

Aspects statistiques de la stabilité en dynamique des populations : application au modèle de Usher en foresterie.

Sous la direction d'Avner Bar-Hen et de Nicolas Picard

Mélanie Zetlaoui

7 décembre 2006

- 1 Introduction
- 2 Modèle de Usher
 - Définition
 - Propriétés du modèle de Usher
 - Applications
- 3 Lois asymptotiques des estimateurs des prédictions
 - Modèle de Usher linéaire
 - Modèle densité-dépendant
- 4 Robustesse des estimateurs des taux de transition
 - Cadre de l'étude
 - Étude de la robustesse
- 5 Conclusions et perspectives

Contexte forestier

- **Gestion** des forêts naturelles tropicales humides : régénération de la ressource et prélèvements.
- Formulation **statistique**.
- **Prédire** la dynamique de la population.

Modèles matriciels

Structure discrète :

- temps : $t = 0, 1, 2, \dots$
- distribution en états (âge, taille, ...) : C_1, \dots, C_I classes d'états.

Modèle : évolution du vecteur d'effectifs

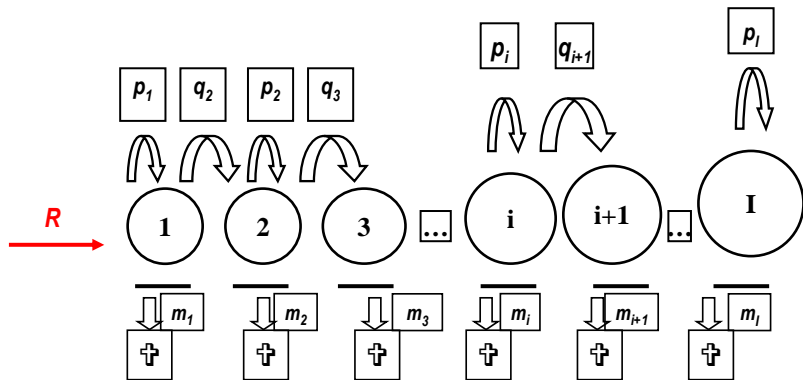
$$N(t+1) = (N_1(t+1), \dots, N_I(t+1)) = U(N, t)N(t)$$

- composantes : croissance, régénération et mortalité.

Différents types de modèles matriciels :

- modèles de Lefkovitch (1965) et de Leslie (1945) ;
- modèle de Usher (1966)

Description du modèle de Usher (1966)



$$p_i + q_{i+1} + m_i = 1$$

Hypothèses du modèle de Usher

Modèle linéaire :

- hypothèse markovienne ;
- hypothèse de Usher ;
- hypothèse d'homogénéité ;
- hypothèse d'indépendance.

Modèle densité-dépendant : la dynamique dépend de l'état courant de la population.

Stochasticités

- Stochasticité démographique :
 - $\mathcal{N}(t)$ chaîne de Markov de \mathbb{N}^I ;
 - Cas du modèle linéaire.
- Stochasticité environnementale (fertilité du sol, changements climatiques,...).
- Stochasticité d'échantillonnage.

Étude

Cadre :

- **prédictions** dans l'état stationnaire du modèle.

$$\text{Exemple : } N_{\infty} = \lim_{t \rightarrow \infty} N(t);$$

- stochasticité d'échantillonnage.

Objectifs :

- 1 construire des intervalles de confiance des prédictions :

$$\mathbb{P}(N_{\infty} \in [a; b]') \geq 1 - \alpha$$

- estimateur du maximum de vraisemblance ;
- modèle linéaire et densité-dépendant.

- 2 « optimiser » les intervalles de confiance asymptotiques :

- étude de la robustesse des estimateurs des paramètres ;
- efficacité des estimateurs.

- 1 Introduction
- 2 **Modèle de Usher**
 - Définition
 - Propriétés du modèle de Usher
 - Applications
- 3 Lois asymptotiques des estimateurs des prédictions
 - Modèle de Usher linéaire
 - Modèle densité-dépendant
- 4 Robustesse des estimateurs des taux de transition
 - Cadre de l'étude
 - Étude de la robustesse
- 5 Conclusions et perspectives

Modèle de Usher linéaire

l classes d'état, $N(t) = (N_i(t))_{i=1,\dots,l}$.

