Anticipated and adaptive prediction in functional discriminant analysis

Cristian PREDA

Polytech'Lille, USTL Lille, France

Gilbert SAPORTA

CNAM Paris, France

Mohamed Hedi BEN MBAREK

Institut Supérieur de Gestion, Sousse, Tunisia

The context

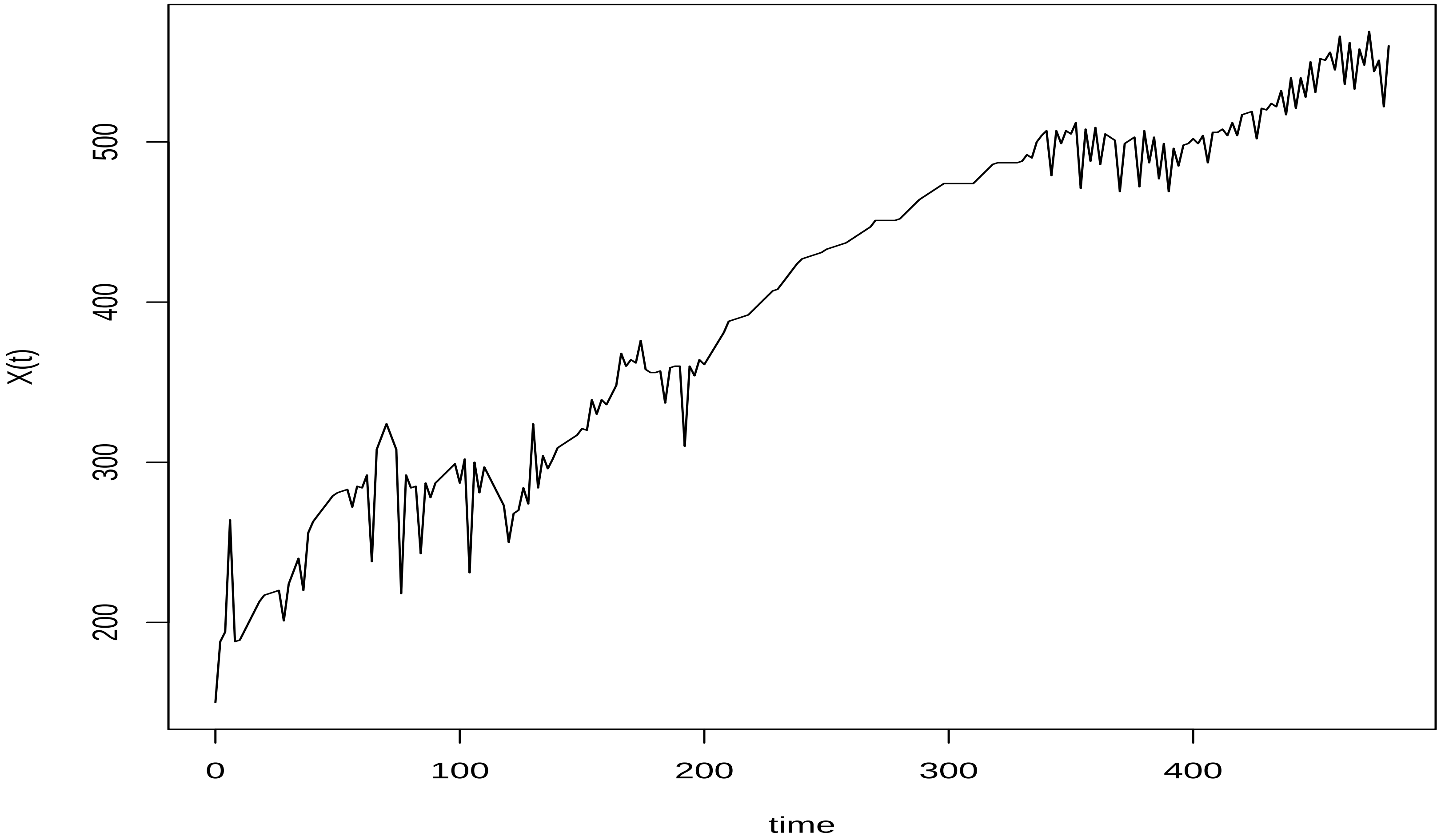
Cookie's quality at Danone and the kneading process

A cookie from Danone :

- choose a type of flour (components, density, etc)
- kneading process ($\approx$ 1h)
- **- Put the dough in form and cook it**
- Evaluate quality of the obtained cookies $\Big\}$ $>$ 2h.

