Nonparametric Sequential Prediction of Time Series.

Extension to quantile prediction.

G. Biau – B. Patra

University Paris VI – University Paris VI - Lokad

ENGREF, JUNE 2009

Outline

Outline

1. Nonparametric mean prediction
   - Context
   - A consistent strategy
   - Experimental results
2. Non parametric quantile prediction
   - Context and quantile regression
   - A similar consistent strategy
   - Experimental results
3. Other contexts
4. Conclusion

Introduction

- Time series prediction has a long history [Yule, 1927].
- Genetics, medecine, climate, finance…
- Until 1970's, parametric approach.
- Recently: nonparametric approaches.

Nonparametric framework

A slightly different spirit

- Consider the sequential (= on-line) prediction of time series.
- Including series that do not necessarily satisfy classical statistical assumptions for bounded, mixing or Markovian processes.

Goal

- Show consistency results under a minimum of hypotheses.

Nonparametric framework

A slightly different spirit

- Consider the sequential (= on-line) prediction of time series.
- Including series that do not necessarily satisfy classical statistical assumptions for bounded, mixing or Markovian processes.

Goal

- Show consistency results under a minimum of hypotheses.

Context

Sequential prediction

- At time $n = 1, 2, \ldots$, predict the next outcome $y_n \in \mathbb{R}$ of sequence $y_1, y_2, \ldots$
- Side information $x_1, x_2, \ldots$, where each $x_i \in \mathbb{R}^d$.

Notation

- $y_1^{n-1} = (y_1, \ldots, y_{n-1})$.
- $x_1^n = (x_1, \ldots, x_n)$.

On-line learning

The elements $y_0, y_1, y_2, \ldots$ and $x_1, x_2, \ldots$ are revealed one at a time, in order, beginning with $(x_1, y_0), (x_2, y_1), \ldots$

Context

Sequential prediction

- At time $n = 1, 2, \ldots$, predict the next outcome $y_n \in \mathbb{R}$ of sequence $y_1, y_2, \ldots$
- Side information $x_1, x_2, \ldots$, where each $x_i \in \mathbb{R}^d$.

Notation

- $y_1^{n-1} = (y_1, \ldots, y_{n-1})$.
- $x_1^n = (x_1, \ldots, x_n)$.

On-line learning

The elements $y_0, y_1, y_2, \ldots$ and $x_1, x_2, \ldots$ are revealed one at a time, in order, beginning with $(x_1, y_0), (x_2, y_1), \ldots$

Context

Sequential prediction

- At time $n = 1, 2, \ldots$, predict the next outcome $y_n \in \mathbb{R}$ of sequence $y_1, y_2, \ldots$
- Side information $x_1, x_2, \ldots$, where each $x_i \in \mathbb{R}^d$.

Notation

- $y_1^{n-1} = (y_1, \ldots, y_{n-1})$.
- $x_1^n = (x_1, \ldots, x_n)$.

On-line learning

The elements $y_0, y_1, y_2, \ldots$ and $x_1, x_2, \ldots$ are revealed one at a time, in order, beginning with $(x_1, y_0), (x_2, y_1), \ldots$

Prediction

Strategy

- Sequence $g = \{g_n\}_{n=1}^{\infty}$ of forecasting functions

$$g_n : (\mathbb{R}^d)^n \times \mathbb{R}^{n-1} \to \mathbb{R}.$$

- Prediction at time n is $g_n(x_1^n, y_1^{n-1})$.

Hypotheses

- $(x_1, y_1), (x_2, y_2), \ldots$ are realizations of random variables (X_1, Y_1), $(X_2, Y_2), \ldots$
- The process $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ is jointly stationary and ergodic.

Prediction

Strategy

- Sequence $g = \{g_n\}_{n=1}^{\infty}$ of forecasting functions

$$g_n : (\mathbb{R}^d)^n \times \mathbb{R}^{n-1} \to \mathbb{R}.$$

- Prediction at time n is $g_n(x_1^n, y_1^{n-1})$.

Hypotheses

- $(x_1, y_1), (x_2, y_2), \ldots$ are realizations of random variables (X_1, Y_1), $(X_2, Y_2), \ldots$
- The process $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ is jointly stationary and ergodic.

Measuring the error

Definition

At time n, the (*normalized*) *cumulative prediction error* on the strings X_1^n and Y_1^n is

$$L_n(g) = \frac{1}{n} \sum_{t=1}^{n} \left(g_t(X_1^t, Y_1^{t-1}) - Y_t \right)^2 .$$

Goal versus reality

- Goal: make $L_n(g)$ small.
- Reality: fundamental limit $L^\star$ [Algoet, 1994].

Measuring the error

Definition

At time n, the (*normalized*) *cumulative prediction error* on the strings X_1^n and Y_1^n is

$$L_n(g) = \frac{1}{n} \sum_{t=1}^{n} \left(g_t(X_1^t, Y_1^{t-1}) - Y_t \right)^2 .$$

Goal versus reality

- Goal: make $L_n(g)$ small.
- Reality: fundamental limit $L^\star$ [Algoet, 1994].

How good can we get?

Fundamental limit [Algoet, 1994]

For any prediction strategy g and jointly stationary ergodic process $\{(X_n, Y_n)\}_{-\infty}^{\infty}$,

$$\liminf_{n \to \infty} L_n(g) \geq L^\star \quad \text{almost surely,}$$

where

$$L^\star = \mathbb{E}\left\{\left(Y_0 - \mathbb{E}\left\{Y_0 | X_{-\infty}^0, Y_{-\infty}^{-1}\right\}\right)^2\right\}$$

is the minimal mean squared error of any prediction for the value of Y_0 based on the infinite past observation sequences

$$Y_{-\infty}^{-1} = (\ldots, Y_{-2}, Y_{-1}) \quad \textit{and} \quad X_{-\infty}^{0} = (\ldots, X_{-1}, X_0).$$

Consistency

Definition

A prediction strategy g is universally consistent with respect to a class $\mathcal{C}$ of stationary and ergodic processes $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ if for each process in the class,

$$\lim_{n\to\infty} L_n(g) = L^\star \quad \text{almost surely.}$$

- There exist universally consistent strategies for the class $\mathcal{C}$ of all bounded, stationary and ergodic processes [Algoet, 1992 and Morvai et al., 1996].
- Very complex or slow-converging algorithms.

Consistency

Definition

A prediction strategy g is universally consistent with respect to a class $\mathcal{C}$ of stationary and ergodic processes $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ if for each process in the class,

$$\lim_{n \to \infty} L_n(g) = L^\star \quad \text{almost surely.}$$

- There exist universally consistent strategies for the class $\mathcal{C}$ of all bounded, stationary and ergodic processes [Algoet,1992 and Morvai et al.,1996].
- Very complex or slow-converging algorithms.

Outline

State of the art

A book on prediction of individual sequences.

Cesa-Bianchi, N. and Lugosi, G. Prediction, Learning, and Games, Cambridge University Press, New York, 2006.

Quantization strategies

- Györfi and Lugosi (2001) and Nobel (2003): bounded processes.
- Györfi and Ottucsák (2007): fourth-finite moment processes.

Biau and al. contribution

Several simple nonparametric strategies for non-necessarily bounded processes:

- Kernel-based strategy.
- Nearest neighbor-based strategy.

State of the art

A book on prediction of individual sequences.

