Echantillonnage optimal et robuste en régression linéaire avec un processus d'erreur Höldérien

Mohamad BELOUNI

Cinquièmes Rencontres des Jeunes Statisticien-ne-s
Aussois, août 26–30, 2013.

Laboratoire Jean Kuntzmann

Equipe IPS

Ecole Doctorale Mathématiques, Sciences et Technologies
de l'Information, Informatique

Université Joseph Fourier

Directeur de thèse : Karim Benhenni

LABORATOIRE
JEAN-KUNTZMANN
MATHÉMATIQUES APPLIQUÉES - INFORMATIQUE

INFÉRENCE
PROCESSUS
STOCHASTIQUES

Plan de l'exposé

Modèle

Pour une fonction de concentration g définie dans un intervalle d'observation $[t_0, t_n]$, on considère le modèle :

$$X(t_i) \quad = \quad g(t_i) + \varepsilon(t_i) \qquad i = 0, 1, \ldots, n,$$

où $\{t_0, t_1, \ldots, t_n\}$ sont les instants d'observation de X dans l'intervalle $[t_0, t_n]$. Le processus stationnaire centré $\varepsilon(t)$ appartient à la classe de Hölder d'indice α et dont la fonction d'autocovariance $R(t, s) = R(t - s)$ peut se développer autour de 0 de la façon suivante pour $0 < \alpha < 2$:

$$R(t - s) = R(0) - \frac{1}{2}\lambda_\alpha |t - s|^\alpha + o(|t - s|^\alpha)$$

avec

$$\lambda_\alpha = R^{(\alpha)}(0-) - R^{(\alpha)}(0+) = \text{constante strictement positive},$$

où $R^{(\alpha)}(0\pm)$ sont les derivées fractionnaires d'ordre α de R à droite et à gauche de 0 :

$$\begin{aligned} R^{(\alpha)}(0+) &= \lim_{\tau \longrightarrow 0+} \frac{R(\tau) - R(0)}{\tau^\alpha} \\ R^{(\alpha)}(0-) &= \lim_{\tau \longrightarrow 0-} \frac{R(0) - R(\tau)}{(-\tau)^\alpha}. \end{aligned}$$

On fixe les points $t_0 = 0$ et $t_n = 1$, et on s'intéresse au choix optimal des points intermédiaires $\{t_1, \dots, t_{n-1}\}$ pour l'estimation de l'intégrale suivant :

$$\mathrm{AUC}(g) = \int_0^1 g(t)\,\mathrm{d}t.$$

On note l'estimateur linéaire de $I(g)$ par $L_n(g)$. Le but est de sélectionner les temps d'observations $\{t_1^*, \dots, t_{n-1}^*\}$ qui minimisent l'erreur quadratique moyenne

$$\mathrm{EQM}(L_n(g)) = \mathbb{E}(L_n(g) - \mathrm{AUC}(g))^2.$$

Plans d'échantillonnage réguliers de taille $n + 1$:

On suppose que $h(t)$ est une densité positive sur $[0, 1]$ avec une fonction de répartition strictement positive

$$H(t) = \int_0^t h(s)\,\mathrm{d}s, \qquad 0 \leq t \leq 1.$$

Le plan d'échantillonnage régulier défini par :

$$T_n = \left\{ t_k = H^{-1}\left(\frac{k}{n}\right), k = 0, \ldots, n \right\},$$

comprend les bornes de l'intervalle $[0, 1]$, $t_0 = 0$ et $t_n = 1$. Pour une densité uniforme sur $[0, 1]$, $h \equiv 1_{[0,1]}$, le plan d'échantillonnage régulier est périodique.