Idea : predict the cookie's quality from elements derived from the kneading process : Danone gets time and money !

Dough resistance during the kneading process : $X = X(t)$**,** $t \in \{0, 2, 4, \ldots 480s\}$



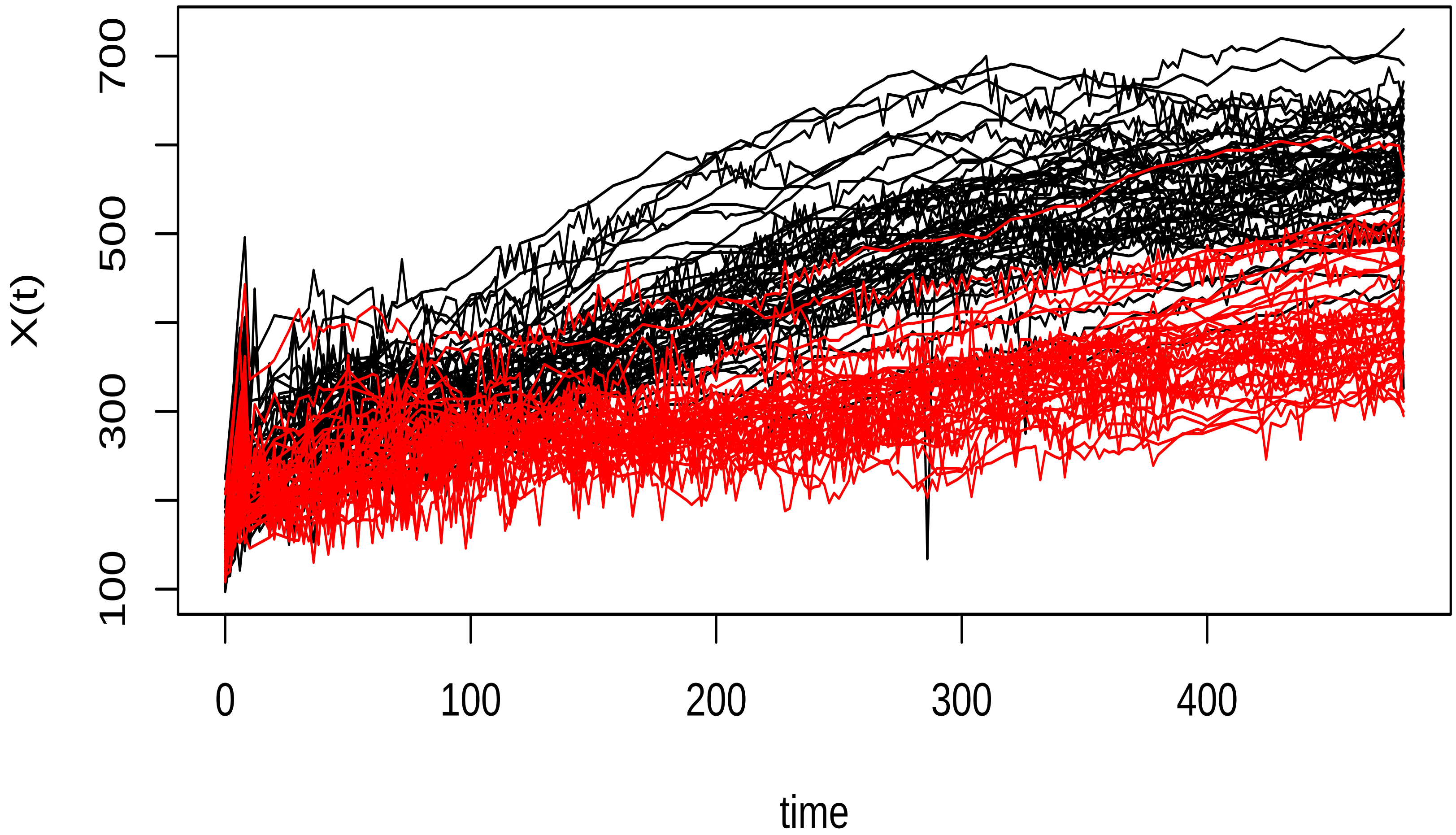
Flour evaluation

The quality of cookies obtained with some type of flour is given by experts.

The response (Y) is : this type of flour is **Good** or **Bad**.

90 flours were evaluated : 50 are good and 40 are bad.

90 flours : quality and dough resistance during 480s.



Good(black)   Bad(red)

Functional discriminant analysis

$$X = \{X_t\}_{t\in[0,T]}, \quad X_t : \mathbf{\Omega} \to \mathbb{R},$$

- $E(X_t^2) < \infty$,
- L_2–continuous,
- $\forall\omega \in \mathbf{\Omega} : (X_t(\omega))_{t\in[0,T]} \in L_2([0,T])$,
- $E(X_t) = 0, \forall t \in [0,T]$.

$$Y : \mathbf{\Omega} \to \{0,1\}.$$

Discriminant score :

$$d_T = \Phi(X)$$

Discriminant score estimation

$$\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$$

- Linear discriminant score :

$$\Phi(X) = \langle \beta, X \rangle_{L_2[0,T]}, \quad \beta \in L_2[0,T].$$

Criterion (Fisher):

$$\max_{\beta \in L_2[0,T]} \frac{\mathbb{V}(\mathbb{E}(\Phi(X)|Y))}{\mathbb{V}(\Phi(X))}$$

Estimation by functional linear regression model (PCR, PLS, etc).

- Nonparametric estimation

$$\hat{\Phi}(X) = \frac{\sum_{i=1}^{n} Y_i K(u(X, X_i)/h)}{\sum_{i=1}^{n} K(u(X, X_i)/h)}$$

