Université Claude Bernard Lyon 1
Institut de Science Financière et d'Assurances

Modélisation avancée en mortalité

Vraisemblance locale & méthodes avancées

Julien Tomas
Institut de Science Financière et d'Assurances
Laboratoire de recherche de Sciences Actuarielle et Financière

Table des matières

Table des matières

Table des matières

Table des matières

Notations

Durée de vie restante

- Soit $T_x(t)$ la durée de vie restante d'un individu âgé x dans l'année calendaire t, i.e.

$$\Pr\big[T_x(t) > \xi\big] = \Pr\big[T(t) > x + \xi | T(t) > x\big] = {}_{\xi}p_x(t).$$

- Donc, un individu âgé x dans l'année calendaire t décèdera à l'âge $x + T_x(t)$.
- La fonction de distribution de $T_x(t)$ est

$${}_{\xi}q_x(t) = 1 - {}_{\xi}p_x(t) = \Pr\big[T_x(t) < \xi\big] = \Pr\big[T(t) \leq x + \xi | T(t) > x\big].$$

Notations

Probabilité de survie et de décès durant l'année

- La probabilité annuelle de décès pour un individu âgé x dans l'année calendaire t est définie par

$$q_x(t) = \Pr\big[T_x(t) \leq 1\big] = \Pr\big[T(t) \leq x+1 | T(t) > x\big].$$

- La probabilité annuelle de survie pour un individu âgé x dans l'année calendaire t est

$$p_x(t) = \Pr\big[T_x(t) > 1\big] = \Pr\big[T(t) > x+1 | T(t) > x\big].$$

- On a donc pour un entier k,

$${}_kp_x(t) = \Pr\big[T_x(t) > k\big] = p_x(t) \times p_{x+1}(t) \times \cdots \times p_{x+k-1}(t).$$

Notations

Nombre d'individus, de décès, exposition et forces de mortalité

- On note $L_{x,t}$, le nombre d'individus (l'effectif) vivants à l'âge x dans l'année calendaire t.
- $D_{x,t} = L_{x,t} - L_{x+1,t+1}$, est le nombre de décès à l'âge x dans l'année calendaire t.
- L'exposition au risque à l'âge x durant l'année calendaire t mesure le temps durant lequel les individus sont exposés au risque (de décès). Il s'agit de la durée totale vécue par ces individus durant la période d'observation,

$$E_{x,t} = \int_{\xi=0}^{1} L_{x+\xi,t+\xi}\, d\xi.$$

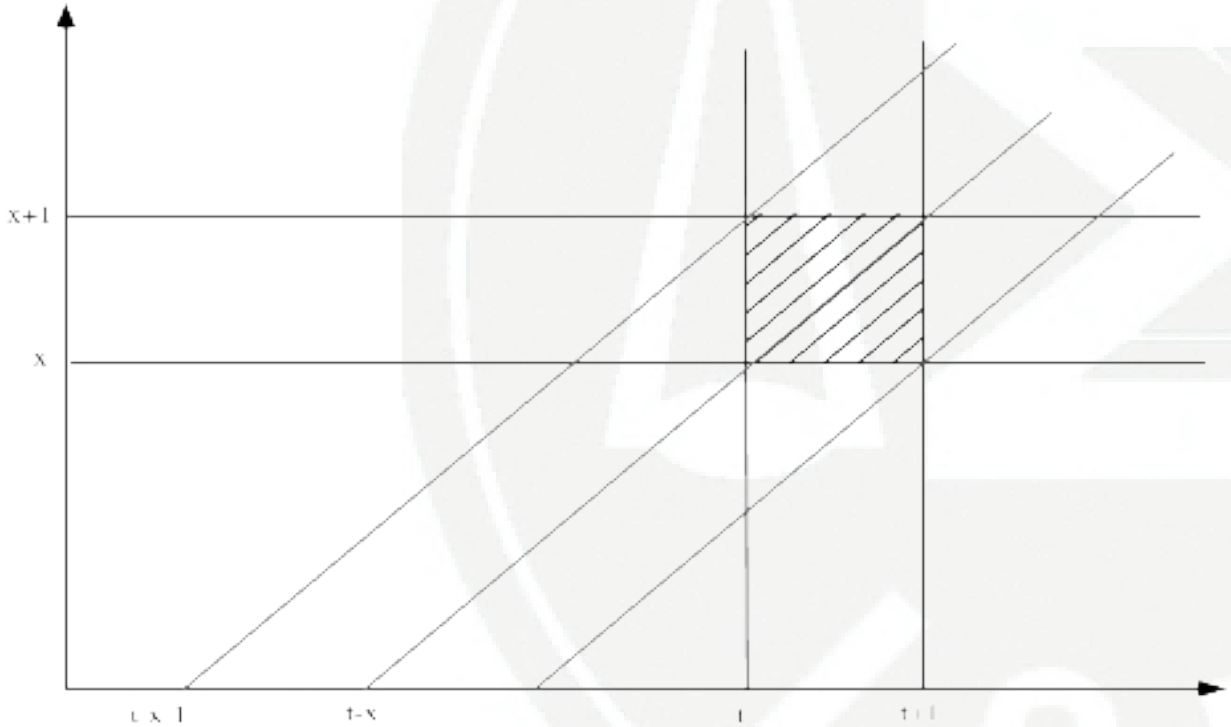
- La force de mortalité à l'âge $x+\tau$ dans l'année calendaire t est définie par

$$\varphi_{x+\tau}(t) = \lim_{\Delta\tau \to 0^+} \frac{\Pr\left[\tau < T_x(t) \leq \tau + \Delta\tau \middle| T_x(t) > \tau\right]}{\Delta\tau}$$

Hypothèse de mortalité constante par morceaux

- Dorénavant, on fait l'hypothèse de constance des forces de mortalité dans chaque carré unitaire du diagramme de Lexis :

$$\varphi_{x+\xi}(t+\zeta) = \varphi_x(t) \text{ pour } 0 \leq \xi < 1 \text{ et } 0 \leq \zeta < 1 \text{ et } x, t \text{ entiers.}$$



- Sous cette hypothèse, $p_x(t) = \exp\Big(-\int_0^1 \underbrace{\varphi_{x+\xi}(t+\xi)}_{=\varphi_x(t)}\Big) = \exp\big(-\varphi_x(t)\big)$.

Lien avec la loi de Poisson

Modélisation

- À chacune des observations i, on associe une indicatrice δ_i indiquant si l'individu est décédé ou non,

$$\delta_i = \begin{cases} 1 & \text{si l'individu } i \text{ est décédé,} \\ 0 & \text{sinon,} \end{cases}$$

  pour $i = 1, \ldots, L_{x,t}$.

- On définit τ_i, le temps durant lequel l'individu i est observé (c'est l'exposition au risque).
- On suppose avoir à disposition pour chacun des $L_{x,t}$ individus les observations (δ_i, τ_i).
- Ainsi,

$$\sum_{i=1}^{L_{x,t}} \tau_i = \mathrm{E}_{x,t} \text{ et } \sum_{i=1}^{L_{x,t}} \delta_i = D_{x,t}.$$

Lien avec la loi de Poisson

Contribution de l'individu i à la vraisemblance

- La contribution du $i$ème individu à la vraisemblance s'écrit,
  - si l'individu survie à son $(x+1)$ème anniversaire $(\delta_i = 0, \tau_i = 1)$ alors : $p_x(t) = \exp\big(-\varphi_x(t)\big)$;
  - si l'individu décède avant son $(x+1)$ème anniversaire $(\delta_i = 1, \tau_i < 1)$ alors : ${}_{\tau_i}p_x(t)\ \varphi_{x+\tau_i}(t) = \exp\big(-\tau_i\varphi_x(t)\big)\varphi_x(t)$.
- La contribution de l'individu i à la vraisemblance vaut donc

$$\exp\big(-\tau_i\varphi_x(t)\big)\big(\varphi_x(t)\big)^{\delta_i}.$$

- La vraisemblance devient

$$\begin{aligned}\mathcal{L}\big(\varphi_x(t)\big) &= \prod_{i=1}^{L_{x,t}} \exp\big(-\tau_i\varphi_x(t)\big)\big(\varphi_x(t)\big)^{\delta_i} \\ &= \exp\big(-E_{x,t}\ \varphi_x(t)\big)\big(\varphi_x(t)\big)^{D_{x,t}},\end{aligned}$$

- et la log vraisemblance associée est

$$\log\mathcal{L}\big(\varphi_x(t)\big) = -E_{x,t}\ \varphi_x(t) + D_{x,t}\log\varphi_x(t).$$

Lien avec la loi de Poisson

Cadre de travail des modèles linéaires généralisés

- En Maximisant la log vraisemblance $\log \mathcal{L}\big(\varphi_x(t)\big)$ on obtient l'estimateur du maximum de vraisemblance de $\varphi_x(t)$, à savoir

$$\widehat{\varphi}_x(t) = D_{x,t}/E_{x,t}$$

  qui coïncide avec le taux de mortalité $\widehat{m}_x(t)$.

- La vraisemblance $\mathcal{L}\big(\varphi_x(t)\big)$ est proportionnelle à la vraisemblance de Poisson basée sur $D_{x,t} \sim \mathcal{P}\big(E_{x,t}\,\varphi_x(t)\big)$.

- Il est équivalent de travailler avec la *vraie* vraisemblance ou à partir d'une vraisemblance de Poisson.

- Ainsi sous l'hypothèse de la mortalité constante par morceaux entre des valeurs non-entières de x et t, on considère

$$D_{x,t} \sim \mathcal{P}\big(E_{x,t}\,\varphi_x(t)\big), \tag{1}$$

  pour pouvoir utiliser le cadre de travail des modèles linéaires généralisés (GLMs).

Le lissage des données

Irrégularités observées dans la progression des taux

- Les estimations brutes, sur lesquelles se basent les tables de mortalité, peuvent être considérées comme un échantillon provenant d'une population plus importante et sont, par conséquent, soumises à des fluctuations aléatoires.
- Toutefois, l'actuaire souhaite la plupart du temps lisser ces quantités afin de mettre en avant les caractéristiques de la mortalité du groupe qu'il pense être régulières.
- Ces irrégularités dans la progression des forces de mortalité pourraient être réduites en augmentant l'exposition $E_{x,t}$ observée.
- Une méthode plus pratique est de graduer les données pour éliminer partiellement ces erreurs aléatoires.

Le lissage des données

Approche paramétrique VS non-paramétrique I

Plusieurs approches de lissage peuvent être adoptées.

- Approches paramétriques, (expression mathématique qui décrit la mortalité en fonction de paramètres).
- Approches non-paramétriques

La relation entre les forces de mortalité, l'âge et l'année calendaire peut être modélisée par

$$\varphi_i = \psi(x_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

où ψ est une fonction de régression et ϵ_i un terme d'erreur représentant les erreurs aléatoires dans les observations ou variations qui ne sont pas incluses dans les x_i.

Le lissage des données

Approche paramétrique VS non-paramétrique II

- Les approches paramétriques supposent que la fonction de réponse ψ a une forme pré-spécifiée. La relation est pleinement décrite par un nombre fini de paramètres. Cela simplifie les calculs et permet par exemple l'extrapolation. Mais pour être utiles, elles doivent reproduire étroitement les données.
- Un modèle paramétrique pré-sélectionné peut être trop restrictif pour modéliser les caractéristiques de la mortalité.
- La modélisation non-paramétrique offre un outil flexible dans l'analyse de la relation. Comme les méthodes paramétriques, elles sont susceptibles de donner des estimations biaisées, mais de telles sorte qu'il est possible d'équilibrer une augmentation du biais avec une diminution de la variation d'échantillonnage.

Le lissage des données

Natura non agit per saltum : L'idée basique du lissage

- Chaque force de mortalité est liée étroitement à ses voisines.
- Les observations φ_j dans le voisinage de φ_i contiennent de l'information à propos de la valeur de ψ à x_i.
- Les forces de la nature opèrent graduellement et leurs effets deviennent visibles de façon continue et non par des sauts brusques
- Cela implique que les observations φ_j, dans un rayon d'un point x_i, peuvent être utilisées pour augmenter l'information que nous avons à x_i et une estimation améliorée de φ_i peut être obtenue en lissant les estimations individuelles φ_j.

Le lissage des données

Natura non agit per saltum : L'idée basique du lissage

- Cette procédure d'approximation de la fonction de réponse ψ est communément appelée lissage.
- Ainsi la mortalité n'est pas résumée par un petit nombre de paramètres mais décrite par les n forces de mortalité.
- Dans la littérature actuarielle, ce procédé est connu sous le terme de graduation.
- Les petites collines et vallées des données brutes sont graduées jusqu'à devenir lisses, comme lorsque l'on construit une route sur un terrain accidenté.

Méthodes de vraisemblance locale

Développement historique I

- Les méthodes de lissages se sont développées à des périodes et dans des pays différents à la fin du 18ème. La plupart sont apparues dans des études actuarielles. Les taux de mortalité et de maladies étaient lissés selon une fonction de l'âge.
- Premières références : Johann Lambert en 1765, John Finlaison en 1823, Wesley Woolhouse en 1866, De Forest en 1873, Thomas Sprague en 1887, John Spencer en 1904, Robert Henderson en 1916 et Edmund Whittaker en 1923.
- Cependant les régressions locales ont reçu que peu d'attention jusqu'à la fin des années 1970, cf. Seal (1982), Haberman (1996) et Loader (1999) pour une revue historique.

Méthodes de vraisemblance locale

Développement historique II

- Les méthodes de régression locale sont originaires des méthodes à noyaux introduites par Rosenblatt (1956) et Parzen (1962).
- La méthode à noyaux est un cas spécial de la régression locale où la famille paramétrique est une fonction constante.
- Applications actuarielles des méthodes à noyaux : Copas and Haberman (1983) et Gavin *et al.* (1993).
- Les régressions locales ont connu un regain d'intérêt après les développements de Stone (1977) et Stone (1982) et la procédure *loess* de Cleveland (1979).
- Tibshirani and Hastie (1987) ont introduit la procédure de vraisemblance locale, et ont étendu le domaine des méthodes de lissage à d'autres distributions que gaussienne. Extensions : Loader (1996), Fan *et al.* (1998) et Loader (1999).

Rappel sur les GLMs

Caractéristiques I

- Durant les trente dernières années, l'utilisation des GLMs (Nelder and Wedderburn (1972)) a reçu beaucoup d'attention depuis les applications de McCullagh and Nelder (1989).
- Les GLMs sont idéalement adaptés à l'analyse de données non-normales que l'on rencontre typiquement lorsque l'on s'intéresse à des sujets relatifs à l'assurance.
- La modélisation diffère des modèles linéaires gaussiens par deux importants aspects :
  - La distribution de la variable dépendante est choisie dans la famille exponentielle et n'est donc pas spécifiquement Normale mais peut être explicitement non-Normale.
  - Une transformation de l'espérance de la variable dépendante est linéairement liée aux variables explicatives.

Rappel sur les GLMs

Caractéristiques II

Les modèles linéaires généralisés possèdent 3 caractéristiques :

- Un élément aléatoire, qui établit que les observations sont des variables aléatoires indépendantes Y_i, $i = 1, \ldots, n$ avec une densité appartenant à la famille exponentielle linéaire.
- Un élément systématique qui attribut à chaque observation un prédicteur linéaire η_i.
- Un troisième élément qui connecte les deux premières composantes : μ_i l'espérance de Y_i est lié au prédicteur linéaire η_i par une fonction de lien.

Rappel sur les GLMs

La famille exponentielle linéaire I

La technique des GLMs s'applique à toutes distributions appartenant à la famille exponentielle, i.e. lorsque la variable dépendante Y_i à une loi de probabilité de la forme

$$f(y_i|\theta_i,\phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right\},$$

pour des fonctions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ spécifiques. Les fonctions a and c sont telles que $a(\phi) = \phi$ and $c = c(y_i,\phi)$. Le paramètre θ_i est le paramètre canonique (ou paramètre naturel) et ϕ est le paramètre de dispersion.

Rappel sur les GLMs

La famille exponentielle linéaire II

Exemple d'une loi de Poisson :

Distribution de y_i	θ_i	$a(\phi)$	$b(\theta_i)$	$c(y_i, \phi)$	$\mathbb{E}[Y_i]$	$\mathbb{V}[\mu_i] = \frac{\mathbb{V}[Y_i]}{a(\phi)}$
Poisson(μ_i)	$\ln(\mu_i)$	1	$\exp(\theta_i)$	$-\log y_i!$	μ_i	μ_i

TABLE: Loi de Poisson appartenant à la famille exponentielle.

L'espérance et la variance sont calculées comme la première et seconde dérivées de $b(\theta_i)$:

$$\mathbb{E}[Y_i] = \frac{\partial}{\partial \theta} b(\theta_i) = \frac{\partial b}{\partial \mu_i} \frac{\partial \mu_i}{\partial \theta_i} = \mu_i$$

$$\mathbb{V}[Y_i] = a(\phi) \frac{\partial^2}{\partial \theta_i^2} b(\theta_i) = a(\phi) \left(\frac{\partial^2 b}{\partial \mu_i^2} \left(\frac{\partial \mu_i}{\partial \theta_i} \right)^2 + \frac{\partial b}{\partial \mu_i} \frac{\partial^2 \mu_i}{\partial \theta_i^2} \right) = a(\phi) \mu_i.$$

Rappel sur les GLMs

Qu'est-ce que veut dire linéaire dans GLMs

- "Linéaire" signifie que les variables explicatives sont combinées linéairement pour modéliser l'espérance.
- Si $x_1, x_2, \ldots, x_p$ sont des variables explicatives alors des combinaisons linéaires de la forme $\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$ servent comme des prédicteurs linéaires de l'espérance de la variable dépendante.
- La linéarité dans les GLMs se réfère seulement à la linéarité dans les coefficients β_j, non dans les variables explicatives.
- Par exemple,

$$\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 \text{ et } \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_3$$

sont "linéaires" au sens définie par les GLMs, mais

$$\beta_0 + \beta_1 x_1 + \exp(\beta_2 x_2)$$

ne l'est pas.

Rappel sur les GLMs

Fonction de lien

- $\mu_i = \mathbb{E}[Y_i|x_i]$, $i = 1, 2, \ldots, n$ est liée au prédicteur linéaire η_i par une fonction de lien $g()$ monotone et différenciable

$$g(\mu_i) = \eta_i \Leftrightarrow \mu_i = g^{-1}(\eta_i)$$

- La fonction de lien est dite canonique lorsque $\theta_i = \eta_i$, où θ_i est le paramètre canonique.
- La fonction de lien canonique assure donc que $g(\mu_i) = \theta_i$ et $g^{-1} = b'()$ (puisque $\mu_i = b'(\theta_i)$).

	Lien canonique
Poisson	$\eta_i = \log(\mu_i)$

TABLE: Lien canonique pour la loi de Poisson

Table des matières

Des GLMs à la vraisemblance locale

- L'approche GLM fait l'hypothèse que θ_i a une forme paramétrique spécifique, e.g.

$$\begin{aligned}\theta_i &= \beta_0 + \beta_1 u_i + \beta_2 v_i + \beta_3 u_i^2 + \beta_4 u_i\, v_i + \beta_5 v_i^2 \\ &\equiv \boldsymbol{x}^T \boldsymbol{\beta},\end{aligned}$$

où $\boldsymbol{x} = (1, u_i, v_i, u_i^2, u_i v_i, v_i^2)^T$ et $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_5)^T$.