1. **Définition du modèle et plans d'échantillonnages**

2. **Estimation de la fonction de concentration et de l'AUC**

    - Modèle de régression linéaire multiple
    - Estimateur BLUE et plan d'échantillonnage optimal
    - Estimateur simple et plan d'échantillonnage optimal

3. **Perspectives**

On pose

$$g(t) = \beta_1 f_1(t) + \cdots + \beta_q f_q(t) = \boldsymbol{f}'(t)\boldsymbol{\beta},$$

où $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_q)$ est un vecteur de paramètres inconnus et

$\boldsymbol{f}'(t) = (f_1(t), \ldots, f_q(t))$ est un vecteur de fonctions de régression

connues et q est un nombre entier fixé positif. Chaque fonction de régression $f_j(t), j = 1, \ldots, q$, peut s'écrire sous la forme :

$$f_j(t) = \int_0^1 R(t,s)\varphi_j(s)\,\mathrm{d}s,\ j = 1, \ldots, q \quad t \in [0,1] \tag{1}$$

où φ_j, $j = 1, \ldots, q$, sont des fonctions continues sur $[0,1]$, φ_j/h, $j = 1, \ldots, q$ sont deux fois continûment différentiable.

1. **Définition du modèle et plans d'échantillonnages**

2. **Estimation de la fonction de concentration et de l'AUC**

   - Modèle de régression linéaire multiple
   - Estimateur BLUE et plan d'échantillonnage optimal
   - Estimateur simple et plan d'échantillonnage optimal

3. **Perspectives**

Estimateur BLUE

L'estimateur BLUE de β sur tout l'intervalle d'observation $[0,1]$ est défini par

$$\hat{\beta} = \boldsymbol{S}^{-1} \int_0^1 X(t)\varphi(t)\,\mathrm{d}t,$$

où $\varphi'(t) = (\varphi_1(t), \ldots, \varphi_q(t))$, et la matrice $\boldsymbol{S}$ de termes généraux

$$s_{ij} = \int_0^1 \int_0^1 \varphi_i(t) R(t-s) \varphi_j(s)\,\mathrm{d}t\,\mathrm{d}s, \quad i,j = 1, \ldots, q$$

est supposée être inversible. On estime alors l'$\mathrm{AUC}(g)$ par :

$$L(g) = \boldsymbol{z}'\hat{\beta},$$

où $\boldsymbol{z}' = (z_1, , \ldots, z_q)$; $z_j = \int_0^1 f_j(t)\,\mathrm{d}t$, $j = 1, \ldots, q$.

Si $X(t)$ est observé en $n+1$ points d'un plan d'échantillonnage régulier $\boldsymbol{T_n} = \{t_k\}_0^n$ dans l'intervalle $[0,1]$, alors l'estimateur BLUE $\hat{\beta}_{\boldsymbol{n}}^{\text{blue}}$ du vecteur paramètre β est défini par :

$$\hat{\beta}_{\boldsymbol{n}}^{\text{blue}} = \boldsymbol{A}_{T_n}^{-1}\boldsymbol{\eta},$$

où

$$\boldsymbol{A}_{T_n} = (a_{rs})_{q\times q},$$

de termes généraux

$a_{rs} = \sum_{i,j=0}^{n} f_r(t_i)\boldsymbol{R}_{\boldsymbol{T_n}}^{-1}(t_i - t_j)f_s(t_j)\ r,s = 1,\ldots,q$ et

$$\boldsymbol{\eta} = (\eta_r)_{q\times 1},$$

où

$$\eta_r = \sum_{i,j=0}^{n} f_r(t_i)\boldsymbol{R}_{\boldsymbol{T_n}}^{-1}(t_i - t_j)X(t_j),\ r = 1,\ldots,q$$

et $\boldsymbol{R}_{\boldsymbol{T_n}} = (\boldsymbol{R}_{\boldsymbol{T_n}}(t_k - t_j))_{(n+1)\times(n+1)}$ est supposée être inversible pour tout n.

L’estimateur BLUE de l’$\mathrm{AUC}(g)$ est défini par :

$$L_n^{\mathrm{blue}}(g) = \boldsymbol{z}'\hat{\beta}_{\boldsymbol{n}}^{\mathrm{blue}} = \boldsymbol{z}'\boldsymbol{A}_{T_n}^{-1}\boldsymbol{\eta}.$$

Définition

Soit ψ une fonction continûment différentiable de l'ensemble des matrices carrées définies positives de dimension q dans $\mathbb{R}$. Un plan d'échantillonnage $\boldsymbol{T}_n^*$ est dit asymptotiquement ψ-optimal si

$$\lim_{n \longrightarrow \infty} \left\{ \psi\left(Var(\hat{\beta}_{\boldsymbol{T}_n^*}) \right) - \psi\left(\boldsymbol{S}^{-1} \right) \right\} \times \left\{ \inf_{\boldsymbol{T}_n} \psi\left(Var(\hat{\beta}_{\boldsymbol{n}}^{blue}) \right) - \psi\left(\boldsymbol{S}^{-1} \right) \right\}^{-1} = 1.$$

où l'infinimum est pris parmi tous les plans d'échantillonnages de taille $n+1$.