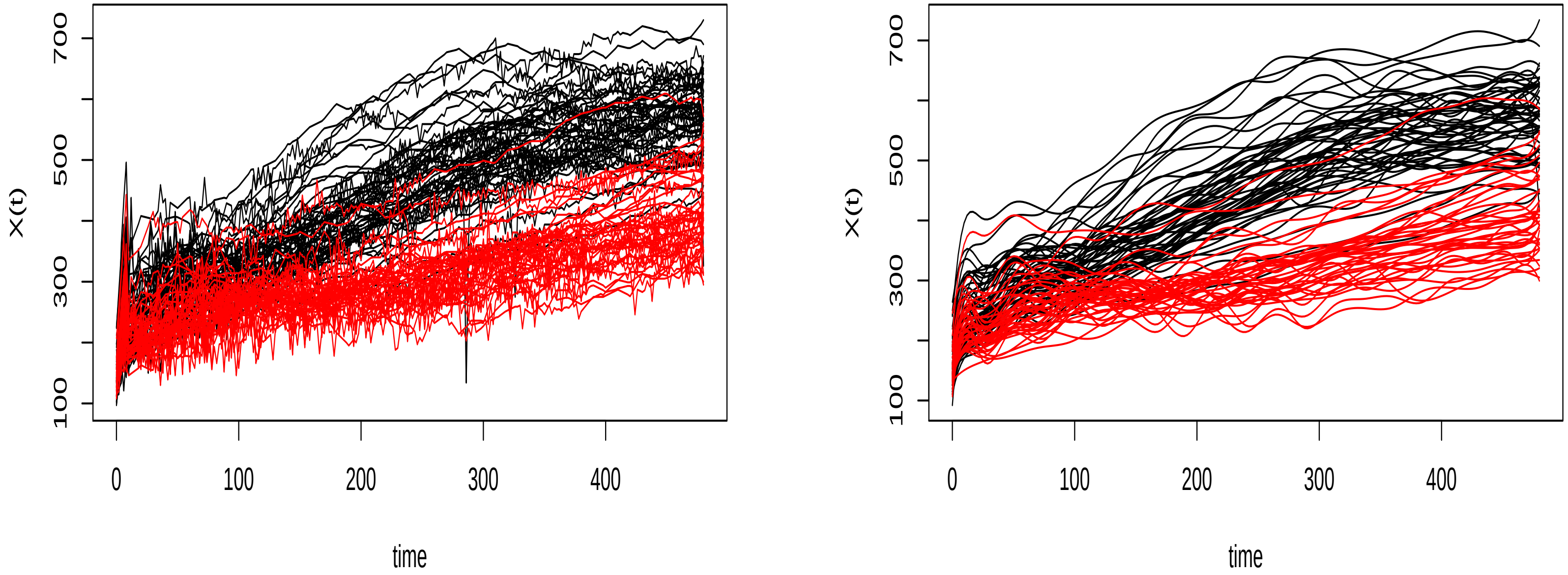
- RKHS approximation

$$\hat{\Phi}(X) = \sum_{i}^{n} \alpha_i K(X_i, X)$$

Criterion (logistic loss) :

$$L(x, y, \Phi) = -y\Phi(x) + \log(1 + e^{\Phi(x)})$$

Kneading data results



Good (black) and bad (red) flours. Left : original data. Right : smoothed data

Model	PLS_FLDA	NP	PC_FLDA	Gaussian(6)	LDA
Error rate	0.112	0.103	0.142	0.108	0.154

Error rate averaged over 100 test samples.

Anticipated prediction

X is observed on $[0, T]$

Problem : *find the smallest* T^*, $T^* < T$, *such that the prediction of* Y *by* X *observed on* $[0, T^*]$ *is "similar" to the prediction obtained with* X *observed on* $[0, T]$.

Discrimination power : ROC curve

– d : the discriminant score.

– threshold r : $Y = 1$ if $d > r$.

– "Sensitivity" : $P(d > r|Y = 1)$

– "Specificity" : $P(d > r|Y = 0)$.

Measure of discrimination : area under the ROC curve (AUC)

– Estimation of AUC

$\{Y=1\}: X_1 = d_{|Y=1}$, sample of size n_1

$\{Y=0\}: X_0 = d_{|Y=0}$, sample of size n_2

$$\widehat{AUC} = \frac{\#\{X_1 > X_0\}}{n_1 n_2}$$

$D = \{d_t\}_{0<t\leq T}, \ \ \{\widehat{AUC}(t)\}_{0<t\leq T}$

Criterion for T^* : compare $AUC(t)$ and $AUC(T)$ for $t < T$ and chose T^* as the largest t such that the test

$$H_0: \quad AUC(t) = AUC(T),$$

$$H_1: \quad AUC(t) < AUC(T)$$

is significant (p-value < 0.05).

Simulation

$$\text{Class } \{Y = 0\} : X_t = \begin{cases} W(1-t), & 0 \le t \le 1 \\ -2\sin(t-1) + W(t-1), & 1 < t \le 2 \end{cases}$$

$$\text{Class } \{Y = 1\} : X_t = \begin{cases} W(1-t), & 0 \le t \le 1 \\ 2\sin(t-1) + W(t-1), & 1 < t \le 2 \end{cases}$$

