Bayesian logistic betting strategy against probability forecasting*

Akimichi Takemura, Univ. Tokyo

(joint with Masayuki Kumon, Jing Li and Kei Takeuchi)

November 12, 2012

*arXiv:1204.3496. To appear in *Stochastic Analysis and Applications*.

Items

(i) **Introduction**

(ii) **Formulation of probability forecasting game**

(iii) **Logistic betting strategy**

(iv) **Forcing of law of large numbers**

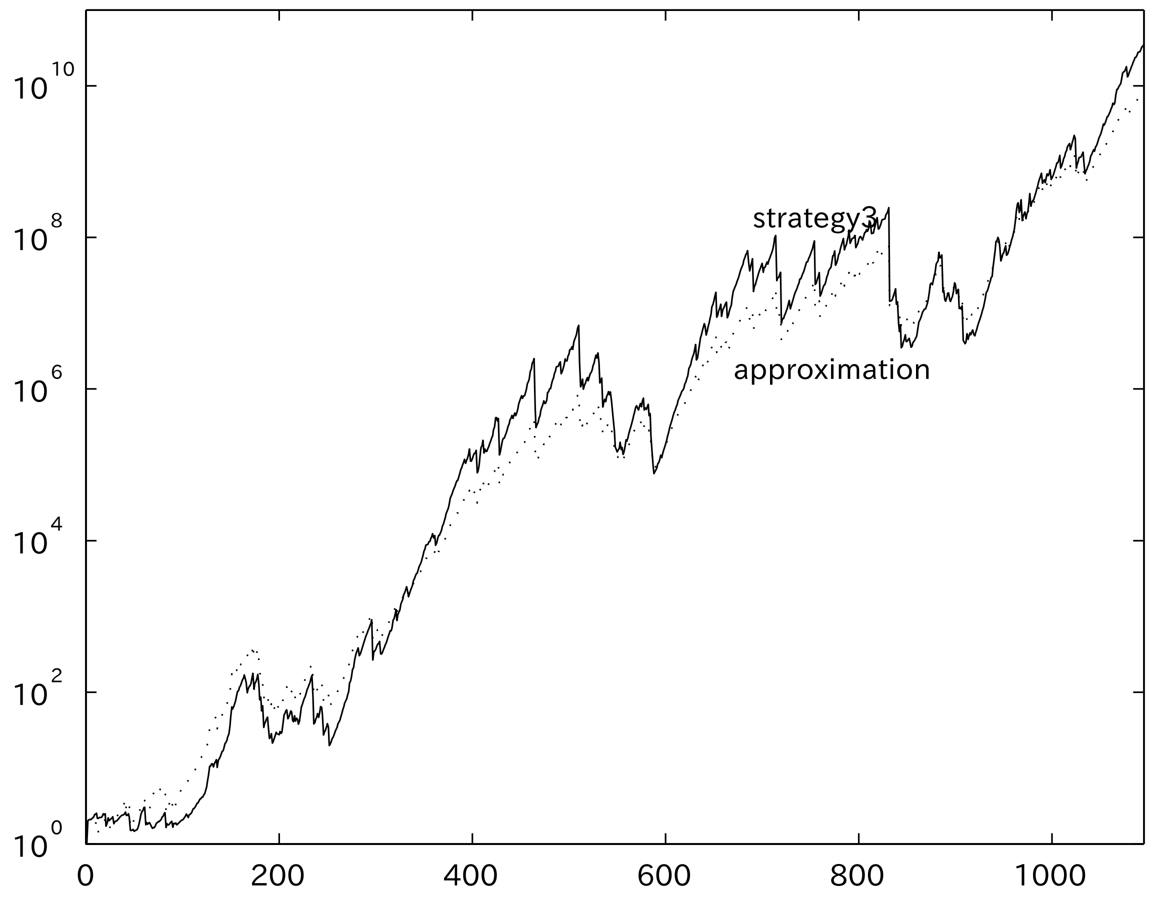
(v) **Experiments**

Introduction

- **Primary example of probability forecasting: probability of precipitation**
- **Japan Meteorological Agency started probability forecasting of "rain" in 1980. (multiples of 10%)**
- **Is the agency doing well?**
  - **Is their prediction consistent with the data?**
  - **Is their prediction "accurate"?**
  - **Is better prediction possible?**

- **Suppose that we are allowed to bet against the agency.**
- **If we can beat the agency in the betting game, then the agency is NOT doing well.**
- **In fact, it is found that we can beat the agency in the hypothetical betting game, as shown later $\rightarrow$ their prediction is not accurate.**

Beating the agency!



