Probability Distribution And Density For Functional Random Variables

E. Cuvelier[1] M. Noirhomme-Fraiture[1]

[1] Institut d'Informatique
Facultés Universitaires Notre-Dame de la paix
Namur

CIL Research Contact Day, 2006

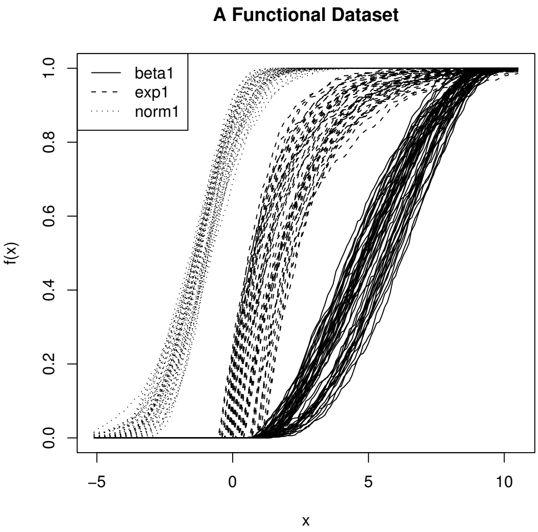
Outline

Symbolic Data Analysis(SDA) In A Nutshell

SDA summarizes datasets using

- singles valued variables
- interval variables
- multi-valued weighted variables
- continuous probability distributions

A Classification Problem...



...Using Dynamical Clustering Extension For Mixture Decomposition

Algorithm(Diday)

Start with a random partition
Repeat:
Step 1 : Find parameters β_i which maximize a criterion
Step 2 : Build classes (P_i) with parameters found at Step 1

$$P_i = \{u : f(u, \beta_i) \geq f(u, \beta_m) \forall m\}$$

until stabilization of the partition.

Choosen criterion : Log-Likelihood

$$lvc(P, \beta) = \sum_{i}^{K} \sum_{u \in P_i} log(f_i, \sigma(u))$$

...Using Dynamical Clustering Extension For Mixture Decomposition

Algorithm(Diday)

Start with a random partition
Repeat:
Step 1 : Find parameters β_i which maximize a criterion
Step 2 : Build classes (P_i) with parameters found at Step 1

$$P_i = \{u : f(u, \beta_i) \geq f(u, \beta_m) \forall m\}$$

until stabilization of the partition.

Choosen criterion : Log-Likelihood

$$lvc(P, \beta) = \sum_{i}^{K} \sum_{u \in P_i} log(f_i, \sigma(u))$$

So we need CDF and PDF for functional random variable!

Define The Framework

Framework

- $\mathcal{D} \subseteq \mathbb{R}$ a closed interval of $\mathbb{R}$,
- $C^0(\mathcal{D})$ the set of continuous, bounded functions of domain $\mathcal{D}$,
- for $u \in C^0(\mathcal{D})$: $\|u\|_2 = \left\{ \int_{\mathcal{D}} |u(x)|^2 \, dx \right\}^{1/2}$,
- $L^2(\mathcal{D}) = \left\{ u \in C^0(\mathcal{D}) \ : \ \|u\|_2 < \infty \right\}$,
- for $u, v \in L^2(\mathcal{D})$: $d_2(u, v) = \|u - v\|_2$,

Definition

Let $f, g \in L^2(\mathcal{D})$, the pointwise order between f and g on $\mathcal{D}$ is defined by:

$$f \leq_{\mathcal{D}} g \iff f(x) \leq g(x), \ \forall x \in \mathcal{D}$$

Functional Random Variable
Functional cumulative distribution function

- Ω set of objects "described" by $u \in L^2(\mathcal{D})$.
- *functional random variable (frv)* is any function from Ω to $L^2(\mathcal{D})$:

$$\underline{X} : \Omega \to L^2(\mathcal{D}) : \omega \mapsto X(\omega)$$

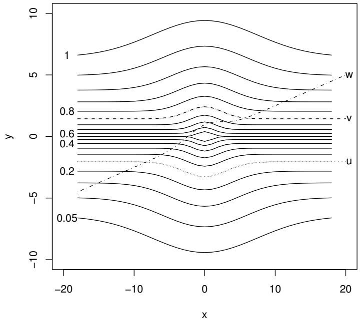
$$X(\omega) : \mathcal{D} \to \mathbb{R} : r \mapsto X(\omega)(r)$$