Modèle

$$N(t+1) = U N(t)$$

Matrice de transition :

$$U = \begin{pmatrix} p_1 + f_1 & f_2 & \cdots & f_l \\ q_2 & p_2 & & 0 \\ & \ddots & \ddots & \\ 0 & & q_l & p_l \end{pmatrix}$$

avec $0 < p_i + q_{i+1} < 1$ et $f_i > 0$.

Cas d'une fécondité moyenne : $f = f_1 = \dots = f_l$.

Autre formulation : $U = QS + R$, où :

- R : matrice de recrutement dans la première classe ;
- S : matrice diagonale de survie ;

$$\bullet Q = \begin{pmatrix} 1 - q_1^\bullet & & 0 & 0 \\ q_1^\bullet & \ddots & & \vdots \\ & \ddots & 1 - q_{l-1}^\bullet & 0 \\ 0 & & q_{l-1}^\bullet & 1 \end{pmatrix}$$

Modèle densité-dépendant

Modèle

$$N(t+1) = U(N(t)) N(t)$$

Forme des paramètres :

- fonction de Riker : $v(N(t)) = c \exp(-b N(t) \cdot a)$

avec $v(N(t))$ taux de vie, c et b constantes ;

- fonction de $N(t) \cdot a$: **indice de compétition**
avec a vecteur constant positif de \mathbb{R}^l .

Modèle agrégé

Lien avec les chaînes de Markov :

- Modèle stochastique : $\mathcal{N}(t)$ est une chaîne de Markov de \mathbb{N}^I .

Relation avec le modèle déterministe

$$\mathbb{E}[\mathcal{N}(t+1)|\mathcal{N}(t)] = U(\mathcal{N}(t))\mathcal{N}(t)$$

- Cas linéaire :
 - ★ $\mathbb{E}[\mathcal{N}(t+1)] = U\mathbb{E}[\mathcal{N}(t)]$;
 - ★ approx. « grande population » du modèle déterministe par le modèle stochastique (TCL).

Lien avec les équations aux dérivées partielles (EDP) :

approx. en temps continu d'une EDP par le modèle déterministe.

Comportement asymptotique du modèle linéaire

Hypothèse : U matrice irréductible aperiodique.

Théorème de Perron-Frobenius :

- première valeur propre unique $\lambda_1 > 0$;
- vecteur propre associé $w_1 > 0$.

État asymptotique :

$$\frac{N(t)}{\lambda_1^t} \xrightarrow{t \rightarrow \infty} c_1 w_1, \quad \text{avec } c_1 \in \mathbb{R}$$

- état d'équilibre ;
- taux de croissance asymptotique : λ_1 ;
- distribution stationnaire : w_1 .

Comportement asymptotique du modèle densité-dépendant

- État stationnaire :

existence d'un point d'équilibre N

$$N = U(N) N$$

- Cycles, pseudo-cycles, chaos.
- Diagrammes de bifurcation.

Applications : données de Paracou

- Site expérimental : Paracou (Guyane Française).
- État : diamètre des arbres.
- Pas de temps : 2 ans.
- Années 1984 et 1986 : pas d'exploitation (45 732 obs.).
- Années 1988 et 1990 : exploitation du site (39 799 obs.).
- Largeur des classes de diamètre : 5cm.

- 1 Introduction
- 2 Modèle de Usher
 - Définition
 - Propriétés du modèle de Usher
 - Applications
- 3 Lois asymptotiques des estimateurs des prédictions
 - Modèle de Usher linéaire
 - Modèle densité-dépendant
- 4 Robustesse des estimateurs des taux de transition
 - Cadre de l'étude
 - Étude de la robustesse
- 5 Conclusions et perspectives

Prédictions du modèle

Fonctions implicites des paramètres :

- *taux de croissance* $\lambda_1 = h(U)$,

h définie par : $\Phi(U, h(U)) = 0$ avec $\Phi(U, \lambda) = \det(U - \lambda I)$;

- *distribution stationnaire* $w_1 = g(U) \in \mathbb{R}^I$,

g déterminée par : $Ug(U, \lambda_1) - \lambda g(U, \lambda_1) = 0$.