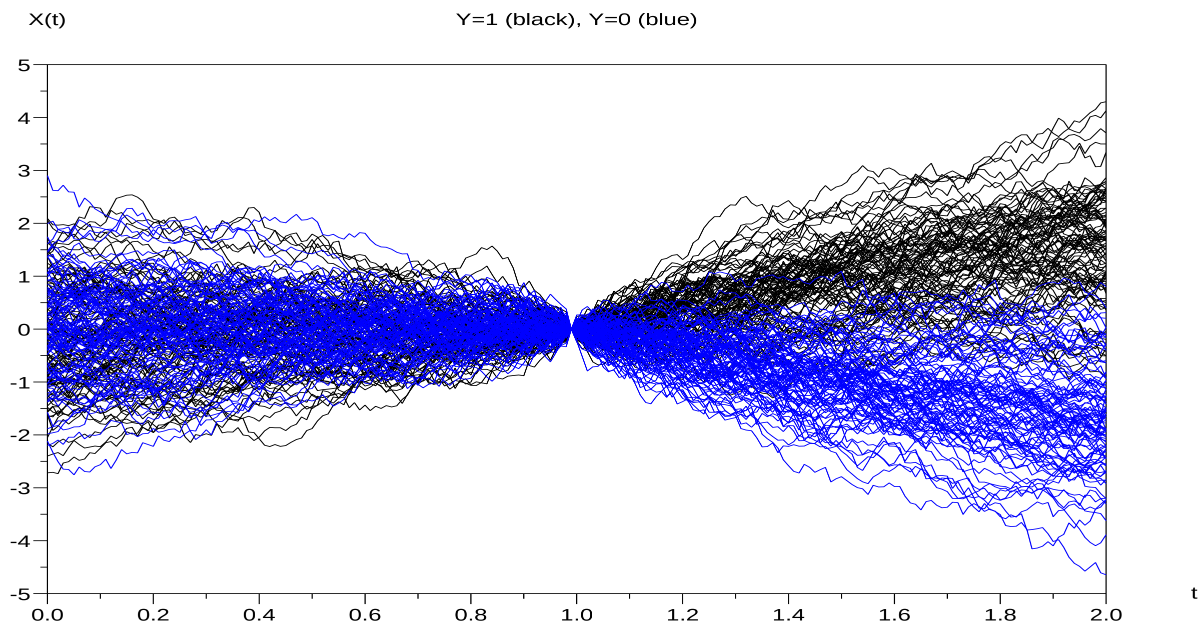


Figure : Sample of size $n = 100$ for each class of Y.

$M = 50$ learning and test samples of size 100.

For each $t \in \{0, 2.00, 1.98, \ldots, 0\}$: sample of size $M = 50$ of $\widehat{AUC}(t)$.

– $T^* = 1.46$.

– test statistic : $S = 1.663$

$\overline{\widehat{AUC}(t^*)} = 0.856$, $\overline{\widehat{AUC}(T)} = 0.872$

Kneading data

$Y \in \{Bad, Good\}$. The sample of 90 flours is randomly divided into a learning sample of size 60 and a test sample of size 30.

Error test rate (PLS estimation) $T = 480 : 0.112$,
– $\overline{\widehat{AUC}}(T) = 0.746$.

Anticipated prediction : $T^* = 186$.
Error test rate (PLS estimation) $T^* = 186 : 0.121$,
– $\overline{\widehat{AUC}}(T^*) = 0.778$.

Conclusion : *the predictive power of the dough curves for the cookies quality is resumed by the first* 186 *seconds of the kneading process.*

