Curry et Bayes sont sur un bateau

Zarour Samy

Le programme

1. Structures algébriques
   a. Du magma à la monade

2. Théorie des probabilités
   a. Les probabilités sont des monades

3. Place au code avec Figaro et Scala
   a. Implémentation, exemple en Scala

Structures algébriques

Du magma à la monade

Magma

(M, *)

- M est un **ensemble**
- * est une **loi de composition interne**:
    - $\forall\ (a,b) \in M^2, a*b \in M$

```
trait Magma[M]{

	def *(a:M, b:M): M

}
```

Exemple (magma)

M = { †, ‡, ◊ }

† * ◊ = ◊

◊ * † = ‡

Par définition x,y dans M,
x * y est dans M

<u>Contre exemple:</u>

(Z^-, *), multiplication usuelle

*	†	‡	◊
†	‡	†	◊
‡	◊	‡	†
◊	‡	◊	◊

Monoïde

$(M, *)$

- $(M, *)$ est un **magma**
- $*$ est **associative**
  - $\forall (a,b,c) \in M^3, (a*b)*c = a*(b*c)$
- M possède un **élément neutre** pour $*$ (**unifère**)
  - $\exists e \in M, \forall a \in M, a*e = e*a = a$

Exemples (monoïde)

- (N, +) muni de l'addition usuelle (élément neutre: 0)
- (R^R, o) muni de la composition de fonction, avec l'élément neutre id: x -> x
    - f : x -> 5x et g : x -> x^3
    - g o f : x -> g[f(x)] = g(5x) = $(5x)^3$ = $125x^3$
    - f o g : x -> f[g(x)] = f(x^3) = $5x^3$

Contre exemple: (Z^-, *) et (2Z, *)

Pourquoi le monoïde

Associativité => parallélisme:

a+b+c+d+e+f = (a+b)+(c+d)+(e+f) = (a+b+c)+(d+e+f)

- Calcul sur plusieurs threads
- Composition des éléments en restant dans un contexte

Utilisation du monoïde

```scala
case class Session(start:Long, end:Long)

trait Monoid[T]{
  def zero:T
  def *(a: Session, b: Session)
}

implicit val monoidSession = new Monoid[Session] {
  override def zero: Session = Session(0,0)
  override def *(a: Session, b: Session): Session = (a,b) match {
    case (Session(0,0), Session(0,0)) => zero
    case (x, Session(0,0)) => x
    case (Session(0,0), x) => x
    case _ => Session(
      start = math.min(a.start,b.start),
      end = math.max(a.end,b.end)
    )
  }
}
```

Les catégories

Une catégorie C est la donnée des éléments suivants:

- Deux collections: Ob(C) (**objets**) et Fl(C) (**flèches**)
- Deux applications: s : Fl(C) -> Ob(C) (**source**) et t : Fl(C) -> Ob(C) (**target**)
- Une application identité 1 : Ob(C) -> Fl(C), X -> 1_X
- Une application g o f : Fl(C) x Fl(C) -> Fl(C), (f,g) -> g o f telle que t(f) = s(g) (f et g composables)

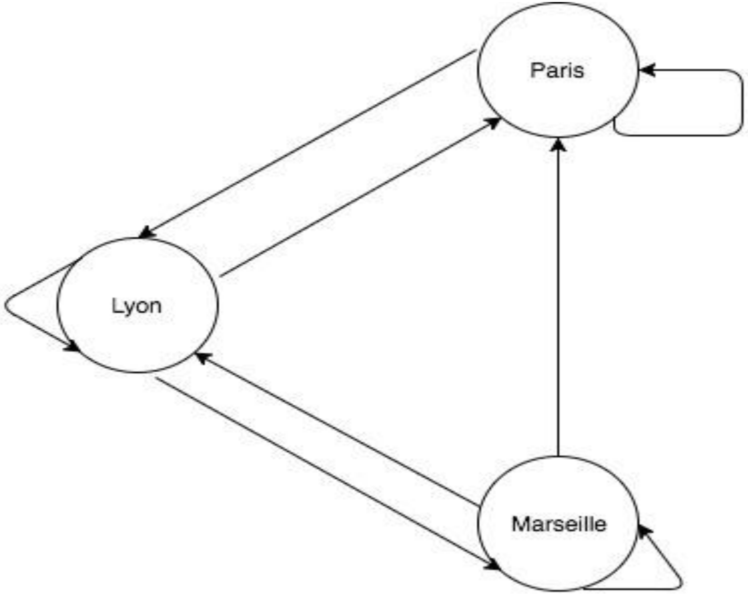
Tels que:

- $\forall X \in Ob(C)$, $s(1_X) = t(1_X) = X$
- $\forall (f,g) \in Fl(C)^2 \cap$ Composable, s(gof) = s(f) & t(gof) = t(g)
- $\forall f : X \to Y \in Fl(C)$, $1_Y \circ f = f \circ 1_X = f$
- $\forall (f,g,h) \in Fl(C)^3$ / $((f,g), (g,h)) \in \text{Composable}^2$, $h \circ (g \circ f) = (h \circ g) \circ f$

Exemple (catégorie)

Ob(C) = { Paris, Lyon, Marseille }

Fl(C) = { P->L, L->P, L->M, M->L, M->P, P->P, L->L, M->M}



Finalement ...

- Généralisation des lois monoïdales pour la composition des flèches: **associativité** + **identité**
- Représentation de haut niveau (les objets peuvent être n’importe quoi)

Foncteur

Soient C et D deux catégories. Un foncteur $F : C \to D$ est la donnée de:

- Deux applications $F_{ob} : Ob(C) \to Ob(D)$ et $F_{fl} : Fl(C) \to Fl(D)$

Telles que:

- $\forall f \in Fl(C)$, $F_{ob}(s(f)) = s(F_{fl}(f))$ & $F_{ob}(t(f)) = t(F_{fl}(f))$
- $\forall X \in C$, $F_{fl}(1_X) = 1_{F(X)}$
- $\forall (f,g) \in Fl(C)^2 \cap$ Composable, $F_{fl}(g \circ f) = F_{fl}(g) \circ F_{fl}(f)$

Exemple (foncteur)

F est un foncteur qui:

- envoie les capitales vers les pays
- les chemins entre villes en chemins entre pays