Cesa-Bianchi, N. and Lugosi, G. Prediction, Learning, and Games, Cambridge University Press, New York, 2006.

Quantization strategies

- Györfi and Lugosi (2001) and Nobel (2003): bounded processes.
- Györfi and Ottucsák (2007): fourth-finite moment processes.

Biau and al. contribution

Several simple nonparametric strategies for non-necessarily bounded processes:

- Kernel-based strategy.
- Nearest neighbor-based strategy.

State of the art

A book on prediction of individual sequences.

Cesa-Bianchi, N. and Lugosi, G. Prediction, Learning, and Games, Cambridge University Press, New York, 2006.

Quantization strategies

- Györfi and Lugosi (2001) and Nobel (2003): bounded processes.
- Györfi and Ottucsák (2007): fourth-finite moment processes.

Biau and al. contribution

Several simple nonparametric strategies for non-necessarily bounded processes:

- Kernel-based strategy.
- Nearest neighbor-based strategy.

Nearest neighbor strategy

- Define infinite array of experts $h^{(k,\ell)} = \{h_n^{(k,\ell)}\}_{n=1}^{+\infty}$. $k, \ell = 1, 2, \ldots$

What are k and $\bar{\ell}$?

- k is the length of the past observation vector we consider.
- $\bar{\ell}$ (simple function of ℓ) is the number of nearest neighbors of length k we consider.
- More precisely, $\bar{\ell} = \lfloor p_\ell n \rfloor$ where $p_\ell \in (0, 1)$ and $\lim_{\ell \to \infty} p_\ell = 0$.

Each expert has a job

- At time n, expert $h_n^{(k,\ell)}$ searches for the $\bar{\ell}$ nearest neighbors of length k.

Nearest neighbor strategy

- Define infinite array of experts $h^{(k,\ell)} = \{h_n^{(k,\ell)}\}_{n=1}^{+\infty}$. $k, \ell = 1, 2, \ldots$

What are k and $\bar{\ell}$?

- k is the length of the past observation vector we consider.
- $\bar{\ell}$ (simple function of ℓ) is the number of nearest neighbors of length k we consider.
- More precisely, $\bar{\ell} = \lfloor p_\ell n \rfloor$ where $p_\ell \in (0, 1)$ and $\lim_{\ell \to \infty} p_\ell = 0$.

Each expert has a job

- At time n, expert $h_n^{(k,\ell)}$ searches for the $\bar{\ell}$ nearest neighbors of length k.

Nearest neighbor strategy

- Define infinite array of experts $h^{(k,\ell)} = \{h_n^{(k,\ell)}\}_{n=1}^{+\infty}$. $k, \ell = 1, 2, \ldots$

What are k and $\bar{\ell}$?

- k is the length of the past observation vector we consider.
- $\bar{\ell}$ (simple function of ℓ) is the number of nearest neighbors of length k we consider.
- More precisely, $\bar{\ell} = \lfloor p_\ell n \rfloor$ where $p_\ell \in (0, 1)$ and $\lim_{\ell\to\infty} p_\ell = 0$.

Each expert has a job

- At time n, expert $h_n^{(k,\ell)}$ searches for the $\bar{\ell}$ nearest neighbors of length k.

Nearest neighbor strategy

- Define infinite array of experts $h^{(k,\ell)} = \{h_n^{(k,\ell)}\}_{n=1}^{+\infty}$. $k, \ell = 1, 2, \ldots$

What are k and $\bar{\ell}$?

- k is the length of the past observation vector we consider.
- $\bar{\ell}$ (simple function of ℓ) is the number of nearest neighbors of length k we consider.
- More precisely, $\bar{\ell} = \lfloor p_\ell n \rfloor$ where $p_\ell \in (0, 1)$ and $\lim_{\ell \to \infty} p_\ell = 0$.

Each expert has a job

- At time n, expert $h_n^{(k,\ell)}$ searches for the $\bar{\ell}$ nearest neighbors of length k.

1,34	1,78	2,56	1,88	0,57	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31

2,27	2,89	2,12	1,78	2,67	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	0,99

0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	0,91	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	**?**

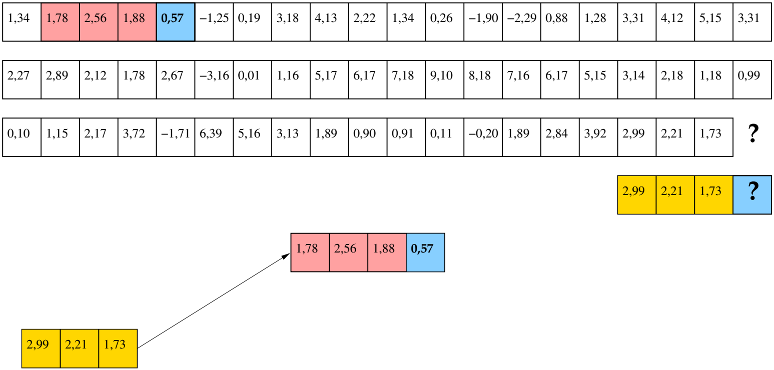
Figure: Work of fundamental expert with $k = 3$ and $\bar{\ell} = 4$.

1,34	1,78	2,56	1,88	0,57	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31

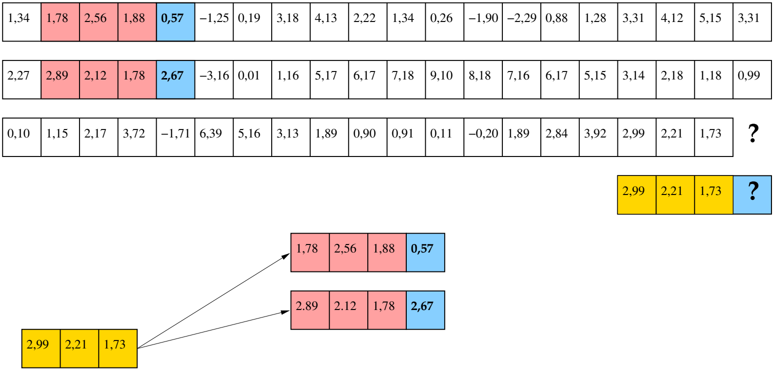
2,27	2,89	2,12	1,78	2,67	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	0,99

0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	0,91	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	?

2,99	2,21	1,73	?