Formulation of probability forecasting game

- **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. (pays nothing if the event does not occur).**
- **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 $M_n \times (x_n - p_n)$.**
- **If the agency's predictions are not good, Skeptic may be able to increase his capital denoted by $\mathcal{K}_n$.**

Binary Probability Forecasting (BPF)

Protocol:

Skeptic announces his initial capital $\mathcal{K}_0 = 1$**.**

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 $\mathcal{K}_n \geq 0$**.**

BPF with Side Information c_n (BPFSI)

Protocol:

$\mathcal{K}_0 := 1, \mathcal{S}_0 := 0, \mathcal{V}_0 := 0$.

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 $x_n \in \{0,1\}$.

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

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

$\mathcal{V}_n := \mathcal{V}_{n-1} + c_n c_n' \, p_n(1 - p_n)$. $\quad$ $'$: **transpose**

Collateral Duty: Skeptic must keep $\mathcal{K}_n \geq 0$.

Role of side information c_n

- $\mathcal{V}_n = \mathrm{Var}(\mathcal{S}_n)$ **is the variance of** $\mathcal{S}_n$**.**
- **If** $c_n \equiv 1$**, we are looking at the behavior of** $\mathcal{S}_n = \sum_{i=1}^{n}(x_i - p_i)$**. In SLLN we are interested in the convergence of** $\mathcal{S}_n/n$ **of** $\mathcal{S}_n/\mathcal{V}_n$**.**
- **In BPFSI, we are assuming that Forecaster discloses relevant side information** c_n **to Skeptic.**
- **However** c_n **can be any information available to Skeptic, just before his move.**

- **We introduced the side information because of the following measure-theoretic consideration.**
  - **Let** $(\Omega, P, \mathcal{F})$ **be a probability space and let** $\mathcal{G} \subset \mathcal{F}$ **be a subfield of** $\mathcal{F}$**.**
  - **Let** X **be a random variable s.t.** $E|X| < \infty$ **and let** $E(X|\mathcal{G})$ **denote the conditional expectation w.r.t.** $\mathcal{G}$**.**
  - **Let** Y **be** $\mathcal{G}$**-measurable. Then**

$$Y = E(X|\mathcal{G}) \textbf{ a.s.} \quad \Leftrightarrow \quad E[C(X - Y)] = 0$$

$$\forall C : \textbf{G-measurable and bounded}.$$

Proof.

($\Rightarrow$) Obvious.

($\Leftarrow$) Let $C = I_{\{E(X|\mathcal{G})-Y>0\}}$**. This** C **is** $\mathcal{G}$**-measurable. Then**

$$\begin{aligned} 0 &= E[I_{\{E(X|\mathcal{G})>Y\}}(X - Y)] \\ &= E[E[I_{\{E(X|\mathcal{G})-Y>0\}}(E(X|\mathcal{G}) - Y)|\mathcal{G}]] \\ &= E[I_{\{E(X|\mathcal{G})-Y>0\}}(E(X|\mathcal{G}) - Y)] \end{aligned}$$

Then

$$P\{E(X|\mathcal{G}) > Y\} = 0.$$

Similarly $P\{E(X|\mathcal{G}) < Y\} = 0.$ **(Q.E.D.)**

- **Therefore we can test whether $\{x_n - p_n\}$'s are martingale differences, by considering the covariance between them and any bounded predictable sequence $\{c_n\}$.**
- **For example, if $c_n = x_{n-1}$ we are testing whether the first-order autocovariance is zero or not.**

Equivalence of betting strategy and Skeptic's probability

- **Define the betting ratio as**

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

- **Suppose that Skeptic himself models the Reality's move as Bernoulli random variable with success probability $\hat{p}_n$.**

- **For a given Forecaster's move p_n, define a one-to-one correspondence between $\hat{p}_n$ and ν_n:**

$$\nu_n = \frac{\hat{p}_n - p_n}{p_n(1-p_n)} = \frac{\hat{p}_n}{p_n} - \frac{1-\hat{p}_n}{1-p_n}$$

- **Then a strategy of Skeptic is equivalent to a formulation of $\hat{p}_n$ in BPFSI.**

- **Hence we describe Skeptic's strategy by a formulation of $\hat{p}_n$**

- **We formulate logistic regression model for $\hat{p}_n$ (standard statistics stuff).**

Regularity conditions

Let $\lambda_{\max,n}$ **and** $\lambda_{\min,n}$ **denote the maximum and the minimum eigenvalues of** $\mathcal{V}_n$**.**

