Statistical modelling of functional data

Philippe BESSE[‡], Hervé CARDOT[†,‡], Robert FAIVRE[†], Michel GOULARD[†]

‡ Lab. Statistique et Probabilités, Université Paul Sabatier, Toulouse

† Unité de Biométrie et Intelligence Artificielle INRA - Toulouse

Functional Data

Deville (1974), Besse & Ramsay (1986), Ramsay & Silverman (1997), ...

An observation is a function (growth curve, temperature curve, ...)

Consider a continuous time process $X(\omega, t), t \in T$ and the associated random function $X(\omega)$ supposed to take values in $H = L^2[T]$.
(more generally H can be separable Hilbert space).

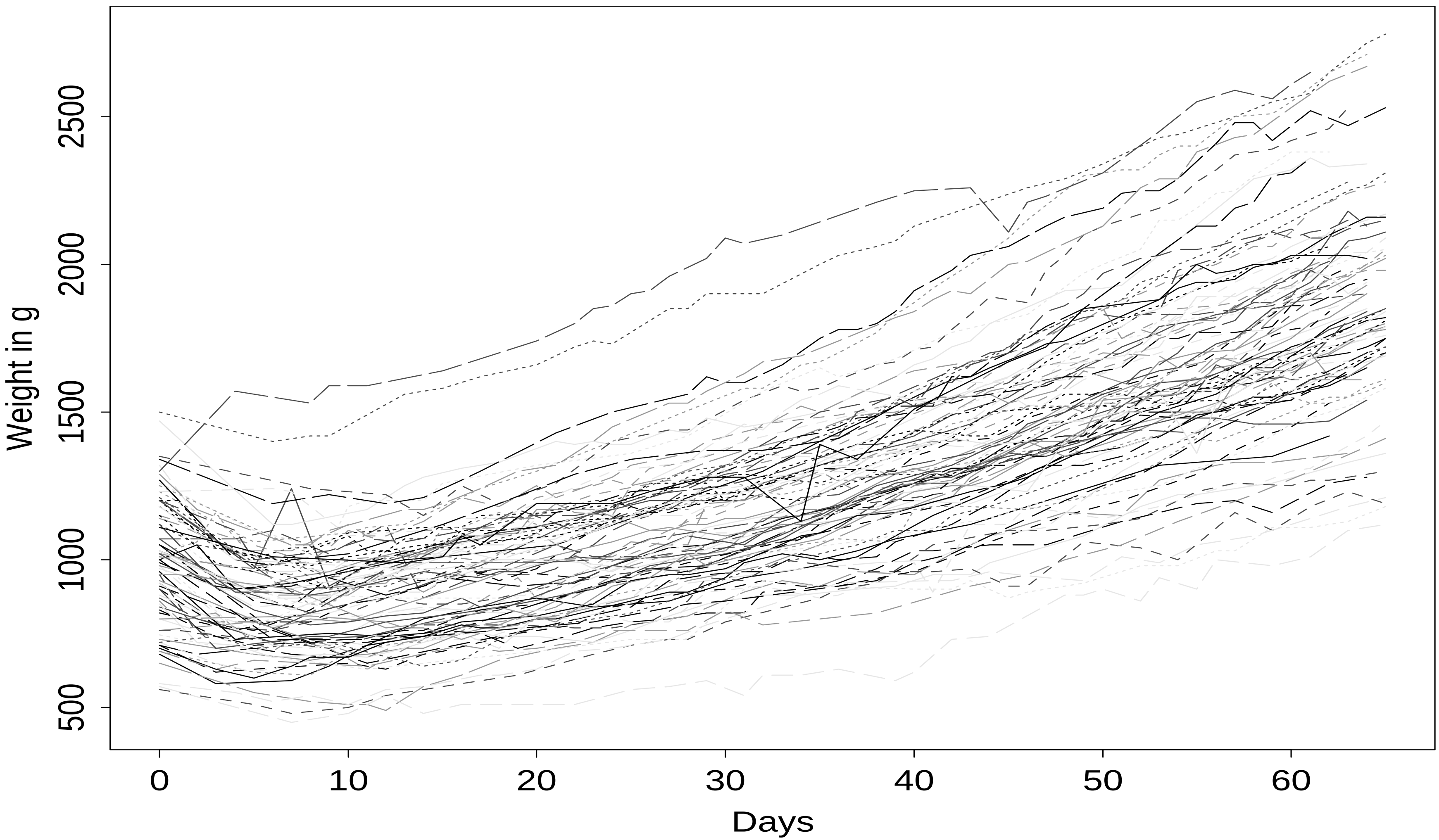
If $\mathbb{E}||X||^2 < \infty$, we can define the expectation $\mu(t) = \mathbb{E}(X(t))$ and the covariance operator $\Gamma : \ H \mapsto H$,