- L'approche de la vraisemblance locale ne fait plus l'hypothèse que θ_i a une forme paramétrique rigide.
- On suppose que θ_i est une fonction lisse non-spécifiée $\psi(x_i)$ qui a $(p+1)$ dérivées continues au point x_i.
- L'idée est d'ajuster un modèle polynomial à l'intérieur d'une fenêtre d'observation.

Des GLMs à la vraisemblance locale

Expansion de Taylor

- Pour x_j dans le voisinage de x_i, on va approximer $\psi(x_j)$ via une expansion de Taylor par un polynôme de degré p, e.g. une approximation quadratique :

$$\begin{aligned}\psi(x_j) &= \psi(u_j, v_j) \\ &\approx \psi(u_i, v_i) + \frac{\partial\psi(u_i, v_i)}{\partial u_i}(u_j - u_i) + \frac{\partial\psi(u_i, v_i)}{\partial v_i}(v_j - v_i) \\ &+ \frac{1}{2}\frac{\partial^2\psi(u_i, v_i)}{\partial u_i^2}(u_j - u_i)^2 + \frac{\partial\psi(u_i, v_i)}{\partial v_i \partial u_i}(u_j - u_i)(v_j - v_i) \\ &+ \frac{1}{2}\frac{\partial^2\psi(u_i, v_i)}{\partial v_i^2}(v_j - v_i)^2 \equiv \boldsymbol{x}^T\boldsymbol{\beta},\end{aligned}$$

où $\boldsymbol{x} = \left(1, u_j - u_i, v_j - v_i, (u_j - u_i)^2, (u_j - u_i)(v_j - v_i), (v_j - v_i)^2\right)^T$

et $\boldsymbol{\beta} = \left(\psi(u_i, v_i), \frac{\partial\psi(u_i,v_i)}{\partial u_i}, \frac{\partial\psi(u_i,v_i)}{\partial v_i}, \frac{1}{2}\frac{\partial^2\psi(u_i,v_i)}{\partial u_i^2}, \frac{\partial\psi(u_i,v_i)}{\partial v_i \partial u_i}\frac{1}{2}\frac{\partial^2\psi(u_i,v_i)}{\partial v_i^2}\right)$.

Des GLMs à la vraisemblance locale

Ecriture de la vraisemblance

- La log vraisemblance d'un GLM s'écrit :

$$l(\beta; \boldsymbol{y}, \phi) = \sum_{i=1}^{n} \ln f(y_i|\theta_i, \phi) = \sum_{i=1}^{n} \frac{y_i\,\theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi).$$

- La fonction de log vraisemblance locale à x_i s'écrit :

$$l(\beta; \boldsymbol{y}, w_j(x_i), \phi) = \sum_{j=1}^{n} w_j \ln f(y_j|\theta_j, \phi) = \sum_{j=1}^{n} w_j \frac{y_j\,\theta_j - b(\theta_j)}{a(\phi)} + \sum_{j=1}^{n} w_j c(y_j, \phi).$$

  - Maximiser la vraisemblance locale par rapport à β donne le vecteur des estimateurs $\widehat{\beta}$.
  - Dans le cas d'une approximation quadratique $\widehat{\beta} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_5)$, et l'estimateur de $\psi(x_i)$ est donné par : $\widehat{\psi}(x_i) = \widehat{\beta}_0$.

- Ainsi le rôle des GLMs est celui d'un modèle en *arrière-plan* qui est ajusté localement.

Des GLMs à la vraisemblance locale

La fonction de poids I

La localisation s'effectue via la fonction de poids :

$$w_j = \begin{cases} W(\rho(x_i, x_j)/h) & \text{if } \rho(x_i, x_j)/h \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

$W(.)$ est une fonction de poids non négative qui dépend de la distance $\rho(x_i, x_j)$, e.g. la distance Euclidienne :

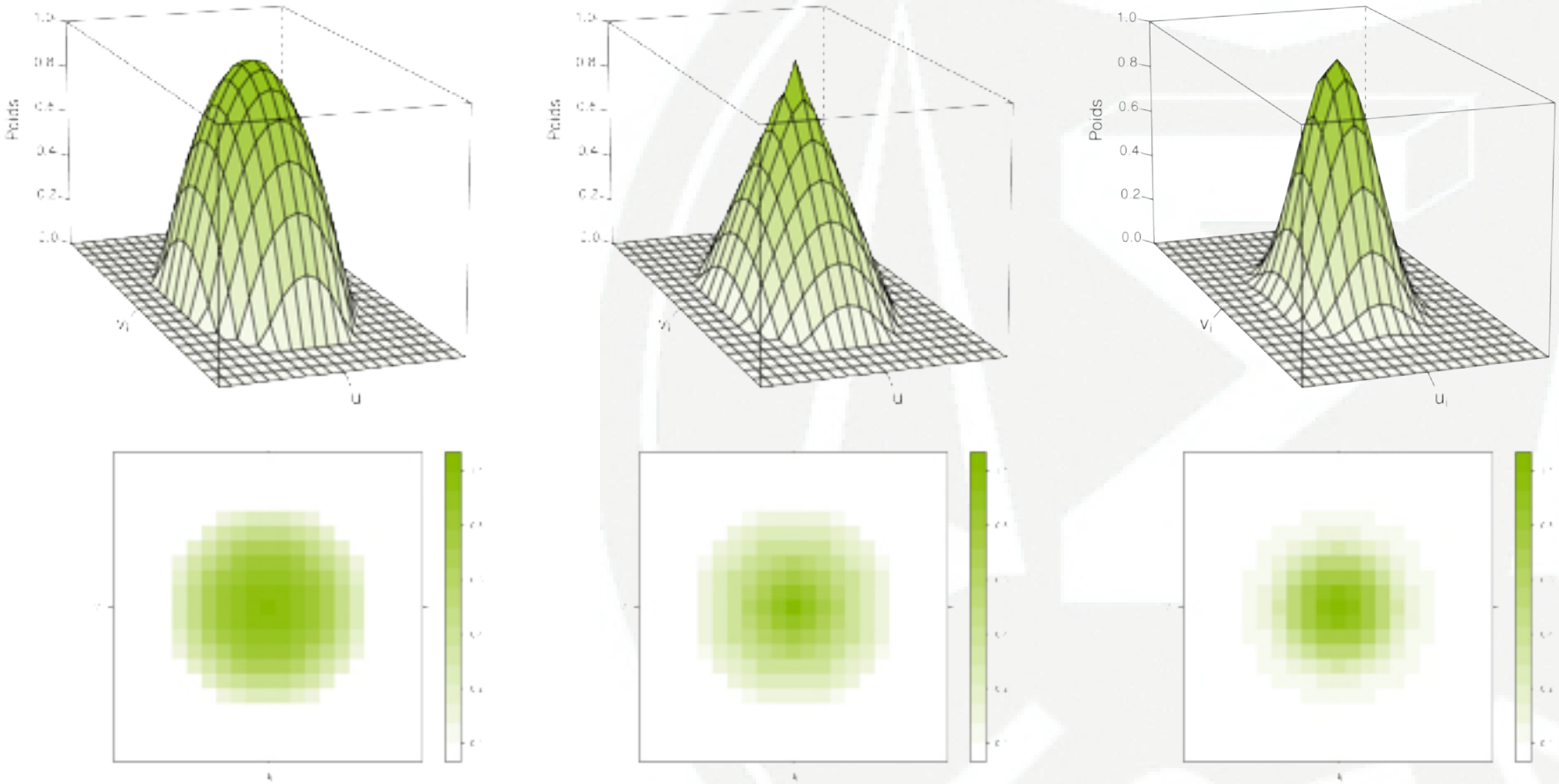
$$\rho(x_i, x_j) = \sqrt{(u_j - u_i)^2 + (v_j - v_i)^2}.$$

Fonction de poids	$W(a)$
Uniforme	$\frac{1}{2} I(\lvert a \rvert \leq 1)$
Triangulaire	$(1 - \lvert u \rvert) I(\lvert a \rvert \leq 1)$
Epanechnikov	$\frac{3}{4}(1 - a^2) I(\lvert a \rvert \leq 1)$
Quartic (Biweight)	$\frac{15}{16}(1 - a^2)^2 I(\lvert a \rvert \leq 1)$
Triweight	$\frac{35}{32}(1 - a^2)^3 I(\lvert a \rvert \leq 1)$
Tricube	$(1 - \lvert u^3 \rvert)^3 I(\lvert a \rvert \leq 1)$
Gaussienne	$\frac{1}{\sqrt{2\pi}} \exp(\frac{1}{2} a^2)$

avec $a = \rho(x_i, x_j)/h$

Des GLMs à la vraisemblance locale

La fonction de poids II



(a) Epanechnikov (b) Triangulaire (c) Triweight

FIGURE: Système de pondération pour des fonctions de poids avec $h = 7$

Illustration de l'idée générale

Comment estimer ψ ? Par un ajustement local...

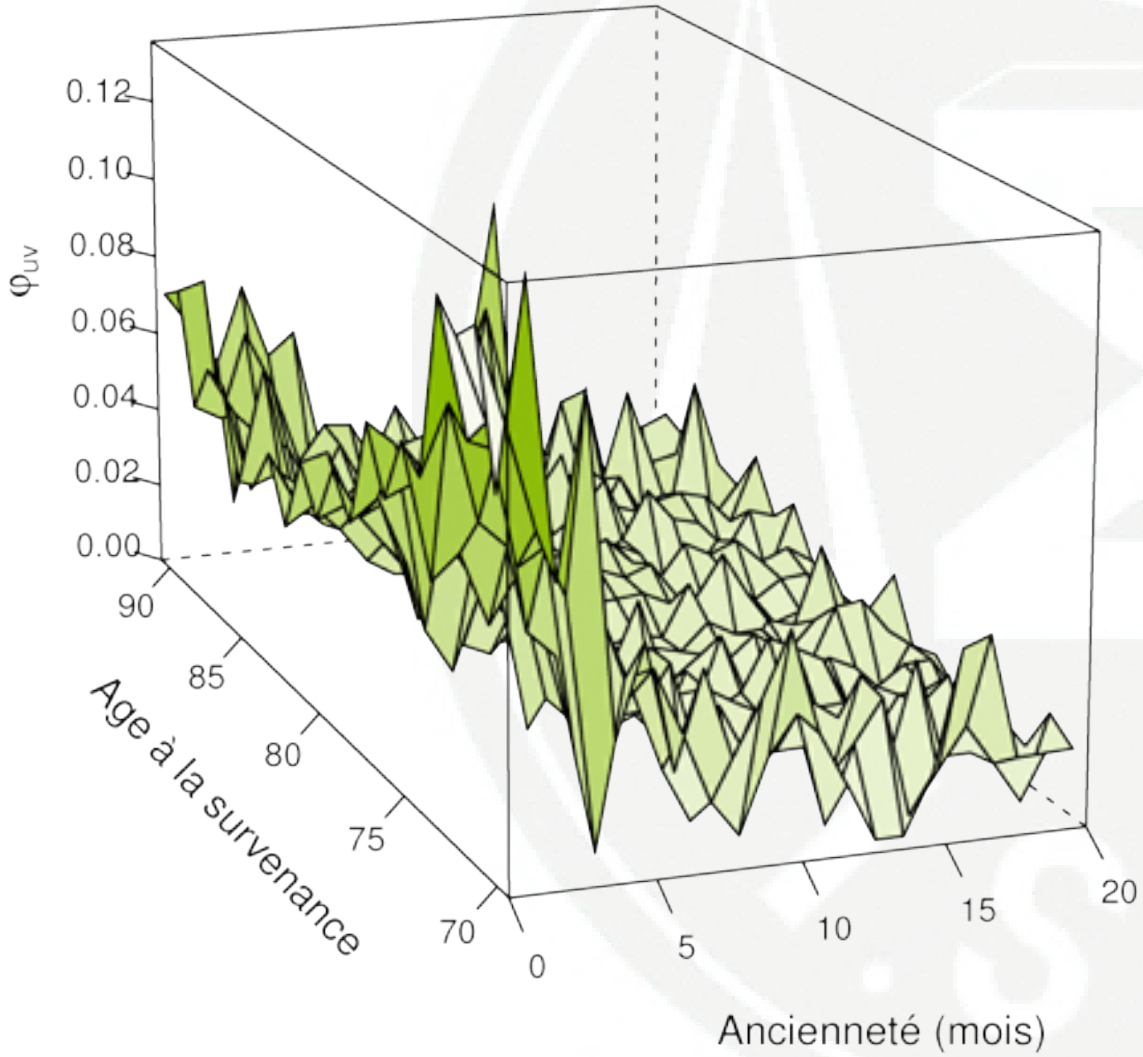


Illustration de l'idée générale

Comment estimer ψ ? Par un ajustement local...

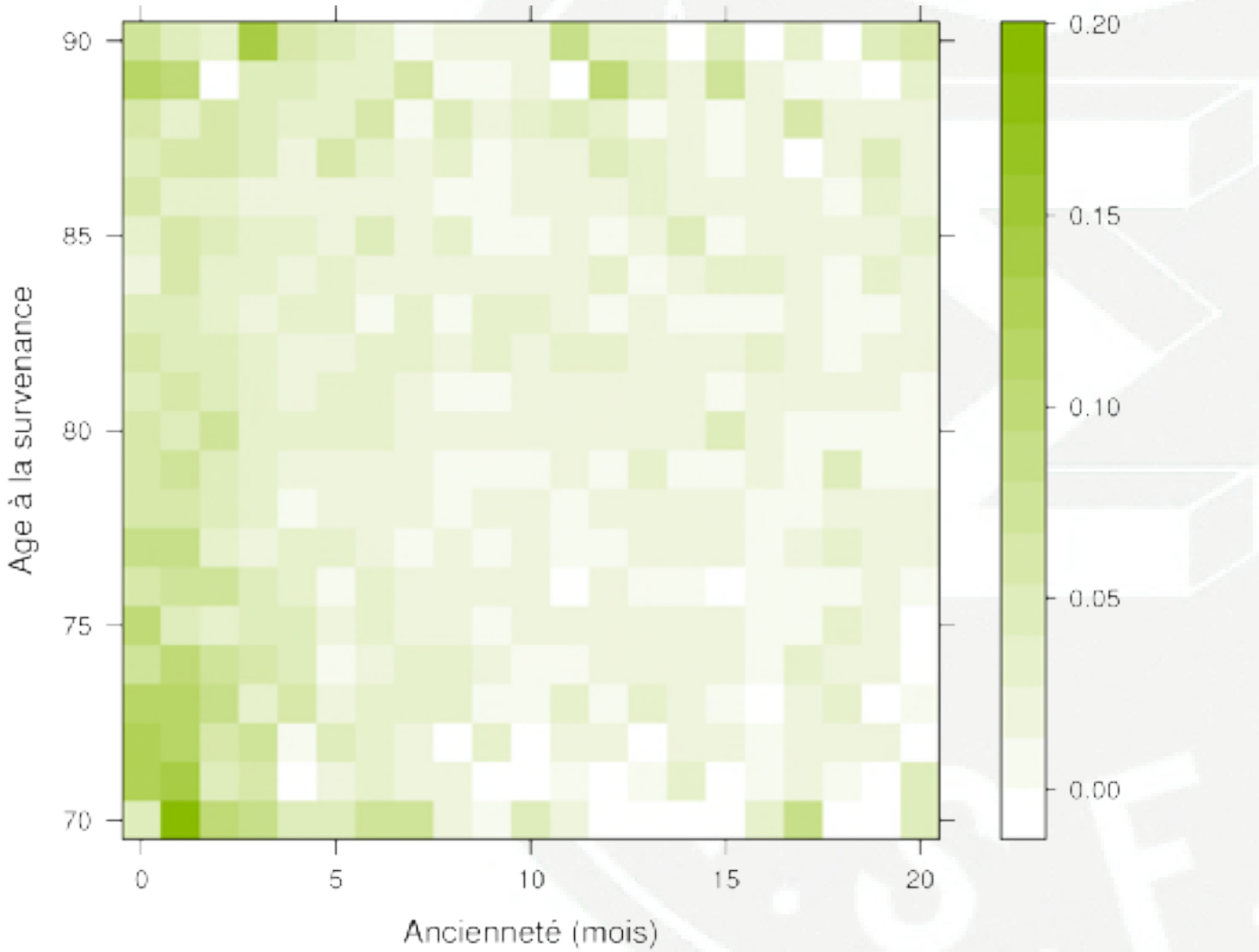


Illustration de l'idée générale

Comment estimer ψ ? Par un ajustement local...

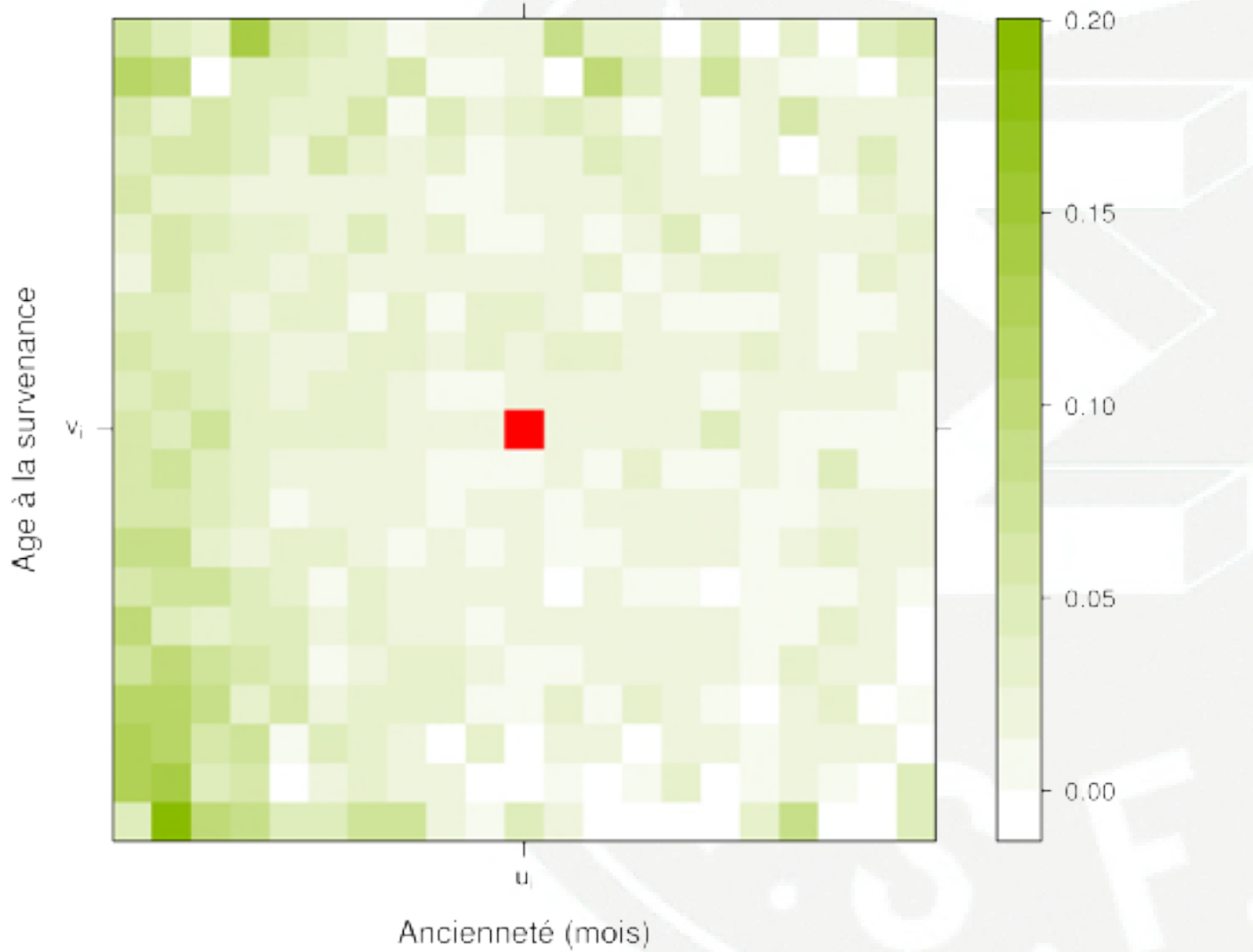


Illustration de l'idée générale

Comment estimer ψ ? Par un ajustement local...