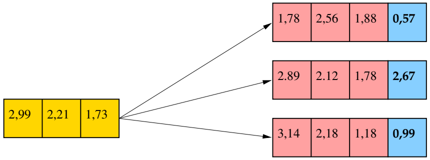


1,34 1,78 2,56 1,88 0,57 −1,25 0,19 3,18 4,13 2,22 1,34 0,26 −1,90 −2,29 0,88 1,28 3,31 4,12 5,15 3,31
2,27 2,89 2,12 1,78 2,67 −3,16 0,01 1,16 5,17 6,17 7,18 9,10 8,18 7,16 6,17 5,15 3,14 2,18 1,18 0,99
0,10 1,15 2,17 3,72 −1,71 6,39 5,16 3,13 1,89 0,90 0,91 0,11 −0,20 1,89 2,84 3,92 2,99 2,21 1,73 ?
2,99 2,21 1,73 ?
1,78 2,56 1,88 0,57
2,99 2,21 1,73


1,34 1,78 2,56 1,88 0,57 −1,25 0,19 3,18 4,13 2,22 1,34 0,26 −1,90 −2,29 0,88 1,28 3,31 4,12 5,15 3,31
2,27 2,89 2,12 1,78 2,67 −3,16 0,01 1,16 5,17 6,17 7,18 9,10 8,18 7,16 6,17 5,15 3,14 2,18 1,18 0,99
0,10 1,15 2,17 3,72 −1,71 6,39 5,16 3,13 1,89 0,90 0,91 0,11 −0,20 1,89 2,84 3,92 2,99 2,21 1,73 ?
2,99 2,21 1,73 ?
1,78 2,56 1,88 0,57
2.89 2.12 1,78 2,67
2,99 2,21 1,73

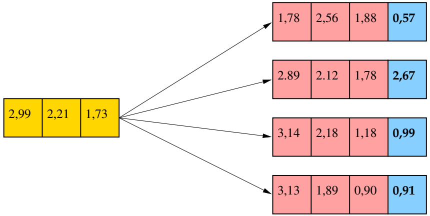
1,34	1,78	2,56	1,88	**0,57**	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31
2,27	2,89	2,12	1,78	**2,67**	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	**0,99**
0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	0,91	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	**?**

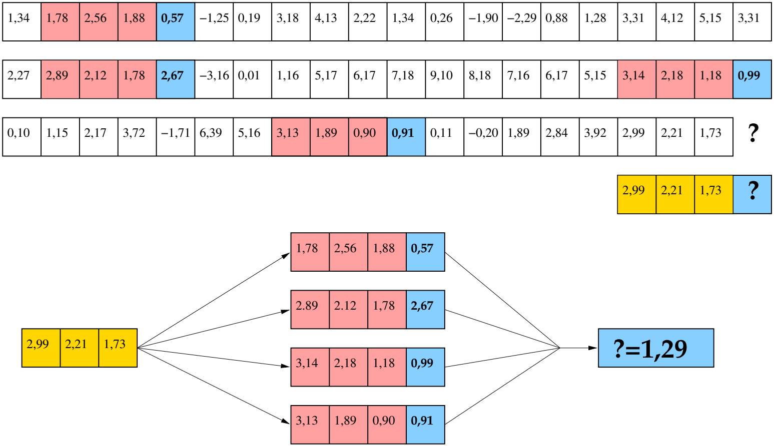
2,99	2,21	1,73	**?**



1,34	1,78	2,56	1,88	**0,57**	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31
2,27	2,89	2,12	1,78	**2,67**	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	**0,99**
0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	**0,91**	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	**?**

2,99	2,21	1,73	**?**




1,34 1,78 2,56 1,88 0,57 −1,25 0,19 3,18 4,13 2,22 1,34 0,26 −1,90 −2,29 0,88 1,28 3,31 4,12 5,15 3,31
2,27 2,89 2,12 1,78 2,67 −3,16 0,01 1,16 5,17 6,17 7,18 9,10 8,18 7,16 6,17 5,15 3,14 2,18 1,18 0,99
0,10 1,15 2,17 3,72 −1,71 6,39 5,16 3,13 1,89 0,90 0,91 0,11 −0,20 1,89 2,84 3,92 2,99 2,21 1,73 ?
2,99 2,21 1,73 ?
2,99 2,21 1,73
1,78 2,56 1,88 0,57
2.89 2.12 1,78 2,67
3,14 2,18 1,18 0,99
3,13 1,89 0,90 0,91
?=1,29

Prediction and Aggregation

Prediction of one expert

$$h_n^{(k,\ell)}(x_1^n, y_1^{n-1}) = T_{\min\{n^\delta,\ell\}}\left(\frac{\sum_{\{t\in J_n^{(k,\ell)}\}} y_t}{|J_n^{(k,\ell)}|}\right).$$

Aggregated prediction of all experts

$$g_n(x_1^n, y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(x_1^n, y_1^{n-1}).$$

Where do the $p_{k,\ell,n}$ come from?

Exponentially weight the experts based on their past performance.

Prediction and Aggregation

Prediction of one expert

$$h_n^{(k,\ell)}(x_1^n, y_1^{n-1}) = T_{\min\{n^\delta,\ell\}}\left(\frac{\sum_{\{t \in J_n^{(k,\ell)}\}} y_t}{|J_n^{(k,\ell)}|}\right).$$

Aggregated prediction of all experts

$$g_n(x_1^n, y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(x_1^n, y_1^{n-1}).$$

Where do the $p_{k,\ell,n}$ come from?

Exponentially weight the experts based on their past performance.

Prediction and Aggregation

Prediction of one expert

$$h_n^{(k,\ell)}(x_1^n, y_1^{n-1}) = T_{\min\{n^\delta,\ell\}}\left(\frac{\sum_{\{t\in J_n^{(k,\ell)}\}} y_t}{\left|J_n^{(k,\ell)}\right|}\right).$$

Aggregated prediction of all experts

$$g_n(x_1^n, y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(x_1^n, y_1^{n-1}).$$

Where do the $p_{k,\ell,n}$ come from?

Exponentially weight the experts based on their past performance.

Definitions

- Let $\{q_{k,\ell}\}$ be a probability distribution over all pairs (k, ℓ) of positive integers such that $q_{k,\ell} > 0$ for all (k, ℓ).
- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) L_{n-1}(h^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction

$$g_n(x_1^n, y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(x_1^n, y_1^{n-1}).$$

Aggregation continued. . ..

Definitions

- Let $\{q_{k,\ell}\}$ be a probability distribution over all pairs (k, ℓ) of positive integers such that $q_{k,\ell} > 0$ for all (k, ℓ).
- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) L_{n-1}(h^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction

$$g_n(x_1^n, y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(x_1^n, y_1^{n-1}).$$

Aggregation continued...

Definitions

- Let $\{q_{k,\ell}\}$ be a probability distribution over all pairs (k,ℓ) of positive integers such that $q_{k,\ell} > 0$ for all (k,ℓ).
- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) L_{n-1}(h^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction

$$g_n(x_1^n, y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(x_1^n, y_1^{n-1}).$$

Aggregation continued. . .

Definitions

- Let $\{q_{k,\ell}\}$ be a probability distribution over all pairs (k, ℓ) of positive integers such that $q_{k,\ell} > 0$ for all (k, ℓ).
- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) L_{n-1}(h^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction

$$g_n(x_1^n, y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(x_1^n, y_1^{n-1}).$$

Result

Theorem [Biau, Bleakley, Györfi, Ottucsák, 2009]

- Let $\mathcal{C}$ be the class of all jointly stationary and ergodic processes $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ such that $\mathbb{E}\{Y_0^4\} < \infty$.
- Choose parameter η_n as

$$\eta_n = \frac{1}{\sqrt{n}}.$$

- Then the nearest neighbor forecasting strategy is universally consistent with respect to the class $\mathcal{C}$, that is

$$\lim_{n\to\infty} L_n(g) = L^\star \quad \text{almost surely.}$$

Result

Theorem [Biau, Bleakley, Györfi, Ottucsák, 2009]

- Let $\mathcal{C}$ be the class of all jointly stationary and ergodic processes $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ such that $\mathbb{E}\{Y_0^4\} < \infty$.
- Choose parameter η_n as

$$\eta_n = \frac{1}{\sqrt{n}} .$$

- Then the nearest neighbor forecasting strategy is universally consistent with respect to the class $\mathcal{C}$, that is

$$\lim_{n\to\infty} L_n(g) = L^\star \quad \text{almost surely.}$$

Result

Theorem [Biau, Bleakley, Györfi, Ottucsák, 2009]