- *functional cumulative distribution function (fcdf)* of a *frv* $\underline{X}$:

$$\begin{aligned} F_{\underline{X},\mathcal{D}}(u) &= P\{\omega \in \Omega : X(\omega)(x) \leq u(x), \forall x \in \mathcal{D}\} \\ &= P[\underline{X}(x) \leq u(x), \forall x \in \mathcal{D}] \\ &= P[\underline{X} \leq_{\mathcal{D}} u] \qquad (1) \\ &= P[\mathcal{A}(u)] \qquad (2) \end{aligned}$$

with $\mathcal{A}(u) = \{\omega \in \Omega : X(\omega) \leq_{\mathcal{D}} u\}$

A simple example with N=20 sample functions

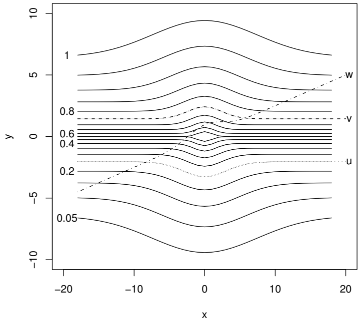


Empirical Estimation ?

$$\widehat{F}_{\underline{X},\mathcal{D}}(u) = \frac{\#\{f \in A : f(x) \leq u(x), \forall x \in \mathcal{D}\}}{\#A}$$

- $\widehat{F}_{\underline{X},\mathcal{D}}(u) =$
- $\widehat{F}_{\underline{X},\mathcal{D}}(v) =$
- $\widehat{F}_{\underline{X},\mathcal{D}}(w) =$

A simple example with N=20 sample functions

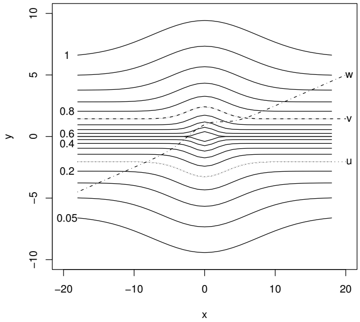


Empirical Estimation ?

$$\widehat{F}_{\underline{X},\mathcal{D}}(u) = \frac{\#\{f \in A : f(x) \leq u(x), \forall x \in \mathcal{D}\}}{\#A}$$

- $\widehat{F}_{\underline{X},\mathcal{D}}(u) = \frac{1}{4}$
- $\widehat{F}_{\underline{X},\mathcal{D}}(v) =$
- $\widehat{F}_{\underline{X},\mathcal{D}}(w) =$

A simple example with N=20 sample functions

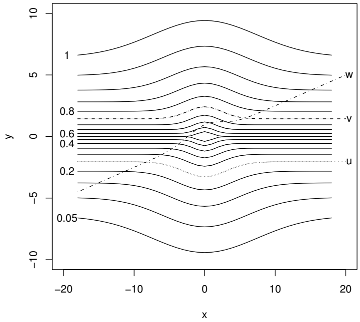


Empirical Estimation ?

$$\widehat{F}_{\underline{X},\mathcal{D}}(u) = \frac{\#\{f \in A : f(x) \leq u(x), \forall x \in \mathcal{D}\}}{\#A}$$

- $\widehat{F}_{\underline{X},\mathcal{D}}(u) = \frac{1}{4}$
- $\widehat{F}_{\underline{X},\mathcal{D}}(v) = \frac{3}{4}$
- $\widehat{F}_{\underline{X},\mathcal{D}}(w) =$

A simple example with N=20 sample functions

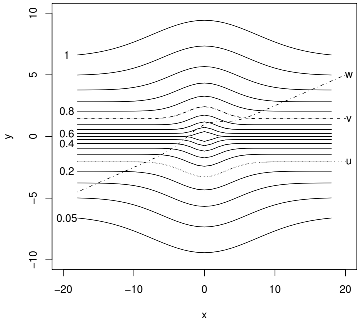


Empirical Estimation ?