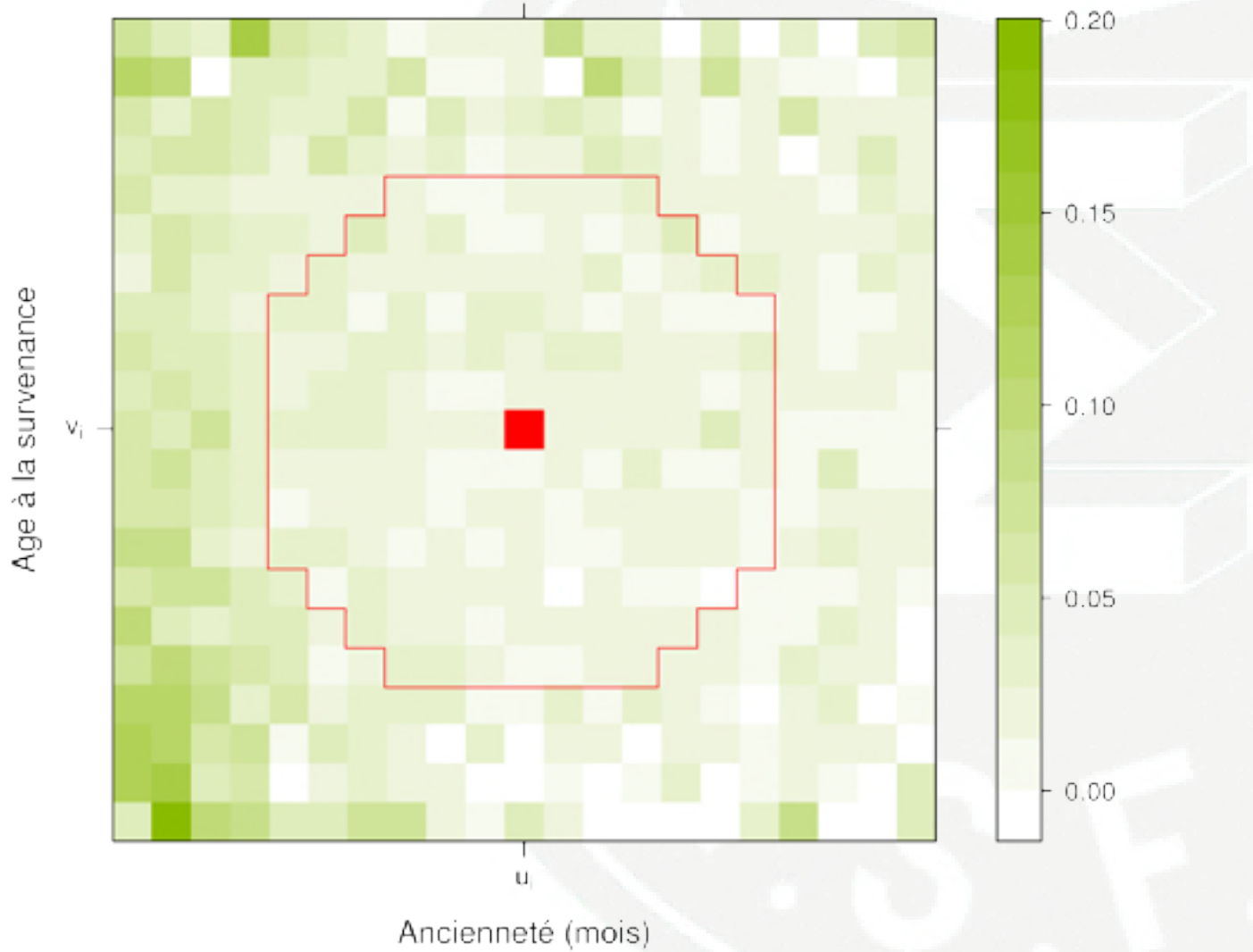


Illustration de l'idée générale

Comment estimer ψ ? Par un ajustement local...

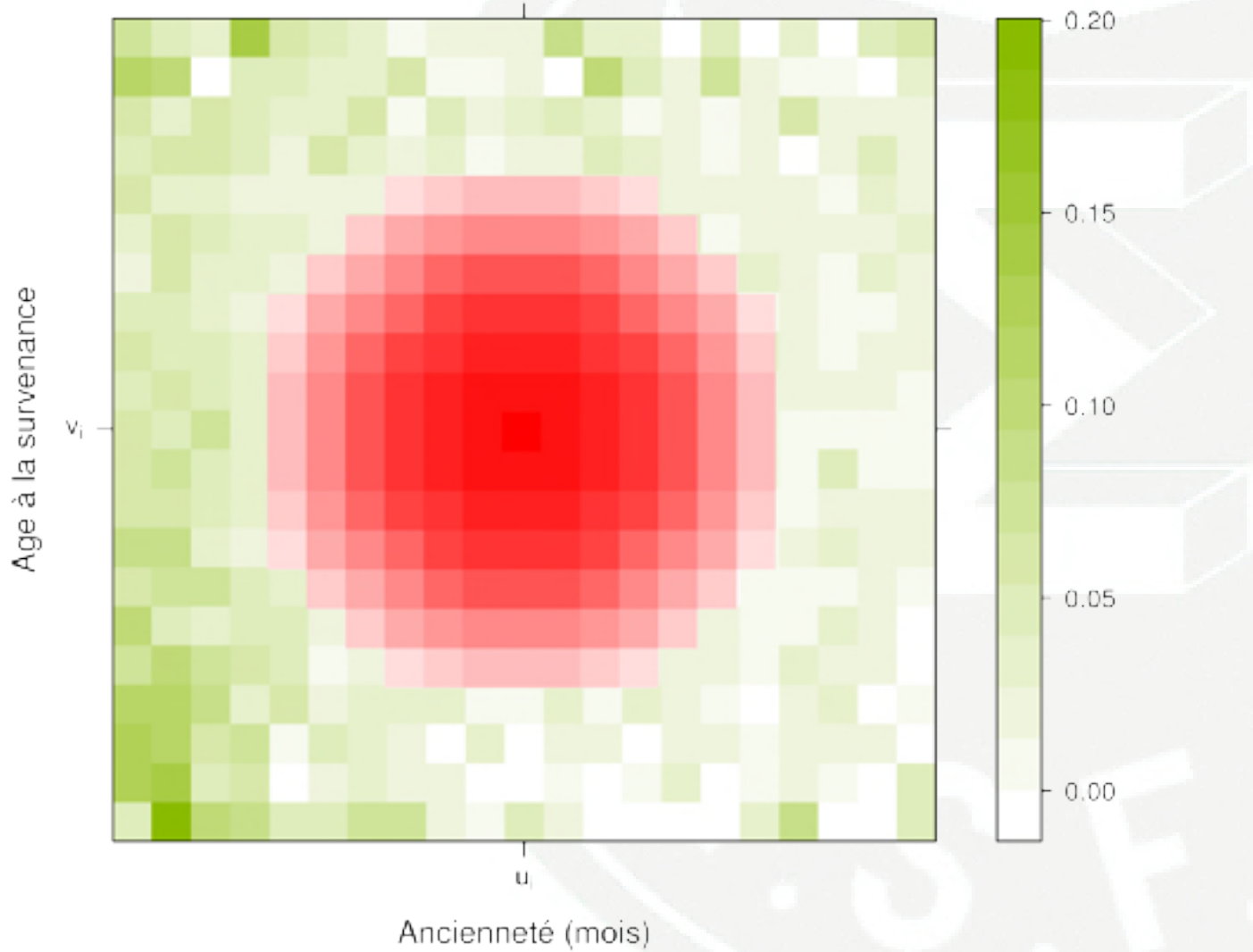


Illustration de l'idée générale

Comment estimer ψ ? Par un ajustement local...

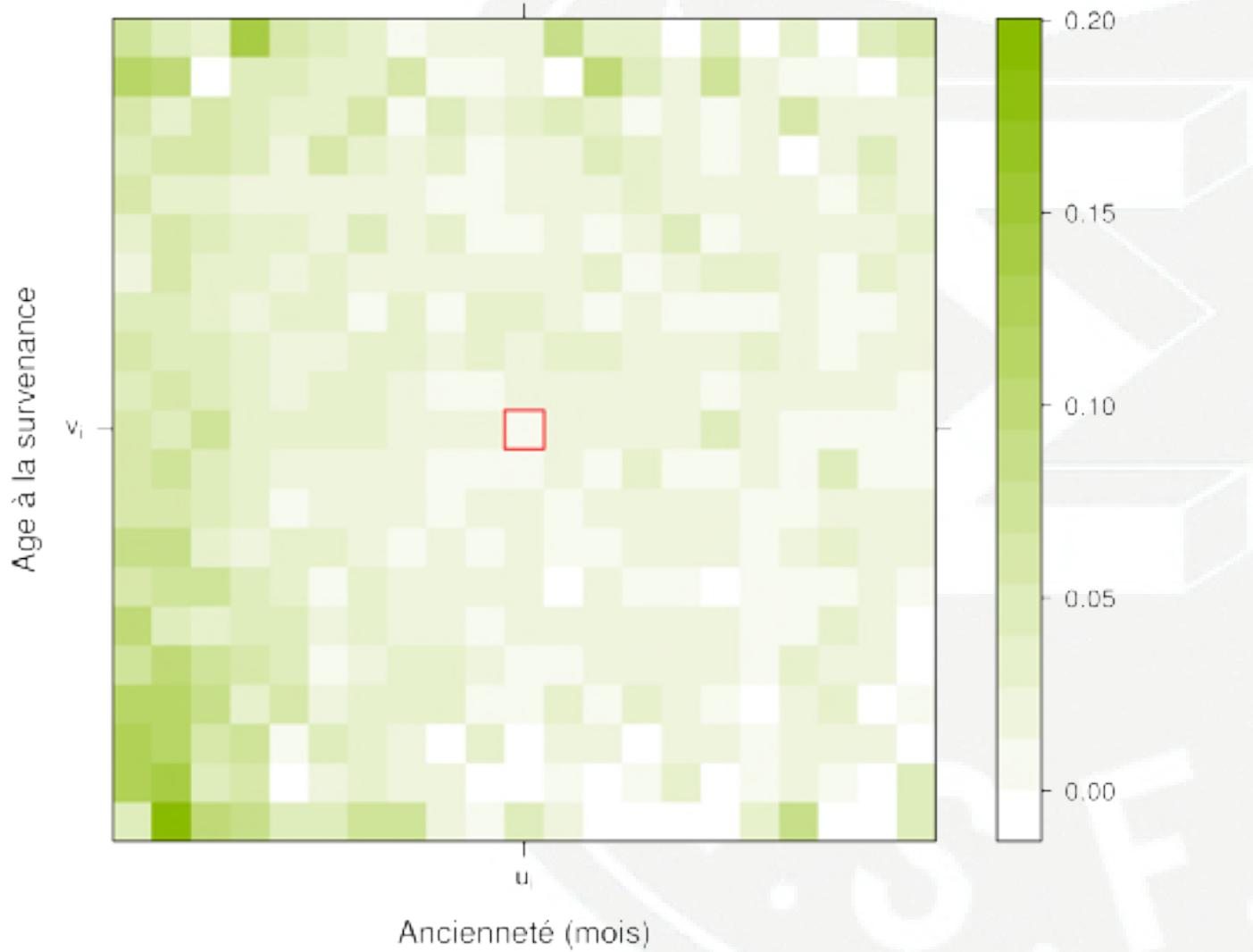


Illustration de l'idée générale

Comment estimer ψ ? Par un ajustement local...

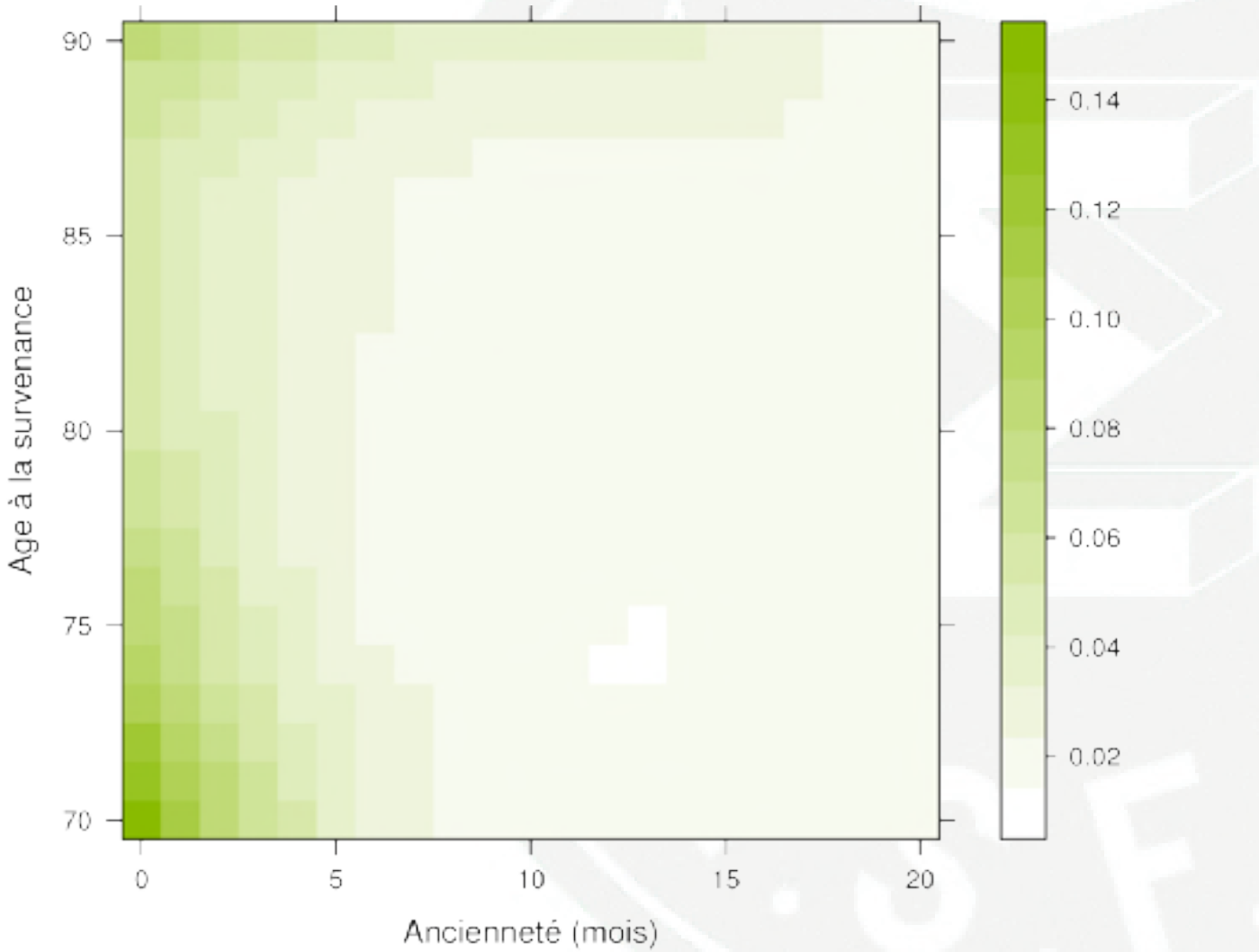


Illustration de l'idée générale

Comment estimer ψ ? Par un ajustement local...

