Analyse de sensibilité:

Théorie des jeux et Tsunamis

Marc Abeille, Benjamin Osmont

ENS Cachan

30 juin 2011

Introduction

L'analyse de sensibilité propose une bibliothèque d'outils que nous allons appliquer à :

- Théorie des jeux
- Étude des tsunamis

Pourquoi ?

- Modélisation du problème $\rightarrow Y = f(X_1, \ldots, X_p)$
- Problème : f est trop complexe : expression analytique ou code numérique difficile à étudier de manière intrusive

Objectif :

- Comprendre comment les entrées agissent sur la sortie

1. Motivations
2. Outils
3. Résultats

1 Motivations
- Étude du regret
- Étude de la distribution des actions
- Inondation des tsunamis

2 Outils

3 Résultats

Étude du regret

On joue à un jeu stratégique répété : à chaque étape, deux joueurs jouent simultanément:

- Choix des actions : $a \in A$ et $b \in B$.
- Obtention d'un paiement : $g(a, b)$.

Objectif:

- Connaître ce qui influe sur le fait de "bien jouer".

Comment quantifier le fait de "bien jouer": r_n regret à l'étape n

$$r_n^a = \frac{1}{n}\left[\sum_{1 \leq i \leq n} g(a, b_i) - g(a_i, b_i)\right] \quad \forall a \in A$$

Choix du modèle:

- $X_1, \ldots, X_p$ = stratégie utilisée (paramétrée), nombre de répétition
- Y = regret

Problème:

- Caractère aléatoire des actions.

Solution proposée:

- ANOVA

Distribution des actions

On joue au même jeu, mais on ne s'intéresse pas à la même quantité.

Objectif :

- Étudier la façon de jouer d'un joueur.

Choix du modèle :

- $X_1, \ldots, X_p =$ paiement, nombre de répétition;
- $Y =$ matrice de distribution empirique des actions.

Solution proposée :

- Introduire une décomposition de la variance dans le cas matriciel.

Tsunamis

Tsunami du 25 octobre 2010 sur les îles Mentawai :

Figure: Mentawai island

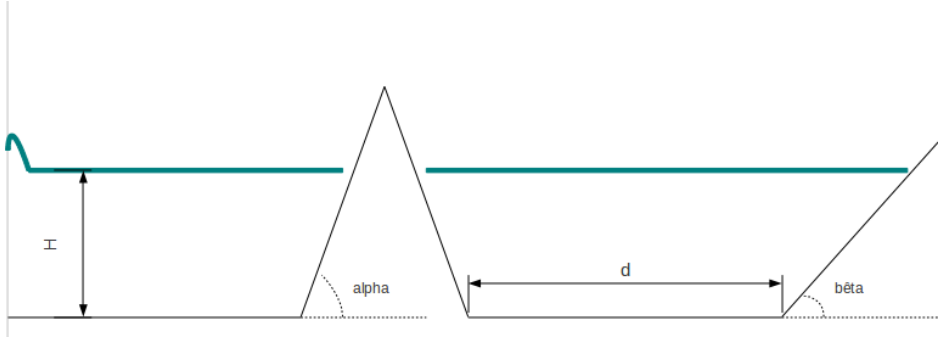

Figure: Cadre d'étude

Choix du modèle :

- Y = run-up
- $X_1, \ldots, X_p$ = paramètres de la topographie et de la vague

Tsunamis

Objectifs :

- Trouver le jeu de paramètres qui engendre des valeurs extrèmes de run-up
- Hiérarchiser les variables par ordre d'importance sur leur influence sur le run-up

Problème :

- Les équations qui régissent le problème ne sont pas résolvables analytiquement

Solution proposée :

- Simuler des expériences à l'aide de codes numériques
- Approximer la fonction f par des méthodes de régression

1 Motivations

2 Outils
- ANOVA
  - Cas réel
  - Cas matriciel
- Régression

3 Résultats

Décomposition de Sobol

L'ANOVA est basée sur le théorème de Sobol:

Décomposition de Sobol [Sensitivity estimates for nonlinear mathematical models, 1993]

Si f est intégrable sur $[0,1]^p$, elle admet une unique décomposition:

$f(x) = f_0 + \sum_{i=1}^{p} f_i(x_i) + \sum_{i<j} f_{i,j}(x_i, x_j) + \cdots + f_{1,2,\ldots,p}(x_1, \ldots, x_p)$ où f_0 est constante et les fonctions $f_{i_1,\ldots,i_s}$ sont orthogonales.