$$\widehat{F}_{\underline{X},\mathcal{D}}(u) = \frac{\#\{f \in A : f(x) \leq u(x), \forall x \in \mathcal{D}\}}{\#A}$$

- $\widehat{F}_{\underline{X},\mathcal{D}}(u) = \frac{1}{4}$
- $\widehat{F}_{\underline{X},\mathcal{D}}(v) = \frac{3}{4}$
- $\widehat{F}_{\underline{X},\mathcal{D}}(w) = 0.1?$

A simple example with N=20 sample functions



Empirical Estimation ?

$$\widehat{F}_{\underline{X},\mathcal{D}}(u) = \frac{\#\{f \in A : f(x) \leq u(x), \forall x \in \mathcal{D}\}}{\#A}$$

- $\widehat{F}_{\underline{X},\mathcal{D}}(u) = \frac{1}{4}$
- $\widehat{F}_{\underline{X},\mathcal{D}}(v) = \frac{3}{4}$
- $\widehat{F}_{\underline{X},\mathcal{D}}(w) = 0.1$?

But w is greater than 20% of the functions of A for most of the values of $\mathcal{D}$?!?

Trying Approximations

Let :

- $n \in \mathbb{N}_0$, $q = 2^n - 1$,
- $\left\{x_1^n, \ldots, x_q^n\right\} \subset \mathcal{D}$,
- $x_1^n = \inf(\mathcal{D})$ & $x_q^n = \sup(\mathcal{D})$,
- $\left|x_{i+1}^n - x_i^n\right| = \frac{|\mathcal{D}|}{q}$

$$\mathcal{A}_n(u) = \bigcap_{i=1}^{q} \{\omega \in \Omega : X(\omega)(x_i^n) \leq u(x_i^n)\}$$

$$\mathcal{A}(u) = \{\omega \in \Omega : X(\omega) \leq_{\mathcal{D}} u\}$$

$$F_{\underline{X},\mathcal{D}}(u) = P[\mathcal{A}(u)] \approx P[\mathcal{A}_n(u)] = H\left(u(x_1^n), \ldots, u(x_q^n)\right)$$

where $H(\cdot, \ldots, \cdot)$ is a joint distribution of dimension q.

Which Joint Distributions Are Suitable ?

$$F_{\underline{X},\mathcal{D}}(u) = P[\mathcal{A}(u)] \approx P[\mathcal{A}_n(u)] = H\left(u(x_1^n), \ldots, u(x_q^n)\right)$$

- Normal ?
- Student ?
- take a matrix as parameter

Which Joint Distributions Are Suitable ?

$$F_{\underline{X},\mathcal{D}}(u) = P[\mathcal{A}(u)] \approx P[\mathcal{A}_n(u)] = H\left(u(x_1^n), \ldots, u(x_q^n)\right)$$

- Normal ?
- Student ?
- take a matrix as parameter
- $(q^2 - q)$ real parameters
- What's the matter when $q \to \infty$?

Which Joint Distributions Are Suitable ?

$$F_{\underline{X},\mathcal{D}}(u) = P[\mathcal{A}(u)] \approx P[\mathcal{A}_n(u)] = H\left(u(x_1^n), \ldots, u(x_q^n)\right)$$

- Normal ?
- Student ?
- take a matrix as parameter
- $(q^2 - q)$ real parameters
- What's the matter when $q \to \infty$?
- And why not a DIY distribution using Copulas ?

What Is A Copula ?

Definition

A copula is a multivariate cumulative distribution function defined on the n-dimensional unit cube $[0, 1]^n$ such that every marginal distribution is uniform on the interval [0, 1] :

$$C : [0, 1]^n \to [0, 1] : (u_1, \dots, u_n) \mapsto C(u_1, \dots, u_n)$$

What Is A Copula ?