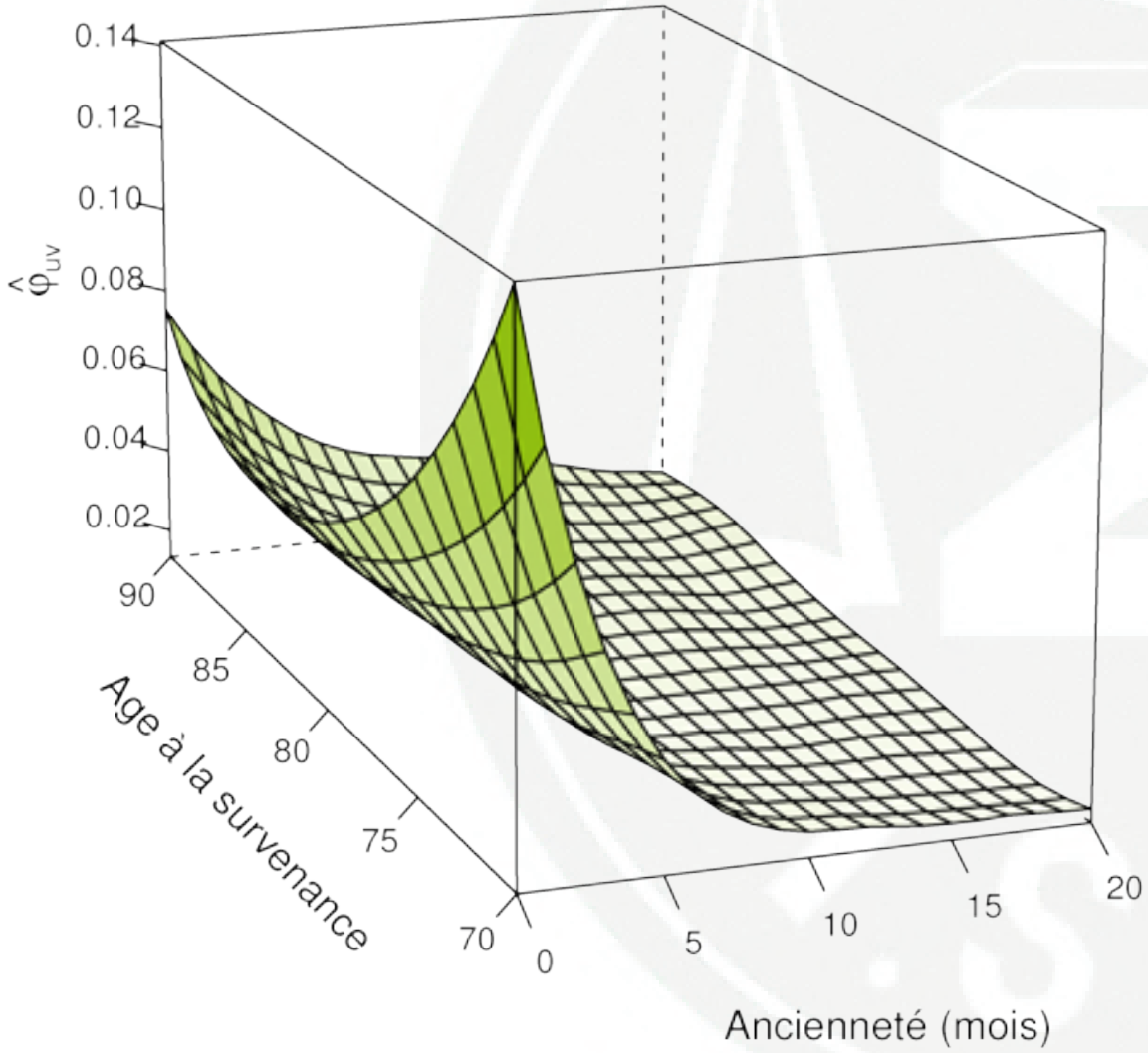
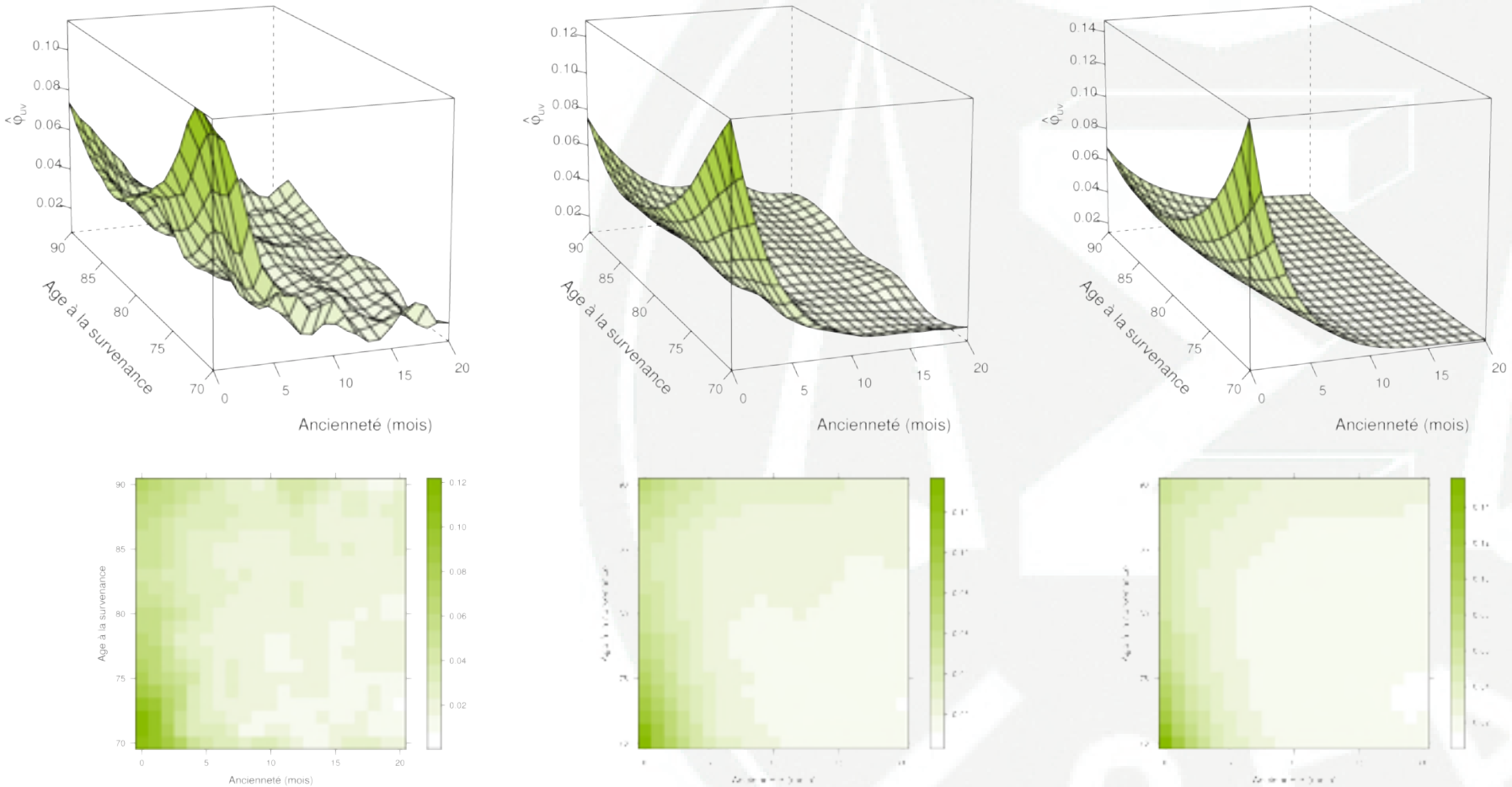


Illustration de l'idée générale

Comparaison de l'ajustement avec différents voisinages



(a) $h = 3$ (29 obs.) (b) $h = 5$ (81 obs.) (c) $h = 10$ (317 obs.)

FIGURE: Comparaison de l'ajustement avec différents voisinages

Equations de vraisemblance

Parallèle entre GLM et vraisemblance locale I

GLMs

- La log vraisemblance pour les observations y_i, $i = 1, \dots, n$, s'écrit :

$$l(\beta_j; y_i) = \sum_{i=1}^{n} \frac{y_i\, \theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi).$$

Vraisemblance locale

- La fonction de log vraisemblance locale à x_i s'écrit :

$$l(\beta_v; \boldsymbol{y}, w_j(x_i)) = \sum_{j=1}^{n} w_j \frac{y_j\, \theta_j - b(\theta_j)}{a(\phi)} + \sum_{j=1}^{n} w_j c(y_j, \phi).$$

Les paramètres β sont estimés par maximum de vraisemblance.

On résout les équations normales :

$$\sum_{i=1}^{n} \frac{\partial}{\partial \beta_j} l(\beta_0, \dots, \beta_p; y_i) = 0, \quad j = 0, \dots, p.$$

$$\sum_{j=1}^{n} w_j \frac{\partial}{\partial \beta_v} l(\beta_0, \dots, \beta_p; y_j) = 0, \quad v = 1 \dots, p.$$

Equations de vraisemblance

Parallèle entre GLM et vraisemblance locale II

GLMs

- La dérivée de l par rapport à β_j est :

$$\frac{\partial l}{\partial \beta_j} = \frac{\partial l}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j},$$

où

$$\frac{\partial l}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)} = \frac{y_i - \mu_i}{a(\phi)},$$

$$\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = \frac{\mathbb{V}[Y_i]}{a(\phi)},$$

$$\frac{\partial \eta_i}{\partial \beta_0} = 1, \qquad \frac{\partial \eta_i}{\partial \beta_1} = u_i,$$

$$\frac{\partial \eta_i}{\partial \beta_2} = v_i, \qquad \frac{\partial \eta_i}{\partial \beta_3} = u_i^2,$$

$$\frac{\partial \eta_i}{\partial \beta_4} = u_i v_j, \qquad \frac{\partial \eta_i}{\partial \beta_5} = v_i^2.$$

Vraisemblance locale

- De même,

$$\frac{\partial l}{\partial \beta_v} = \frac{\partial l}{\partial \theta_j}\frac{\partial \theta_j}{\partial \mu_j}\frac{\partial \mu_j}{\partial \eta_j}\frac{\partial \eta_j}{\partial \beta_v},$$

où

$$\frac{\partial l}{\partial \theta_j} = \frac{y_j - \mu_j}{a(\phi)},$$

$$\frac{\partial \mu_j}{\partial \theta_j} = b''(\theta_j) = \frac{\mathbb{V}[Y_j]}{a(\phi)},$$

$$\frac{\partial \eta_j}{\partial \beta_0} = 1, \qquad \frac{\partial \eta_j}{\partial \beta_1} = u_j - u_i,$$

$$\frac{\partial \eta_j}{\partial \beta_2} = v_j - v_i, \qquad \frac{\partial \eta_j}{\partial \beta_3} = (u_j - u_i)^2,$$

$$\frac{\partial \eta_j}{\partial \beta_4} = (u_j - u_i)(v_j - v_i), \qquad \frac{\partial \eta_j}{\partial \beta_5} = (v_j - v_i)^2.$$

Equations de vraisemblance

Parallèle entre GLM et vraisemblance locale III

GLMs

On obtient les équations de vraisemblance suivantes :

$$\sum_{i=1}^{n} \frac{\partial l}{\partial \beta_0} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\mathbb{V}[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} = 0$$

$$\vdots$$

$$\sum_{i=1}^{n} \frac{\partial l}{\partial \beta_5} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\mathbb{V}[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} v_i^2 = 0.$$

Vraisemblance locale

$$\sum_{j=1}^{n} \frac{\partial l}{\partial \beta_0} = \sum_{j=1}^{n} w_j \frac{(y_j - \mu_j)}{\mathbb{V}[Y_j]} \frac{\partial \mu_j}{\partial \eta_j} = 0$$

$$\vdots$$

$$\sum_{j=1}^{n} \frac{\partial l}{\partial \beta_5} = \sum_{j=1}^{n} w_j \frac{(y_j - \mu_j)}{\mathbb{V}[Y_j]} \frac{\partial \mu_j}{\partial \eta_j} (v_j - v_i)^2 = 0.$$

Equations de vraisemblance

Parallèle entre GLM et vraisemblance locale IV

GLMs

$\boldsymbol{X}^T \boldsymbol{V}(\boldsymbol{y} - \boldsymbol{\mu}) = 0$, où

$$\boldsymbol{X} = \begin{bmatrix} 1 & u_1 & v_1 & u_1^2 & u_1 v_1 & v_1^2 \\ 1 & u_2 & v_2 & u_2^2 & u_2 v_2 & v_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & u_n & v_n & u_n^2 & u_n v_n & v_n^2 \end{bmatrix},$$

et $\boldsymbol{V}$ est une matrice diagonale avec $\frac{1}{\mathbb{V}[Y_i]} \frac{\partial \mu_i}{\partial \eta_i}$ comme $i$ème élément sur sa diagonale.

Vraisemblance locale

En notation matricielle :

$\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{V}(\boldsymbol{y} - \boldsymbol{\mu}) = 0$, où

$$\boldsymbol{X} = \begin{bmatrix} 1 & u_1 - u_i & v_1 - v_i & \dots & (v_1 - v_i)^2 \\ 1 & u_2 - u_i & v_2 - v_i & \dots & (v_2 - v_i)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & u_n - u_i & v_n - v_i & \dots & (v_n - v_i)^2 \end{bmatrix},$$

et $\boldsymbol{W}$ est une matrice diagonale avec w_j comme $j$ème élément sur sa diagonale.

La fonction de lien $\eta = g(\mu)$ détermine $\partial \mu_i / \partial \eta_i = \partial g^{-1}(\eta_i) / \partial \eta_i$.
Ces équations ne possèdent en général pas de solutions explicites et doivent être résolues numériquement. On utilise une modification de l'algorithme de Newton-Raphson qui porte le nom de méthode de scoring de Fisher.

Résolution des équations de vraisemblance

Parallèle entre GLM et vraisemblance locale V

Dans le cadre d'un GLM,

- chaque étape de l'algorithme constitue un ajustement de type moindres carrées pondérés,
- c'est une généralisation des MCO qui prend en compte la non-constance de la variance des observations.
- les observations recueillies en des points où la variabilité est plus faible sont affectées d'un poids plus important dans la détermination des paramètres.
- à chaque itération, les poids sont remis à jours.
- on emploie le terme de moindres carrées itérativement re-pondérés ou IRWLS.

Dans le cadre de la vraisemblance locale, on emploiera une version localisée de l'IRWLS.

Résolution des équations de vraisemblance

Parallèle entre GLM et vraisemblance locale VI

- L'algorithme de Newton-Raphson utilise la matrice d'information de Fisher.
- Cette matrice contient l'information concernant la courbure de la fonction de log vraisemblance au point d'estimation.
- Plus grande est la courbure, plus l'information apportée au sujet des paramètres du modèle est importante.
- (En effet les écarts-types des estimateurs sont les racines carrées des éléments diagonaux de l'inverse de la matrice d'information de Fisher. Plus la courbure de la fonction de vraisemblance est importante, plus les écarts-types sont petits.)
- Pour un GLM, la dérivée partielle seconde est :

$$\frac{\partial^2 l}{\partial \beta_j \beta_k} = \frac{\partial^2 l}{\partial \eta_i^2} x_{ij} x_{ik}.$$

Résolution des équations de vraisemblance

Parallèle entre GLM et vraisemblance locale VII

- En utilisant la règle de dérivation en chaine, on obtient :

$$\begin{aligned}\frac{\partial^2 l}{\partial \eta_i^2} &= \frac{\partial}{\partial \eta_i}\left(\frac{\partial l}{\partial \theta_i}\frac{\partial \theta_i}{\partial \eta_i}\right)\\ &= \frac{\partial^2 l}{\partial \theta_i^2}\left(\frac{\partial \theta_i}{\partial \eta_i}\right)^2 + \frac{\partial l}{\partial \theta_i}\frac{\partial^2 \theta_i}{\partial \eta_i^2}\end{aligned}$$

- Comme $\partial l/\partial \theta_i = (y_i - \mu_i)/a(\phi)$, sa dérivée est $\partial^2 l/\partial \theta_i^2 = -1/a(\phi)\partial \mu_i/\partial \theta_i$. De plus $\partial \mu_i/\partial \theta_i = b''(\theta_i)$, on obtient

$$\begin{aligned}\frac{\partial^2 l}{\partial \eta_i^2} &= \frac{1}{a(\phi)}\left[-b''(\theta_i)\left(\frac{\partial \theta_i}{\partial \mu_i}\right)^2\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 + (y_i - \mu_i)\frac{\partial^2 \theta_i}{\partial \eta_i^2}\right]\\ &= \frac{1}{a(\phi)}\left[-\frac{1}{b''(\theta_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 + (y_i - \mu_i)\frac{\partial^2 \theta_i}{\partial \eta_i^2}\right].\end{aligned}$$

- SI $\theta \equiv \eta$, alors $\partial^2 \theta_i/\partial \eta_i^2 = 0$.

Résolution des équations de vraisemblance

Parallèle entre GLM et vraisemblance locale VIII

- Dans la méthode de scoring de Fisher, l'actuelle matrice hessienne dans l'itération de Newton-Raphson est remplacée par sa valeur espérée, qui est la négative de la matrice d'information de Fisher $\mathcal{I}$. Dans ce cas là aussi le second terme disparait. On obtient

$$\mathcal{I}_{jk} = \mathbb{E}\left[-\frac{\partial^2 l}{\partial\beta_j\beta_k}\right] = \sum_{i=1}^{n}\frac{1}{\phi}\frac{1}{b''(\theta_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 x_{ij}x_{ik}$$
$$= \sum_{i=1}^{n}\frac{\omega_{ii}x_{ij}x_{ik}}{\phi} = \frac{1}{\phi}(\boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X})_{jk}$$

  où $\boldsymbol{\Omega}$ est un matrice diagonale avec $\omega_{ii} = \frac{1}{b''(\theta_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2$ qui dépend de μ_i. Car $\eta_i = g(\mu_i)$, on a $\partial\eta_i/\partial\mu_i = g'(\mu_i)$.
- Dans le cadre de la vraisemblance locale, on obtient

$$\mathcal{I}_{vk} = \frac{1}{\phi}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{\Omega}\boldsymbol{X})_{vk}$$

Résolution des équations de vraisemblance

Parallèle entre GLM et vraisemblance locale IX

GLMs

La liste d'équations normale $\partial l/\partial \beta_j = 0$, $j = 0, \ldots, 5$, peut s'écrire comme :

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\omega_{ii} x_{ij} (y_i - \mu_i)}{\phi} g'(\mu_i)$$

ou $$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{1}{\phi} \boldsymbol{X}^T \boldsymbol{\Omega} \boldsymbol{u},$$

avec $u_i = (y_i - \mu_i) g'(\mu_i)$.

Vraisemblance locale

$\partial l/\partial \beta_v = 0$, $v = 0, \ldots, 5$, et

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{1}{\phi} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{\Omega} \boldsymbol{u},$$

avec $u_j = (y_j - \mu_j) g'(\mu_j)$.

On trouve alors une estimation *améliorée* $\boldsymbol{b}^*$ de β à partir d'un précédent $\boldsymbol{b}$ comme suit :

$$\boldsymbol{b}^* = \boldsymbol{b} + \mathcal{I}^{-1} \frac{\partial l}{\partial \beta} \Leftrightarrow \mathcal{I}(\boldsymbol{b}^* - \boldsymbol{b}) = \frac{\partial l}{\partial \beta}.$$

Résolution des équations de vraisemblance

Parallèle entre GLM et vraisemblance locale X

GLMs

Soit $\widehat{\boldsymbol{\eta}}$ et $\widehat{\boldsymbol{\mu}}$, les vecteurs des prédicteurs linéaires et des réponses estimées lorsque le paramètre est égale à $\boldsymbol{b}$, donc

$$\widehat{\boldsymbol{\eta}} = \boldsymbol{X}\boldsymbol{b} \text{ et } \widehat{\boldsymbol{\mu}} = g^{-1}(\widehat{\boldsymbol{\eta}}),$$

alors

$$\mathcal{I}\boldsymbol{b} = \frac{1}{\phi}\boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X}\boldsymbol{b} = \frac{1}{\phi}\boldsymbol{X}^T\boldsymbol{\Omega}\widehat{\boldsymbol{\eta}}.$$

On peut réécrire les équations itératives de Fisher comme suit :

$$\mathcal{I}\boldsymbol{b}^* = \frac{1}{\phi}\boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X}\boldsymbol{z},$$

où $z_i = \widehat{\eta}_i + (y_i - \mu_i)g'(\widehat{\mu}_i)$.

Vraisemblance locale

alors

$$\mathcal{I}\boldsymbol{b} = \frac{1}{\phi}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{\Omega}\widehat{\boldsymbol{\eta}}.$$

On peut réécrire les équations itératives de Fisher comme suit :

$$\mathcal{I}\boldsymbol{b}^* = \frac{1}{\phi}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{\Omega}\boldsymbol{X}\boldsymbol{z},$$

où $z_j = \widehat{\eta}_j + (y_j - \mu_j)g'(\widehat{\mu}_j)$.

Les éléments z_i sont appelés *pseudo réponses*.

Résolution des équations de vraisemblance

Parallèle entre GLM et vraisemblance locale XI

GLMs

Enfin, on trouve une estimation du maximum de vraisemblance de β par la procédure itéative suivante :

On répète

$$\boldsymbol{b}^* := \left(\boldsymbol{X}^T \boldsymbol{\Omega} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{\Omega} \boldsymbol{z};$$

en utilisant $\boldsymbol{b}^*$, on révise les *pseudo poids* $\boldsymbol{\Omega}$, ainsi que les *pseudo réponses* $\boldsymbol{z}$ jusqu'à convergence.

Vraisemblance locale

On répète

$$\boldsymbol{b}^* := \left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{\Omega} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{\Omega} \boldsymbol{z};$$

en utilisant $\boldsymbol{b}^*$, on révise les *pseudo poids* $\boldsymbol{\Omega}$, ainsi que les *pseudo réponses* $\boldsymbol{z}$ jusqu'à convergence.

• Ainsi dans le cadre de la vraisemblance locale, l'estimation de β est accomplie par la méthode du score de Fisher dans chaque voisinage, en allant dans l'ordre où i va de 1 à n.