ANOVA

Par orthogonalité des composantes, on obtient en passant à la variance la décomposition ANOVA:

ANOVA

Si f est de carré intégrable sur $[0,1]^p$
$\mathbb{V}(f(X)) =$
$$\sum_{i=1}^{p} \mathbb{V}[f_i(X_i)] + \sum_{1 \leq i < j \leq p} \mathbb{V}[f_{i,j}(X_i, X_j)] + \cdots + \mathbb{V}[f_{1,\ldots,p}(X_1, \ldots, X_p)]$$

Que l'on réécrit en divisant par $\mathbb{V}[Y]$:

$$1 = \sum_{i=1}^{p} S_i + \sum_{i<j} S_{i,j} + \ldots + S_{1,\ldots,p}$$

$S_{i_1,\ldots,i_s}$: les indices de sensibilité par rapport à $x_{i_1}, \ldots, x_{i_s}$.

Décomposition de Sobol matricielle

La décomposition de Sobol n'a été introduite que dans le cas où la sortie est réelle. Nous avons étendu ce résultat.

Décomposition de Sobol dans le cas matriciel

Il existe une unique décomposition de f de la forme

$$f(X) = f_0 + \sum_{i=1}^{p} f_i(X_i) + \sum_{i<j} f_{i,j}(X_i, X_j) + \cdots + f_{1,2,\ldots,p}(X_1, \ldots, X_p) \text{ où } f_0$$

est constante, et vérifiant pour toute matrice B

$$\int_{\Omega} Tr({}^t B f_{i_1,\ldots,i_s}(X_{i_1}, \ldots, X_{i_s})) dP_{X_{i_k}} = 0 \quad \forall 1 \leq k \leq s$$

Idée de la preuve

$B \to \int_\Omega Tr({}^tBf(X))dP_X$ est une forme linéaire continue. Par le théorème de représentation de Riesz, $\exists! G$,
$\forall B \quad \int_\Omega Tr({}^tBf(X))dP_X = \langle B, G\rangle = Tr({}^tBG)$ On pose $f_0 = G$.
Soit $i \in \{1, \dots, p\}$ et $X_i \in \Omega$. Soit $h_{X_i} : B \to \int_\Omega Tr({}^tBf(X))\frac{dP_X}{dP_{X_i}}$. Par le théorème de représentation de Riesz, $\exists! G_i(X_i)$,
$\forall B \quad h_{X_i}(B) = \langle B, G_i(X_i)\rangle$. On construit ainsi une fonction G_i vérifiant:

$$\forall X_i \in \Omega, \forall B, \ h_{X_i}(B) = \langle B, G_i(X_i)\rangle$$

On pose alors $f_i(X_i) = G_i(X_i) - f_0$. On définit ainsi tous les termes de la somme à un indice.

ANOVA matricielle

On pose, sur l'espace des matrices aléatoires, la variance:

$$\mathbb{V}(X) = \mathbb{E}[Tr({}^tXX) - Tr({}^t\mathbb{E}(X)\mathbb{E}(X))]$$

Par la condition imposée à la décomposition:

$$\mathbb{V}(f(X)) = \sum_{i=1}^{p} \mathbb{V}[f_i(X_i)] + \sum_{1 \leq i < j \leq p} \mathbb{V}[f_{i,j}(X_i, X_j)] + \cdots + \mathbb{V}[f_{1,\ldots,p}(X_1, \ldots, X_p)]$$

$$\Leftrightarrow 1 = \sum_{i=1}^{p} S_i + \sum_{i<j} S_{i,j} + \ldots + S_{1,\ldots,p}$$

Méthodes de régression

Données :

- $X \in \mathbb{M}_{n,p}$: matrice des n valeurs prises par les p paramètres
- $Y \in \mathbb{R}^n$: vecteur des observations du run-up pour les n simulations

Objectif : Trouver $\tilde{f} = \underset{f \in \mathbb{F}}{\operatorname{argmin}}(||Y - f(X)||_2^2 + \lambda J(f))$ avec

- $\mathbb{F} = \{f(X) = X\beta, \beta \in \mathbb{R}^p\}$
- J la fonction de pénalité
- λ le coefficient de pénalité

Remarque: Possibilité de complexifier le modèle en ajoutant des variables $X_i = h(X_1, \ldots, X_p)$.

