \( p_n \) are inversely proportional when \( p_n \) is small.

5 Experiments

In this section we give some numerical studies of our strategy. In Section 5.1 we present some simulation results and in Section 5.2 we apply our strategy against probability forecasting by the Japan Meteorological Agency.

5.1 Some simulation studies

We consider three cases and apply three strategies to these examples. In our simulation studies Reality chooses her moves probabilistically, either by Bernoulli trials or by a Markov chain model.

- Case 1: \( x_n \) is a Bernoulli variable with the success probability 0.7 and \( p_n \) alternates between 0.4 and 0.6 (i.e. 0.4 = \( p_1 \) = \( p_3 \) = \( \cdots \) and 0.6 = \( p_2 \) = \( p_4 \) = \( \cdots \)).

- Case 2: \( x_n \) is a Bernoulli variable with the success probability 0.5 and \( p_n \) alternates between 0.4 and 0.6.

- Case 3: \( p_n = 0.5 \) and \( x_n \) is generated by a Markov chain model with transition probabilities shown in Figure 2.

- Strategy 1: \( \theta \) is a scalar and \( c_n = 1 \) in (5). Assume that the prior density for \( \theta \) is given as uniform distribution for \([0, 1]\). The capital process is written as

\[
\mathcal{K}_n^\pi = \int_0^1 \frac{e^{\theta \sum_{i=1}^n x_i}}{\prod_{i=1}^n (1 + p_i(e^{\theta} - 1))} \, d\theta.
\]
Figure 2: Transition probabilities for $x_n$

- Strategy 2: $\theta' = [\theta_1, \beta - 1]$ and $c'_n = [1, \log \frac{p_n}{1 - p_n}]$. Assume independent priors for $\theta_1$ and $\beta$, which are uniform distributions over $[0,1]$. The capital process is written as

$$K_n^\pi = \int_0^1 \int_0^1 \frac{e^{\theta_1 \sum_{i=1}^n x_i + (\beta - 1) \sum_{i=1}^n x_i \log \frac{p_i}{1 - p_i}}}{\prod_{i=1}^n (1 + p_i(e^{\theta_1 + (\beta - 1) \log \frac{p_i}{1 - p_i}} - 1)))} d\theta_1 d\beta.$$

- Strategy 3: $\theta' = [\theta_1, \beta - 1, \theta_3]$ and $c'_n = [1, \log \frac{p_n}{1 - p_n}, x_{n-1}]$. Assume independent priors for $\theta_1$, $\beta$ and $\theta_3$, which are uniform distributions over $[0,1]$. The capital process is written as

$$K_n^\pi = \int_0^1 \int_0^1 \int_0^1 \frac{e^{\theta_1 \sum_{i=1}^n x_i + (\beta - 1) \sum_{i=1}^n x_i \log \frac{p_i}{1 - p_i} + \theta_3 \sum_{i=1}^n x_{i-1}}}{\prod_{i=1}^n (1 + p_i(e^{\theta_1 + (\beta - 1) \log \frac{p_i}{1 - p_i} + \theta_3 x_{i-1}} - 1)))} d\theta_1 d\beta d\theta_3.$$

Figure 3: Case 1 with strategy 1

Figure 4: Case 2 with strategy 1

As shown in Figure 3 and Figure 4, we can beat Reality by strategy 1 only in case 1. So we improve our strategy and apply strategy 2 to case 2. We can see from Figure 5 and Figure 6 that
strategy 2 can work well in case 2 but still not effective in case 3. Finally, we use strategy 3 in case 3 and observe that it shows a good result for Skeptic in Figure 7.

From these simulations, we see that Skeptic can beat Reality with more flexible strategy utilizing more side information.

5.2 Betting against probability of precipitation by the Japan Meteorological Agency

Now we apply our strategy to probability of precipitation provided by the Japan Meteorological Agency. We collected the forecast probabilities for the Tokyo area from archives of the morning
edition of the Mainichi Daily News and the actual weather data on 9:00 and 15:00 of each
day for Tokyo area from http://www.weather-eye.com/ for the period of three years from
January 1, 2009 to December 31, 2011. We counted a day as rainy if the data on this site records
rain on 9:00 or on 15:00 of that day in Tokyo area.

The forecast probability $p_n$ is only announced as multiples of 10% (i.e. 0%, 10%, . . . , 90%,
100%) by JMA. The data are summarized in Table 1. $p_n$ represents the probability of precipitation
on day $n$ and $x_n$ indicates the actual precipitation. Actual ratio is calculated from the ratio
of the number of rainy days to all days for a given value of $p_n$.