Implémentation de l'algorithme sous R

R Development Core Team (2013)

- On considère le cas d'un Poisson GLM pour estimer les forces de mortalité comme dans le modèle (1) : $D_i \sim \mathcal{P}(E_i \, \varphi_i)$.
- Lorsque l'on modèle des données d'expérience provenant de l'assurance vie, on souhaite généralement prendre en compte l'exposition. Le prédicteur linéaire η devient

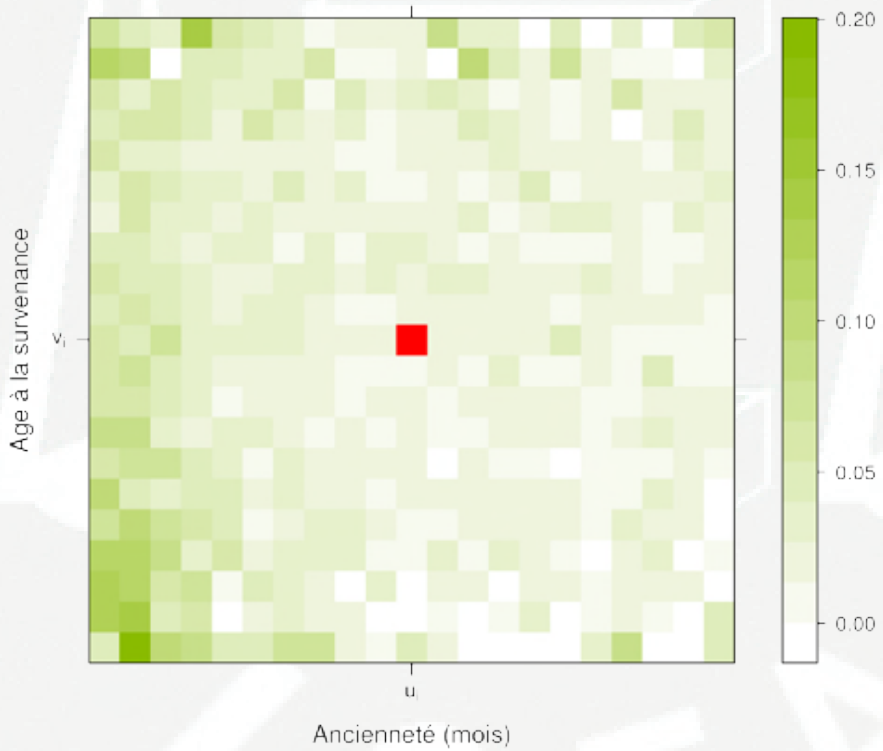
$$\eta_j = \log(\mathbb{E}[Y|X = x_j]) = \log(\mu_j) = \log(E_j \, \varphi_j) = \log(E_j) + \log(\varphi_j).$$

- Le terme E_j est appelé l'*offset* et peut-être facilement incorporé dans la régression.
- C'est un élément du prédicteur linéaire dont son coefficient est contraint à 1.

Implémentation de l'algorithme sous R

R Development Core Team (2013)

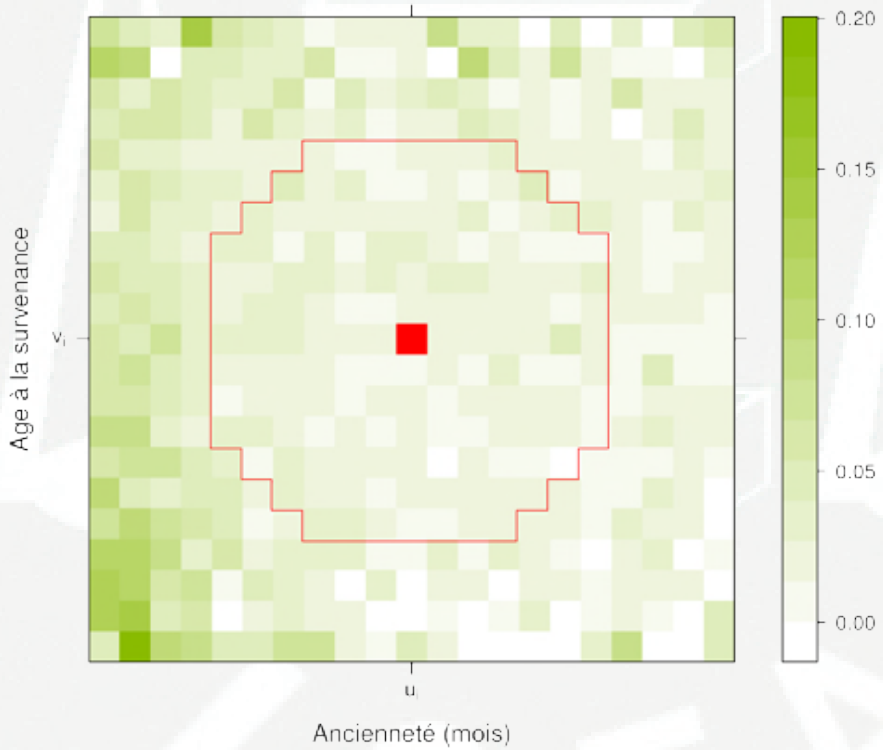
- Pour cet exemple, on reprend les données utilisées lors de l'illustration de l'idée générale.
- On va estimer φ_i pour $i = 221$, i.e. $\varphi_{221} = 0.01983471$.



Implémentation de l'algorithme sous R

R Development Core Team (2013)

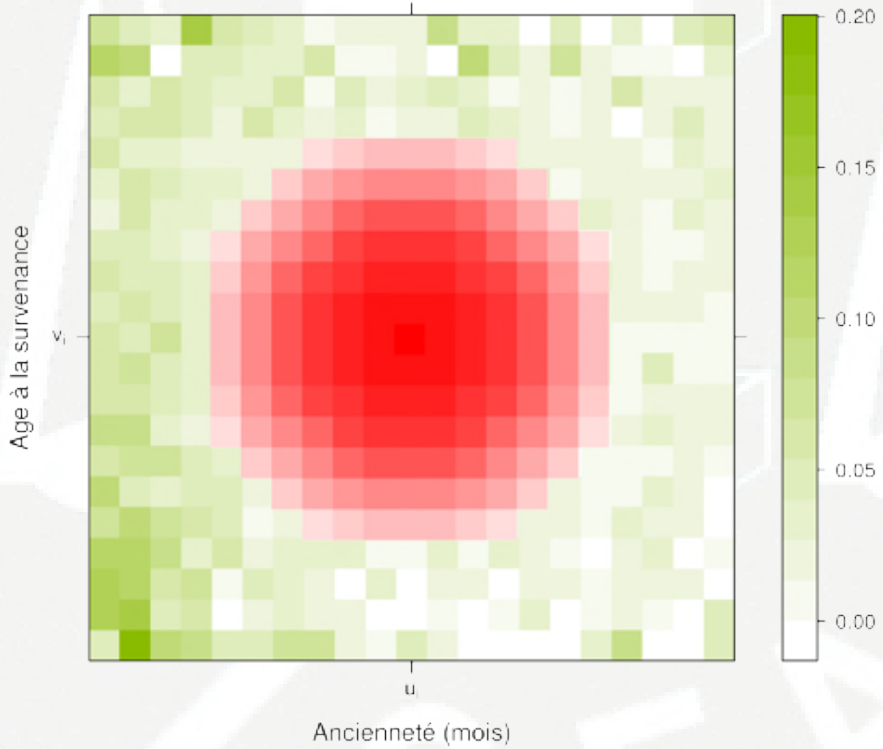
- Pour cet exemple, on reprend les données utilisées lors de l'illustration de l'idée générale.
- On va estimer φ_i pour $i = 221$, i.e. $\varphi_{221} = 0.01983471$.
- On choisit une fenêtre d'observation dont le rayon $h = 7$.



Implémentation de l'algorithme sous R

R Development Core Team (2013)

- Pour cet exemple, on reprend les données utilisées lors de l'illustration de l'idée générale.
- On va estimer φ_i pour $i = 221$, i.e. $\varphi_{221} = 0.01983471$.
- On choisit une fenêtre d'observation dont le rayon $h = 7$.
- On applique une fonction de poids Epanechnikov, i.e. $W(a) = \frac{3}{4}(1 - a^2)I(|a| \leq 1)$ pour $a = \rho(x_i, x_j)/h$, où $\rho(x_i, x_j)$ est la distance euclidienne.



Implémentation de l'algorithme sous R

R Development Core Team (2013)

- Fonction qui donne la matrice du design, $\boldsymbol{X}$, avec un ajustement quadratique :

```
1  Xmat = function(u_j, v_j, u_i, v_i) {
2    MAT <- matrix(1, length(u_j) * length(v_j), 6)
3    u_j.vec <- rep(u_j, times = length(v_j))
4    v_j.vec <- rep(v_j, each = length(u_j))
5    MAT[, 2] <- u_j.vec - u_i
6    MAT[, 3] <- v_j.vec - v_i
7    MAT[, 4] <- (u_j.vec - u_i) * (v_j.vec - v_i)
8    MAT[, 5] <- (u_j.vec - u_i)^2
9    MAT[, 6] <- (v_j.vec - v_i)^2
10   return(MAT) }
```

$$\boldsymbol{X} = \begin{bmatrix} 1 & u_1 - u_i & v_1 - v_i & \dots & (v_1 - v_i)^2 \\ 1 & u_2 - u_i & v_2 - v_i & \dots & (v_2 - v_i)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & u_n - u_i & v_n - v_i & \dots & (v_n - v_i)^2 \end{bmatrix}.$$

- Fonction qui donne la distance euclidienne :

```
1  euc.dist = function(u_j, v_j, u_i, v_i) {
2    u_j.vec <- rep(u_j, times = length(v_j))
3    v_j.vec <- rep(v_j, each = length(u_j))
4    VEC <- sqrt((u_j.vec - u_i)^2 + (v_j.vec - v_i)^2)
5    return(VEC) }
```

$$\rho(x_i, x_j) = \sqrt{(u_j - u_i)^2 + (v_j - v_i)^2}.$$

Implémentation de l'algorithme sous R

R Development Core Team (2013)

- Fonction qui donne le système de pondération (Epanechnikov) :

```
1 kernel = function(e.d, h) {
2   VEC <- as.vector(e.d[e.d <= h] / h)
3   VAL <- (3/4)*(1 - VEC^2)
4   return(VAL) }
```

$W(a) = \frac{3}{4}(1 - a^2)I(|a| \leq 1)$ pour $a = \rho(x_i, x_j)/h$, où $\rho(x_i, x_j)$ est la distance euclidienne.

- Fonction qui donne la matrice de poids, $\boldsymbol{W}$:

```
1 wmat = function(h, u_j, v_j, u_i, v_i) {
2   n <- length(u_j) * length(v_j)
3   e.d <- euc.dist(u_j, v_j, u_i, v_i)
4   kr <- vector(, n)
5   kr[1 : n] <- 0
6   kr[which(e.d <= h)] <- kernel(e.d, h)
7   Wmat <- diag(kr, n, n)
8   return(Wmat) }
```

Matrice diagonale avec comme $j$ème élément

$$w_j = \begin{cases} W(\rho(x_i, x_j)/h) & \text{if } \rho(x_i, x_j)/h \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Implémentation de l'algorithme sous R

R Development Core Team (2013)

- Fonction qui donne les moindres carrées itérativement re-pondérés ou IRWLS :

```
1  irwls = function(d, u_j, v_j, h, u_i, v_i, Wmat,
2                                                offset){
3    XM <- Xmat(u_j, v_j, u_i, v_i)
4    z <- log(d + ( d == 0 ))
5    offset <- offset + (offset == 0)
6    b <- solve(crossprod(XM, Wmat) %*% XM) %*%
7                              crossprod(XM, Wmat) %*% z
8    n <- length(u_j) * length(v_j)
9    g.prim <- OM <- matrix(0,n,n)
10   cat("Start", b , "\n")
11   for (it in 1 : 200) {
12     eta <- (XM %*% b) + log(offset)
13     mu <- exp(eta)
14     g.prim <- 1/mu
15     diag(OM) <- diag(mu)
16     z <- XM %*% b + g.prim * (d - mu)
17     F1 <- solve(crossprod(XM, Wmat) %*% OM %*% XM)
18     b.star <- tcrossprod(F1,XM) %*% Wmat %*% OM %*% z
19     cat("it =", it, b.star , "\n")
20     VAL <- abs(b - b.star)
21     if(all(VAL < 1e-6) == TRUE){ break }
22     b <- b.star }
23   return(b.star) }
```

Valeurs initiales :

- $\boldsymbol{X}$, matrice de design.
- $\boldsymbol{z} = g(\boldsymbol{y})$, (fonction de lien).
- $\beta = (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{z}$.

Procédure itérative :

- $\boldsymbol{\eta} = \boldsymbol{X}\beta + \log(\text{offset})$.
- $\boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta})$, (i.e. $\exp(\boldsymbol{\eta})$).
- $\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}} = g'(\boldsymbol{\mu}) = \frac{1}{mu}$.
- $\boldsymbol{\Omega}$, matrice diagonale avec élément $\omega_{jj} = \frac{1}{\mathbb{V}[\mu_j]} \left(\frac{\partial \mu_j}{\partial \eta_j} \right)^2 = \mu_j$.
- $\boldsymbol{z} = \boldsymbol{X}\beta + \frac{\boldsymbol{y}-\boldsymbol{\mu}}{\boldsymbol{\mu}}$.
- $\beta = (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{\Omega} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{\Omega} \boldsymbol{z}$.

Résultats :

- $\boldsymbol{b}^*$, estimation de β.

Implémentation de l'algorithme sous R (suite)

R Development Core Team (2013)

```
load("Data/Original data/dij.RData")
load("Data/Original data/eij.RData")

d <- D_i.j[1:21,1:21]; offset <- E_i.j[1:21,1:21]
u_j <- 0:20; v_j <- 70:90; u_i <- 10; v_i <- 80
h <- 7

Wmat <- wmat(h, u_j, v_j, u_i, v_i)

> b <- irwls(c(d), u_j, v_j, h, u_i, v_i, Wmat, c(offset))
Start 2.283744 -0.0616927 0.08938239 0.003371329 0.005613695 -0.0148457
it = 1 1.285481 -0.06162731 0.08921188 0.003371619 0.005615199 -0.01479154
it = 2 0.290194 -0.06145029 0.08875103 0.003372459 0.005619259 -0.01464519
it = 3 -0.6970451 -0.06097459 0.08751762 0.003375117 0.005630085 -0.01425376
it = 4 -1.662737 -0.05972105 0.08430269 0.003384938 0.005658006 -0.01323542
it = 5 -2.572236 -0.05658603 0.07649214 0.003427692 0.005723854 -0.01077386
it = 6 -3.345498 -0.04974172 0.06065755 0.003615231 0.00584605 -0.005850047
it = 7 -3.8479 -0.0390258 0.03980308 0.004193585 0.005966555 0.000408992
it = 8 -4.018391 -0.03094496 0.02870773 0.004888365 0.006001288 0.003412775
it = 9 -4.033062 -0.02944871 0.02775906 0.00500892 0.006075062 0.003527398
it = 10 -4.033277 -0.02941503 0.02776782 0.005007262 0.00610412 0.003512404
it = 11 -4.033292 -0.02940931 0.02776473 0.005008167 0.006105367 0.003513004
it = 12 -4.033293 -0.02940906 0.02776484 0.00500816 0.006105498 0.003512944

> c(exp(b))
[1] 0.0177159 0.9710192 1.0281539 1.0050207 1.0061242 1.0035191
```

- Charge les données.
- Fixe les valeurs de x_j, t_j, x_i, t_i, ainsi que le rayon de fenêtre d'observation h.
- Les valeurs de $\boldsymbol{b}^*$ produit par le code R.
- On obtient $\widehat{\psi}(x_{221}) = \exp(\widehat{\beta}_0) = 0.0177159$

Variance et Intervalles de confiance

Approximation de la variance I

- Soit $\boldsymbol{A}(\beta)$ le vecteur gradient de la log vraisemblance dont la $v$ème composante est

$$A_v(\beta) = \frac{\partial}{\partial \beta_v} l(\beta|\boldsymbol{y}),$$

- et $\boldsymbol{H}(\beta)$ la matrice hessienne de $l(\beta|\boldsymbol{y})$, dont l'élément (v, k) est

$$\frac{\partial^2}{\beta_v \beta_k} l(\beta|\boldsymbol{y}).$$

- Une approximation linéaire donne

$$0 = \boldsymbol{A}(\widehat{\beta}) \approx \boldsymbol{A}(\beta) + \boldsymbol{H}(\beta)(\widehat{\beta} - \beta),$$

- ce qui entraine

$$\widehat{\beta} - \beta \approx -\boldsymbol{H}(\beta)^{-1}\boldsymbol{A}(\beta).$$

Variance et Intervalles de confiance

Approximation de la variance II

- Ainsi une approximation de la variance de $\widehat{\beta}$ est :

$$\begin{aligned}\mathbb{V}\text{ar}\left[\widehat{\beta}\right] &\approx \mathbb{E}\left[-\boldsymbol{H}(\beta)^{-1}\right]_{\beta=\widehat{\beta}}\mathbb{E}\left[\boldsymbol{A}(\beta)\boldsymbol{A}(\beta)^T\right]_{\beta=\widehat{\beta}}\mathbb{E}\left[-\boldsymbol{H}(\beta)^{-1}\right]_{\beta=\widehat{\beta}}\\ &= \mathcal{I}_{\widehat{\beta}}^{-1}\left(\sum_{j=1}^{n} w_j\,\mathbb{E}\left[\frac{\partial l(\beta|\boldsymbol{y})}{\partial \beta^T}\frac{\partial l(\beta|\boldsymbol{y})}{\partial \beta}\right]\right)\mathcal{I}_{\widehat{\beta}}^{-1}.\end{aligned}$$

- En notation matricielle, on obtient :

$$\mathbb{V}\text{ar}\left[\widehat{\beta}\right] = \left(\boldsymbol{X}^T\boldsymbol{W}\,\boldsymbol{\Omega}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{V}\mathbb{E}\left[\boldsymbol{y}-\boldsymbol{\mu}\right]^2\boldsymbol{V}\boldsymbol{W}\boldsymbol{X}\right)\left(\boldsymbol{X}^T\boldsymbol{W}\,\boldsymbol{\Omega}\boldsymbol{X}\right)^{-1},$$

- On sait que $\mathbb{E}\left[y_j-\mu_j\right]^2 \equiv b''(\theta_j)$, donc

$$\boldsymbol{V}\mathbb{E}\left[\boldsymbol{y}-\boldsymbol{\mu}\right]^2\boldsymbol{V} = \boldsymbol{\Omega},$$

- en conséquence,

$$\mathbb{V}\text{ar}\left[\widehat{\beta}\right] \approx \left(\boldsymbol{X}^T\boldsymbol{W}\,\boldsymbol{\Omega}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{W}^2\boldsymbol{\Omega}\boldsymbol{X}\right)\left(\boldsymbol{X}^T\boldsymbol{W}\,\boldsymbol{\Omega}\boldsymbol{X}\right)^{-1}.$$

Variance et Intervalles de confiance

Approximation de la variance III

Une estimation de $\mathbb{V}\text{ar}\left[\widehat{\eta}\right]$ est obtenue de la façon suivante : Pour $\widehat{\eta}(x_i) = \widehat{\beta}_0(x_i)$,

$$\widehat{\eta}(x_i) = \eta(x_i) + \frac{\partial k(\beta)}{\partial \beta_0^T}\left(\widehat{\beta} - \beta\right) + o\left\|\widehat{\beta} - \beta\right\|,$$

et,

$$\mathbb{E}\left[\widehat{\eta}(x_i)\right] = \eta(x_i) + \frac{\partial k(\beta)}{\partial \beta_0^T}\mathbb{E}\left[\widehat{\beta} - \beta\right].$$

On obtient l'estimation de la variance suivante

$$\mathbb{V}\text{ar}\left[\widehat{\eta}(x_i)\right] = \frac{\partial k(\beta)}{\partial \beta_0^T}\mathbb{E}\left[\left(\widehat{\beta}(x_i) - \beta(x_i)\right)^2\right]\frac{\partial k(\beta)}{\partial \beta_0}.$$

Sachant que $\partial k(\beta)/\partial \beta_0^T = \boldsymbol{e}_1^T$, on a

$$\mathbb{V}\text{ar}\left[\widehat{\eta}(x_i)\right] \approx \boldsymbol{e}_1^T\left(\boldsymbol{X}^T\boldsymbol{W}\,\boldsymbol{\Omega}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{W}^2\boldsymbol{\Omega}\boldsymbol{X}\right)\left(\boldsymbol{X}^T\boldsymbol{W}\,\boldsymbol{\Omega}\boldsymbol{X}\right)^{-1}\boldsymbol{e}_1,$$

où $\boldsymbol{e}_v$ est un vecteur de longueur $p+1$ ayant 1 à la $v$ème composante et zéro à toutes les autres.