Adaptive prediction

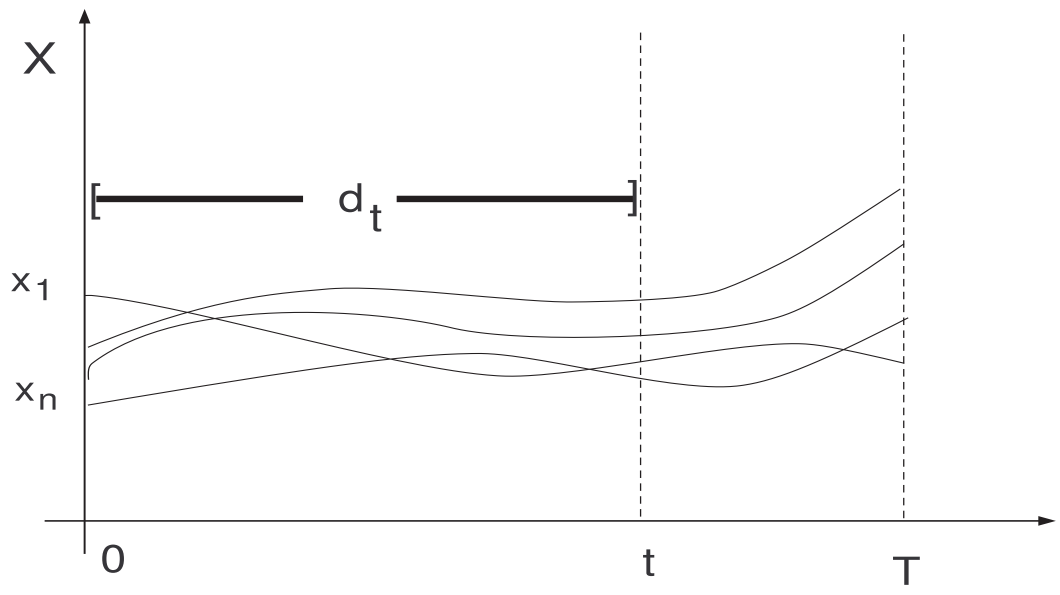
Remark : in anticipated prediction T^* is a constant.

Let $\omega \in \Omega$ be a new observation for which one wants to predict Y from X. Suppose that X is observed in a sequential way. The problem addressed by the adaptive prediction is :

Problem : *find the smallest* $T^*(\omega)$, $T^*(\omega) < T$, *such that* X *observed on* $[0, T^*(\omega)]$ *provides similar prediction as it is observed on* $[0, T]$.

Remark :
- here T^* is a random variable.
- to observe $X(\omega)$ on $[T^*, T]$ will not change the prediction for $Y(\omega)$ obtained with X on $[0, T^*(\omega)]$.

Conservative index for prediction :

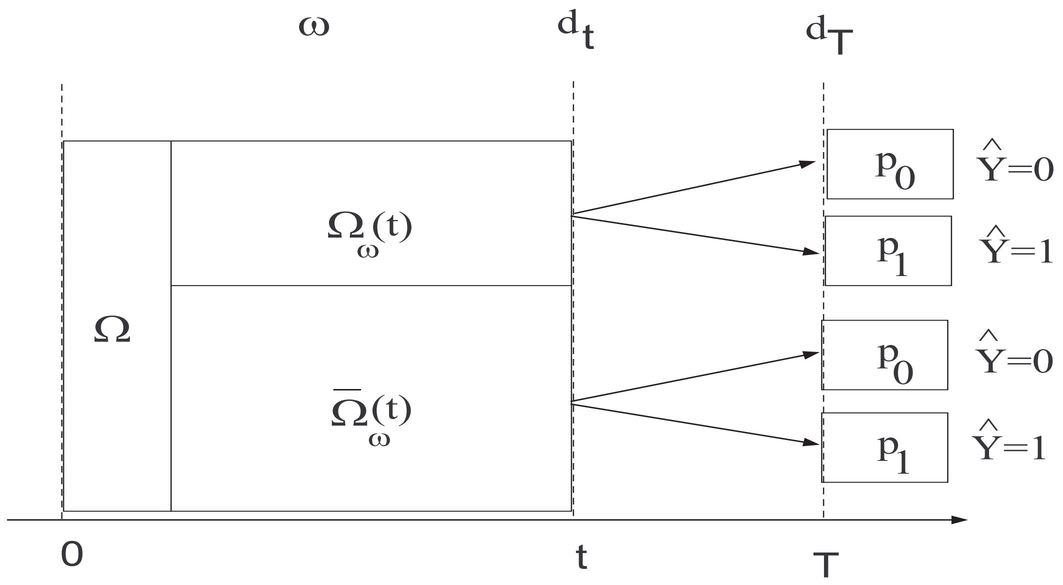


The discriminant score d_t.