$$\Gamma f(s) = \int_T \gamma(s,t) f(t) \, dt, \quad f \in H$$

where $\gamma(s,t) = \text{cov}(X(s), X(t))$.

Γ is nonnegative and nuclear (and thus compact).

Growth curves



PCA for functional Data

The first studies on that issue were dealing with the statistical description (representation in small dimension space), extending Principal Components Analysis to function spaces.

We would like to approximate

$$X(t,\omega) \quad \approx \quad X_q(\omega,t) \; = \; \mu(t) + \sum_{j=1}^{q} \eta_j(\omega) v_j(t)$$

If v_j is the jth eigenfunction of Γ associated to the jth largest eigenvalue λ_j, we have, with $\eta_j(\omega) =< X, v_j >$,

$$\mathbb{E}\|X - X_q\|^2 = \sum_{j=q+1}^{\infty} \lambda_j .$$

Empirical Orthogonal Functions, Obhukov (1960),
Karhunen-Loeve expansion, Karhunen (1947), Loeve (1945).

Smoothing and Penalization

- From an “aesthetic” point of view: smooth representation can lead to better interpretation (Ramsay, Besse, ...)
- From a mathematical point of view: improves accuracy (Pezzulli & Silverman 1993, Silverman 1996, Cardot 2000, ...).

Suppose we observe a sample of curves $X_1(t), \ldots, X_n(t)$, $t \in T$.
$\mu_n(t)$ the empirical mean and Γ_n the empirical covariance operator.

Let $J(.) : \ S \subset H \mapsto R^+$ be a penalty operator (such as $J(\alpha) = \|\alpha^{(2)}\|^2$).

$\sup\limits_{v \in S}$	$< \Gamma_n v, v > -\ell J(v)$	Rice & Silverman (1991)
$\sup\limits_{v \in S}$	$\dfrac{< \Gamma_n v, v >}{\|v\|^2 + \ell J(v)}$	Silverman (1996)
$\min\limits_{z_i \in H_q}$	$\dfrac{1}{n} \sum\limits_{i=1}^{n} \left(\|X_i - z_i\|^2 + \ell J(z_i) \right)$	Besse *et al.* (1997)

Models with functional covariates

- Linear Model, (Ramsay & Dalzell 1993, Goutis 1998, Cardot *et al.* 1999)
  a sample $(Y_i, X_i)_{i=1,\dots,n}$, where Y_i is real and X_i is functional

$$Y_i = \mu + \int_T \beta(t) X_i(t)\ dt + \epsilon_i$$

- Functional Autoregressive Process (Bosq 1991, Bosq 2000),

$$X_i(t) = \mu(t) + \int_T \rho(s,t) X_{i-1}(s)\ ds + \epsilon_i(t)$$

- GLM for functional data (Marx & Eilers 1999, Cardot *et al.* 2000, ...,

$$Y_i | X_i = x \quad \sim \quad P(<\beta, x>)$$

Estimation procedures

Covariance matrices are compact operators in an infinite dimension space and so estimation is an ill-posed problem.
In the linear model, the score equation

$$\frac{1}{n}\sum_{i=1}^{n} < X_i, \widehat{\beta} > X_i \quad = \quad \frac{1}{n}\sum_{i=1}^{n} Y_i X_i.$$

has an infinite number of solutions.