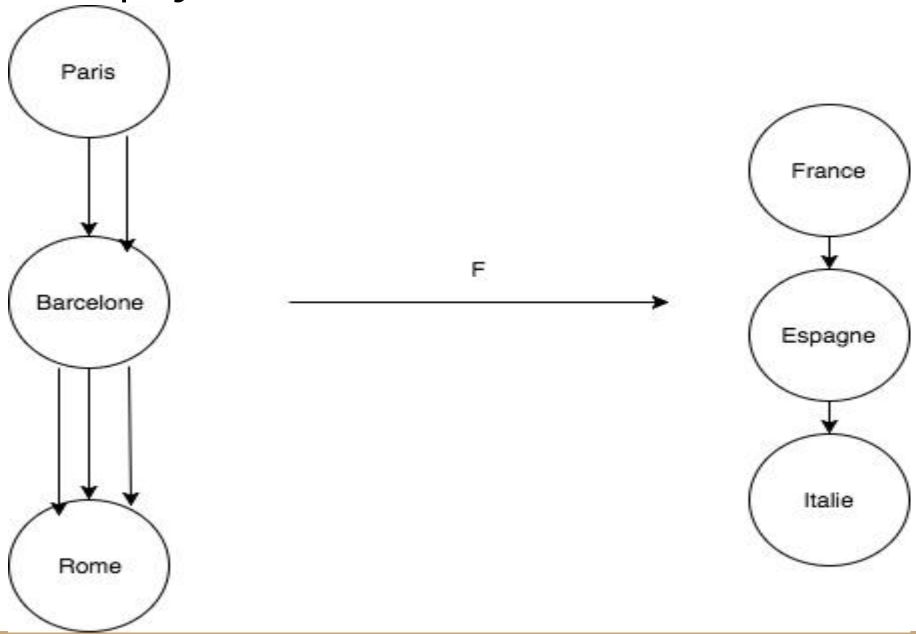
Ob(C) = { Paris, Barcelone, Rome }

Ob(D) = { France, Espagne, Italie }

F_{fl}(Paris->Barcelone) = France->Espagne

F_{ob}(Paris) = France

On a une perte d'information



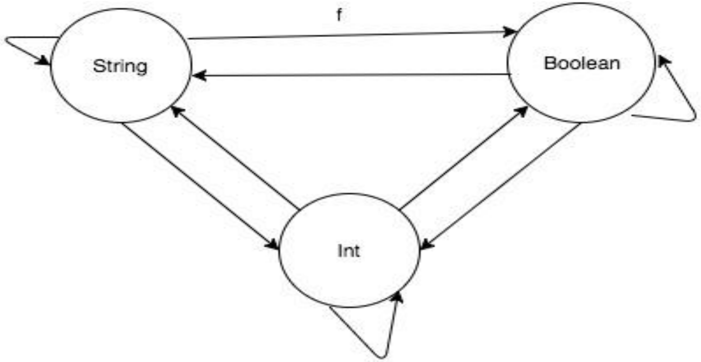
Exemple FP (foncteur)

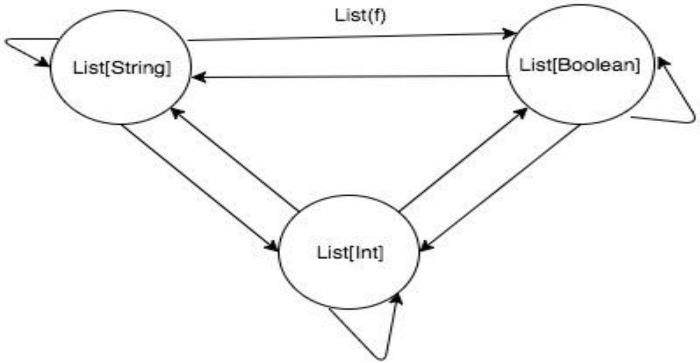
List: T -> List[T]

List : (f: T -> U) -> List(f) : List[T] -> List[U]

Ob(C) = { String, Double, Int, List[String], ... }

Fl(C) = { Function[String, Double], Function[List[Int], String], ... }





Finalement ...

Les foncteurs représentent des **fonctions d’ordre supérieur**

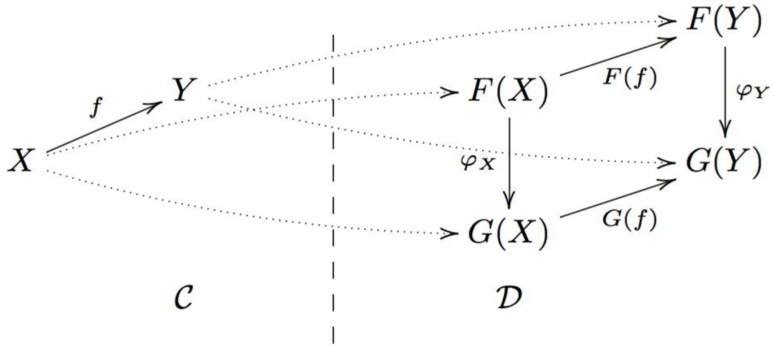
Un foncteur est l’image d’une catégorie dans une autre catégorie (perte d’info)

Transformation naturelle

Soient C et D deux catégories.

Soient F,G : C -> D deux foncteurs parallèles et f : X -> Y, une fonction, pour tous X et Y dans Ob(C).

Il est naturel d'imposer une condition de commutativité dans D via une fonction φ:



Transformation naturelle

Soient $F, G : C \to D$ deux foncteurs parallèles.

Une transformation naturelle $\varphi : F \to G$ est une application $\varphi : Ob(C) \to Fl(D)$ telle que:

- $\forall X, Y \in Ob(C)$
- $\forall f : X \to Y$,

$$\varphi_Y \circ F(f) = G(f) \circ \varphi_X$$

Monade

Soit C une catégorie.