Denote by

$$\Omega_\omega(t) = \{\omega_i \in \Omega | \hat{Y}_t(\omega) = \hat{Y}_{t,i}\} \text{ and } \overline{\Omega}_\omega(t) = \Omega - \Omega_\omega(t)$$

the class of elements having the same prediction as ω, respectively its complement with respect to Ω.



Conservation rate of the prediction for ω and t.

Let

$$p_{0|\Omega_\omega(t)} = \frac{\left|\{\omega' \in \Omega | \hat{Y}_T(\omega') = 0\} \cap \Omega_\omega(t)\}\right|}{|\Omega_\omega(t)|}.$$

be the observed rate of elements in $\Omega_\omega(t)$ predicted in the class $Y = 0$ at the time T using the score d_T. Similarly, let $p_{1|\Omega_\omega(t)}$, $p_{0|\overline{\Omega}_\omega(t)}$and $p_{1|\overline{\Omega}_\omega(t)}$
Let define by $C_{\Omega_\omega(t)} = \max\{p_{0|\Omega_\omega(t)}, p_{1|\Omega_\omega(t)}\}$, respectively by $C_{\overline{\Omega}_\omega(t)} = \max\{p_{0|\overline{\Omega_\omega}(t)}, p_{1|\overline{\Omega}_\omega(t)}\}$ the *conservation* rate of the prediction at the time t with respect to the time T for the elements of $\Omega_\omega(t)$, respectively of $\overline{\Omega}_\omega(t)$.
As a global measure of conservation we consider

$$C_\Omega(\omega, t) = \min\{C_{\Omega_\omega(t)}, C_{\overline{\Omega}_\omega(t)}\}.$$

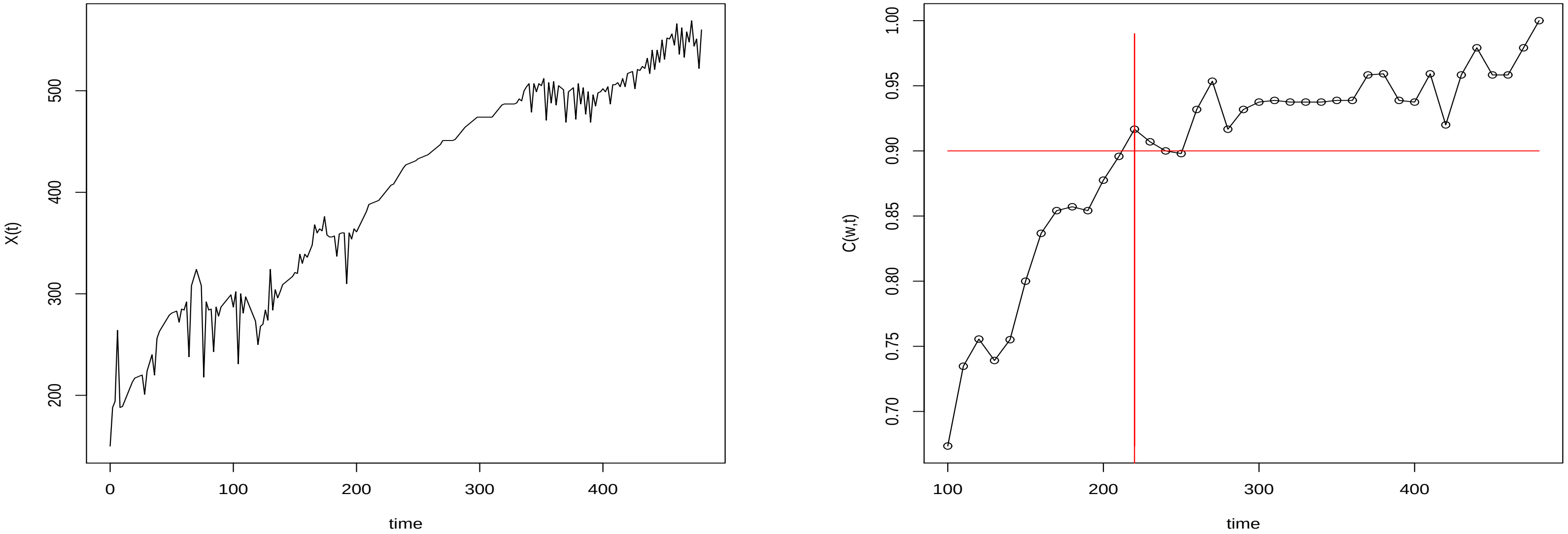
Remark : For each $t \in [0, T]$, $C_\Omega(\omega, t)$ is such that $0.5 \leq C_\Omega(\omega, t) \leq 1$ and $C_\Omega(\omega, T) \ = \ 1.$