RIDGE

Méthodes d'estimation :

- **Méthode des moindres carrés** :
  - $J(f) = 0$
  - $\tilde{f}(X) = X\tilde{\beta}_{MC}$ avec $\tilde{\beta}_{MC} = (X^T X)^{-1} X^T Y$

  D'après le théorème de Gauss - Markov : $\tilde{\beta}_{MC}$ est le meilleur estimateur sans biais, mais peut avoir une grande variance.
- **Méthode RIDGE** :
  - $J(f) = ||\beta||_2^2$
  - $\tilde{\beta}_{RIDGE} = (X^T X + \lambda I_p)^{-1} X^T Y$

  Remarque:
  - $X^T X$ n'a plus besoin d'être inversible
  - Introduction de biais mais diminution de variance
  - Nécessité de determiner la valeur du coefficient λ

LASSO

- **Méthode LASSO** :
  - $J(f) = ||\beta||_1$
  - Pas d'expression analytique pour $\tilde{\beta}_{LASSO}$, nécessité de recourir à des algorithmes (LARS)

  Remarque: L'utilisation de la norme L_1 permet de rendre le système plus parcimonieux.

Nous utiliserons pour notre étude les méthodes RIDGE et LASSO, qui incluent l'estimateur des moindre carrés ($\lambda = 0$).

1 Motivations

2 Outils

3 Résultats
- Théorie des jeux
  - Étude du regret
  - Distribution des actions
- Étude des tsunamis
  - Jeu de paramètres optimum
  - Hiérarchisation des paramètres

Estimation des indices

Réalisation d'algorithme de simulation de jeu pour deux joueurs minimisant leur regret:

- Stratégie exponential weight de paramètre β
- Stratégie de Hart et Mas-Colell

Estimation numérique des indices de sensibilité

Étude du regret

- Stratégie exponential weight de paramètre β;
- $S_T = 0.027 \quad S_\beta = 0.906 \quad S_{T,\beta} = 0.067$;

$\rightarrow$ le regret est fortement dépendant de β et faiblement dépendant de T et de l'intéraction entre T et β.

Distribution des actions

Distribution empirique des actions pour un jeu de gain g:

- Stratégie de Hart et Mas-Colell;
- $S_T = 0.039 \quad S_g = 0.721 \quad S_{T,g} = 0.24$

$\rightarrow$ la façon de jouer du joueur 1 est fortement dépendante du paiement, et de l'interaction entre le paiement et le nombre de tour.

Estimation de la fonction

- 108 observations obtenues par simulations (VOLNA, Code 1D)
- 3 modèles (ajout de nouvelles variables de type X_1X_2)
- 2 méthodes (LASSO, RIDGE)

Validation croisée

Données					
K0	K1	K2	K3	K4	K5
Test	Expérience				

Figure: échantillonage

Choix du coefficient de pénalité

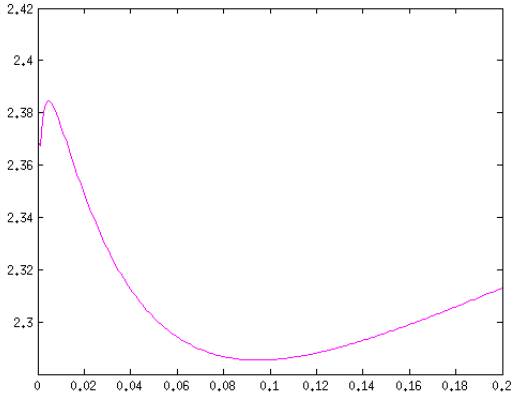

Figure: Erreur en fonction des valeurs de λ

Comparaison des méthodes et des modèles

Erreurs	Modèle 1	Modèle 2	Modèle 3
LASSO	0.8148	0.6333	0.8780
RIDGE	0.8112	0.6344	0.8642

$$f_1(h, d, \alpha, \beta, w) = a_0 + a_1 h + a_2 \alpha + a_3 \beta + a_4 d + a_5 w \tag{1}$$

$$f_2(h, d, \alpha, \beta, w) = a_0 + a_1 h + a_2 \alpha + a_3 \beta + a_4 d + a_5 w + a_6 w^2 + a_7 w^3 + a_8 w^4 \tag{2}$$

$$f_3(h, d, \alpha, \beta, w) = a_0 + a_1 h + a_2 \alpha + a_3 \beta + a_4 d + a_5 wh + a_6 w\beta + a_7 w\alpha + a_8 wd + a_9 w + a_{10} w^2 + a_{11} w^3 + a_{12} w^4 \tag{3}$$