Modèle statistique

$d = (d_1, \dots, d_l)$ distribution en taille à l'instant initial.

- X vecteur aléatoire dans $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$ avec :

$$\mathcal{A}_1 = \{(i, j, l); 1 \leq i \leq l-1; j = i, i+1; l \in \mathbb{N}\}$$

$$\mathcal{A}_2 = \{(l, l, l); l \in \mathbb{N}\}$$

$$\mathcal{A}_3 = \{(i, \dagger, l); 1 \leq i \leq l; l \in \mathbb{N}\}$$

où \dagger : état mort

- Loi F de X :

$$\mathbb{P}[X = (i, i, l)] = p_i d_i g_{il}$$

$$\mathbb{P}[X = (i, i+1, l)] = q_{i+1} d_i g_{il}$$

$$\mathbb{P}[X = (i, \dagger, l)] = (1 - p_i - q_{i+1}) d_i g_{il}$$

où $g_{il} = \mathbb{P}(Y = l)$ avec $Y \sim \mathcal{P}(f_i)$.

Loi asymptotique des estimateurs des prédictions

(X_1, \dots, X_n) n -échantillon de loi F ; $\theta = (p_i; q_i; f_i; d_i) \in \mathbb{R}^{4I-1}$.

- θ identifiable; EMV : estimateur empirique.
- Comportement asymptotique de $\hat{\theta}_n$:

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N} (0, I(\theta)^{-1})$$

avec $I(\theta)$ l'information de Fisher du modèle statistique.

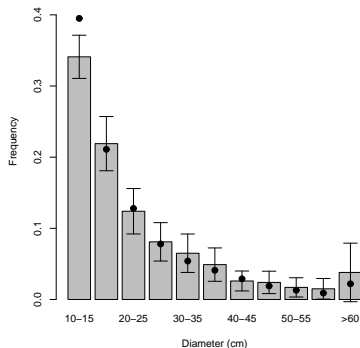
Delta-méthode \Rightarrow comportement asymptotique de $\hat{\lambda}_{1n}$ et \hat{w}_{1n}

$$\sqrt{n} (\hat{\lambda}_{1n} - \lambda_1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2) \quad \text{avec } \sigma^2 = (d_\theta h)^t \cdot I(\theta)^{-1} \cdot (d_\theta h)$$

$$\sqrt{n} (\hat{w}_{1n} - w_1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma) \quad \text{avec } \Gamma = (d_\theta g)^t \cdot I(\theta)^{-1} \cdot (d_\theta g)$$

Application

- $\hat{\lambda}_1 = 1.002$, $\hat{\sigma}^2 = 0.035$. Test sur λ_1 à 5% : $\lambda_1 > 1$.



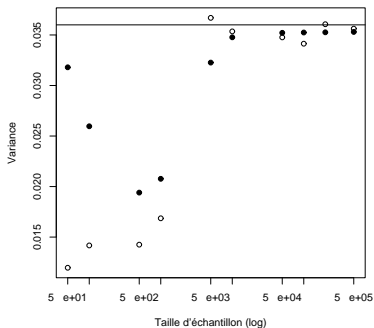
Matrice de corrélation de \hat{w}_1 :

$$\begin{pmatrix} 1.00 & -0.76 & - \\ -0.76 & 1.00 & - \\ - & - & + \end{pmatrix}$$

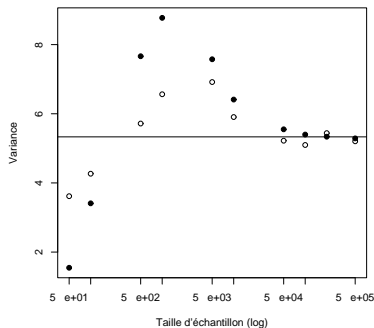
Valeurs de \hat{w}_1 ; • distribution en 1984.

- Précision de w_1 en fonction de la taille des données.

Simulations



(a)



(b)

Fig.: Comparaison de la variance empirique et de l'estimation de la variance asymptotique pour (a) λ_1 et (b) w_1 ; 1000 simulations.