Pour la suite, pour une matrice fixée symétrique $\boldsymbol{M}$ définie positive, on considère $\psi(\boldsymbol{B}) = \mathrm{trace}\,(\boldsymbol{BM})$, où $\boldsymbol{B}$ une matrice carrée définie positive de dimension q.

Sacks et Ylvisaker (1968) ont montré le résultat suivant pour $\alpha = 1$.

Théorème

L'estimateur $\hat{\beta}_{\boldsymbol{n}}^{blue}$, *satisfait*

$$\lim_{n \longrightarrow \infty} n^2 \left(\inf_{\boldsymbol{T_n}} \psi \left(Var(\hat{\beta}_{\boldsymbol{n}}^{blue}) \right) - \psi(\boldsymbol{S}^{-1}) \right) = \frac{\lambda_1}{12} \left(\int_0^1 \left\{ \varphi'(t) \left(\boldsymbol{S}^{-1} \boldsymbol{M} \boldsymbol{S}^{-1} \right) \varphi(t) \right\}^{1/3} \mathrm{d}t \right)^3 .$$

Par ailleurs, un plan d'échantillonnage $\boldsymbol{T}_{\boldsymbol{n}}^{*}$ *engendré par la densité*

$$h^*(t) = \left\{ \varphi'(t) \left(\boldsymbol{S}^{-1} \boldsymbol{M} \boldsymbol{S}^{-1} \right) \varphi(t) \right\}^{1/3} \Big/ \int_0^1 \left\{ \varphi'(u) \left(\boldsymbol{S}^{-1} \boldsymbol{M} \boldsymbol{S}^{-1} \right) \varphi(u) \right\}^{1/3} \mathrm{d}u,$$

est asymptotiquement ψ*-optimal.*

Corollaire

L'estimateur $\hat{\beta}^{blue}_{\boldsymbol{T}_n^*}$*, construit à partir de* $\boldsymbol{T_n^*}$ *et engendré par la densité* $h^*(t)$*, vérifie*

$$\lim_{n\longrightarrow\infty} n^2\left(\psi\left(Var(\hat{\beta}^{blue}_{\boldsymbol{T}_n^*})\right) - \psi\left(\boldsymbol{S}^{-1}\right)\right)$$

$$= \quad \frac{\lambda_1}{12}\left(\int_0^1 \left\{\varphi'(t)\left(\boldsymbol{S}^{-1}\boldsymbol{M}\boldsymbol{S}^{-1}\right)\varphi(t)\right\}^{1/3}\mathrm{d}t\right)^3 .$$

Remarque : d'autres critères d'optimalité sont définis dans Sacks et Ylvisaker (1968) mais les résultats asymptotiques obtenus ici peuvent être généralisés à ces critères.

Inconvenients :

Les estimateurs BLUE, $\hat{\beta}_{n}^{\text{blue}}, L_n^{\text{blue}}(g)$ sont instables et non robustes car ils utilisent l'inverse de la matrice d'autocovariance (inconnue en pratique).

Remède :

On construit deux estimateurs $\hat{\beta}_{n}^{\text{trap}}$ et $L_n^{\text{trap}}(g)$ simples et stables qui ne dépendent pas de la matrice d'autocovariance et qui utilisent seulement les observations aux points du plan d'échantillonnage régulier.