Une monade sur C est un triplet (T, μ, η) où:

- T : C -> C est un foncteur
- μ : T o T -> T et η : 1 -> T sont des transformations naturelles (1 : C -> C est le foncteur identité)

Tel que les diagrammes suivants commutent:

$$
\begin{tikzcd}
T \circ T \circ T \arrow[r, "\mu_T"] \arrow[d, "T\mu"'] & T \circ T \arrow[d, "\mu"] \\
T \circ T \arrow[r, "\mu"'] & T
\end{tikzcd}
\qquad
\begin{tikzcd}
T \arrow[r, "T_\eta"] \arrow[dr, "1"'] & T \circ T \arrow[d, "\mu"] & T \arrow[l, "\eta_T"'] \arrow[dl, "1"] \\
& T &
\end{tikzcd}
$$

Avec les notations: $\mu_T = \mu$x1, 1xμ=Tμ, ηx1=η_T, 1xη=Tη

Exemple en FP (monade)

Soit List le foncteur défini par:

- List_{ob} : T -> List[T]
- List_{fl} : (f : T -> U) -> (List(f) : List[T] -> List[U])

On choisit μ = *flat* : List[List[T]] -> List[T]

$$
\begin{tikzcd}
\mathbf{List}(\mathbf{List}(\mathbf{List}(T))) \arrow[r, "\mathbf{List}\ \mathtt{flat}"] \arrow[d, "\mathtt{flat}\ \mathbf{List}"'] & \mathbf{List}(\mathbf{List}(T)) \arrow[d, "\mathtt{flat}"] \\
\mathbf{List}(\mathbf{List}(T)) \arrow[r, "\mathtt{flat}"'] & \mathbf{List}(T)
\end{tikzcd}
$$

Exemple en FP (monade)

val myList = List(List(List(a,b), List(c)),List(List(d,e)))

μ_T(myList) = List(List(a,b,c), List(d,e))

Tμ(myList) = List(List(a,b), List(c), List(d,e))

$$\begin{tikzcd}
{[[[a,b],[c]],[[d,e]]]} \arrow[r, mapsto] \arrow[d, mapsto] & {[[a,b,c],[d,e]]} \arrow[d, mapsto] \\
{[[a,b],[c],[d,e]]} \arrow[r, mapsto] & {[a,b,c,d,e]}
\end{tikzcd}$$

Exemple en FP (monade)

La commutativité de l'autre diagramme signifie que:

flat(List(List(a,b,c))) = List(a,b,c)

flat(List(List(a), List(b), List(c))) = List(a,b,c)

Finalement ...

On peut définir une monade par plusieurs fonctions:

- id + map + flatten
- id + flatMap

Une monade est un endofoncteur muni de deux applications naturelles

=> monoïde dans la catégorie des endofoncteurs

Théorie des probabilités

De la monade aux distributions

Espace mesurable

Un espace $(X, \mathcal{X})$ est dit mesurable lorsque X est un ensemble et $\mathcal{X}$ une **tribu** sur X:

- $\mathcal{X}$ est non vide
- $\mathcal{X}$ est stable par complémentaire
- $\mathcal{X}$ est stable par union dénombrable

Exemple: lancé de pièce

$X = \{ 0, 1 \}$

$\mathcal{X} = \mathcal{P}(X) = \{ \{0,1\}, \{1\}, \{0\}, \varnothing \}$ est appelée *tribu discrète*

$\mathcal{P}(X)$=l'ensemble des parties de X. Les éléments de $\mathcal{X}$ sont appelés événements

Mesure

Soit (X, A) un espace mesurable.

Une mesure est un application μ : A -> R^+ vérifiant:

- $\mu(\emptyset) = 0$
- Pour toute famille dénombrable disjointe deux à deux $(E_i)_{i \in I}$, $\mu(\cup E_i) = \sum \mu(E_i)$

Probabilité

Soit (Ω, A) un espace mesurable.

Une probabilité est une mesure $\mathbb{p} : A \to [0,1]$ telle que :

- $\mathbb{p}(\Omega) = 1$
- Pour toute famille finie ou dénombrable $(A_i)_{i\in I}$ de A, $\mathbb{p}(\cup_{i\in I} A_i) = \sum_{i\in I} \mathbb{p}(A_i)$

<u>Exemple</u>: lancé de dé

$\Omega = \{ 1, 2, 3, 4, 5, 6 \}$

$\mathbb{p}(1) = \mathbb{p}(2) = \dots = \mathbb{p}(6) = \frac{1}{6}$

$\mathbb{p}(\{1\} \cup \{2\}) = \mathbb{p}(1) + \mathbb{p}(2) - \mathbb{p}(\{1\} \cap \{2\}) = \frac{1}{6} + \frac{1}{6} - 0 = \frac{1}{3}$

Catégorie d'espaces mesurables

Soit ***Mes*** la catégorie des espaces mesurables.