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.**

Logistic betting strategy

- **Model for $\hat{p}_n$:**

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

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

- $p_n/(1-p_n)$ **may be one component of c_n.**

- **Skeptic's model is, at round n,**

$$P(X_n = 1) = \hat{p}_n, \qquad P(X_n = 0) = 1 - \hat{p}_n.$$

- **As in the usual statistical logistic regression, θ can be estimated by maximum likelihood estimation or Bayes estimation.**

- **Bayes estimation corresponds to static mixture of betting strategies for fixed θ.**

- **Let $\mathcal{K}_n^\theta$ denote the capital process for a given θ.**

- **Let $\pi(\theta)$ denote a "prior" probability density for θ.**

- **Basic arguments in Shafer-Vovk use discrete π. Discete π is more clear, but continus π is more convenient for numerical analysis.**

- **“Bayesian logistic betting strategy”:**

$$\mathcal{K}_n^{\pi} = \int_{\mathbb{R}^d} \mathcal{K}_n^{\theta} \pi(\theta) d\theta$$

is the capital process of a Bayesian logistic betting strategy with the prior density π.

- $\mathcal{K}_n^{\theta}$ **can be explicitly written for fixed θ. $\mathcal{K}_n^{\pi}$ needs numerical integration.**

Forcing of law of large numbers

Theorem 1 (Usual form of SLLN) In BPFSI, by a Bayesian logistic strategy with a prior distribution π supporting a neighborhood of the origin, Skeptic can weakly force

$$\text{i}), \text{ii}), \text{iii}) \quad \Rightarrow \quad \lim_n \mathcal{V}_n^{-1} \mathcal{S}_n = 0,$$

where i),ii),iii) are the regularity conditions.

Stronger form of SLLN

Theorem 2 In BPFSI, by a Bayesian logistic strategy with a prior distribution π supporting a neighborhood of the origin, Skeptic can weakly force

$$\text{i}), \text{ii}), \text{iii}) \quad \Rightarrow \quad \limsup_n \frac{\mathcal{S}_n' \mathcal{V}_n^{-1} \mathcal{S}_n}{\log \det \mathcal{V}_n} \leq 1,$$

where i), ii), iii) are the regularity conditions.

Approximation for $\log \mathcal{K}_n^{\pi}$

- **Theorem 2 is the consequence of the following approximation:**

$$\log \mathcal{K}_n^{\pi} = \frac{1}{2}\mathcal{S}_n' \mathcal{V}_n^{-1} \mathcal{S}_n - \frac{1}{2} \log\det \mathcal{V}_n + o(\log\det \mathcal{V}_n)$$

- $\mathcal{S}_n' \mathcal{V}_n^{-1} \mathcal{S}_n / 2$ **corresponds to the maximized log likelihood, i.e. the log likelihood at the maximum likelihood estimate (MLE)** $\hat{\theta}_n$ **of** θ **at round** n **:**

$$\hat{\theta}_n^* = \text{argmax}_{\theta}\, \mathcal{K}_n^{\theta}$$

- $\hat{\theta}_n^*$**: "hindsight best strategy" at round** n**.**

Experiments

- **Let** $c'_n = (1, \log p_n/(1 - p_n), x_{n-1})$ **and** $\theta = (\theta_1, \beta - 1, \theta_3)$**.**

- **We model the probability of precipitation** $\hat{p}_n$ **by**

$$\log \frac{\hat{p}_n}{1 - \hat{p}_n} = \log \frac{p_n}{1 - p_n} + c'_n \theta = \theta_1 + \beta \log \frac{p_n}{1 - p_n} + \theta_3 x_{n-1}$$

- **We first tried** π**, which is the uniform distribution on** $[0, 1]^3$**. This did not work. (Actually we should have considered** $[-1, 1]^3$**.)**

- **We then used the uniform distribution on** $[0, 1] \times [0, 2] \times [0, 1]$ **(**$\Rightarrow$ **p.4.** **!!)**

Why $[0, 2]$ for β works?

- **Because the hindsight best β_n^* is about 1.5.**
- **Look at the following table.**
- **It shows that Japan Meteorological Agency has the tendency of avoiding clear-cut forecasts.**
- **We only found this fact after the uniform prior on $[0, 1]^3$ did not work.**

Table 1: Actual ratio of rainy days

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

Future works

- **Our logistic betting strategy is nice, because the strategy directly depends Forecaster's current forecast p_n. In many papers in game-theoretic probability, Skeptic's strategies often ignore Forecaster's current forecast.**
- **We would like to extend our strategy for the binary case $x_n \in \{0, 1\}$ to the continuous case $x_n \in \mathbb{R}$.**
- **Such a strategy should be very useful for defensive forecasting.**