Given a confidence conservation threshold $\gamma \in (0, 1)$, e.g. $\gamma = 0.90$, we define the following *adaptive prediction rule* for ω and t :

(1) if $C_\Omega(\omega, t) \geq \gamma$ then the observation of X for ω on the time interval $[0, t]$ is sufficient for the prediction of $Y(\omega)$. $\hat{Y}(\omega)$ is then the same as the prediction at time T of the subgroup of $\Omega_\omega(t)$ corresponding to $C_{\Omega_\omega(t)}$.

(2) if $C_\Omega(\omega, t) < \gamma$ then the observation process of X for ω should continue after t. Put $t = t + h$ and repeat the adaptive prediction procedure.

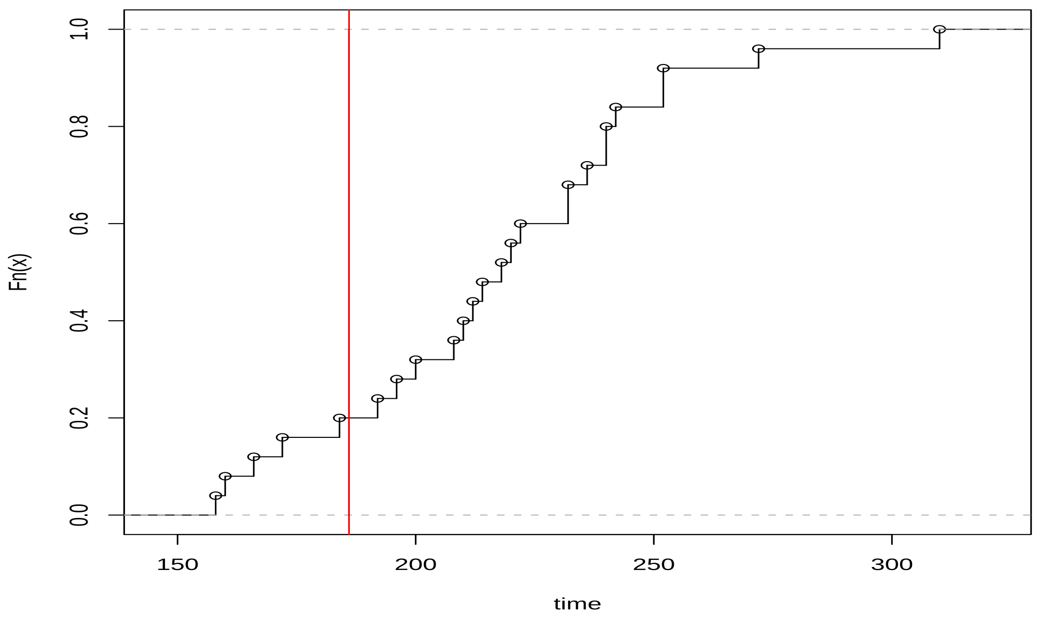
Then, $T^*(\omega)$ is the smallest t such that the condition (1) of the adaptive prediction rule is satisfied.



Left : new flour ω.   Right : $C_{\Omega}(\omega, t)$, $t \in [100, 480]$, $\gamma = 0.90$.

$$T^*(\omega) = 220.$$

For 25 new flours, the adaptive procedure is applied.



Empirical cumulative distribution function of T^* (in red, the time point t=186).