Ob(***Mes***) = { (X, $\mathcal{X}$) , X est un ensemble et $\mathcal{X}$ une tribu sur X}

Fl(***Mes***) = { f : (X,$\mathcal{X}$) -> (Y,$\mathcal{Y}$), (X,$\mathcal{X}$),(Y,$\mathcal{Y}$) $\in$ Ob(***Mes***) }

Soit M = (M,$\mathcal{M}$) $\in$ ***Mes***.

On définit $\mathcal{P}(M)$ = { μ : $\mathcal{M}$ -> [0,1] } l'ensemble des mesures de probabilités sur M

$\mathcal{P}(M)$ est un espace mesurable.

$\mathcal{P}$ est un foncteur

Soient $M,N \in$ ***Mes***

Soit $f : M \to N$ un morphisme d'espaces mesurables et $\mu \in \mathcal{P}(M)$

En définissant $\mathcal{P}(f)(\mu) = \mu \circ f^{-1}$ on a:

$\mathcal{P}(f) : \mathcal{P}(M) \to \mathcal{P}(N)$

<u>$\mathcal{P}$ est un endofoncteur sur ***Mes***</u>

Exemple ($\mathcal{P}$ est un foncteur)

M = { pile, face, tranche }, N = { 0, 1 }

μ : $\mathcal{M}$ -> [0,1] / μ(pile)=μ(face)=0.499 et μ(tranche)=0.002

f : *M* -> *N* telle que f(pile) = 1 et f(tranche) = f(face) = 0

$\mathcal{P}(f)(\mu) = \mu \circ f^{-1}$

avec : $f^{-1}(0) = \{\text{face}\} \cup \{\text{tranche}\}$ et $f^{-1}(1) = \{\text{pile}\}$:

$\mu(\{\text{face}\} \cup \{\text{tranche}\}) = 0.501$ et $\mu(\{\text{pile}\}) = 0.499$

$\mathcal{P}(f)$ envoie μ sur ν : N -> {0.501, 0.499}

$\mathcal{P}$ est une monade

Soit M = (M, $\mathcal{M}$) ∈ ***Mes*** et A⊆M

($\mathcal{P}$, μ, η) tel que:

- η : 1(M) -> $\mathcal{P}$(M) tel que η(x)(A) = 1 si x∈A et 0 sinon
- μ : $\mathcal{P}^2$ -> $\mathcal{P}$ tel que $\mu(\varrho)$: A -> $\int_{\mathcal{P}(M)} \tau_A d\varrho$,

avec $\tau_A : \nu \to \nu(A) = \int_{\mathcal{M}} 1_A d\nu$

Exemple ($\mathcal{P}$ est une monade)

Soit M = (M, $\mathcal{M}$) ∈ ***Mes*** tel que M = {pile, face}

Soit $\mathbb{p}_1$, $\mathbb{p}_2$ ∈ $\mathcal{P}(M)$ / $\mathbb{p}_1(\text{pile})=\mathbb{p}_1(\text{face})=\frac{1}{2}$ et $\mathbb{p}_2(\text{pile})=\frac{1}{3}$ $\mathbb{p}_2(\text{face})=\frac{2}{3}$

Soit $\mathbb{p}_3$ ∈ $\mathcal{P}(\mathcal{P}(M)) : \mathcal{P}(M) \to [0, 1]$ tel que:

$\mathbb{p}_3(\mathbb{p}_1) = \frac{1}{4}$ et $\mathbb{p}_3(\mathbb{p}_2) = \frac{3}{4}$

$\mu(\mathbb{p}_3)(\text{pile}) = \int_{\mathcal{P}(M)} \tau_{\{\text{pile}\}} d\mathbb{p}_3 = \int_{\mathcal{P}(M)} (\int_{\mathcal{M}} 1_{\{\text{pile}\}} d\nu) d\mathbb{p}_3 = \frac{1}{4} * \frac{1}{2} + \frac{3}{4} * \frac{1}{3} \approx 0.37$

$\mu(\mathbb{p}_3)(\text{face}) = \frac{1}{4} * \frac{1}{2} + \frac{3}{4} * \frac{2}{3} \approx 0.63$

Scala et Figaro

Probabilistic programming

Probabilistic programming

Moyen de créer un système aidant à la décision face à l’incertain. Il utilise le raisonnement probabiliste pour répondre à des questions.