Regularization is needed

- by a dimension reduction approach
- by adding a penalty

Dimension reduction

Denoting by $\widehat{v}_1, \ldots, \widehat{v}_q$ the orthonormal eigenfunctions of the empirical covariance operator associated to the eigenvalues $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_q$, we get a solution in the function space $\widehat{V}_q$ span by $\widehat{v}_1, \ldots, \widehat{v}_q$,

$$\widehat{\beta}_q = \sum_{j=1}^{q} \frac{< \Delta_n, \widehat{v}_j >}{\widehat{\lambda}_j} \, \widehat{v}_j,$$

where $\Delta_n = \frac{1}{n} \sum_{i=1}^{n} Y_i X_i$.
Smoothing steps can also be combined with this approach and generally give better results on real data sets (Besse and Cardot 1996, Besse *et al.* 2000).

Asymptotic properties of such estimators have been derived (Bosq, 1991, Bosq 2000, Cardot *et al.* 1999, ...):

q must not tend too rapidly to infinity as the sample size n increases.

Penalty approach

Adding a penalty allows to get a stable and a unique solution (Leurgans *et al.* 1993, Marx and Eilers 1999, Mas 1999, ...).

For the linear model we are seeking for $\widehat{\beta}_\ell$ satisfying

$$\frac{1}{n}\sum_{i=1}^{n} < X_i, \widehat{\beta} > X_i + \ell \, \nabla J(\widehat{\beta}) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} Y_i X_i,$$

ℓ being a tuning parameter that controls the trade off between the fidelity to the data and the “regularity” of the solution.
We have

$$\widehat{\beta}_\ell \quad = \quad (\Gamma_n + \ell \, \nabla J)^{-1} \, \Delta_n .$$

The solution is generally expanded in a basis of B-splines or Fourier series.

A more general framework

More generally, we can consider a loss function $\mathcal{L}$ such as a least squares criterion or the opposite of the log-likelihood.

The regularization approach consists in minimizing

$$\min_{\beta \in S} \quad \mathcal{L}(Y_1, \ldots, Y_n, X_1, \ldots, X_n, \beta) + \ell J(\beta)$$

in a space $S \subset H$.

The dimension reduction approach leads to find the optimum of

$$\min_{\beta \in \widehat{S}_q} \quad \mathcal{L}(Y_1, \ldots, Y_n, \widehat{\Pi}_q X_1, \ldots, \widehat{\Pi}_q X_n, \beta)$$

where $\widehat{S}_q$ is a q-dimensional function space and $\widehat{\Pi}_q$ is a projector onto $\widehat{S}_q$.

The values of the tuning parameters q and ℓ are chosen with criterions based on cross-validation.

Remarks on the convergence

For both approaches some consistency results have been obtained.
In general rates of convergence are rather slow compared to parametric and usual nonparametric ones.

For instance, in the context of Generalized Linear Models, we get the bound (Cardot & Sarda 2002), for $\ell \geq n^{-(1-\delta)/2}, \delta > 0$

$$\mathbb{E}\left(< \beta - \widehat{\beta}_\ell, X >^2\right) \quad = \quad O\left(n^{-2p/(4p+1)}\right),$$

$\beta \in C^p[T]$

Are these rates the best ones ?

What is the influence of the eigenvalues ?

Remote sensing application

Satellite SPOT4 with sensor Végétation launched in March 1998

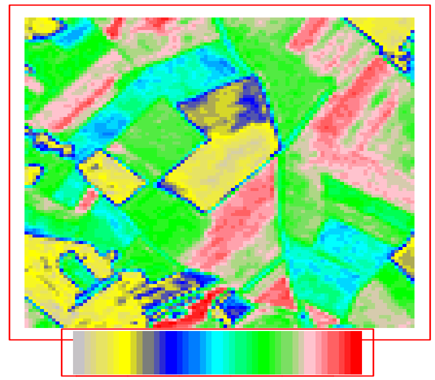
Gives **daily** measures of relected radiations for 4 wavelengths at a **low spatial resolution**, a pixel representing an area of 1km^2.

The size of plots being much less than 1 km^2, a pixel corresponds to different crops (maize, wheat, forest, ...), and the observed reflectances are a mixture of “pure” reflectances.