Variance et Intervalles de confiance

Approximation de la variance IV

En notant $\boldsymbol{S} = \left(\boldsymbol{X}^T \boldsymbol{W}\, \boldsymbol{\Omega} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{W}\, \boldsymbol{\Omega}$, on obtient

$$\mathbb{V}\mathrm{ar}\left[\widehat{\eta}(x_i)\right] \approx \boldsymbol{\Omega}^{-1} \boldsymbol{S}\boldsymbol{S}^T.$$

La matrice $\boldsymbol{S}$ est appelée *smooth weight diagram*, dont les lignes sont

$$\boldsymbol{s}(x_i)^T = (s_1(x_i), s_2(x_i), \ldots, s_n(x_i)) = \boldsymbol{e}_1^T \left(\boldsymbol{X}^T \boldsymbol{W}\, \boldsymbol{\Omega} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{W}\, \boldsymbol{\Omega}.$$

Ainsi l'approximation de la variance peut être écrite sous une forme compacte :

$$\mathbb{V}\mathrm{ar}\left[\widehat{\eta}(x_i)\right] = \left[\omega_{ii}\right]^{-1}_{\mu=\widehat{\mu}} \sum_{j=1}^{n} s_j^2(x_i) = b''(\theta_i)\left(g'(\widehat{\mu}_i)\right)^2 \|\boldsymbol{s}(x_i)\|^2.$$

Par la *delta method* et comme $\widehat{\mu}_i = g^{-1}(\widehat{\eta}_i)$, on obtient une estimation de la variance de μ_i,

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\widehat{\mu}(x_i)\right] &\approx \left[\partial g^{-1}(\eta_i)/\partial \eta_i\right]_{\eta=\widehat{\eta}} \mathbb{V}\mathrm{ar}\left[\widehat{\eta}(x_i)\right] \\ &= \left[\partial g^{-1}(\eta_i)/\partial \eta_i\right]_{\eta=\widehat{\eta}} b''(\theta_i)\left(g'(\widehat{\mu}_i)\right)^2 \|\boldsymbol{s}(x_i)\|^2.\end{aligned}$$

Variance et Intervalles de confiance

Les degrés de liberté

- Les degrés de liberté fournissent une mesure du montant de lissage. Cette mesure est comparable avec différents paramètres appliqués à un même ensemble de données.
  Par exemple, 10 degrés de liberté représentent un modèle lisse avec peu de flexibilité, alors que 30 degrés de liberté représentent un modèle montrant beaucoup de caractéristiques.
- Pour les modèles de vraisemblance locale, on définit υ comme

$$\upsilon = \text{tr}(\boldsymbol{S}) = \sum_{i=1}^{n} \widehat{\text{infl}}(x_i)\, \omega_{ii},$$

  où $\text{infl}(x_i) = \boldsymbol{e}_1^T \left(\boldsymbol{X}^T \boldsymbol{W}\boldsymbol{\Omega}\boldsymbol{X}\right)^{-1} \boldsymbol{e}_1$ est l'influence.
- L'influence mesure la sensibilité de la courbe ajustée aux points individuels. Par exemple, si $\text{infl}(x_i)$ approche 1, les résidus correspondants approchent zéro.

Sélection des paramètres

Equilibre entre le biais et la variance I

- On doit garder à l'esprit que l'ajustement et donc la sélection des paramètres de lissage est un compromis entre deux objectifs :
  - L'élimination des irrégularités, et
  - l'ajustement désiré dans la progression des forces de mortalité.
- En pratique on va choisir les paramètres de lissage pour équilibrer le compromis entre le biais et la variance.
- Pour trouver une telle constellation, la stratégie est d'évaluer un nombre d'ajustements *candidats* et d'utiliser un critère pour sélectionner parmi les ajustements celui qui aura le score le plus faible.

Sélection des paramètres

Equilibre entre le biais et la variance II

- Une possible fonction de perte est la déviance pour une observation individuelle (x_i, y_i), définie par

$$D(y_i, \widehat{\theta}(x_i)) = 2 \left(\sup_{\theta} l(y_i, \theta(y_i)) - l(y_i, \theta(\widehat{\mu}_i)) \right)$$
$$= 2 \left(y_i(\theta(y_i) - \theta(\widehat{\mu}_i)) - b\{\theta(y_i)\} + b\{\theta(\widehat{\mu}_i)\} \right).$$

- La déviance totale est $\sum_{i=1}^{n} D(y_i, \widehat{\theta}(x_i))$.
- Cas de Poisson $D(y_i, \widehat{\theta}(x_i)) = 2 \left(y_i \log(y_i/\widehat{\mu}_i) - (y_i - \widehat{\mu}_i) \right)$.
- Cela entraine une généralisation de l'Akaike information criterion (AIC) basée sur la deviance

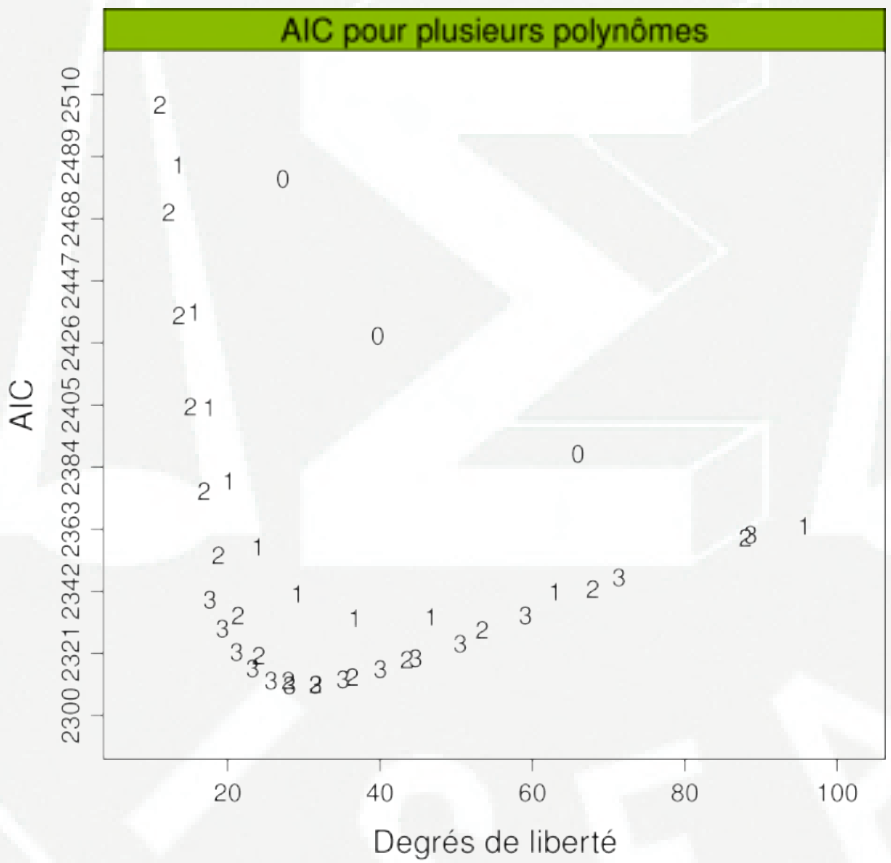
$$AIC = \sum_{i=1}^{n} D(y_i, (\theta(\widehat{\mu}_i))) + 2\,\upsilon,$$

où υ sont les degrés de liberté.

Sélection des paramètres

AIC-plot

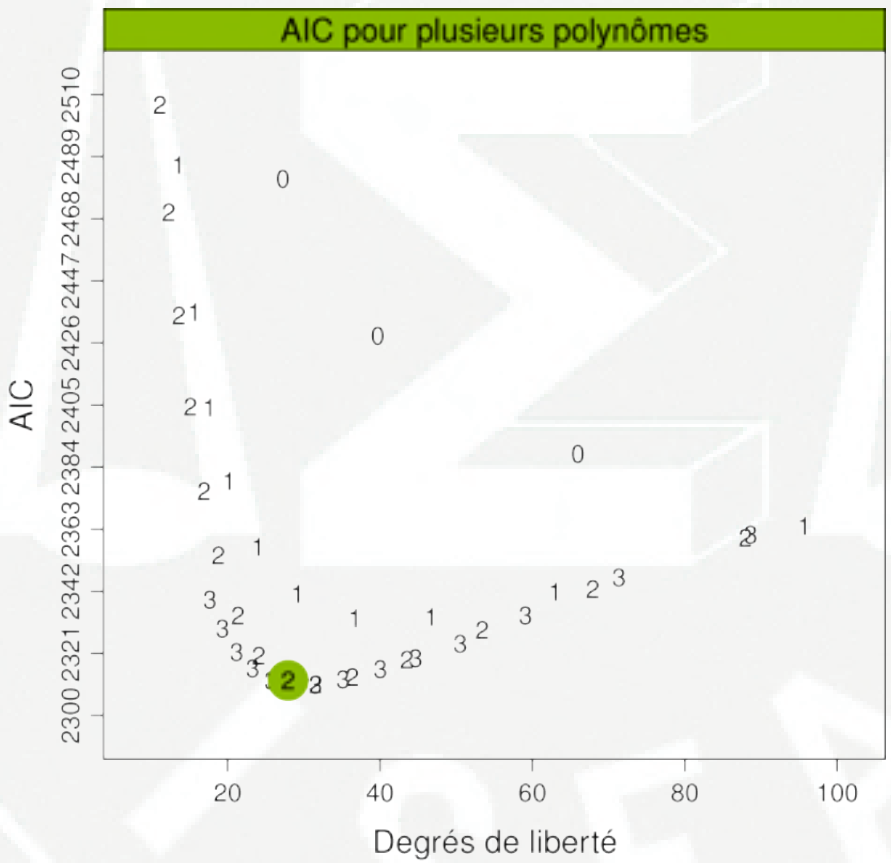
- Utiliser un graphique : AIC-plot.
- On utilise les degrés de liberté plutôt que la fenêtre d'observation comme l'axe horizontal.
- Cela permet de comparer des ajustements obtenus avec différents paramètres de lissage.
- Cleveland and Devlin (1988) ont recommandé de choisir les paramètres de lissage au point lorsque le critère atteint un plateau après une descente abrupte.



Sélection des paramètres

AIC-plot

- Utiliser un graphique : AIC-plot.
- On utilise les degrés de liberté plutôt que la fenêtre d'observation comme l'axe horizontal.
- Cela permet de comparer des ajustements obtenus avec différents paramètres de lissage.
- Cleveland and Devlin (1988) ont recommandé de choisir les paramètres de lissage au point lorsque le critère atteint un plateau après une descente abrupte.



Sélection des paramètres

Graphique des résidus

- En conjonction, on va toujours regarder le graphique des résidus. L'analyse des résidus est aussi important pour les procédures non-paramétriques que pour les modèles paramétriques.
- Dans les cas des GLMs, on dénote
  i. Résidus de la réponse : $r_i = y_i - \widehat{\mu}_i$;
  ii. Résidus de Pearson : $r_i = (y_i - \widehat{\mu}_i)/\sqrt{\mathbb{V}\text{ar}\,[\widehat{\mu}_i]}$;
  iii. Résidus de la déviance : $r_i = \text{sign}(y_i - \widehat{\mu}_i)D(y_i, \theta(\widehat{\mu}_i))^{1/2}$.
- Ces graphiques des résidus fournissent un diagnostic qui complète le critère de sélection.
- Le but est de déterminer si de larges résidus correspondent à des caractéristiques qui ont été mal modélisées.
- Aucune tendance forte devrait apparaître dans les résidus de la réponse et de Pearson. De plus, si les résidus de la déviance exhibent plusieurs résidus successifs ayant un même signe, ceci indique que les données sont sur-lissées

Conclusion

- La vraisemblance locale combine d'excellentes propriétés théoriques avec simplicité et flexibilité.
- Malheureusement, comme nous faisons face à une situation où les données ont une structure importante aux bordures, cette simplicité a des défauts.
- Aux bordures, la fonction de poids est asymétrique et l'estimation peut avoir un biais important. Cela force les critères à sélectionner une fenêtre plus petite pour réduire le biais, mais cela conduit à un sous-lissage au centre de la table.
- En conséquence, dans certains cas, aucun paramètre global de lissage fournit un ajustement adéquat des données.
- Plutôt que de restreindre les paramètres de lissage à une valeur fixe, une approche plus souple est de permettre aux paramètres de lissage de varier selon l'emplacement.

Table des matières

Méthodes avancées

Méthodes adaptatives de vraisemblance locale

- On évalue une constellation différente pour chaque point auquel la courbe est estimée.
- On traite les choix de la fenêtre d'observation, du degré d'approximation et de la fonction de poids comme si on modélise des données.
- On choisit ces paramètres pour équilibrer le compromis entre biais et variance.

Méthodes avancées

Méthodes adaptatives de vraisemblance locale

- On distingue la méthode adaptative ponctuelle qui utilise la règle de l'intersection des intervalles de confiance et une méthode globale qui utilise des facteurs de correction de la fenêtre d'observation.
- On varie le montant de lissage selon l'emplacement et permet des ajustements selon la fiabilité des données.
- Parmi les paramètres de lissage, la fonction de poids a beaucoup moins d'influence sur le compromis biais/variance que la fenêtre d'observation ou le degré d'approximation. Le choix n'est pas déterminant. Par convenance, on utilise la fonction de poids Epanechnikov.

Intersection des intervalles de confiance

Introduction

- La règle de l'intersection des intervalles de confiance a été introduite par Goldenshulger and Nemirovski (1997) et développée par Katkovnik (1999).
- Application à la vraisemblance locale de Poisson pour la restauration adaptative des images a été étudiée par Katkovnik *et al.* (2005).
- Voir aussi la thèse de doctorat de Chichignoud (2010).
- La règle de l'intersection des intervalles de confiance donne une méthode alternative pour évaluer un ajustement locale.

Intersection des intervalles de confiance

Principes I

- On définit un ensemble de fenêtres d'observation :

$$H = \{h_k < h_{k-1} < \ldots < h_0\}.$$

- On détermine la fenêtre d'observation optimale en évaluant les ajustements obtenus.
- Soit $\widehat{\psi}(x_i, h_k)$ l'estimation à x_i pour la fenêtre d'observation h_k.
- Pour sélectionner la fenêtre d'observation, la règle d'intersection des intervalles de confiances examine une séquence des intervalles de confiance des estimations $\widehat{\psi}(x_i, h_k)$:

$$\mathcal{D}(x_i, h_k) = \left[\widehat{L}(x_i, h_k), \widehat{U}(x_i, h_k)\right],$$
$$\widehat{U}(x_i, h_k) = \widehat{\psi}(x_i, h_k) + c\,\widehat{\sigma}(x_i)\|\boldsymbol{s}(x_i, h_k)\|,$$
$$\widehat{L}(x_i, h_k) = \widehat{\psi}(x_i, h_k) - c\,\widehat{\sigma}(x_i)\|\boldsymbol{s}(x_i, h_k)\|,$$

où c est le paramètre du seuil de l'intervalle de confiance.

Intersection des intervalles de confiance

Principes II

- A partir des intervalles de confiance, on définie

$$\overline{\widehat{L}}(x_i, h_{k-n}) = \max\left[\widehat{L}(x_i, h_k), \widehat{L}(x_i, h_{k-n})\right],$$
$$\underline{\widehat{U}}(x_i, h_{k-n}) = \min\left[\widehat{U}(x_i, h_k), \widehat{U}(x_i, h_{k-n})\right],$$
$$n = 1, 2, \ldots, k.$$

- La valeur la plus large de $k - n$ de telle sorte que $\underline{\widehat{U}}(x_i, h_{k-n}) \geq \overline{\widehat{L}}(x_i, h_{k-n})$ donne k^*.
- Cela entraine la fenêtre h_k^*, qui est la fenêtre optimale selon la règle des intervalles de confiance.

Intersection des intervalles de confiance

Principes III

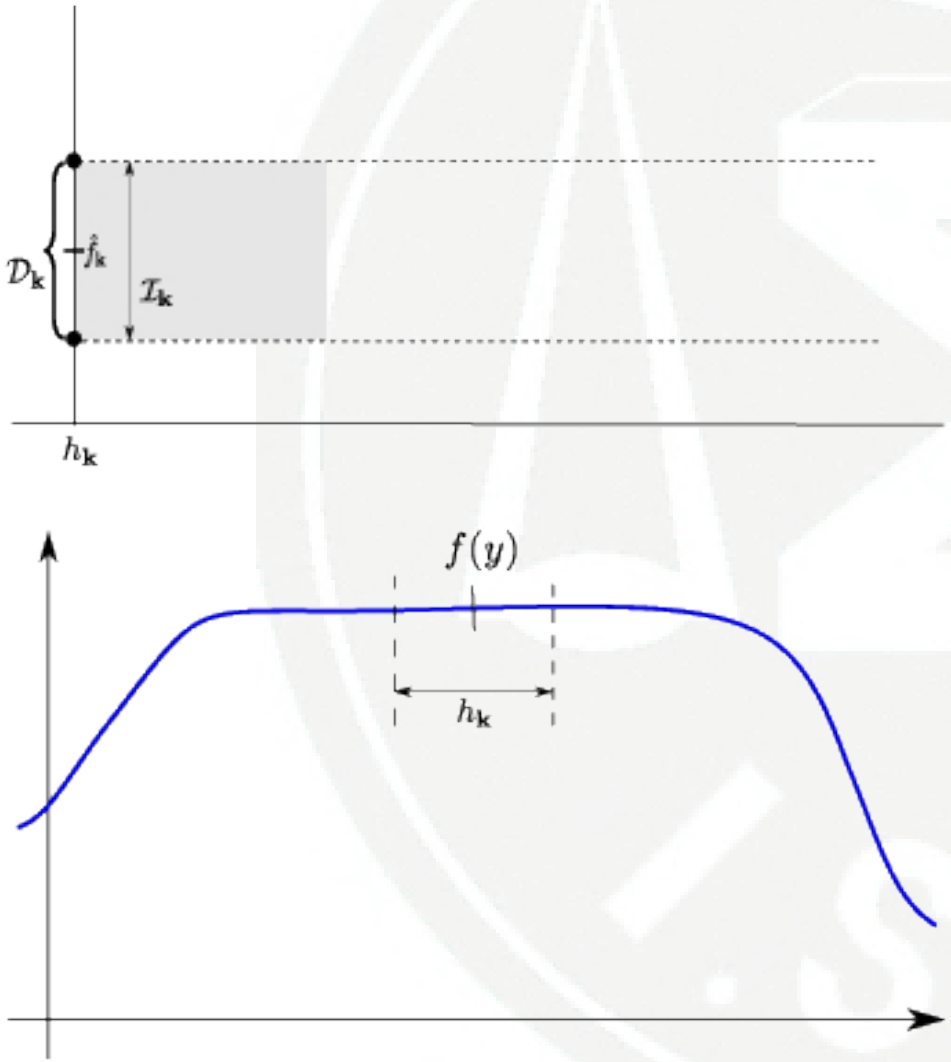
- En d'autres termes, si on note $\mathcal{I}_j = \bigcap_{l=j}^{k} \mathcal{D}(x_i, h_l)$ pour $j = k, k-1, \ldots, , 0$, on choisit k^* de telle sorte que

$$\begin{cases} \mathcal{I}_j \neq \emptyset, & \forall j \leq k^*, \\ \mathcal{I}_{k^*-1} = \emptyset. \end{cases}$$

- Comme la fenêtre h_k augmente, la déviation standard de $\widehat{\psi}(x_i, h_k)$, et donc $\|\boldsymbol{s}(x_i, h_k)\|$, diminue.
- L'intervalle de confiance devient plus petit. Si h_k est trop large, l'estimation $\widehat{\psi}(x_i, h_k)$ sera largement biaisée et les intervalles de confiance seront inconsistants dans le sens que les intervalles construits à des fenêtres différentes n'auront pas d'intersections communes.