Modèle densité-dépendant

- Cas d'une fécondité moyenne.
- Modèle général :

$$p_i(N(t)) = \mu_i \exp(-\xi_i N(t).a) \quad (i = 1...l)$$

$$q_i(N(t)) = \nu_i \exp(-\kappa_i N(t).a) \quad (i = 2...l)$$

$$f(N(t)) = \alpha \exp(-\beta N(t).a)$$

- Modèle « simplifié » (Favrichon, 1998) :

$$p_i(N(t)) = (a_0 + a_1 i + a_2 i^2) \exp(-\xi N(t).a) \quad (i = 1...l)$$

$$q_i(N(t)) = (b_0 + b_1 i + b_2 i^2) \exp(-\kappa N(t).a) \quad (i = 2...l)$$

$$f(N(t)) = \alpha \exp(-\beta N(t).a)$$

Prédiction du modèle

- **Hypothèse** : existence et unicité du point d'équilibre N_∞ , tel que $N_\infty = U(N_\infty)N_\infty$.
- N_∞ défini comme fonction implicite des paramètres du modèle :

$$\Psi(N_\infty, \theta) = 0$$

avec θ vecteur des paramètres et $\Psi(N, \theta) = N - U(N)N$.

Modèle statistique

Hypothèses :

- données non longitudinales ;
- K parcelles indépendantes ;
- $n = \sum_{k=1}^K n_k$ taille de l'échantillon.

Modèle statistique :

- modèle de mélange ;
- proportions du mélange fixées égales à $n_1/n, \dots, n_K/n$.

Conditionnellement à la parcelle k , la loi de X est :

$$\begin{aligned} \Pr[X = (i, i, k)] &= (1 - f_k^*) p_i(N(k)) d_{ik} \\ \Pr[X = (i, i + 1, k)] &= (1 - f_k^*) q_i(N(k)) d_{ik} \\ \Pr[X = (i, \dagger, k)] &= (1 - f_k^*) (1 - p_i(N(k)) - q_i(N(k))) d_{ik} \\ \Pr[X = (0, 1, k)] &= f_k^* \end{aligned}$$

Estimation des paramètres

- Vraisemblance :

$$\ell_{\theta}(x) = \sum_{k=1}^K n_k/n \left\{ \sum_{i=1}^I (1 - f_k^*) p_{ik} d_{ik} \mathbf{1}_{x=(i,i,k)} + (1 - f_k^*) q_{ik} d_{ik} \mathbf{1}_{x=(i,i+1,k)} \right. \\ \left. + (1 - f_k^*) (1 - p_{ik} - q_{ik}) d_{ik} \mathbf{1}_{x=(i,\dagger,k)} + f_k^* \mathbf{1}_{x=(0,1,k)} \right\}$$

- EMV : maximisation numérique de la log-vraisemblance.
- Normalité asymptotique de l'EMV.
- Test sur la densité-dépendance et sélection de modèle.

Application

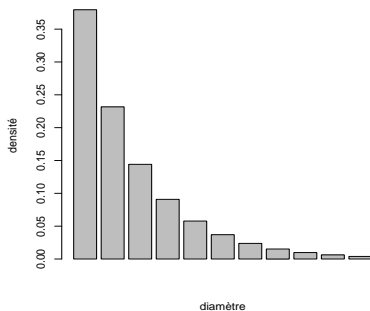
- Maximisation de la vraisemblance : maxima locaux.

Paramètre	Estimation	p -value
a_0	0.7445	
a_1	$-4.7152 \cdot 10^{-3}$	
a_2	$4.0030 \cdot 10^{-5}$	$> 10^{-10}$
ξ	$-2.3091 \cdot 10^{-4}$	$> 10^{-10}$
b_0	0.3370	
b_1	$2.4462 \cdot 10^{-2}$	
b_2	$-7.9992 \cdot 10^{-4}$	$2 \cdot 10^{-11}$
κ	$2.0721 \cdot 10^{-3}$	$> 10^{-10}$

Estimation dans le modèle simplifié.

$$p_i(N(t)) = (a_0 + a_1 i + a_2 i^2) \exp(-\xi N(t) \cdot \mathbf{a})$$

$$q_i(N(t)) = (b_0 + b_1 i + b_2 i^2) \exp(-\kappa N(t) \cdot \mathbf{a})$$



Valeurs de N_∞ .