- Let $\mathcal{C}$ be the class of all jointly stationary and ergodic processes $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ such that $\mathbb{E}\{Y_0^4\} < \infty$.
- Choose parameter η_n as

$$\eta_n = \frac{1}{\sqrt{n}}.$$

- Then the nearest neighbor forecasting strategy is universally consistent with respect to the class $\mathcal{C}$, that is

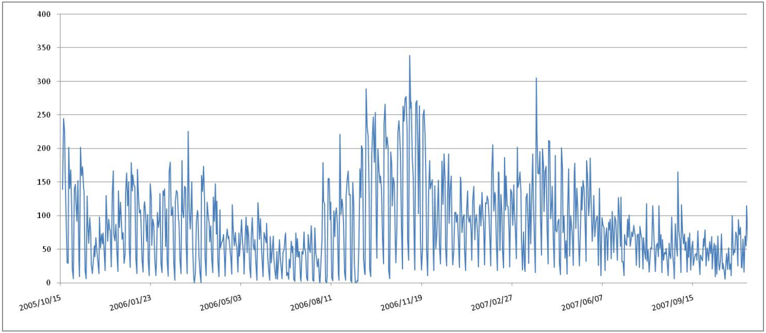
$$\lim_{n\to\infty} L_n(g) = L^\star \quad \text{almost surely.}$$

Outline

Experimental results

Call center data set

- Daily call volumes entering a call center.
- Long series between 382 and 826 time values. 21 series.



Future outcome predictions results

Model Name	Avg Abs Error	Avg Sqr Error	Mape (%)
AR(7)	65.80	9738	31.6
DayOfTheWeekMean	53.95	7099	22.8
HoltWinters	49.84	6025	**21.5**
MeanExpertMixture	52.37	6536	22.3
MA	179	62448	52.0

Figure: Forecasting future outcomes

Outline

Quantile forecasting

Quantile forecasting

Given a stochastic process $Y_1, Y_2, \ldots$.

- Previously, estimate the conditional mean of Y_n given $Y_1, \ldots, Y_{n-1}$.
- Now: the conditional τth quantile of Y_n given $Y_1, \ldots, Y_{n-1}$.

Quantile forecasting

Quantile forecasting

Given a stochastic process $Y_1, Y_2, \ldots$.

- Previously, estimate the conditional mean of Y_n given $Y_1, \ldots, Y_{n-1}$.
- Now: the conditional τth quantile of Y_n given $Y_1, \ldots, Y_{n-1}$.

Quantile forecasting

What for?

- Understand conditional distributions.
- $\tau = 0.5$ robust forecasting.
- Build confidence interval.

Applications fields

- Finance: CVAR. Also biology, medecine, telecoms...
- Here: call volumes (optimize staff in a call center).

Quantile forecasting

What for?

- Understand conditional distributions.
- $\tau = 0.5$ robust forecasting.
- Build confidence interval.

Applications fields

- Finance: CVAR. Also biology, medecine, telecoms...
- Here: call volumes (optimize staff in a call center).

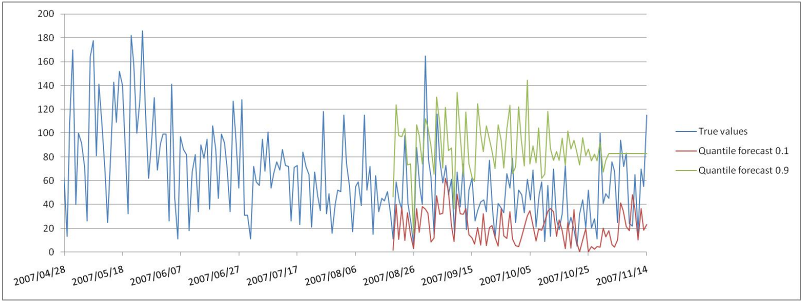


Figure: Quantile forecast with $\tau = 0.1, 0.9$.

Quantile Regression

Conditional quantiles

X multivariate random variable, Y real valued random variable,

$$q_\tau(X) \triangleq F^{\leftarrow}_{Y|X}(\tau) = \inf\{t \in \mathbb{R} : F_{Y|X}(t) \geq \tau\}.$$

$F_{Y|X}$ conditional cumulative distribution function.

Proposition [Koenker, 2005]

$$q_\tau(X) \in \underset{q(.)\in\mathbb{R}}{\operatorname{argmin}} \mathbb{E}_{\mathbb{P}_{Y|X}} \left[\rho_\tau\left(Y - q(X)\right)\right].$$

If $F_{Y|X}$ is (strictly) increasing then $q_\tau(X)$ is the unique minimizer.

Quantile Regression

Conditional quantiles

X multivariate random variable, Y real valued random variable,

$$q_\tau(X) \triangleq F^{\leftarrow}_{Y|X}(\tau) = \inf\{t \in \mathbb{R} : F_{Y|X}(t) \geq \tau\}.$$

$F_{Y|X}$ conditional cumulative distribution function.

Proposition [Koenker, 2005]

$$q_\tau(X) \in \underset{q(.)\in\mathbb{R}}{\operatorname{argmin}} \mathbb{E}_{\mathbb{P}_{Y|X}} \left[\rho_\tau \left(Y - q(X)\right)\right].$$

If $F_{Y|X}$ is (strictly) increasing then $q_\tau(X)$ is the unique minimizer.

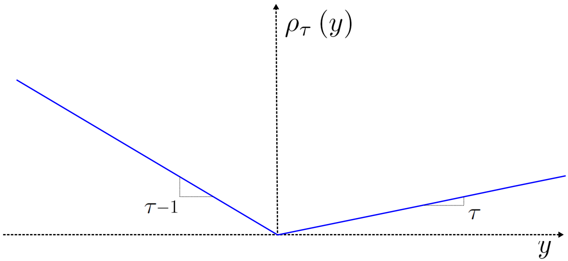


Figure: Pinball function graph.

Non parametric framework

Framework

- Here, we observe a string realization y_1^{n-1} of a stationary and ergodic process $\{Y_n\}_{n=-\infty}^{\infty}\ldots$
- $\ldots$ and try to estimate $q_\tau(y_1^{n-1}) = F^{\leftarrow}_{Y_n|Y_1^{n-1}=y_1^{n-1}}(\tau)$, the conditional quantile at time n.

Strategy

Sequence $g = \{g_n\}_{n=1}^{\infty}$ of τth quantile forecasting functions

$$g_n : \mathbb{R}^{n-1} \longrightarrow \mathbb{R}.$$

Quantile estimation at time n is $g_n(y_1^{n-1})$.

Non parametric framework

Framework

- Here, we observe a string realization y_1^{n-1} of a stationary and ergodic process $\{Y_n\}_{n=-\infty}^{\infty}\ldots$
- ... and try to estimate $q_\tau(y_1^{n-1}) = F^{\leftarrow}_{Y_n|Y_1^{n-1}=y_1^{n-1}}(\tau)$, the conditional quantile at time n.

Strategy

Sequence $g = \{g_n\}_{n=1}^{\infty}$ of τth quantile forecasting functions

$$g_n : \mathbb{R}^{n-1} \longrightarrow \mathbb{R}.$$

Quantile estimation at time n is $g_n(y_1^{n-1})$.