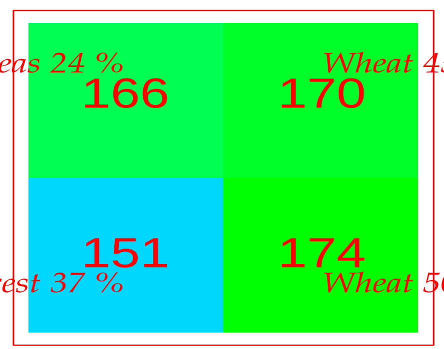
$$\left.\begin{array}{l}\text{Land use}\\ \text{Digital Elevation Model}\\ \text{Agronomic growth models}\\ \text{Weather data}\end{array}\right\} \Rightarrow \begin{array}{l}\text{Crops yield prediction}\\ \text{at a regional scale}\end{array}$$

Scale change and mixed pixels

Résolution 20 m x 20 m



Resolution 1 km x 1 km



The data

$\pi_{ij},\ j = 1, \ldots, p,$ is the proportion of land use of crop j in pixel i of 1 km^2.
In our application, the observed area is about 40km $\times 40$km,
we have $n = 1554$ pixels at each date.
Ten ($p = 10$) different classes of crops were present
(winter crops 49 %, forest 13.5 %, rapeseed 12 %, urban 6 %, ...)

The curves of reflectance of a pixel i is denoted by

$$\boldsymbol{X_i} = [X_i(t_1), \ldots, X_i(t_K)]^T$$

where $t_1 < \ldots < t_k < \ldots < t_K$ are the instants of measure.

The images in which the clouds were too important were removed.

We finally get $K = 32$ different images from March to August 1998.

The multilogit model

We suppose that the proportions π_{ij} given the temporal evolution of the reflectance $\{X_i(t), t \in T\}$ can be modelled as resulting from a multinomial distribution whose parameters satisfy

$$\mathbb{E}(\pi_{ij}|X_i) \quad = \quad \frac{\exp\left(\delta_j + \displaystyle\int_T \beta_j(t)X_i(t)\ dt\right)}{\displaystyle\sum_{\ell=1}^{p} \exp\left(\delta_\ell + \displaystyle\int_T \beta_\ell(t)X_i(t)\ dt\right)}\ .$$

For identifiability reasons we took $\beta_p = 0$ and $\delta_p = 0$.

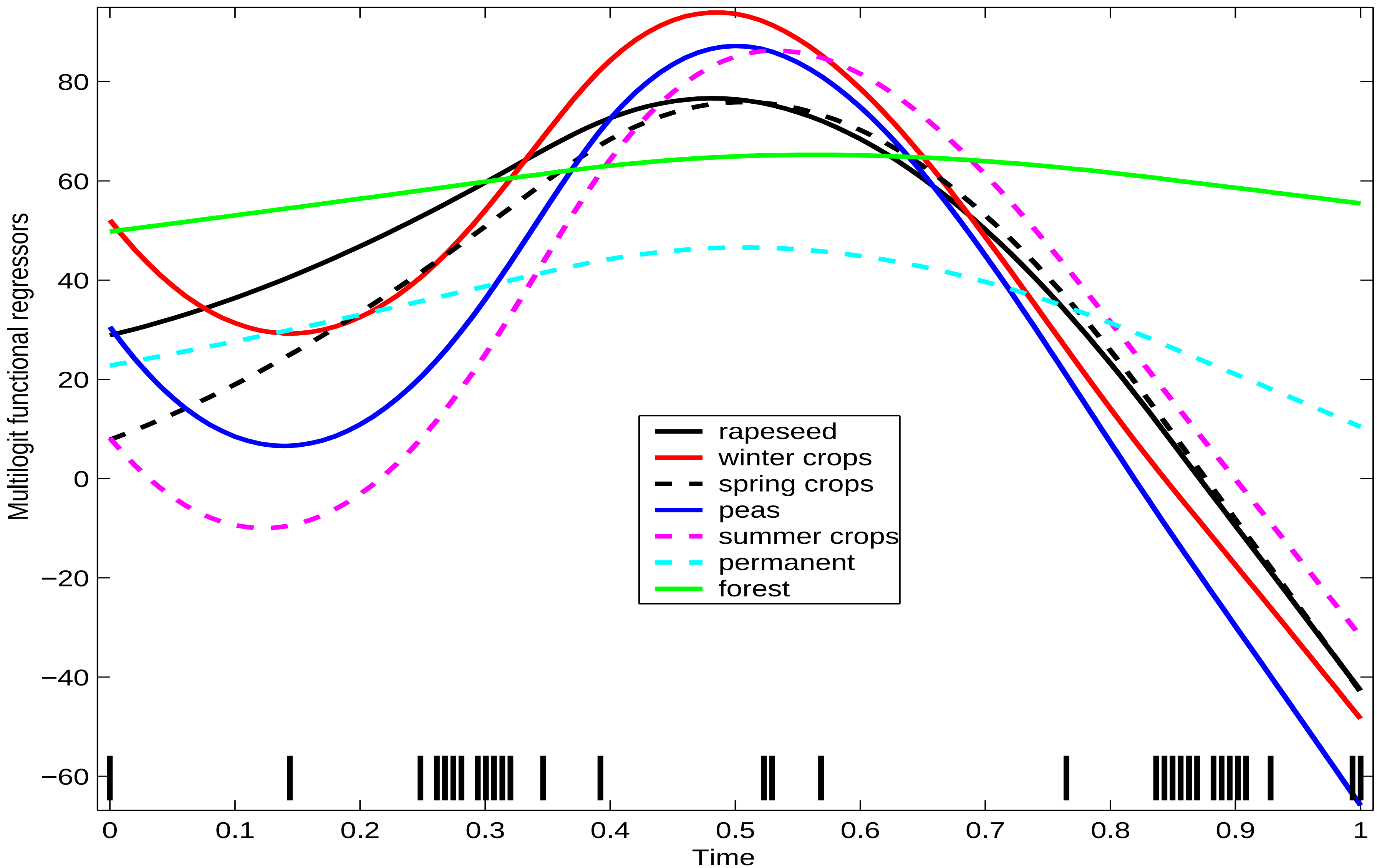
We aim at estimating the vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{p-1})^T$ and the functional coefficients $\beta_j(t),\ j = 1, \ldots, p-1$.