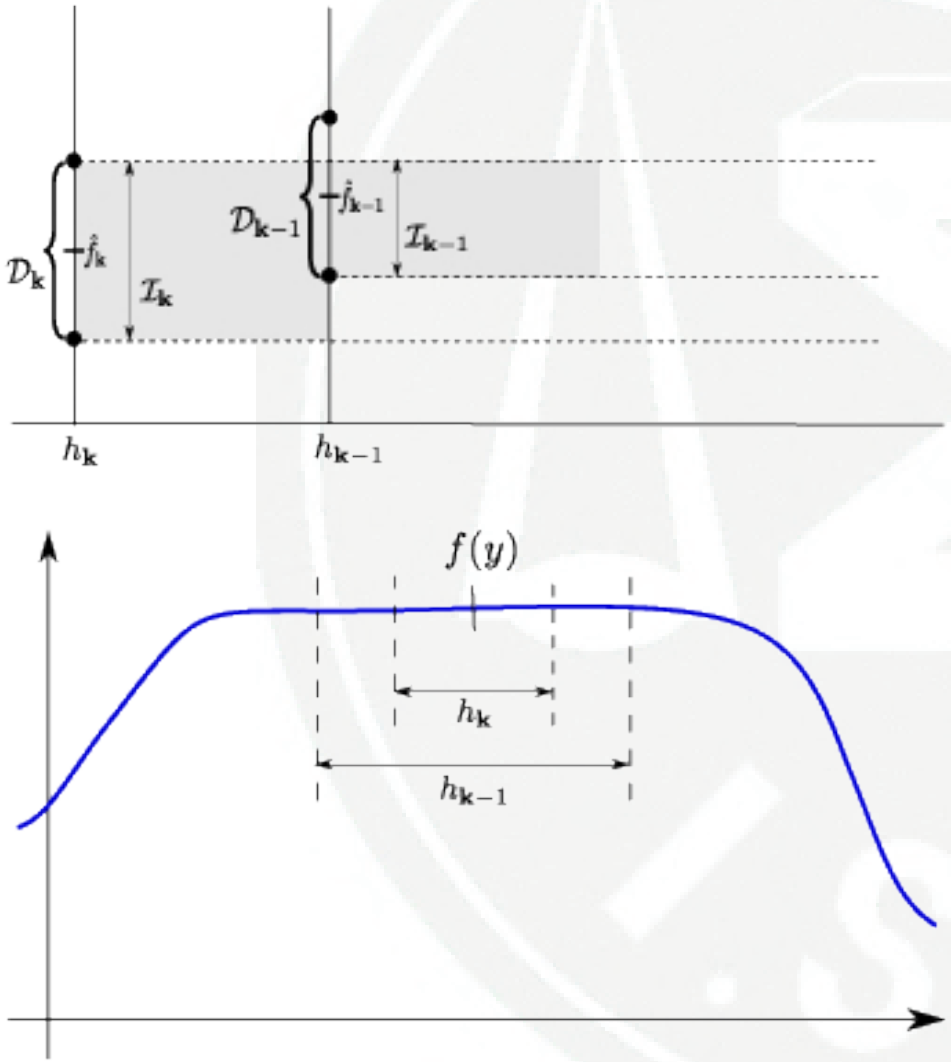
Intersection des intervalles de confiance

Illustration I



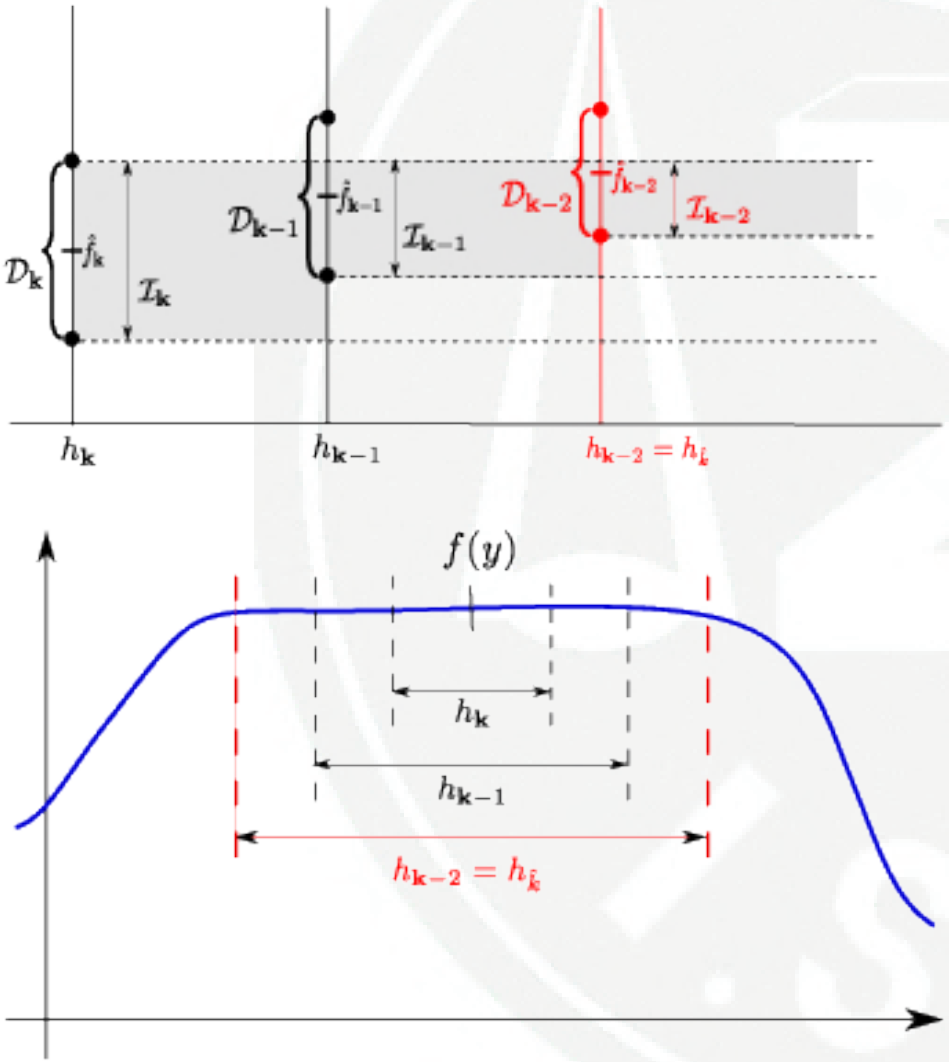
Intersection des intervalles de confiance

Illustration II



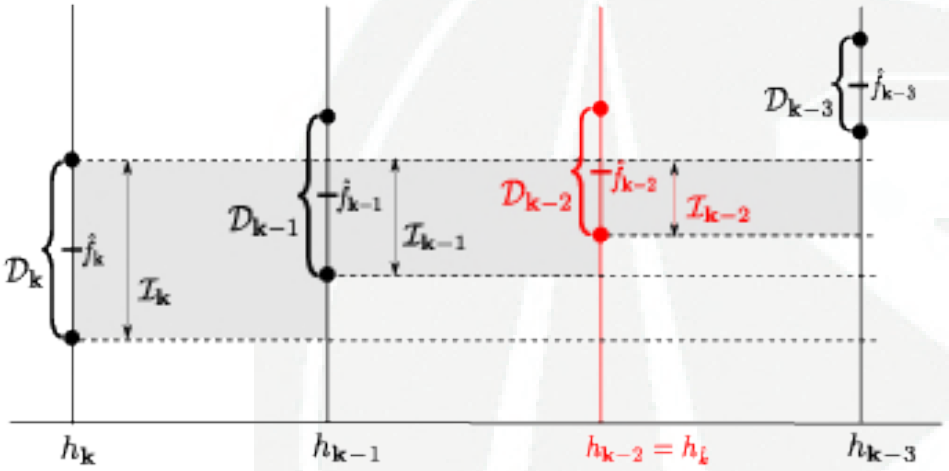
Intersection des intervalles de confiance

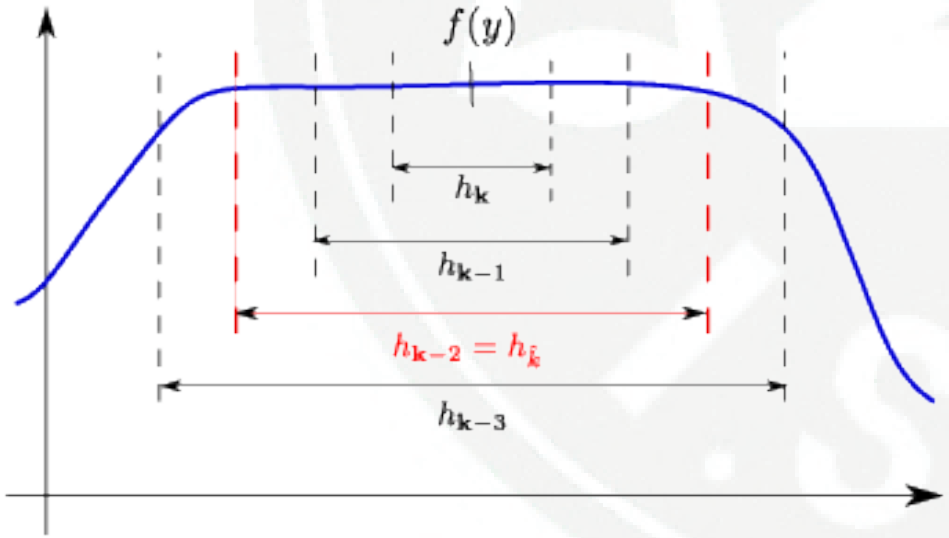
Illustration III



Intersection des intervalles de confiance

Illustration IV





Intersection des intervalles de confiance

Principes

- Le paramètre c joue une part importante dans la performance de l'algorithme car il décide de la fenêtre d'observation optimale.
- Lorsque c est grand, le segment $\mathcal{D}(x_i, h_k)$ devient large et h_k^* aussi. On obtient un sur-lissage.
- Au contraire, lorsque c est petit, le segment $\mathcal{D}(x_i, h_k)$ devient petit et tout comme h_k^*. Les irrégularités ne peuvent pas être éliminées de manière effective.
- En théorie, on peut appliquer le critère AIC pour déterminer une valeur raisonnable de c.

Facteurs de correction de la fenêtre d'observation

Introduction I

- On va incorporer de l'information additionnelle dans une procédure globale en permettant à la fenêtre d'observation de varier selon la fiabilité des données.
- Extension de l'estimateur à noyaux variable proposé par Gavin *et al.* (1995, pp.190-193).
- La fenêtre locale à chaque point est simplement la fenêtre globale multipliée par un facteur de correction.
- Comme on a obtenu les facteurs de corrections de la fenêtre d'observation, on utilise un critère globale, e.g. AIC, pour déterminer la fenêtre globale.

Facteurs de correction de la fenêtre d'observation

Introduction II

- Les facteurs de correction dépendent de l'exposition ou du nombre de décès (dans le cas d'annuités ou de prestations de décès).
- Pour des régions où l'exposition est large, une petite fenêtre d'observation induit une estimation qui reflète plus précisément les taux de mortalité.
- Pour des régions où l'exposition est faible, une fenêtre plus large permet aux estimations des forces de mortalité de progresser de manière plus lisse. On utilise un plus grand nombre d'observations ce qui réduit la variance des forces de mortalité graduées mais avec un biais potentiel plus large.

Facteurs de correction de la fenêtre d'observation

Principes I

- La fenêtre d'observation à chaque point est la fenêtre globale multipliée par un facteur de correction : $h_i = h \times \delta_i^s$ for $i = 1, \ldots, n$.
- La variation de l'exposition ou du nombre de décès dans une base de données peut être importante. Pour atténuer l'effet de cette variation, on choisit :

$$\delta_i^s \propto \widehat{\xi}_i^{-s}, \qquad \text{pour } i = 1, \ldots, n \text{ et } 0 \leq s \leq 1,$$

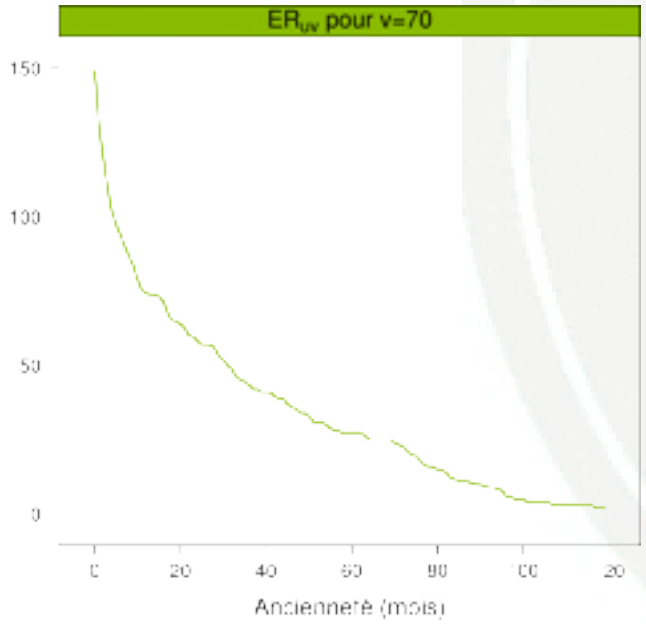
où s est un paramètre de sensibilité et

$$\widehat{\xi}_i = \begin{cases} E_i / \sum_{j=1}^n E_j \text{ pour } i = 1, \ldots, n & \text{pour des annuités,} \\ D_i / \sum_{j=1}^n D_j \text{ pour } i = 1, \ldots, n & \text{pour des prestations décès.} \end{cases}$$

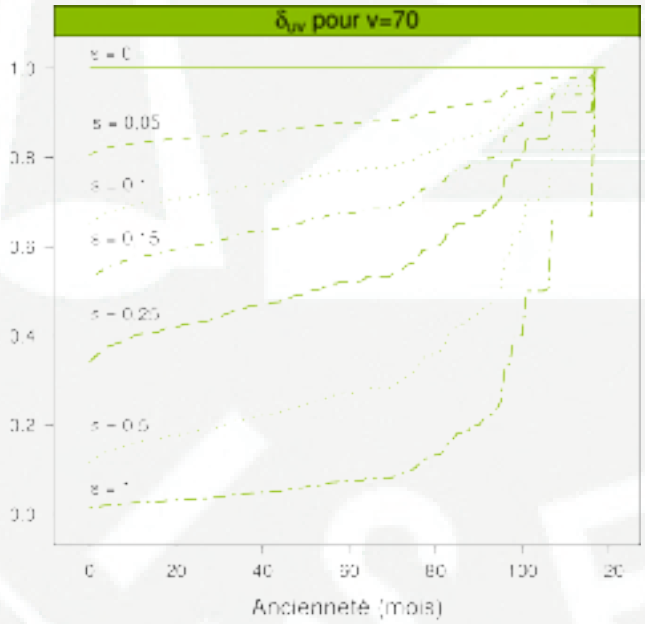
Facteurs de correction de la fenêtre d'observation

Principes II

- Choisir $s = 0$ donne un modèle avec une fenêtre fixe, alors que $s = 1$ peut entrainer une grande variation du montant de lissage.
- L'exposition observée décide de la forme des facteurs de correction, et le paramètre de sensibilité s détermine l'amplitude de cette forme, qui devient plus prononcée lorsque s se rapproche de 1.



(a) Exposition $E_{u,70}$

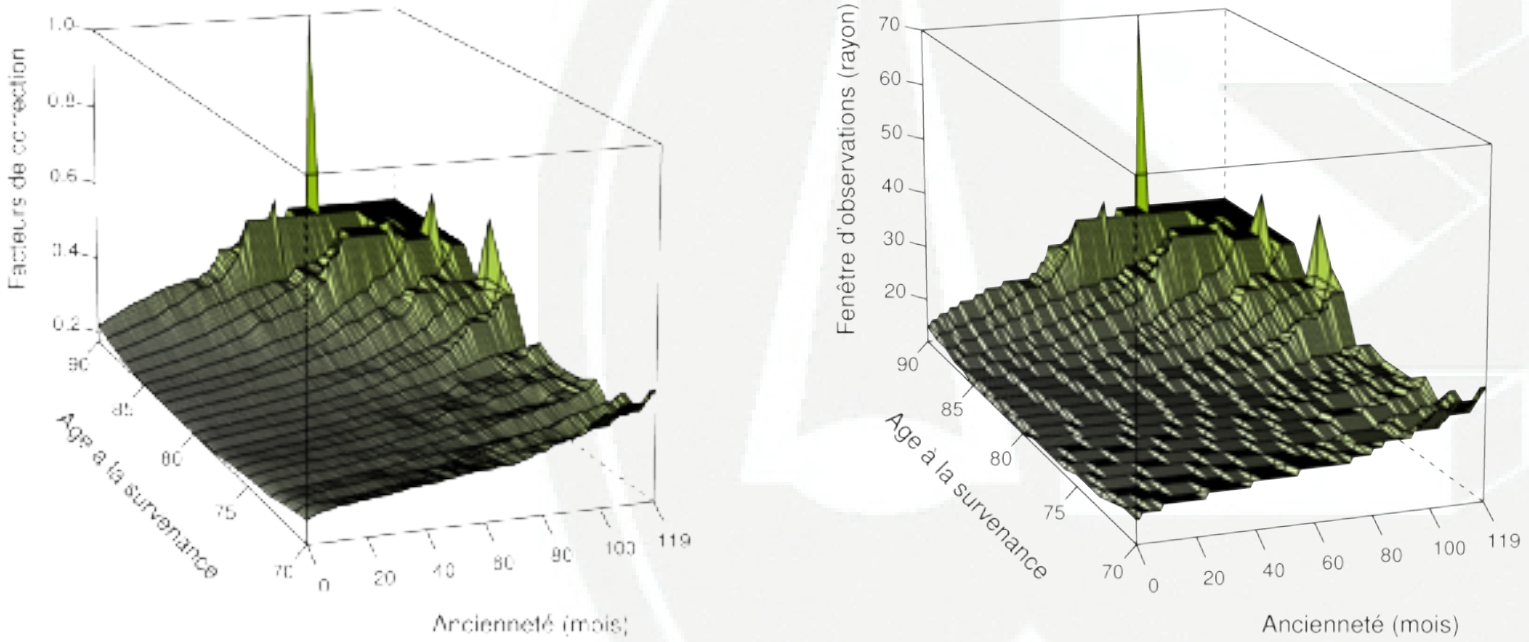


(b) Facteurs de corrections $\delta_{u,70}$

Facteurs de correction de la fenêtre d'observation

Principes III

Si l'exposition est faible, alors $\delta^s_{u,v}$ est large. Cela permet d'appliquer un lissage plus important. Inversement si l'exposition est large.