Definition

A copula is a multivariate cumulative distribution function defined on the n-dimensional unit cube $[0,1]^n$ such that every marginal distribution is uniform on the interval [0, 1] :

$$C : [0,1]^n \to [0,1] : (u_1,\ldots,u_n) \mapsto C(u_1,\ldots,u_n)$$

The interest of copulas comes from the following theorem:

Theorem (Sklar's theorem)

Let H be an n-dimensional distribution function with margins $F_1, ..., F_n$. Then there exists an n-copula C such that for all $x \in \bar{\mathbb{R}}^n$,

$$H(x_1,...,x_n) = C(F_1(x_1),...,F_n(x_n)). \quad (3)$$

If $F_1, ..., F_n$ are all continuous, then C is unique; otherwise, C is uniquely determined on Range of $F_1 \times ... \times$ Range of F_n

Archimedean Copulas

Definition

An Archimedean copula is a function $[0,1]^n \to [0,1]$ given by

$$C(u_1, ..., u_n) = \psi\left[\sum_{i=1}^{n} \phi(u_i)\right] \quad (4)$$

where ϕ, called the generator, is a function $\phi : [0,1] \to [0,\infty]$ such:

- ϕ is a continuous strictly decreasing function
- $\phi(0) = \infty$ & $\phi(1) = 0$
- $\psi = \phi^{-1}$ is completely monotonic on $[0,\infty[$ i.e.

$$(-1)^k \frac{d^k}{dt^k}\psi(t) \geq 0 \quad (5)$$

for all t in $[0,\infty[$ and for all k.

Families of Archimedean Copulas

Name	Generator ϕ	Domain of θ
Clayton	$\phi_\theta(t) = t^\theta - 1$	$\theta > 0$
Frank	$\phi_\theta(t) = -\ln \frac{e^{-\theta \cdot t} - 1}{e^{-\theta} - 1}$	$\theta > 0$
Gumbel-Hougaard	$\phi_\theta(t) = (-\ln t)^\theta$	$\theta \geq 1$

Surfaces Of Distributions And Densities - All You Need Is Margins

Definition

Let G and g the *surface of distributions* and the *surface of densities*, that give the distribution and the density of $\underline{X}(x)$ for a chosen $x \in \mathcal{D}$:

$$G : \mathcal{D} \times \mathbb{R} \rightarrow [0,1] : (x,y) \mapsto G(x,y) = P[\underline{X}(x) \leq y]$$

$$g : \mathcal{D} \times \mathbb{R} \rightarrow [0,1] : (x,y) \mapsto g(x,y) = \frac{\partial}{\partial x} G(x,y)$$

Surfaces Of Distributions And Densities - All You Need Is Margins

Definition

Let G and g the *surface of distributions* and the *surface of densities*, that give the distribution and the density of $\underline{X}(x)$ for a chosen $x \in \mathcal{D}$:

$$G : \mathcal{D} \times \mathbb{R} \to [0,1] : (x,y) \mapsto G(x,y) = P[\underline{X}(x) \leq y]$$

$$g : \mathcal{D} \times \mathbb{R} \to [0,1] : (x,y) \mapsto g(x,y) = \frac{\partial}{\partial x} G(x,y)$$

Normal Case

$$\hat{G}(x,y) = F_{\mathcal{N}(\mu(x),\sigma(x))}(y)$$

$$\hat{g}(x,y) = f_{\mathcal{N}(\mu(x),\sigma(x))}(y)$$

Surfaces Of Distributions And Densities - All You Need Is Margins

Definition

Let G and g the *surface of distributions* and the *surface of densities*, that give the distribution and the density of $\underline{X}(x)$ for a chosen $x \in \mathcal{D}$:

$$G : \mathcal{D} \times \mathbb{R} \to [0,1] : (x,y) \mapsto G(x,y) = P[\underline{X}(x) \leq y]$$

$$g : \mathcal{D} \times \mathbb{R} \to [0,1] : (x,y) \mapsto g(x,y) = \frac{\partial}{\partial x} G(x,y)$$