- 1 Introduction
- 2 **Modèle de Usher**
 - Définition
 - Propriétés du modèle de Usher
 - Applications
- 3 **Lois asymptotiques des estimateurs des prédictions**
 - Modèle de Usher linéaire
 - Modèle densité-dépendant
- 4 **Robustesse des estimateurs des taux de transition**
 - Cadre de l'étude
 - Étude de la robustesse
- 5 Conclusions et perspectives

Étude

Procédure robuste (Huber, 1981) :

- rendre peu sensible les estimateurs des paramètres d'intérêt aux erreurs de mesures ;
- perturber le modèle d'échantillonnage initial.

Robustesse dans le modèle de Usher :

- contraintes du modèle :
 - ★ structure discrète ;
 - ★ hypothèse de Usher.
- influence sur les prédictions.

Objectifs

- alléger l'hypothèse de Usher ;
- minimiser les impacts des erreurs de mesure sur la dynamique.

Traitement des données ne respectant pas l'hypothèse de Usher :

- ★ les éliminer ;
- ★ les lisser.

Minimiser les erreurs de mesure :

- ★ troncature sur les accroissements ;
- ★ troncature sur les bornes des classes.

⇒ **L-estimateurs**

Modèle statistique

Classe $C = [l_0, l_1[$ et $C' = [l_1, l_2[$ classe supérieure.

- X taille au temps t de C .
- ΔX accroissement en taille entre t et $t + 1$.

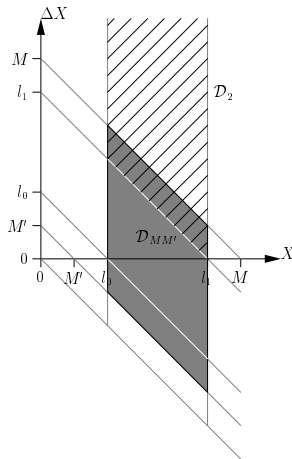
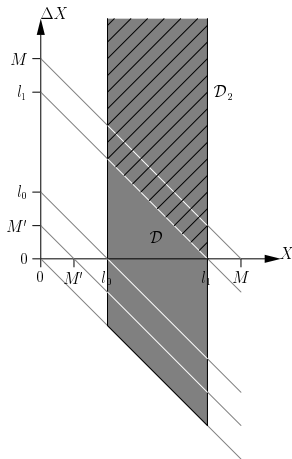
Loi de $(X, \Delta X)$

H de densité $h(x, y) = g(y|x)f(x)$

avec :

- ★ f densité de la loi de X ;
 - ★ $g(.|X)$ densité de la loi de $\Delta X|X$.
- Étude de q^\bullet : paramètre de transition sachant que l'individu reste vivant.

Données ne respectant pas l'hypothèse de Usher



Minimiser les erreurs de mesure

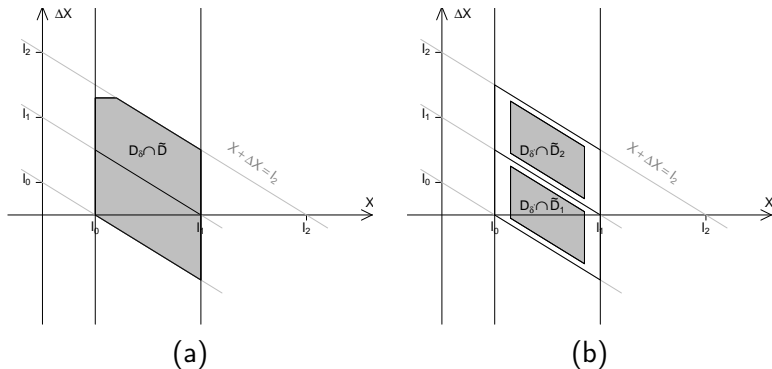


Fig.: Troncature sur la loi des (a) accroissements et (b) bornes des classes.