Errors

Empirical measure criterion.

At time n the cumulative pinball error of the strategy g is

$$G_n(g) = \frac{1}{n}\sum_{t=1}^{n} \rho_\tau \left(y_t - g_t(y_1^{t-1}) \right).$$

Adapted result of [Algoët, 1994]

For any stationary and ergodic process $\{Y_n\}_{n=-\infty}^{+\infty}$,

$$\liminf_{n\to\infty} G_n(g) \geq G^\star \quad \text{a.s.},$$

where

$$G^\star = \mathbb{E}\left[\min_{q(.)} \mathbb{E}_{\mathbb{P}_{Y_0 | Y_{-\infty}^{-1}}} \left[\rho_\tau\left(Y_0 - q(Y_{-\infty}^{-1})\right)\right]\right].$$

Errors

Empirical measure criterion.

At time n the cumulative pinball error of the strategy g is

$$G_n(g) = \frac{1}{n}\sum_{t=1}^{n} \rho_\tau \left(y_t - g_t(y_1^{t-1}) \right).$$

Adapted result of [Algoët, 1994]

For any stationary and ergodic process $\{Y_n\}_{n=-\infty}^{+\infty}$,

$$\liminf_{n\to\infty} G_n(g) \geq G^\star \quad \text{a.s.},$$

where

$$G^\star = \mathbb{E}\left[\min_{q(.)} \mathbb{E}_{\mathbb{P}_{Y_0|Y_{-\infty}^{-1}}}\left[\rho_\tau\left(Y_0 - q(Y_{-\infty}^{-1})\right)\right]\right].$$

Outline

Nearest neighbors strategy

Elementary predictors

- Define infinite array of experts $h^{(k,\ell)} = \{h_n^{(k,\ell)}\}_{n=1}^{+\infty}$. $k, \ell = 1, 2, \ldots$

Each expert has a job

- At time n, expert $h_n^{(k,\ell)}$ finds the $\bar{\ell}$ nearest neighbors of length k.

Nearest neighbors strategy

Elementary predictors

- Define infinite array of experts $h^{(k,\ell)} = \{h_n^{(k,\ell)}\}_{n=1}^{+\infty}$. $k, \ell = 1, 2, \ldots$

Each expert has a job

- At time n, expert $h_n^{(k,\ell)}$ finds the $\bar{\ell}$ nearest neighbors of length k.

1,34	1,78	2,56	1,88	0,57	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31

2,27	2,89	2,12	1,78	2,67	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	0,99

0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	0,91	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	**?**

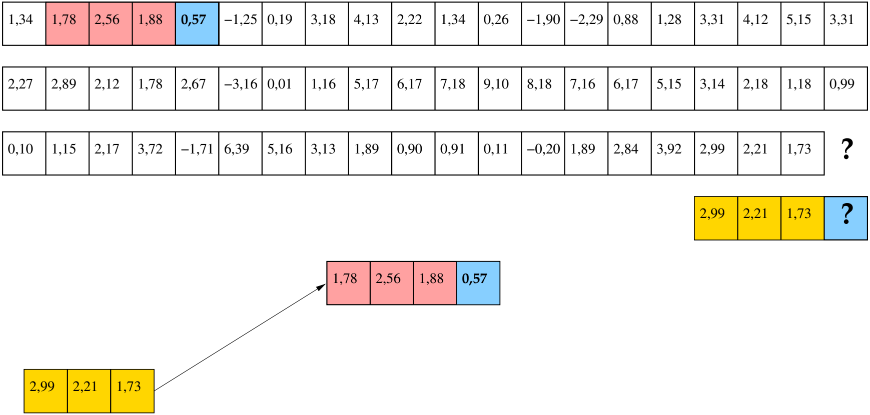
Figure: Work of fundamental expert with $k = 3$ and $\bar{\ell} = 4$.

1,34	1,78	2,56	1,88	0,57	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31

2,27	2,89	2,12	1,78	2,67	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	0,99

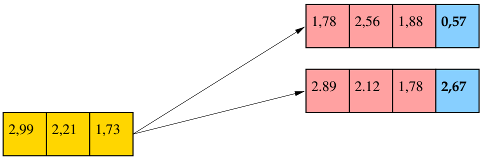
0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	0,91	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	?

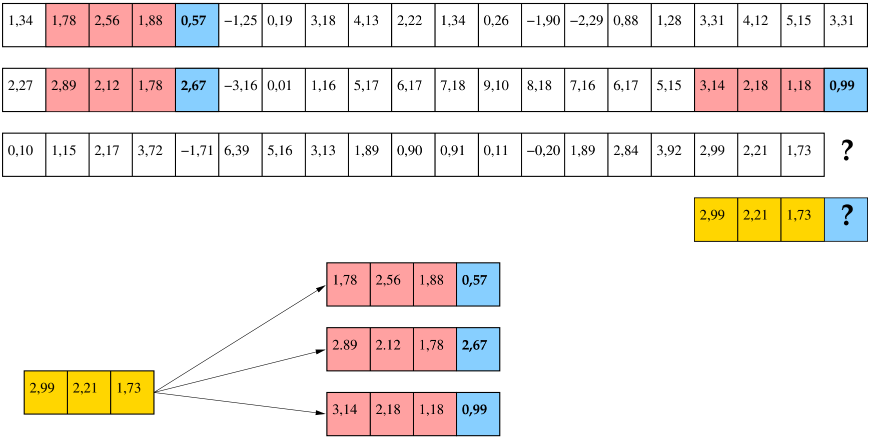
2,99	2,21	1,73	?


1,34 1,78 2,56 1,88 0,57 −1,25 0,19 3,18 4,13 2,22 1,34 0,26 −1,90 −2,29 0,88 1,28 3,31 4,12 5,15 3,31
2,27 2,89 2,12 1,78 2,67 −3,16 0,01 1,16 5,17 6,17 7,18 9,10 8,18 7,16 6,17 5,15 3,14 2,18 1,18 0,99
0,10 1,15 2,17 3,72 −1,71 6,39 5,16 3,13 1,89 0,90 0,91 0,11 −0,20 1,89 2,84 3,92 2,99 2,21 1,73 ?
2,99 2,21 1,73 ?
1,78 2,56 1,88 0,57
2,99 2,21 1,73

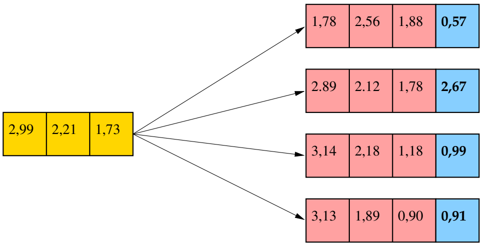
1,34	1,78	2,56	1,88	**0,57**	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31
2,27	2,89	2,12	1,78	**2,67**	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	0,99
0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	0,91	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	**?**