Normal Case

$$\hat{G}(x,y) = F_{\mathcal{N}(\mu(x),\sigma(x))}(y)$$

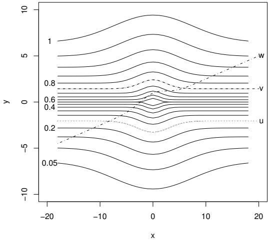
$$\hat{g}(x,y) = f_{\mathcal{N}(\mu(x),\sigma(x))}(y)$$

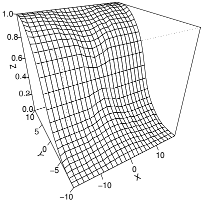
Estimation Case

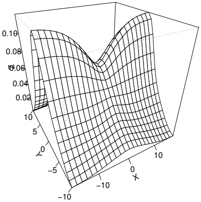
$$\hat{G}(x,y) = \frac{\#\{X_i(x) \leq y\}}{N}$$

$$\hat{g}(x,y) = \frac{\sum_{i=1}^{N} K\left(\frac{y - X_i(x)}{h(x)}\right)}{N \cdot h(x)}$$

Surfaces Of Distributions And Densities - Example







Copula + Surface of Distributions : FCDF ?

$$\mathcal{A}(u) = \{\omega \in \Omega : X(\omega) \leq_{\mathcal{D}} u\}$$

$$\mathcal{A}_n(u) = \bigcap_{i=1}^{q} \{\omega \in \Omega : X(\omega)(x_i^n) \leq u(x_i^n)\}$$

$$\begin{aligned}
F_{\underline{X},\mathcal{D}}(u) &= P[\mathcal{A}(u)] \\
&\approx P[\mathcal{A}_n(u)] \\
&= H(u(x_1^n), \ldots, u(x_q^n)) \\
&= C\left(G\left[x_1^n, u\left(x_1^n\right)\right], \ldots, G\left[x_q^n, u\left(x_q^n\right)\right]\right) && (6) \\
&= \psi\left(\sum_{i=1}^{q} \phi\left(G\left[x_i^n, u\left(x_i^n\right)\right]\right)\right) && (7)
\end{aligned}$$

Bad News - Volumetric Behavior

Objection : A Kind Of Volumetric Behavior

Let's call a function $u \in L^2(\mathcal{D})$ such

$$G(x, u(x)) = p, \ \forall x \in \mathcal{D} \tag{8}$$

a *functional quantile* of value $p(Q_p)$. Then

$$\begin{aligned} P[\mathcal{A}_n(u)] &= \psi\left[\sum_{i=1}^{q} \phi\left(G\left[x_i^n, u(x_i^n)\right]\right)\right] \\ &= \psi\left[\sum_{i=1}^{q} \phi(p)\right] \\ &= \psi\left(q \cdot \phi\left(p\right)\right) < p \end{aligned}$$

More you try, more you move away !!!

Bad News Update - Null Limit

Theorem

If for $u \in L^2(\mathcal{D}) : G(x, u(x)) < 1,\ \forall x \in \mathcal{D}$
Then

$$\lim_{q \to \infty} \psi \left[\sum_{i=1}^{q} \phi \left(G \left[x_i^n, u \left(x_i^n \right) \right] \right) \right] = 0 \quad (9)$$

A Simple Transformation To Margins

Lemma

Let $q \in \mathbb{N}_0$, F be a one dimensional cdf, and ϕ a generator of Archimedean copula, then the following expression is also a cdf.

$$F^*(x) = \psi\left(\frac{1}{q} \cdot \phi(F(x))\right) \tag{10}$$

Theorem

Let $q \in \mathbb{N}_0$, $\{F_i | 1 \leq i \leq q\}$ be a set of one dimensional cdf, and ϕ a generator of Archimedean copula, then the following expression is a multivariate cdf.

$$H(x_1, \ldots, x_q) = \psi\left(\sum_{i=1}^{q} \phi(F_i^*(x_i))\right) = \psi\left(\frac{1}{q}\sum_{i=1}^{q} \phi(F_i(x_i))\right) \tag{11}$$

Quasi-Arithmetic Means

Definition

Let $[a, b]$ be a closed real interval, and $q \in \mathbb{N}_0$. A quasi-arithmetic mean is a function $M : [a, b]^q \to [a, b]$ defined as follows:

$$M(x_1, \ldots, x_q) = \psi\left(\frac{1}{q}\sum_{i=1}^{q}\phi(x_i)\right) \tag{12}$$

where ϕ is a continuous strictly monotonic real function.