1. **Définition du modèle et plans d'échantillonnages**

2. **Estimation de la fonction de concentration et de l'AUC**
   - Modèle de régression linéaire multiple
   - Estimateur BLUE et plan d'échantillonnage optimal
   - Estimateur simple et plan d'échantillonnage optimal

3. **Perspectives**

Estimateur simple et plan d'échantillonnage optimal

Si $X(t)$ est observé en $n+1$ points d'un plan d'échantillonnage régulier $\boldsymbol{T_n} = \{t_k\}_0^n$ dans l'intervalle $[0, 1]$, alors on construit l'estimateur simple de β par :

$$\hat{\beta}_{\boldsymbol{n}}^{\text{trap}} = \boldsymbol{S}_n^{*\,-1} \boldsymbol{L_{n,q}}(X),$$

où

$$\boldsymbol{L_{n,q}}(X) = \frac{1}{2n} \sum_{k=0}^{n-1} \left((\frac{\varphi X}{h})(t_k) + (\frac{\varphi X}{h})(t_{k+1}) \right),$$

est un vecteur colonne

$$\boldsymbol{S_n^*} = \left(s_{ij}^*\right)_{q \times q},$$

est supposée être inversible,

$$s_{ij}^* = \frac{1}{2n} \sum_{k=0}^{n-1} \left((\frac{\varphi_i}{h} f_j)(t_k) + (\frac{\varphi_i}{h} f_j)(t_{k+1}) \right), \quad i, j = 1, \ldots, q$$

où $f_j(t), j = 1, \ldots, q$ sont définies par (1),

Proposition

L'estimateur $\hat{\beta}_{\boldsymbol{n}}^{trap}$, vérifie les propriétés suivantes :

(i) $\quad Biais(\hat{\beta}_{\boldsymbol{n}}^{trap}) = 0 \quad pour\ tout\ n$

(ii) $\quad Var(\hat{\beta}_{\boldsymbol{n}}^{trap}) \longrightarrow Var(\hat{\beta}) = \boldsymbol{S}^{-1}\ quand\ n \to \infty$

alors on construit l'estimateur simple de l'$\mathrm{AUC}(g)$ par :

$$L_n^{\text{trap}}(g) = \boldsymbol{z}'\hat{\beta}_{\boldsymbol{n}}^{\text{trap}} = \boldsymbol{z}'\boldsymbol{S}_n^{*-1}\boldsymbol{L}_{\boldsymbol{n,q}}(X),$$

où $\boldsymbol{z}' = (z_1, , \ldots, z_q)$; $z_j = \int_0^1 f_j(t)\,\mathrm{d}t$, $j = 1, \ldots, q$.

Proposition

L'estimateur $L_n^{trap}(g)$, vérifie les propriétés suivantes :

(i) $\quad Biais(L_n^{trap}(g)) = 0 \quad pour\ tout\ n$

(ii) $\quad Var(L_n^{trap}(g)) \longrightarrow Var(L(g)) = \boldsymbol{z}'\boldsymbol{S}^{-1}\boldsymbol{z}\ quand\ n \to \infty$

Théorème

Pour $\alpha = 1$, *l'estimateur simple* $\hat{\beta}_n^{trap}$, *vérifie aussi*

$$\lim_{n \longrightarrow \infty} n^2 \left(\psi(Var(\hat{\beta}_n^{trap})) - \psi(S^{-1}) \right) = \frac{\lambda_1}{12} \int_0^1 \frac{\varphi'(t) \left(S^{-1} M S^{-1}\right) \varphi(t)}{h^2(t)} \, dt.$$

Par ailleurs, le plan d'échantillonnage T_n^* *engendré par la densité*

$$h^*(t) = \left\{\varphi'(t) \left(S^{-1} M S^{-1}\right) \varphi(t)\right\}^{1/3} \Big/ \int_0^1 \left\{\varphi'(u) \left(S^{-1} M S^{-1}\right) \varphi(u)\right\}^{1/3} du,$$

est asymptotiquement ψ*-optimal.*

Corollaire

L'estimateur $\hat{\beta}^{trap}_{\boldsymbol{T}^*_n}$, *construit à partir de* $\boldsymbol{T}^*_n$ *et engendré par la densité* $h^*(t)$, *vérifie*

$$\lim_{n \longrightarrow \infty} n^2 \left(\psi(Var(\hat{\beta}^{trap}_{\boldsymbol{T}^*_n})) - \psi(\boldsymbol{S}^{-1}) \right)$$

$$= \frac{\lambda_1}{12} \left(\int_0^1 \left\{ \varphi'(t) \left(\boldsymbol{S}^{-1} \boldsymbol{M} \boldsymbol{S}^{-1} \right) \varphi(t) \right\}^{1/3} \mathrm{d}t \right)^3 .$$