Définition des taux de transitions

Données violant l'hypothèse de Usher : $0 \leq M' \leq l_0$ et $l_1 < M$

$$q_{MM'}^\bullet = \mathbb{P}(l_1 \leq X + \Delta X | M' \leq X + \Delta X < M)$$

$$q_\infty^\bullet = \mathbb{P}(l_1 \leq X + \Delta X)$$

Impacts des erreurs de mesure :

$$q_1^\bullet = \mathbb{P}(l_1 \leq X + \Delta X | l_0 \leq X + \Delta X < l_2 \text{ et } \Delta X < l_2 - l_0 - \delta)$$

$$q_2^\bullet = \mathbb{P}(l_1 \leq X + \Delta X | X \in I_{\delta'} \text{ et } X + \Delta X \in J_{\delta'})$$

avec : $I_{\delta'} =]l_0 + \delta'; l_1 - \delta'[, J_{\delta'} =]l_0 + \delta'; l_1 - \delta' [\cup]l_1 + \delta'; l_2 - \delta' [$.

δ et $\delta' \Leftrightarrow$ enlever une proportion α et α' d'individus

Fonction d'influence

$(X_1 \dots X_n)$ de loi F , de loi empirique F_n .
 T_n estimateur tel que : $T_n = T(F_n)$.

Fonction d'influence (Hampel, 1974)

T Gâteaux-différentiable

$$IC_{T,F}(x) = \lim_{\epsilon \rightarrow 0} \frac{T((1 - \epsilon)F + \epsilon\delta_x) - T(F)}{\epsilon}$$

- Sensibilité globale : $\gamma = \sup_x \|IC_{T,F}(x)\|_{\infty}$.
- Normalité asymptotique (*Fréchet- ou Hadamard-différentiabilité de T*).

Fonctionnelles

$\mathcal{D}_{MM'}$, \mathcal{D}_δ et $\mathcal{D}_{\delta'}$:= domaines de troncature.

\mathcal{D}_1 , \mathcal{D}_2 := domaines d'évolution.

$$T_{MM'}(H) = \frac{\iint_{\mathcal{D}_{MM'} \cap \mathcal{D}_2} h(x, y) dx dy}{\iint_{\mathcal{D}_{MM'}} h(x, y) dx dy}$$

$$T(H) = \iint_{\mathcal{D}_2} h(x, y) dy dx$$

$$T_1(H) = \frac{\iint_{\mathcal{D}_\delta \cap \tilde{\mathcal{D}}_2} h(x, y) dx dy}{\iint_{\mathcal{D}_\delta \cap \tilde{\mathcal{D}}} h(x, y) dx dy}$$

$$T_2(H) = \frac{\iint_{\mathcal{D}_{\delta'} \cap \tilde{\mathcal{D}}_2} h(x, y) dx dy}{\iint_{\mathcal{D}_{\delta'} \cap \tilde{\mathcal{D}}} h(x, y) dx dy}$$

Influence des taux de transition

Proposition :

$$IC_{T_{MM'}, H}((u, v)) = \begin{cases} -\frac{T(H) - \alpha_M}{(1 - \alpha_{MM'})^2} & (u, v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_1 \\ \frac{1 - \alpha_{M'} - T(H)}{(1 - \alpha_{MM'})^2} & (u, v) \in \mathcal{D}_{MM'} \cap \mathcal{D}_2 \\ 0 & (u, v) \in \mathcal{D} \setminus \mathcal{D}_{MM'} \end{cases}$$

$$IC_{T, H}((u, v)) = \begin{cases} -T(H) & (u, v) \in \mathcal{D}_1 \\ 1 - T(H) & (u, v) \in \mathcal{D}_2 \end{cases}$$

Proposition :

$$IC_{T_1, H}((u, v)) = \begin{cases} -\frac{T(H) - \alpha_{l_2}}{(1 - \alpha)(1 - \alpha_{l_2 l_0})^2} & \text{pour } (u, v) \in \tilde{\mathcal{D}}_1 \\ \frac{1 - \alpha_{l_0} - T(H)}{(1 - \alpha)(1 - \alpha_{l_2 l_0})^2} & \text{pour } (u, v) \in \tilde{\mathcal{D}}_2 \\ \text{et } v \neq l_2 - l_0 - \delta \\ -\frac{T(H) - \alpha_{l_2} - \alpha(1 - \alpha_{l_2 l_0})}{(1 - \alpha)(1 - \alpha_{l_2 l_0})^2} & \text{pour } (u, v) \in \tilde{\mathcal{D}}_2 \\ \text{et } v = l_2 - l_0 - \delta \\ 0 & \text{sinon} \end{cases}$$

Fonction d'influence de T_2 : constante sur plusieurs domaines différents.