The *Quasi-Arithmetic Mean of Margins (QAMM)* distribution

$$H(x_1, \ldots, x_q) = \psi\left(\frac{1}{q}\sum_{i=1}^{q}\phi(F_i(x_i))\right) \tag{13}$$

QAMM avoid us the bad news!

OK, let us use QAMM

$$
\begin{aligned}
F_{\underline{X},\mathcal{D}}(u) &= P[\mathcal{A}(u)] \\
&\approx P[\mathcal{A}_n(u)] \\
&= \psi\left[\frac{1}{q}\sum_{i=1}^{q}\phi\left(G\left[x_i, u(x_i)\right]\right)\right] \\
&= \psi\left[\frac{1}{|\mathcal{D}|}\sum_{i=1}^{q}\frac{|\mathcal{D}|}{q}\cdot\phi\left(G\left[x_i, u(x_i)\right]\right)\right]
\end{aligned} \tag{14}
$$

And like we have $\frac{|\mathcal{D}|}{q} = \left|x_{i+1}^n - x_i^n\right| = \Delta_x \ \forall i \in \{1, \ldots, q\}$, we can now take the limit

$$
\lim_{q\to\infty} P[\mathcal{A}_n(u)] = \lim_{q\to\infty} \psi\left[\frac{1}{|\mathcal{D}|}\sum_{i=1}^{q}\phi\left(G\left[x_i, u(x_i)\right]\right)\cdot\Delta_x\right] \tag{15}
$$

Quasi-Arithmetic Mean of Margins Limit (QAMML)

Definition

Let :

- $\underline{X}$ be a frv and $u \in L^2(\mathcal{D})$,
- $G : \mathcal{D} \times \mathbb{R} \to [0, 1] : (x, y) \mapsto G(x, y) = P[\underline{X}(x) \leq y]$
- ϕ is a generator of Archimedean copula

We define the *Quasi-Arithmetic Mean of Margins Limit (QAMML)* distribution of $\underline{X}$ by :

$$F_{\underline{X},\mathcal{D}}(u) = \psi \left[\frac{1}{|\mathcal{D}|} \cdot \int_{\mathcal{D}} \phi \left(G \left[x, u \left(x \right) \right] \right) dx \right] \tag{16}$$

Proposition

If $Q_p \in L^2(\mathcal{D})$ is a *functional quantile* of value p, then $F_{\underline{X},\mathcal{D}}(Q_p) = p$

QAMML - Example

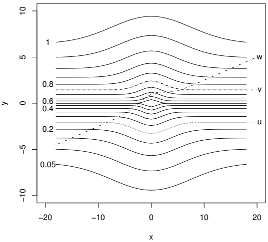


Table: *fcdf* with several values for the parameter of Clayton's generator

Par.	u	v	w
0.5	0.2382	0.7010	0.3732
2	0.2380	0.7010	0.2545
8	0.2373	0.7007	0.1408
16	0.2362	0.7003	0.1144

And What About The Density ?

The Density Problem In An Infinite Dimension Space

- A *fcdf* is a incomplete tool without an associate density.
- $\lim_{q \to \infty} \frac{\partial^q}{\partial x_1 \ldots \partial x_q} H(x_1, \ldots, x_q)$?