Théorème

Pour $\alpha = 1$, *alors l'estimateur* $L_n^{trap}(g)$ *de l'AUC*(g), *vérifie*

$$\lim_{n \longrightarrow \infty} n^2 \left(\psi\left(Var(L_n^{trap}(g))\right) - \psi\left(Var(L(g))\right) \right)$$

$$= \frac{\lambda_1 tr(\boldsymbol{M})}{12} \int_0^1 \frac{\varphi^{'}(t)\left(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}^{-1}\right)\varphi(t)}{h^2(t)}\,\mathrm{d}t,$$

où $A = \boldsymbol{z}\boldsymbol{z}^{'}$.

Par ailleurs, le plan d'échantillonnage $\boldsymbol{T}_n^*$ *engendré par la densité*

$$h^*(t) = \left\{\varphi^{'}(t)\left(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}^{-1}\right)\varphi(t)\right\}^{1/3} \Big/ \int_0^1 \left\{\varphi^{'}(u)\left(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}^{-1}\right)\varphi(u)\right\}^{1/3}\mathrm{d}u,$$

est asymptotiquement ψ*-optimal.*

Corollaire

L'estimateur $L^{trap}_{T^*_n}(g)$*, construit à partir de* $\boldsymbol{T^*_n}$ *et engendré par la densité* $h^*(t)$*, vérifie*

$$\lim_{n\longrightarrow\infty} n^2\left(\psi(Var(L^{trap}_{T^*_n}(g))) - \psi(Var(L(g)))\right)$$

$$= \quad \frac{\lambda_1 tr(\boldsymbol{M})}{12}\left(\int_0^1 \left\{\varphi^{'}(t)\left(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}^{-1}\right)\varphi(t)\right\}^{1/3} \mathrm{d}t\right)^3,$$

où $A = \boldsymbol{zz}^{'}$.

Exemple

On considère le modèle de régression simple ($q = 1$)

$$X(t) = \beta t + \varepsilon(t), \qquad t \in [0, 1]$$

où $\varepsilon(t)$ est un processus d'erreur de type "Ornstein - Uhlenbeck" (Hölder d'indice $\alpha = 1$) et sa fonction d'autocovariance

$$R(t, s) = \sigma^2 e^{-\lambda|t-s|}, \lambda > 0.$$

On montre que $f(t) = t$ peut s'écrire sous la forme suivante

$$f(t) = \int_0^1 R(t, s)\varphi(s)\,\mathrm{d}s + \sum_{l=1}^{L} b_l R(t, a_l),$$

avec $\varphi(s) = \lambda s / \left(2\sigma^2\right), L = 2, a_1 = 0, a_2 = 1, b_1 = -1/\left(2\lambda\sigma^2\right),$ $b_2 = (\lambda + 1) / \left(2\lambda\sigma^2\right).$

La fonction de sauts est :

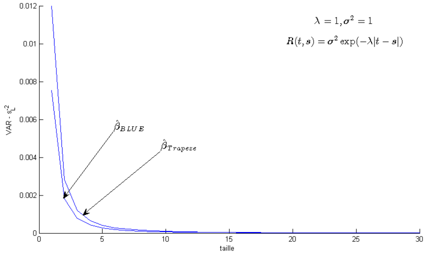
$$\lambda_1 = R^{(1)}(0-) - R^{(1)}(0+) = 2\lambda\sigma^2 > 0$$

La densité du plan d'échantillonnage (asymptotiquement) optimal $h^*(t)$ est

$$h^*(t) = (5/3)t^{2/3}$$

et les points optimaux sont

$$t_k^* = \{k/n\}^{3/5}, k = 0, \ldots, n.$$



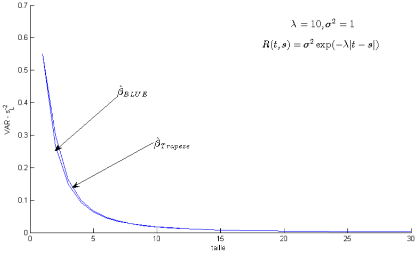


FIG.: $\left(\mathrm{Var}(\hat{\beta}_n) - \mathrm{Var}(\hat{\beta})\right)$ contre la taille de l'échantillon