Sensibilités et variabilité

- Sensibilité globale :

Proposition :

$$\gamma_{MM'} \geq \gamma \quad \forall 0 \leq M' \leq l_0, \forall M > l_1$$
$$\gamma_1 \geq \gamma_{l_2 l_0}$$

- Variance asymptotique :

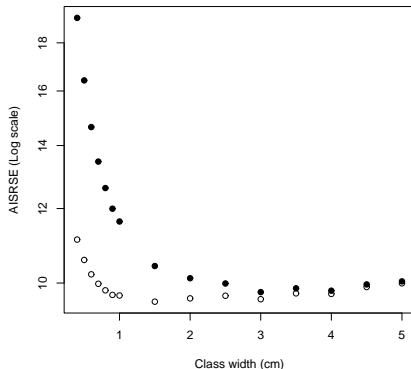
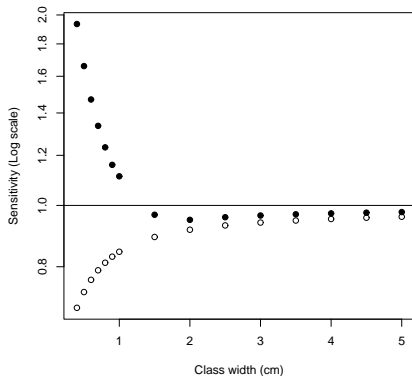
$$V_1 \leq V_{l_2 l_0}$$

Comparaison de $V_{MM'}$ et V : dépend de la proportion des données violant l'hypothèse de Usher.

Application

- Ajustement des lois :
 - ★ distribution diamétrique exponentielle ;
 - ★ accroissement diamétrique log-normale.
- Comparaison de l'estimateur classique $q_{l_2}^\bullet$ et de l'estimateur robuste q_∞^\bullet .

Résultats



- estimateur classique
- estimateur robuste

Conclusions

- Estimation des paramètres : prise en compte de la variabilité d'échantillonnage.
- Cadre d'étude :
 - ★ échantillon de grande taille ;
 - ★ variables aléatoires décrivant les évolutions individuelles indépendantes.
- Estimateurs :
 - ★ EMV ;
 - ★ estimateurs robustes : L -estimateurs.
- Méthodes :
 - ★ efficacité des estimateurs ;
 - ★ delta-méthode.

Perspectives

- Extensions du travail actuel :
 - ★ étendre à d'autres types de prédictions du modèle matriciel ;
 - ★ étendre à d'autres types d'estimateurs des paramètres ;
 - ★ étendre à d'autres types de modèles matriciels (Leslie, Lefkovich...).
- Extension à un cadre d'étude plus général :
 - ★ estimation dans un cadre d'observations longitudinales ;
 - ★ prendre en compte d'autres types de stochasticité ;
 - ★ cas de petits échantillons.

Références

- Zetlaoui, M., Picard, N. and Bar-Hen, A., 2006. Asymptotic distribution of stage-grouped population models. *Mathematical Biosciences* 200, 76-89.
- Zetlaoui, M., Picard, N. and Bar-Hen, A., 2006. Robustness of the estimators of transition rates for stage-classified matrix models. *Comp. Stat. & Dat. Analys.* (Sous presse).
- Zetlaoui, M., Picard, N. and Bar-Hen, A.. Asymptotic distribution of density-dependant stage-grouped population dynamics models (Soumis).

Remerciements

Rapporteurs

Patrice Bertail, Jérôme Saracco

Présidente

Elisabeth Gassiat

Directeurs de thèse

Avner Bar-Hen, Nicolas Picard

Examineurs

Alain Franc, Sylvie Gourlet-Fleury