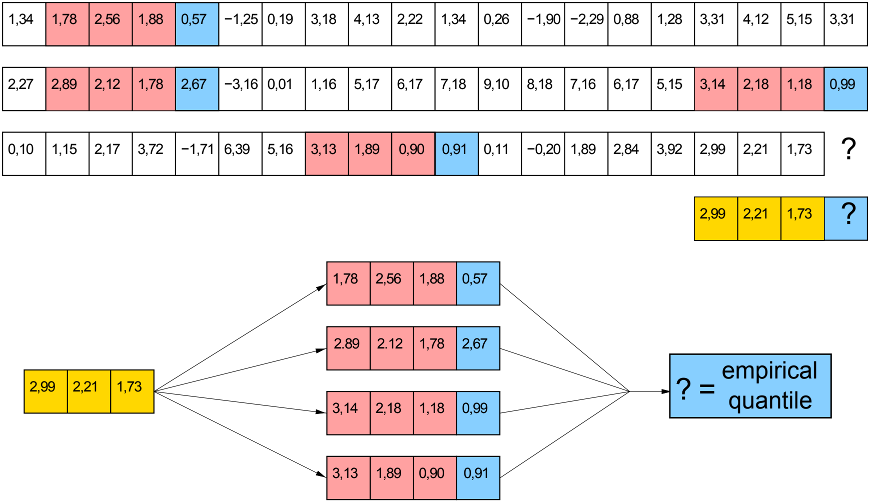
2,99	2,21	1,73	**?**




1,34 1,78 2,56 1,88 0,57 −1,25 0,19 3,18 4,13 2,22 1,34 0,26 −1,90 −2,29 0,88 1,28 3,31 4,12 5,15 3,31
2,27 2,89 2,12 1,78 2,67 −3,16 0,01 1,16 5,17 6,17 7,18 9,10 8,18 7,16 6,17 5,15 3,14 2,18 1,18 0,99
0,10 1,15 2,17 3,72 −1,71 6,39 5,16 3,13 1,89 0,90 0,91 0,11 −0,20 1,89 2,84 3,92 2,99 2,21 1,73 ?
2,99 2,21 1,73 ?
1,78 2,56 1,88 0,57
2.89 2.12 1,78 2,67
2,99 2,21 1,73
3,14 2,18 1,18 0,99

1,34	1,78	2,56	1,88	**0,57**	−1,25	0,19	3,18	4,13	2,22	1,34	0,26	−1,90	−2,29	0,88	1,28	3,31	4,12	5,15	3,31	
2,27	2,89	2,12	1,78	**2,67**	−3,16	0,01	1,16	5,17	6,17	7,18	9,10	8,18	7,16	6,17	5,15	3,14	2,18	1,18	**0,99**	
0,10	1,15	2,17	3,72	−1,71	6,39	5,16	3,13	1,89	0,90	**0,91**	0,11	−0,20	1,89	2,84	3,92	2,99	2,21	1,73	**?**	
																2,99	2,21	1,73	**?**	




1,34 1,78 2,56 1,88 0,57 −1,25 0,19 3,18 4,13 2,22 1,34 0,26 −1,90 −2,29 0,88 1,28 3,31 4,12 5,15 3,31
2,27 2,89 2,12 1,78 2,67 −3,16 0,01 1,16 5,17 6,17 7,18 9,10 8,18 7,16 6,17 5,15 3,14 2,18 1,18 0,99
0,10 1,15 2,17 3,72 −1,71 6,39 5,16 3,13 1,89 0,90 0,91 0,11 −0,20 1,89 2,84 3,92 2,99 2,21 1,73 ?
2,99 2,21 1,73 ?
2,99 2,21 1,73
1,78 2,56 1,88 0,57
2.89 2.12 1,78 2,67
3,14 2,18 1,18 0,99
3,13 1,89 0,90 0,91
? = empirical quantile

Prediction and Aggregation

Prediction of one expert

$$h_n^{(k,\ell)}(y_1^{n-1}) = T_{\min\{n^\delta,\ell\}}\left(\operatorname*{argmin}_{q\in\mathbb{R}} \sum_{\{t\in J_n^{(k,\ell)}\}} \rho_\tau(y_t - q)\right).$$

[Can be easily computed by sorting the sample.]

Aggregated prediction of all experts

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Prediction and Aggregation

Prediction of one expert

$$h_n^{(k,\ell)}(y_1^{n-1}) = T_{\min\{n^\delta,\ell\}}\left(\operatorname*{argmin}_{q\in\mathbb{R}} \sum_{\{t\in J_n^{(k,\ell)}\}} \rho_\tau(y_t - q)\right).$$

[Can be easily computed by sorting the sample.]

Aggregated prediction of all experts

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Aggregation continued. . .

Definitions

- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) G_{n-1}(h^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Aggregation continued. . .

Definitions

- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) G_{n-1}(h^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Aggregation continued. . .

Definitions

- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) G_{n-1}(h^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Aggregation continued. . .

Definitions

- For $\eta_n > 0$, we define the weights

$$w_{k,\ell,n} = q_{k,\ell} e^{-\eta_n (n-1) G_{n-1}(h^{(k,\ell)})}.$$

- We normalize these weights:

$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j=1}^{\infty} w_{i,j,n}}.$$

Global prediction

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}).$$

Theoretical Results

Theorem

- *Let $\mathcal{C}$ be the class of all jointly stationary and ergodic processes $\{Y_n\}_{n=-\infty}^{\infty}$ such that $\mathbb{E}\{Y_0^2\} < \infty$ and $F_{Y_0|Y_{-\infty}^{-1}}$ is a.s. increasing.*
- *Then the nearest neighbor quantile forecasting strategy is universally consistent with respect to the class $\mathcal{C}$, that is, for all process $Y \in \mathcal{C}$*

$$\lim_{n\to\infty} G_n(g) = G^\star \quad \text{almost surely.}$$

Theoretical Results

Theorem

- *Let $\mathcal{C}$ be the class of all jointly stationary and ergodic processes $\{Y_n\}_{n=-\infty}^{\infty}$ such that $\mathbb{E}\{Y_0^2\} < \infty$ and $F_{Y_0|Y_{-\infty}^{-1}}$ is a.s. increasing.*
- *Then the nearest neighbor quantile forecasting strategy is universally consistent with respect to the class $\mathcal{C}$, that is, for all process $Y \in \mathcal{C}$*

$$\lim_{n\to\infty} G_n(g) = G^\star \quad \textit{almost surely.}$$

Mathematical demonstration

Difficulty

- We can not apply ergodic theorem on fundamental experts.
- Ergodicity provides weak convergence of random probability measure (almost surely).

Example

Let $J_n^{(k,\ell)}$ the set of the indices of the neighbors.
We have, almost surely in term of weak convergence,

$$\mathbb{P}_n^{(k,\ell)} \underset{n\to\infty}{\longrightarrow} \mathbb{P}_\infty^{(k,\ell)},$$

where $\mathbb{P}_n^{(k,\ell)} \triangleq \frac{1}{|J_n^{(k,\ell)}|}\sum_{i\in J_n^{(k,\ell)}} \delta_{Y_i}$.

Mathematical demonstration

Difficulty

- We can not apply ergodic theorem on fundamental experts.
- Ergodicity provides weak convergence of random probability measure (almost surely).

Example

Let $J_n^{(k,\ell)}$ the set of the indices of the neighbors.
We have, almost surely in term of weak convergence,

$$\mathbb{P}_n^{(k,\ell)} \underset{n\to\infty}{\longrightarrow} \mathbb{P}_\infty^{(k,\ell)},$$

where $\mathbb{P}_n^{(k,\ell)} \triangleq \frac{1}{|J_n^{(k,\ell)}|} \sum_{i\in J_n^{(k,\ell)}} \delta_{Y_i}$.