On choisit le modèle 2 et la méthode LASSO ce qui nous donne :

$$\tilde{f}(h, d, \alpha, \beta, w) = 1.54 + 1.82h + 1.42\alpha - 3.62\beta - 0.96d + 2.42w - 1.44w^2 + 0.38w^3 - 0.04w^4 \tag{4}$$

Paramètres optimum

On utilise une méthode de Monte Carlo pour trouver le jeu de paramètres qui maximise la fonction $\tilde{f}$:

$\tan\beta_{max} =$	0.0387
$\tan\alpha_{max} =$	0.1971
$w_{max} =$	34.821 km
$h_{max} =$	1957 m
$d_{max} =$	31.63 m

Paramètres	Plage de valeur
$\tan\beta$	$[0.02, 0.25]$
$\tan\alpha$	$[0.02, 0.25]$
w	$[1000, 50000]$ m
h	$[100, 2000]$ m
d	$[0, 10000]$ m

qui entraine un run-up $\frac{R}{a_{init}} = 6.7197$

Formule analytique de C.E. Synolakis [10] $R = 2.831 d \sqrt{\cot\beta} \left(\frac{H}{d}\right)^{\frac{5}{4}} \Rightarrow$
$\frac{R}{a_{init}} = 0.9090$

Hiérarchisation des paramètres

On utilise une analyse de variance pour quantifier la variabilité de la sortie par rapport aux paramètres :

- $S_\beta = 62.1\%$
- $S_h = 15.7\%$
- $S_\alpha = 9.6\%$
- $S_w = 8.1\%$
- $S_d = 4.3\%$

Remarque:

- Les indices sont calculés de manière analytique puisque la fonction est maintenant connue
- On ne calcule pas les indices de corrélation, la somme étant déjà égale à 99.8% il seront négligeables

Conclusion

L'application de méthodes d'analyse de sensibilité à différents problèmes nous a permis de :

- Trouver le jeu de paramètre qui engendre des run-up extrêmes
- Hiérarchiser les paramètres influant sur le run-up
- Déterminer les paramètres sensibles pour :
  - La stratégie exponential weight
  - Les jeux de Hart et Mas-Colell

Bilan :

- Jeux à information partielle
- Amélioration de la fonction de régression

frame I
Bibliographie

D. Dutykh.
Modélisation mathématique des tsunamis.
PhD thesis, Ecole Nationale Supérieure de Cachan, 2007.

D. Dutykh, R. Poncet, and F. Dias.
The VOLNA code for the numerical modelling of tsunami waves: generation, propagation and inundation.
ArXiv e-prints, 2010.

J. Hannan.
Approximation to bayes risk in repeated play.
Princeton University Press, 1957.

S. Hart and A. Mas-Colell.
A simple adaptive procedure leading to correlated equilibrium.
Econometrica, 2000.

E. Hoerl and R. Kennard.
Ridge Regression biased estimation for nonotrhogonal problems.
Technometrics, 1970.

frame II
Bibliographie

M. Holtz.
Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance.
Springer, 2011.

J.F. Nash.
Equilibrium points in n-person games.
Proceedings of the National Academy of Sciences, 1949.

I.M. Sobol.
Sensitivity estimates for nonlinear mathematical models.

S. Sorin.
Exponential weight algorithm in continuous time.

C. E. Synolakis.
The run-up of solitary waves.
Journal of Fluid Mechanics, 1987.

R. Tibshirani.
Regression shrinkage and selection via the lasso.
Journal of the Royal Statistical Society Series B Methodological, 1996.

frame III
Bibliographie

P. Zhao, G. Rocha, and B. Yu.
The composite absolute penalties family for grouped and hierarchical variable selection.
ArXiv e-prints, 2009.