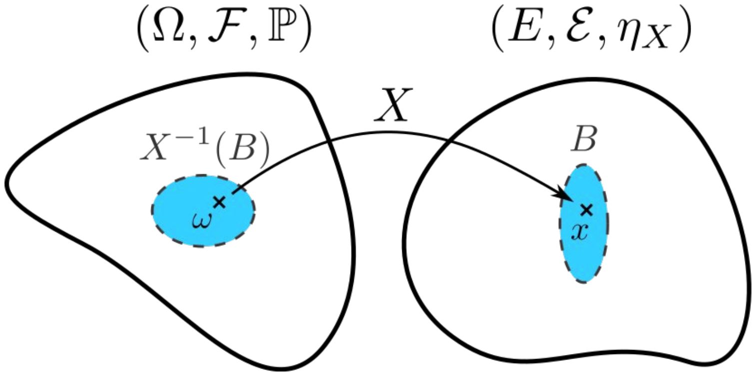
Le raisonnement probabiliste combine la connaissance et la logique. La connaissance est encodée dans un modèle probabiliste.

Variables aléatoires

Soit (Ω, F, $\mathbb{p}$) un espace probabilisé et (E, $\boldsymbol{\varepsilon}$) un espace mesurable. Une variable aléatoire est une fonction mesurable X : Ω -> E

On appelle loi de probabilité de X la fonction $\mathbb{p}_X$: E -> [0,1] définie par:

$\mathbb{p}$: A -> $\mathbb{p}(\{\omega, X(\omega) \in A\})$



<u>Exemple</u>: lancé de 2 pièces

Ω = { (0,1), (1,1), (1,0), (0,0) }

X : (ω_1, ω_2) -> $\omega_1 + \omega_2$

$\mathbb{p}(X=2) = \mathbb{p}(\{(\omega_1, \omega_2), \omega_1+\omega_2=2\} = \mathbb{p}((1,1)) = ¼$

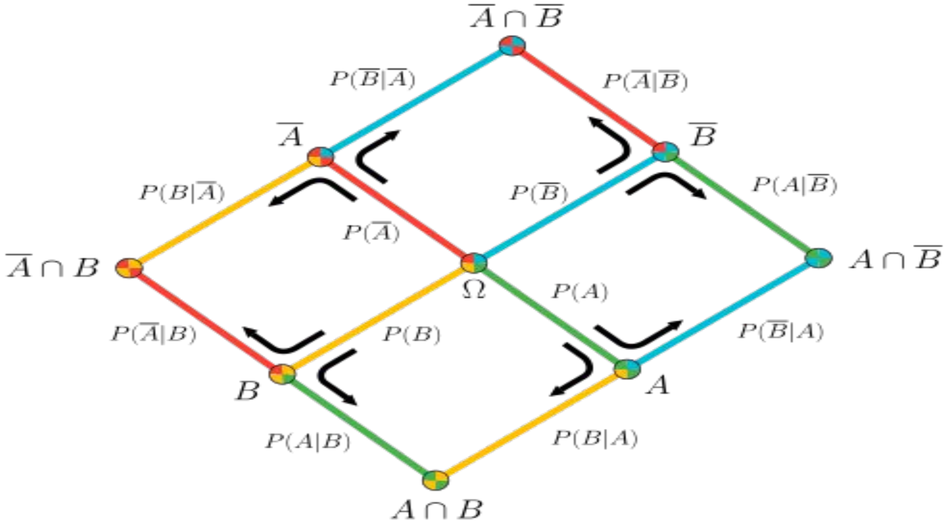
Probabilité conditionnelle

Soient A et B deux événements (*i.e* deux éléments d'une tribu).

P(A|B) (ou $P_B(A)$) est la probabilité de l'évènement A sachant l'évènement B

<u>Théorème de Bayes:</u>

$P_B(A) = P_A(B)P(A) / P(B)$



$P(A|B) \cdot P(B) = P(A \cap B) = P(B|A) \cdot P(A)$