Mathematical demonstration

Lemma

Let $\{\mu_n\}_{n=1}^{\infty}$ be a uniformly integrable sequence of real probability measures, and let μ_∞ be a probability measure with (strictly) increasing distribution function. Suppose that $\{\mu_n\}_{n=1}^{\infty}$ converges weakly to μ_∞. Then, for all $\tau \in (0,1)$,

$$q_{\tau,n} \to q_{\tau,\infty} \quad \text{as } n \to \infty,$$

where $q_{\tau,n} \in \operatorname{argmin}_q \mathbb{E}_{\mu_n}[\rho_\tau(Y-q)]$ for all $n \geq 1$ and $\{q_{\tau,\infty}\} = \operatorname{argmin}_q \mathbb{E}_{\mu_\infty}[\rho_\tau(Y-q)]$.

Outline

Future outcome predictions results

$\tau = 0.5$ median base forecaster : robustness.

Model Name	Avg Abs Error	Avg Sqr Error	Mape (%)
$AR(7)$	65.80	9738	31.6
$QAR(8)_{0.5}$	57.8	9594	24.9
DayOfTheWeekMean	53.95	7099	22.8
HoltWinters	49.84	6025	**21.5**
$QuantileExpertMixture_{0.5}$	**48.1**	**5731**	21.6
MeanExpertMixture	52.37	6536	22.3
MA	179	62448	52.0

Figure: Forecasting future outcomes.

Quantile forecasting

Model Name	PinBall Loss (0.1)	Ramp Loss
$QuantileExpertMixture_{0.1}$	13.71	0.80
$QAR(7)_{0.1}$	**13.22**	**0.88**

Figure: $\tau = 0.1$

Model Name	PinBall Loss (0.9)	Ramp Loss
$QuantileExpertMixture_{0.9}$	**12.27**	**0.07**
$QAR(7)_{0.9}$	19.31	**0.07**

Figure: $\tau = 0.9$

Quantile forecasting

Model Name	PinBall Loss (0.1)	Ramp Loss
$QuantileExpertMixture_{0.1}$	13.71	0.80
$QAR(7)_{0.1}$	**13.22**	**0.88**

Figure: $\tau = 0.1$

Model Name	PinBall Loss (0.9)	Ramp Loss
$QuantileExpertMixture_{0.9}$	**12.27**	**0.07**
$QAR(7)_{0.9}$	19.31	**0.07**

Figure: $\tau = 0.9$

Extensions

Binary prediction

- Predict the next outcome $y_n \in \{0,1\}$ of a sequence of binary numbers $y_1, y_2, \ldots$
- We know the past sequence $y_1^{n-1} = (y_1, \ldots, y_{n-1})$.
- The whole theory carries over, with

$$H_n(g) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{[g_t(Y_1^{t-1}) \neq Y_t]}$$

and

$$g_n^*(Y_1^{n-1}) = \begin{cases} 1 & \text{if } \mathbb{P}\{Y_n = 1 | Y_1^{n-1}\} \geq 1/2 \\ 0 & \text{otherwise.} \end{cases}$$

Extensions

Binary prediction

- Predict the next outcome $y_n \in \{0, 1\}$ of a sequence of binary numbers $y_1, y_2, \ldots$
- We know the past sequence $y_1^{n-1} = (y_1, \ldots, y_{n-1})$.
- The whole theory carries over, with

$$H_n(g) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{[g_t(Y_1^{t-1}) \neq Y_t]}$$

and

$$g_n^*(Y_1^{n-1}) = \begin{cases} 1 & \text{if } \mathbb{P}\{Y_n = 1 | Y_1^{n-1}\} \geq 1/2 \\ 0 & \text{otherwise.} \end{cases}$$

Extensions

Binary prediction

- Predict the next outcome $y_n \in \{0, 1\}$ of a sequence of binary numbers $y_1, y_2, \ldots$
- We know the past sequence $y_1^{n-1} = (y_1, \ldots, y_{n-1})$.
- The whole theory carries over, with

$$H_n(g) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{[g_t(Y_1^{t-1}) \neq Y_t]}$$

and

$$g_n^*(Y_1^{n-1}) = \begin{cases} 1 & \text{if } \mathbb{P}\{Y_n = 1 | Y_1^{n-1}\} \geq 1/2 \\ 0 & \text{otherwise.} \end{cases}$$

Portfolio selection

Mathematical model

- A market with d assets.
- $\mathbf{y}_1, \mathbf{y}_2, \ldots \in \mathbb{R}_+^d$ represent the evolution of the market in time.
- The j-th component of $\mathbf{y}_n$ represents the amount obtained after investing a unit capital in the j-th asset, on the n-th training period.
- The investor is allowed to diversify his capital according to a portfolio vector $\mathbf{b}_n = (b_n^{(1)}, \ldots, b_n^{(d)})$.

Wealth

- S_0 is the investor initial capital and $\mathbf{b}_1 = (1/d, \ldots, 1/d)$.
- At the end of the first training period,

$$S_1 = S_0 \sum_{j=1}^{d} b_1^{(j)} y_1^{(j)} = S_0 \langle \mathbf{b}_1, \mathbf{y}_1 \rangle.$$

Portfolio selection

Mathematical model

- A market with d assets.
- $\mathbf{y}_1, \mathbf{y}_2, \ldots \in \mathbb{R}_+^d$ represent the evolution of the market in time.
- The j-th component of $\mathbf{y}_n$ represents the amount obtained after investing a unit capital in the j-th asset, on the n-th training period.
- The investor is allowed to diversify his capital according to a portfolio vector $\mathbf{b}_n = (b_n^{(1)}, \ldots, b_n^{(d)})$.

Wealth

- S_0 is the investor initial capital and $\mathbf{b}_1 = (1/d, \ldots, 1/d)$.
- At the end of the first training period,

$$S_1 = S_0 \sum_{j=1}^{d} b_1^{(j)} y_1^{(j)} = S_0 \langle \mathbf{b}_1, \mathbf{y}_1 \rangle.$$

Portfolio selection

Mathematical model

- A market with d assets.
- $\mathbf{y}_1, \mathbf{y}_2, \ldots \in \mathbb{R}_+^d$ represent the evolution of the market in time.
- The j-th component of $\mathbf{y}_n$ represents the amount obtained after investing a unit capital in the j-th asset, on the n-th training period.
- The investor is allowed to diversify his capital according to a portfolio vector $\mathbf{b}_n = (b_n^{(1)}, \ldots, b_n^{(d)})$.

Wealth

- S_0 is the investor initial capital and $\mathbf{b}_1 = (1/d, \ldots, 1/d)$.
- At the end of the first training period,

$$S_1 = S_0 \sum_{j=1}^{d} b_1^{(j)} y_1^{(j)} = S_0 \langle \mathbf{b}_1, \mathbf{y}_1 \rangle.$$

Portfolio selection

Mathematical model

- A market with d assets.
- $\mathbf{y}_1, \mathbf{y}_2, \ldots \in \mathbb{R}_+^d$ represent the evolution of the market in time.
- The j-th component of $\mathbf{y}_n$ represents the amount obtained after investing a unit capital in the j-th asset, on the n-th training period.
- The investor is allowed to diversify his capital according to a portfolio vector $\mathbf{b}_n = (b_n^{(1)}, \ldots, b_n^{(d)})$.