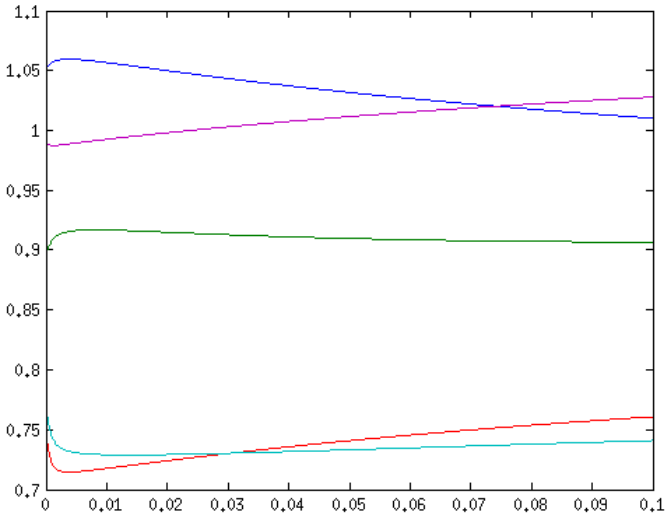

Figure: Erreurs sur K_i en fonction de λ

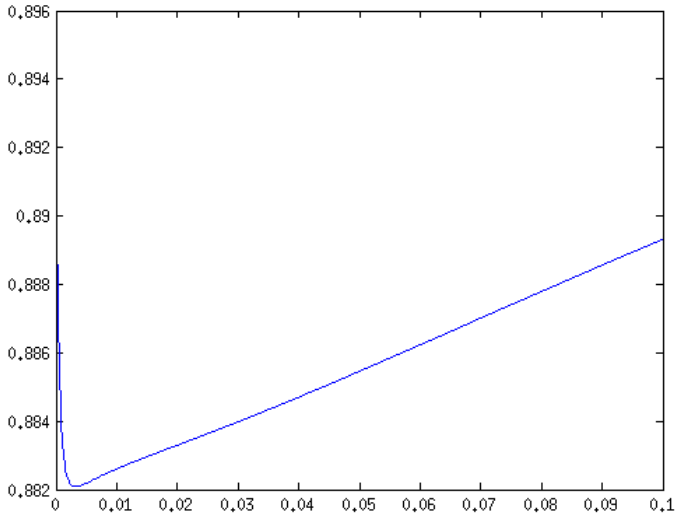

Figure: Moyenne des erreurs en fonction de λ

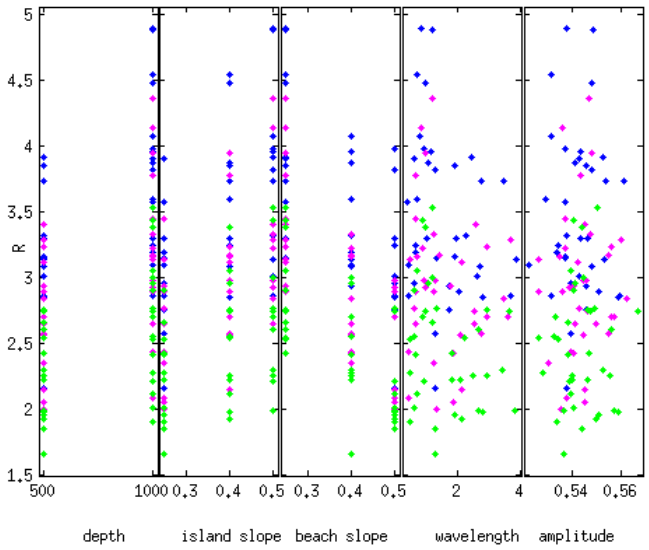

Figure: Choix du modèle

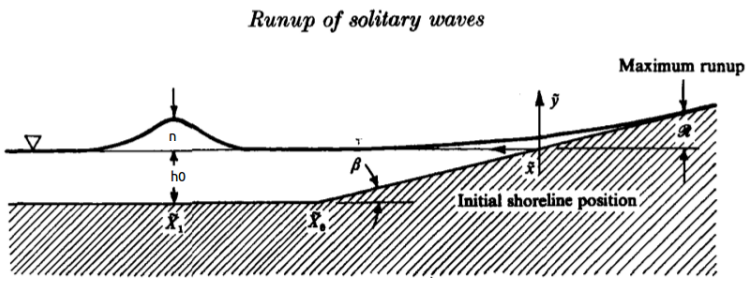

Figure: Schéma de run-up

Equations shallow-water :

$$H_t + \nabla \cdot (H\vec{u}) = 0 \tag{5}$$

$$(H\vec{u})_t + \nabla \cdot (H\vec{u} \otimes \vec{u} + \frac{g}{2}H^2) = gH\nabla h \tag{6}$$

(5) = Conservation de la masse, (6) = Conservation du moment.
où

- $H = h + \eta$
- $h =$ `modélise le fond marin`
- $\eta =$ `modélise la surface de l'eau`