- A learning sample composed of 1055 pixels.
- A test sample composed of 499 pixels.

For computational purposes, we prefered the dimension reduction approach based on a functional principal components analysis.

The number of covariates (the principal components) still may be large.

We decided to select the most significative parameters by means of the likelihood ratio test with an ascendent procedure.



Multilogit functional regressors
80
60
40
20
0
−20
−40
−60
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Time
rapeseed
winter crops
spring crops
peas
summer crops
permanent
forest

Results

Themes	NDVI	PVI	Blue	Red	NIR	SWIR	M_0
urban	0.49	**0.36**	0.47	0.54	0.41	0.51	0.86
water	0.43	**0.29**	0.78	0.62	0.61	0.31	1.30
rapeseed	0.48	0.46	**0.45**	0.50	0.47	0.47	0.59
winter crops	0.20	0.21	**0.19**	0.20	0.22	**0.19**	0.30
spring crops	0.58	**0.56**	0.60	0.61	0.65	0.61	0.69
peas	0.50	**0.43**	0.45	**0.43**	0.48	0.46	0.63
summer crops	0.61	0.68	0.61	0.60	0.76	**0.53**	0.88
permanent crops	0.47	**0.46**	0.52	0.49	**0.46**	0.50	0.61
forest	0.34	0.36	0.34	**0.31**	0.45	0.35	0.98
potatoes	0.90	0.93	0.94	0.90	1.06	**0.85**	1.31

The criterion is $C_{ij} = \frac{|\pi_{ij} - \hat{\pi}_{ij}|}{\pi_{.j}}$ and we take the median in the test sample.

Model M_0 consists in predicting by the mean value of the learning sample.

Statistical Learning for functional data

Many links with Statistical Learning: problems with high dimension, regularization techniques, common applications, ...

A large community (hundreds of papers published), particularly in France.

In Toulouse, a working group “**Staph**”
statistics for functional data and/or statistics for Hilbert.
Nonparametric models for functional data:
regression, supervised learning, additive models, ...
Decomposition of stationary processes (PCA in the frequency domain, ...)
Statistics on operators and non-commutative probabilities,
.....

A web site: `http://www.lsp.ups-tlse.fr/Fp/Ferraty/staph.html`