Try To Preserve The Functional Quantile

The *QAMML* distribution "preserve" the value of a *functional quantile*. We need to find a derivative operator which preserves this property, i.e. suppose that's $0 \leq p < q \leq 1$, we search an operator D such :

$$F_{\underline{X},\mathcal{D}}(Q_q) = F_{\underline{X},\mathcal{D}}(Q_p) + DF_{\underline{X},\mathcal{D}}(Q_p) \cdot d_2(Q_q, Q_p) + o(|\epsilon|) \quad (17)$$

The Gâteaux differential

Definition

Suppose V and W are normed vector spaces, and an operator $F : V \rightarrow W$. The *Gâteaux differential* $DF(u; s)$ of F at u in the direction $s \in V$ is given by:

$$\begin{aligned} DF(u; s) &= \lim_{\epsilon \rightarrow 0} \frac{F(u + \epsilon \cdot s) - F(u)}{\epsilon} && (18) \\ &= F'(u) \cdot s && (19) \end{aligned}$$

If $DF(u; s)$ exists $\forall s \in L^2(\mathcal{D})$ then F is *Gâteaux differentiable* and the map $F'(u)$ is the *Gâteaux derivative* of F at u.

Which Direction Choose ?

- Find the direction between two *functional quantiles*.
- Suppose that G is a distribution with location and scale parameters, with $l(x)$ and $s(x)$ which give these parameters for a value x, then

$$Q_p(x) = Q\left(p; l(x), s(x)\right) = s(x) \cdot Q\left(\alpha; 0, 1\right) + l(x) \qquad (20)$$

- And then the searched direction come easily, since :

$$\begin{aligned} Q_q(x) - Q_p(x) &= s(x) \cdot [Q(q; 0, 1) - Q(p; 0, 1)] \\ &= s(x) \cdot \epsilon \end{aligned}$$

  Where $\epsilon = Q(q; 0, 1) - Q(p; 0, 1)$ is a constant,
- So we have

$$Q_q(x) = Q_p(x) + s(x) \cdot \epsilon \qquad (21)$$

So, we use as direction a scale parameter, or more generally any statistical dispersion function s, like σ.

The Gâteaux Density

Definition

Let $\underline{X}$ be a *frv*, $F_{\underline{X},\mathcal{D}}$ its *fcdf* and u a function of $L^2(\mathcal{D})$. If $s \in L^2(\mathcal{D})$ is a function such $s(x)$ measure the statistical dispersion of the values $\underline{X}(x)$, then we define the *Gâteaux density of $F_{\underline{X},\mathcal{D}}$ at u and in direction of s* by:

$$\begin{aligned} f_{\underline{X},\mathcal{D},s}(u) &= \lim_{\epsilon \longrightarrow 0} \frac{F_{\underline{X},\mathcal{D}}(u + s \cdot \epsilon) - F_{\underline{X},\mathcal{D}}(u)}{d_2(u + s.\epsilon, u)} \\ &= \frac{DF_{\underline{X},\mathcal{D}}(u; s)}{\|s\|_2} \end{aligned} \tag{22}$$

Where $DF_{\underline{X},\mathcal{D}}(u; s)$ is the *Gâteaux differential* of $F_{\underline{X},\mathcal{D}}$ at u in the direction $s \in V$.

A functional Analysis Result

Theorem

Let the following integral transform :

$$\mathcal{T}(f) = \int_a^b K(t,s) \cdot g[s, f(s)]\ ds \tag{23}$$

where the kernel $K(s,t)$ is continuous on $[a,b]^2$, and $g(s,t)$ is a function of two variables, defined and continuous on $[a,b] \times]-\infty, +\infty[$. Then for any function $h \in C[a,b]$ we have

$$D\mathcal{T}(f,h) = \int_a^b K(t,s) \cdot g'_v[s, f(s)] \cdot \mathbf{h}(\mathbf{s})\ ds \tag{24}$$

Where $D\mathcal{T}(f,h)$ is the Gâteaux differential of $\mathcal{T}$ at f in the direction h.