(a) Facteurs de correction $\delta_{u,v}$ (b) Fenêtre d'observation (rayon)

FIGURE: $\delta_{u,v}$ pour $s = 0.15$ et les fenêtres d'observation correspondantes.

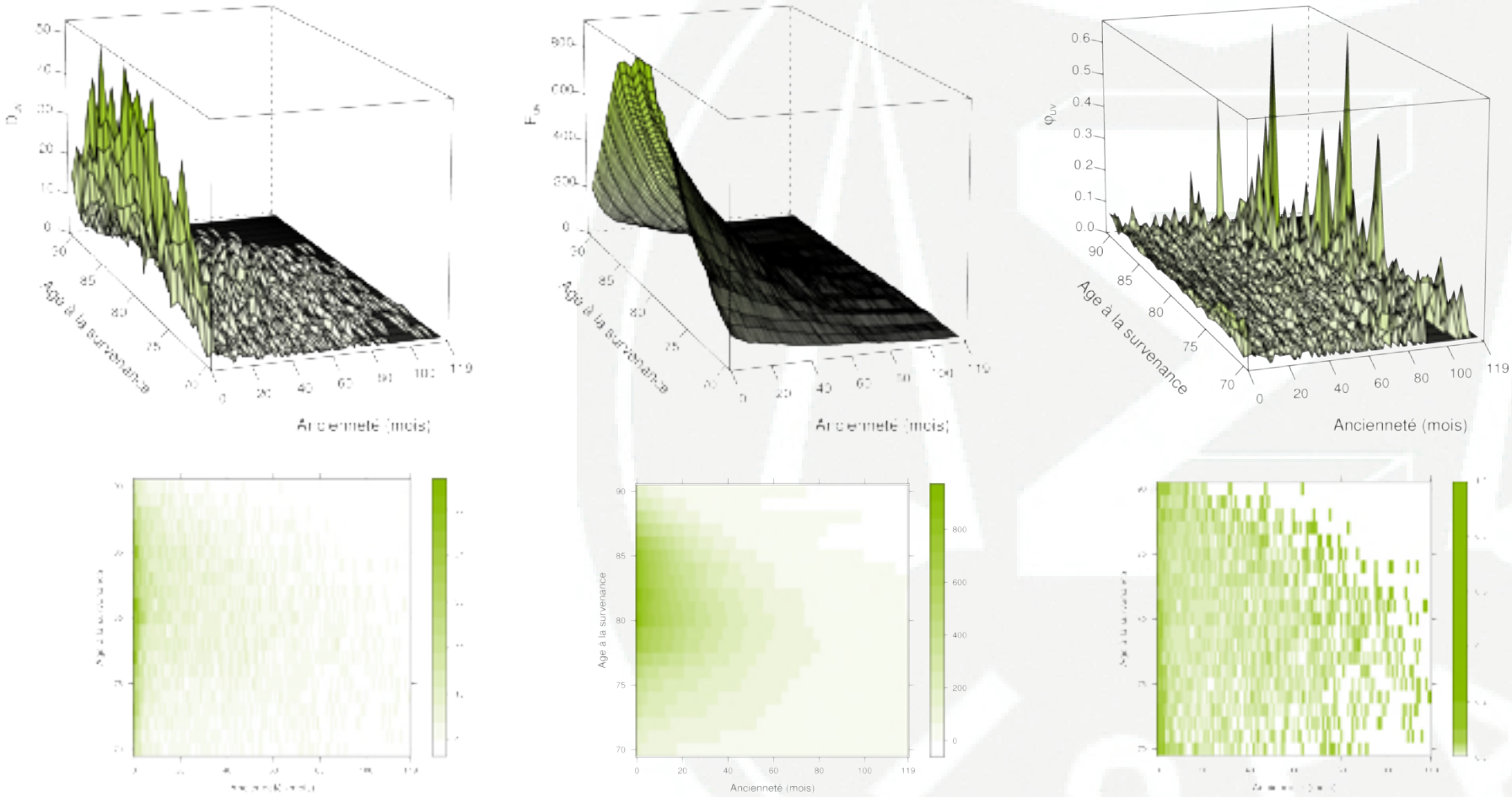
Application à l'assurance dépendance

Présentation des données I

- Analyse les variations de la mortalité en fonction de l'âge de survenance de la pathologie v et de l'ancienneté u (mois). On observe :
  - la plage d'âge de survenance : v $\in$ [70, 90] ans pour la survenance,
  - la durée de dépendance (l'ancienneté) : u $\in$ [0, 119] mois,
  - sur la période : 01/01/1998 à 31/12/2010.
- Les données on été agrégées selon l'âge de survenance et l'ancienneté. Les pathologies sont composées entre autres de démences, maladies neurologiques et de cancers en phase terminale. Les données sont composées de 2/3 de femmes et 1/3 d'hommes.

Application à l'assurance dépendance

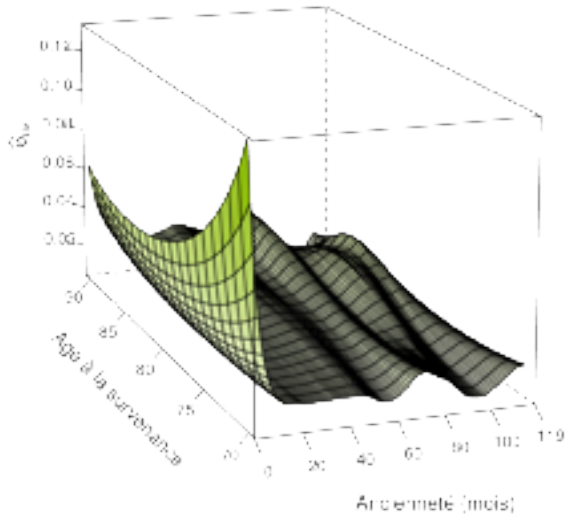
Présentation des données II



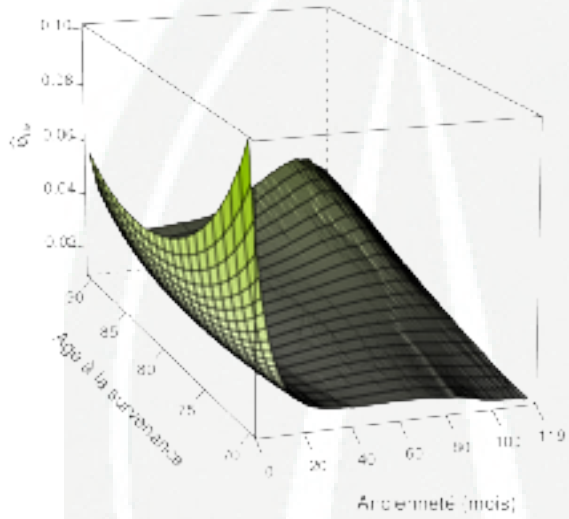
(a) Nombre de décés, $D_{u,v}$ (b) Exposition au risque, $E_{u,v}$ (c) Forces de mortalité, $\varphi_u(v)$

Application à l'assurance dépendance

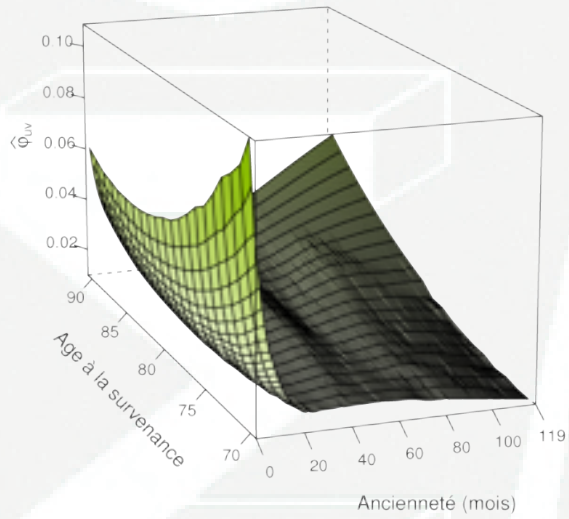
Surfaces ajustées I



(d) Vraisemblance locale



(e) ICI



(f) Facteurs de correction

FIGURE: Surface ajustées $\widehat{\varphi}_{u,v}$

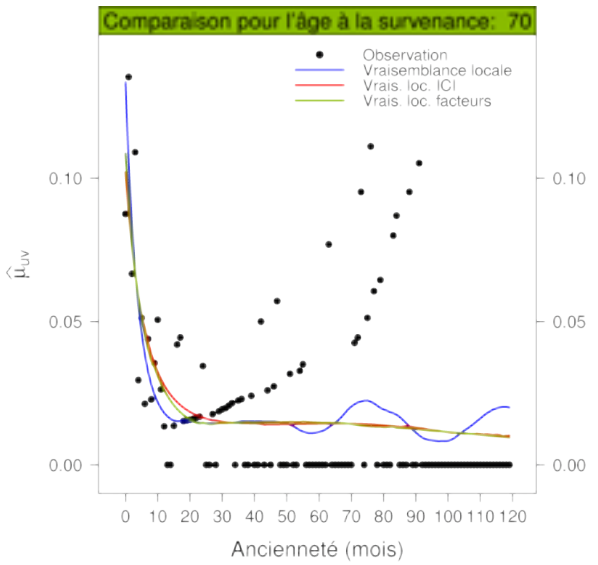
(a) Fonction de poids Epanechnikov, polynôme de degré 2 et fenêtre (rayon) de 13 observations.

(b) Ordre d'approximation fixe (quadratique), et fonction de poids Epanechnikov et $c = 0.1$.

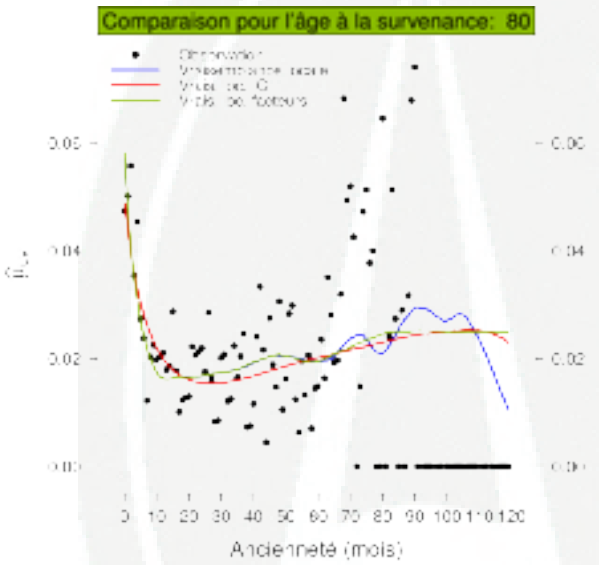
(c) idem, et $s = 0.15$.

Application à l'assurance dépendance

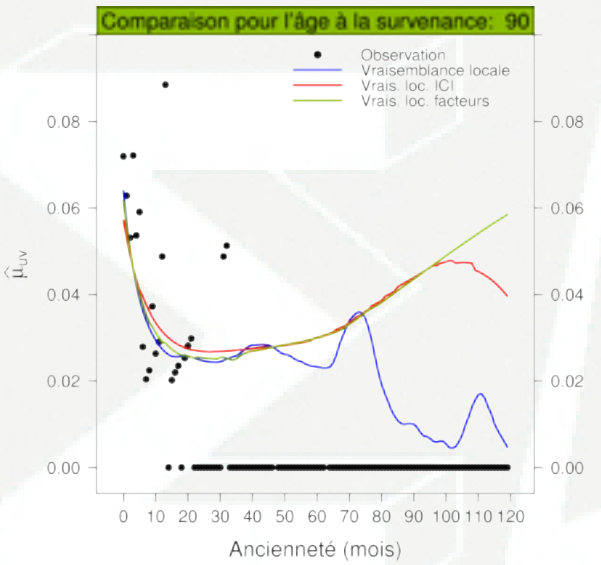
Surfaces ajustées II



(a) $v = 70$



(b) $v = 80$

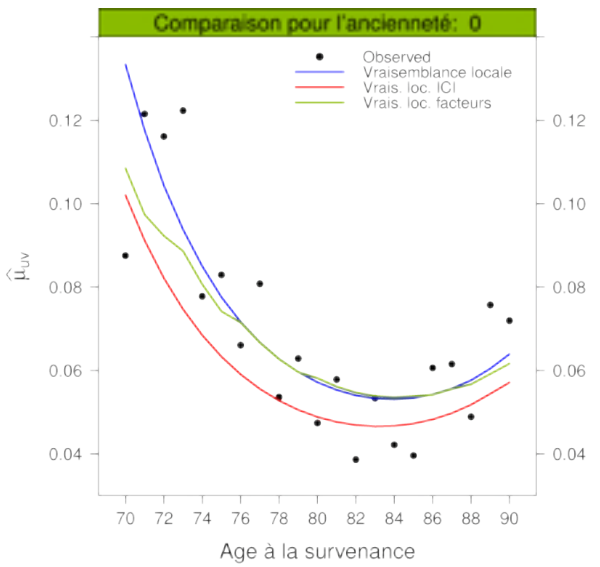


(c) $v = 90$

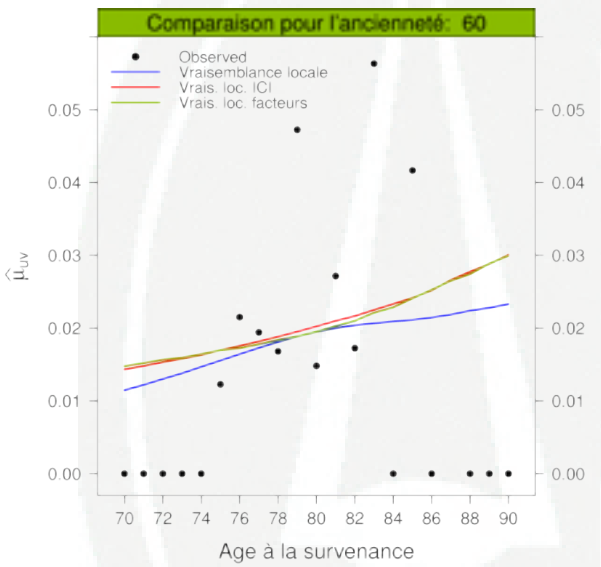
FIGURE: Forces de mortalité observées et ajustées pour plusieurs âges à la survenance.

Application à l'assurance dépendance

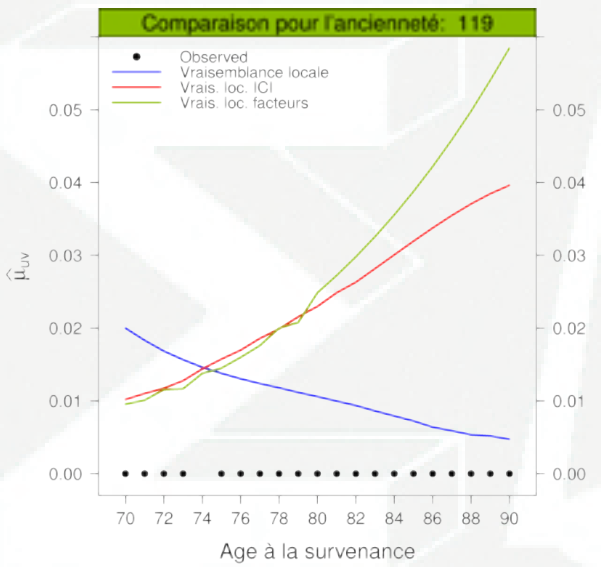
Surfaces ajustées III



(a) $u = 0$



(b) $u = 60$



(c) $u = 119$

FIGURE: Forces de mortalité observées et ajustées pour plusieurs anciennetés.

Application à l'assurance dépendance

Remarques

- ICI : Parce qu'une classe plus large d'estimateurs est disponible, cela peut affecter l'élimination des irrégularités. Pour y remédier, on choisit des fenêtres d'observation relativement large.
- ICI : Procédure demande un temps de calcul plus élevé : l'effort est multiplié par le nombre d'observations.
- Facteurs de corrections : permet d'inclure dans une procédure globale l'information additionnelle obtenue par le changement de l'exposition. Le changement d'exposition détermine la forme des facteurs, et l'amplitude est donnée par un paramètre de sensibilité.
- Il apparait que la méthode permet un ajustement satisfaisant où les autres méthodes ne parviennent pas à modéliser ces caractéristiques, cf. Tomas and Planchet (2013).

Références I

Chichignoud, M. (2010). *Performances statistiques d'estimateurs non-linéaires*. Ph.D. thesis, Université de Provence - U.F.R Mathématiques.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**(368), 829–836.

Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression : An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, **83**, 596–610.

Copas, J. B. and Haberman, S. (1983). Non-parametric graduation using kernel methods. *Journal of the Institute of Actuaries*, **110**, 135–156.

Fan, J., Farmen, M., and Gijbels, I. (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society*, **60**(3), 591–608.

Gavin, J. B., Haberman, S., and Verrall, R. J. (1993). Moving weighted graduation using kernel estimation. *Insurance : Mathematics & Economics*, **12**(2), 113–126.

Gavin, J. B., Haberman, S., and Verrall, R. J. (1995). Graduation by kernel and adaptive kernel methods with a boundary correction. *Transactions of the Society of Actuaries*, **47**, 173–209.

Références II

Goldenshulger, A. and Nemirovski, A. (1997). On spatially adaptive estimation of nonparametric regression. *Mathematical methods of statistics*, **6**(2), 135–170.

Haberman, S. (1996). Landmarks in the history of actuarial science (up to 1919). *Actuarial Research Paper No. 84, Dept. of Actuarial Science and Statistics, City University, London*.

Katkovnik, V. (1999). A new method for varying adaptive bandwidth selection. *IEEE Transactions on signal processing*, **47**(9), 2567–2571.

Katkovnik, V., Foi, A., Egiazarian, K. O., and Astola, J. T. (2005). Anisotropic local likelihood approximations : Theory, algorithm, applications. In E. R. Dougherty, J. T. Astola, and K. O. Egiazarian, editors, *Proceedings of the SPIE - Image processing algorithms and systems IV*, volume 5672, pages 181–192.

Loader, C. R. (1996). Local likelihood density estimation. *The Annals of Statistics*, **24**(4), 1602–1618.

Loader, C. R. (1999). *Local Regression and Likelihood*. Statistics and Computing Series. New York : Springer Verlag.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, volume 37 of *Monographs on Statistics and Applied Probability*. Boca Raton : Chapman & Hall / CRC Press, second edition.

Références III

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, **135**, 370–384.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, **33**(3), 1065–1076.

R Development Core Team (2013). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http ://www.R-project.org.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**(3), 832–837.

Seal, H. L. (1982). Graduation by piecewise cubic polynomials : a historical review. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik*, **15**, 89–114.

Stone, C. J. (1977). Consistent nonparametric regression (with discussion). *Annals of Statistics*, **5**(4), 595–645.

Stone, C. J. (1982). Optimal rates of convergence for nonparametric regression. *The Annals of Statistics*, **10**(4), 1040–1053.

Tibshirani, R. J. and Hastie, T. J. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, **82**(398), 559–567.

Références IV

Tomas, J. and Planchet, F. (2013). Multidimensional smoothing by adaptive local kernel-weighted log-likelihood with application to long-term care insurance. *Insurance : Mathematics & Economics*, **52**(3), 573–589.