Wealth

- S_0 is the investor initial capital and $\mathbf{b}_1 = (1/d, \ldots, 1/d)$.
- At the end of the first training period,

$$S_1 = S_0 \sum_{j=1}^{d} b_1^{(j)} y_1^{(j)} = S_0 \langle \mathbf{b}_1, \mathbf{y}_1 \rangle.$$

Portfolio selection

Mathematical model

- A market with d assets.
- $\mathbf{y}_1, \mathbf{y}_2, \ldots \in \mathbb{R}_+^d$ represent the evolution of the market in time.
- The j-th component of $\mathbf{y}_n$ represents the amount obtained after investing a unit capital in the j-th asset, on the n-th training period.
- The investor is allowed to diversify his capital according to a portfolio vector $\mathbf{b}_n = (b_n^{(1)}, \ldots, b_n^{(d)})$.

Wealth

- S_0 is the investor initial capital and $\mathbf{b}_1 = (1/d, \ldots, 1/d)$.
- At the end of the first training period,

$$S_1 = S_0 \sum_{j=1}^{d} b_1^{(j)} y_1^{(j)} = S_0 \langle \mathbf{b}_1, \mathbf{y}_1 \rangle.$$

Portfolio selection

Mathematical model

- A market with d assets.
- $\mathbf{y}_1, \mathbf{y}_2, \ldots \in \mathbb{R}_+^d$ represent the evolution of the market in time.
- The j-th component of $\mathbf{y}_n$ represents the amount obtained after investing a unit capital in the j-th asset, on the n-th training period.
- The investor is allowed to diversify his capital according to a portfolio vector $\mathbf{b}_n = (b_n^{(1)}, \ldots, b_n^{(d)})$.

Wealth

- S_0 is the investor initial capital and $\mathbf{b}_1 = (1/d, \ldots, 1/d)$.
- At the end of the first training period,

$$S_1 = S_0 \sum_{j=1}^{d} b_1^{(j)} y_1^{(j)} = S_0 \langle \mathbf{b}_1, \mathbf{y}_1 \rangle.$$

Portfolio selection

Mathematical model

- A market with d assets.
- $\mathbf{y}_1, \mathbf{y}_2, \ldots \in \mathbb{R}_+^d$ represent the evolution of the market in time.
- The j-th component of $\mathbf{y}_n$ represents the amount obtained after investing a unit capital in the j-th asset, on the n-th training period.
- The investor is allowed to diversify his capital according to a portfolio vector $\mathbf{b}_n = (b_n^{(1)}, \ldots, b_n^{(d)})$.

Wealth

- S_0 is the investor initial capital and $\mathbf{b}_1 = (1/d, \ldots, 1/d)$.
- At the end of the first training period,

$$S_1 = S_0 \sum_{j=1}^{d} b_1^{(j)} y_1^{(j)} = S_0 \langle \mathbf{b}_1, \mathbf{y}_1 \rangle.$$

Wealth.

- By induction,

$$S_n = S_{n-1}\langle \mathbf{b}_n(\mathbf{y}_1^{n-1}), \mathbf{y}_n\rangle = S_0 \exp\left\{\sum_{t=1}^{n} \log\langle \mathbf{b}_t(\mathbf{y}_1^{t-1}), \mathbf{y}_t\rangle\right\}.$$

- Goal: find the best investment strategy $\{b_n\}_{n=1}^{\infty}$ to maximize the wealth S_n.
- Equivalent to maximize the criterion $W_n(\mathbf{b}) = \frac{1}{n}\sum_{t=1}^{n} \log\langle \mathbf{b}_t(\mathbf{y}_1^{t-1}), \mathbf{y}_t\rangle$.
- Strategies so far: quantization, kernel and nearest neighbor on-line prediction with expert aggregation.

- References: Györfi and Schäfer (2003), Györfi et al. (2006, 2008).

Wealth.

- By induction,

$$S_n = S_{n-1} \langle \mathbf{b}_n(\mathbf{y}_1^{n-1}), \mathbf{y}_n \rangle = S_0 \exp \left\{ \sum_{t=1}^{n} \log \langle \mathbf{b}_t(\mathbf{y}_1^{t-1}), \mathbf{y}_t \rangle \right\}.$$

- Goal: find the best investment strategy $\{b_n\}_{n=1}^{\infty}$ to maximize the wealth S_n.
- Equivalent to maximize the criterion $W_n(\mathbf{b}) = \frac{1}{n} \sum_{t=1}^{n} \log \langle \mathbf{b}_t(\mathbf{y}_1^{t-1}), \mathbf{y}_t \rangle$.
- Strategies so far: quantization, kernel and nearest neighbor on-line prediction with expert aggregation.

- References: Györfi and Schäfer (2003), Györfi et al. (2006, 2008).

Wealth.

- By induction,

$$S_n = S_{n-1} \langle \mathbf{b}_n(\mathbf{y}_1^{n-1}), \mathbf{y}_n \rangle = S_0 \exp \left\{ \sum_{t=1}^{n} \log \langle \mathbf{b}_t(\mathbf{y}_1^{t-1}), \mathbf{y}_t \rangle \right\}.$$

- Goal: find the best investment strategy $\{b_n\}_{n=1}^{\infty}$ to maximize the wealth S_n.
- Equivalent to maximize the criterion $W_n(\mathbf{b}) = \frac{1}{n} \sum_{t=1}^{n} \log \langle \mathbf{b}_t(\mathbf{y}_1^{t-1}), \mathbf{y}_t \rangle$.
- Strategies so far: quantization, kernel and nearest neighbor on-line prediction with expert aggregation.

- References: Györfi and Schäfer (2003), Györfi et al. (2006, 2008).

Wealth.

- By induction,

$$S_n = S_{n-1}\langle \mathbf{b}_n(\mathbf{y}_1^{n-1}), \mathbf{y}_n\rangle = S_0 \exp\left\{\sum_{t=1}^{n} \log\langle \mathbf{b}_t(\mathbf{y}_1^{t-1}), \mathbf{y}_t\rangle\right\}.$$

- Goal: find the best investment strategy $\{b_n\}_{n=1}^{\infty}$ to maximize the wealth S_n.
- Equivalent to maximize the criterion $W_n(\mathbf{b}) = \frac{1}{n}\sum_{t=1}^{n} \log\langle \mathbf{b}_t(\mathbf{y}_1^{t-1}), \mathbf{y}_t\rangle$.
- Strategies so far: quantization, kernel and nearest neighbor on-line prediction with expert aggregation.
- References: Györfi and Schäfer (2003), Györfi et al. (2006, 2008).

- Algoet, P. The strong law of large numbers for sequential decisions under uncertainty, *IEEE Trans. Inform. Theory*, **40**, 609-633, 1994.
- Biau, G., Bleakley, K., Györfi, L. and Ottucsák, G. Non-parametric sequential prediction of time series, *J. Nonparametr. Stat.*, 2009, in press.
- Cesa-Bianchi, N. and Lugosi, G. *Prediction, Learning, and Games*, Cambridge University Press, New York, 2006.
- Györfi, L., Udina, F. and Walk, H. Nonparametric nearest neighbor based empirical portfolio selection strategies, *Statist. Decis.*, **26**, 145-157, 2008.
- Koenker, R. *Quantile Regression*, Cambridge University Press, Cambridge, 2005.

Thank you for your attention.