Gâteaux Density for QAMML

Theorem

Let $F_{\underline{X},\mathcal{D}}$ a fcdf, u a function of $L^2(\mathcal{D})$. If $s \in L^2(\mathcal{D})$ is a functional measure of the statistical dispersion of the values $\underline{X}(x)$, then the Gâteaux density of $F_{\underline{X},\mathcal{D}}$ in u and in direction of s is given by:

$$f_{\underline{X},\mathcal{D},s}(u) = \frac{1}{\|s\|_2 \cdot |\mathcal{D}|} \cdot \psi' \left[\frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} \phi\left(G\left[t, u(t)\right]\right)\, dt \right] \cdot \left\{ \int_{\mathcal{D}} \phi'\left(G\left[t, u(t)\right]\right) \cdot g\left[t, u(t)\right] \cdot s(t)\, dt \right\} \quad (25)$$

Theorem

If $Q_p \in L^2(\mathcal{D})$ is a functional quantile of value p, such $g\left(x, Q_p(x)\right) \cdot s(x) = \pi, \ \forall x \in \mathcal{D}$ then

$$f_{\underline{X},\mathcal{D},s}(Q_p) = \frac{\pi}{\|s\|_2} \quad (26)$$

Recall Our Classification Problem...

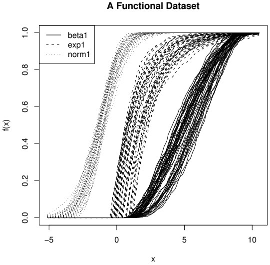


Table: Range of the parameters of the functional data

Name	Size	Location	Scale
Normal	45	$[-1.52, -0.86]$	$[0.79, 1.85]$
Beta	45	$[-2.74, -1.80]$	$[8.07, 10.79]$
Exp	45	$[-5.19, -3.89]$	$[0.51, 0.95]$

...Using Dynamical Clustering Extension For Mixture Decomposition

Algorithm(Diday)

Start with a random partition
Repeat:
Step 1 : Find parameters β_i which maximizes a criterion
Step 2 : Build classes (P_i) with parameters found at Step 1

$$P_i = \{u : f_{\underline{X},\mathcal{D}_i,\sigma}(u, \beta_i) \geq f_{\underline{X},\mathcal{D}_m,\sigma}(u, \beta_m) \forall m\}$$

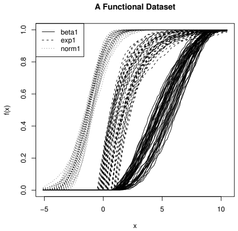
until stabilization of the partition.

Choosen criterion : Log-Likelihood Classifier

$$lvc(P, \beta) = \sum_{i}^{K} \sum_{u \in P_i} log(f_{\underline{X},\mathcal{D}_i,\sigma}(u)) \qquad (27)$$

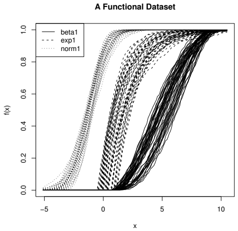
where $\mathcal{D}_i$ the *model's domain* of the ith cluster.

Classification Results



- Methods runs 5 times,
- We keep result with the best criterion,
- Misclassification rate : 0.71%

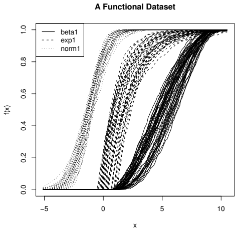
Classification Results



- Methods runs 5 times,
- We keep result with the best criterion,
- Misclassification rate : 0.71%

But your functional clusters are obviously very well separated!

Classification Results



- Methods runs 5 times,
- We keep result with the best criterion,
- Misclassification rate : 0.71%

But your functional clusters are obviously very well separated!
Yes, but if we don't find them here : forget all!

Summary & Future Works

Summary

- *QAMML* is a new mathematical tool which can be used with existing probabilistic methods,
- Validation with the *Dynamical clustering* on functional data coming from the *Symbolic Data Analysis* framework.

Future Works

- $F_{\underline{X},\mathcal{D}}(u) = \psi\left[\int_{\mathcal{D}} \phi\left(G\left[x, u\left(x\right)\right]\right) d\mathcal{F}\right]$ where $\mathcal{F}$ is a uniform cdf over $\mathcal{D}$, other cdf can be considered,
- We look also towards the *Sobolev spaces*, and for a *frv* X consider the *QAMML* distributions of $X', X'', ...X^{(p)}$ and joint them with a copula.

Many developments are still possibles.