<table>
<thead>
<tr>
<th>$p_n$ (%)</th>
<th>$x_n = 1$</th>
<th>$x_n = 0$</th>
<th>Actual Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>61</td>
<td>1.6</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>324</td>
<td>3.0</td>
</tr>
<tr>
<td>20</td>
<td>24</td>
<td>193</td>
<td>11.1</td>
</tr>
<tr>
<td>30</td>
<td>36</td>
<td>117</td>
<td>23.5</td>
</tr>
<tr>
<td>40</td>
<td>20</td>
<td>26</td>
<td>43.5</td>
</tr>
<tr>
<td>50</td>
<td>67</td>
<td>56</td>
<td>54.5</td>
</tr>
<tr>
<td>60</td>
<td>38</td>
<td>14</td>
<td>73.1</td>
</tr>
<tr>
<td>70</td>
<td>36</td>
<td>7</td>
<td>85.7</td>
</tr>
<tr>
<td>80</td>
<td>36</td>
<td>4</td>
<td>90.0</td>
</tr>
<tr>
<td>90</td>
<td>22</td>
<td>1</td>
<td>95.6</td>
</tr>
<tr>
<td>100</td>
<td>3</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

The distinct feature of the prediction by JMA is that that $p_n$ tends to be closer to 50% than the actual ratio. For example, when JMA announces $p_n = 20\%$, the actual ratio is only 11.1\%. Similarly when JMA announces $p_n = 80\%$, the actual ratio is 90\%. Hence JMA has the tendency of avoiding clear-cut forecasts.

In the hindsight strategy, the value of $\beta$, which is a coefficient for $\log(p_n/(1-p_n))$ in strategy
3 is close to 1.5. Hence we modified strategy 3 of the previous section, so that the prior for $\beta$
is uniform between 0 and 2. We also substituted $p_n = 1\%$ and $p_n = 99\%$ for $p_n = 0\%$ and
$p_n = 100\%$, respectively, because our strategy is not defined for $p_n = 0\%$ or 100\%. Figure 8
shows the behavior of strategy 3 and the approximation $S_n^\alpha V_n^{-1} S_n/2$. We see that our strategy
works very well against JMA by exploiting its tendency of avoiding clear-cut forecasts. It is
also of interest that the capital process shows a seasonal fluctuation and it does not perform well
for the rainy season (June and July) in Tokyo area.

6 Summary and discussion

In this paper we proposed a Bayesian logistic betting strategy in the binary probability forecasting game with side information (BPFSI). We proved some theoretical results and showed
Figure 8: Beating JMA by strategy 3 with $\beta$ uniform over $[0, 2]$

good performance of our strategy against probability forecasting by Japanese Meteorological Agency.

Here we discuss some topics for further investigation.

We considered implications of a single Bayesian logistic betting strategy in BPFSI. We can also take a look at the sequential optimizing strategy (SOS) of [11] in BPFSI. Under the condition $\hat{\theta}_n \to 0$, Bayesian strategy and SOS should behave in the same way. However we could not succeed to prove weak forcing of $\hat{\theta}_n \to 0$ by SOS alone.

For the case of $d = 1$ we could employ approaches of [14] to prove results similar to Theorem 4.2. In [14] we also discussed Reality’s strategies. It is of interest to study strategies of Forecaster or Reality in the binary probability forecasting game with side information. Defensive forecasting ([20], [18]) can be considered as a strategy of Forecaster.

We extended the binary probability forecasting game by including side information. In our formulation side information $c_n$ is announced by Forecaster and in our logistic betting strategy $c_n$ is used as regressors in a logistic regression. However Skeptic can use any transformation of $c_n$ in his strategy. In this sense, it might be more natural to formulate the game, where $c_n$ is announced by Skeptic. Binary probability forecasting game is often considered from the viewpoint of prequential probability ([7]) and leads to the notion of randomness of the sequence $p_1, x_1 p_2, x_2 \ldots$ ([19], [13]). From the viewpoint of prequential probability it might also be natural to consider side information $c_n$ as a part of moves by Skeptic for testing the randomness of $p_1, x_1 p_2, x_2 \ldots$.

We assumed multidimensional $c_n$. However from the viewpoint of game-theoretic probabil-
ity, we do not lose much generality by restricting $c_n$ to be a scalar, since if Skeptic can weakly force events $E_1, \ldots, E_d$ then he can weakly force $E_1 \cap \cdots \cap E_d$. By the same reasoning we can also consider $d = \infty$, because if Skeptic can weakly force $E_1, E_2, \ldots$, then he can weakly force $\bigcap_{i=1}^{\infty} E_i$. Interpretation and formulation of side information in game-theoretic probability needs further investigation.

References