Théorème

Si le processus d'erreur est Hölderien d'indice $0 < \alpha < 2$, *l'estimateur* $\hat{\beta}_{\boldsymbol{n}}^{trap}$, *vérifie*

pour $0 < \alpha < 1$

$$\begin{aligned}\lim_{n\longrightarrow\infty} n^{1+\alpha}\left(\psi(Var(\hat{\beta}_{\boldsymbol{n}}^{trap})) - \psi(\boldsymbol{S}^{-1})\right) \\ = -\zeta(-\alpha)\lambda_\alpha \int_0^1 \frac{\varphi'(t)\left(\boldsymbol{S}^{-1}\boldsymbol{M}\boldsymbol{S}^{-1}\right)\varphi(t)}{h^{1+\alpha}(t)}\,\mathrm{d}t.\end{aligned}$$

pour $1 < \alpha < 2$

$$\lim_{n\longrightarrow\infty} n^{2}\left(\psi(Var(\hat{\beta}_{\boldsymbol{n}}^{trap})) - \psi(\boldsymbol{S}^{-1})\right) = 0,$$

où ζ *est la fonction zêta de Riemann.*

La densité optimale qui minimise la constante asymptotique pour $0 < \alpha < 1$ *est :*

$$\begin{aligned} h^*(t) &= \left\{\varphi'(t)\left(\boldsymbol{S}^{-1}\boldsymbol{M}\boldsymbol{S}^{-1}\right)\varphi(t)\right\}^{1/(\alpha+2)} \\ &\quad / \int_0^1 \left\{\varphi'(u)\left(\boldsymbol{S}^{-1}\boldsymbol{M}\boldsymbol{S}^{-1}\right)\varphi(u)\right\}^{1/(\alpha+2)}\,\mathrm{d}u.\end{aligned}$$

Corollaire

L'estimateur $\hat{\beta}^{trap}_{\boldsymbol{T}^*_n}$ *, construit à partir de* $\boldsymbol{T}^*_n$ *et engendré par la densité* $h^*(t)$ *pour* $0 < \alpha < 1$ *vérifie*

$$\lim_{n \longrightarrow \infty} n^{1+\alpha} \left(\psi(Var(\hat{\beta}^{trap}_{\boldsymbol{T}^*_n})) - \psi(\boldsymbol{S}^{-1}) \right)$$

$$= -\zeta(-\alpha)\lambda_\alpha \left(\int_0^1 \left\{ \varphi^{'}(t) \left(\boldsymbol{S}^{-1} \boldsymbol{M} \boldsymbol{S}^{-1} \right) \varphi(t) \right\}^{1/(\alpha+2)} \mathrm{d}t \right)^{\alpha+2} .$$

Théorème

Si le processus d'erreur est Hölderien d'indice $0 < \alpha < 2$, *l'estimateur* $L_n^{trap}(g)$ *de l'AUC*(g) *vérifie*

lorsque $0 < \alpha < 1$,

$$\lim_{n \longrightarrow \infty} n^{1+\alpha} \left(\psi \left(Var(L_n^{trap}(g)) \right) - \psi \left(Var(L(g)) \right) \right)$$
$$= -\zeta(-\alpha)\lambda_\alpha tr(\boldsymbol{M}) \int_0^1 \frac{\varphi^{'}(t) \left(\boldsymbol{S}^{-1} \boldsymbol{A} \boldsymbol{S}^{-1} \right) \varphi(t)}{h^{1+\alpha}(t)} \, \mathrm{d}t,$$

lorsque $1 < \alpha < 2$,

$$\lim_{n \longrightarrow \infty} n^2 \left(\psi \left(Var(L_n^{trap}(g)) \right) - \psi \left(Var(L(g)) \right) \right) = 0,$$

où $A = \boldsymbol{z}\boldsymbol{z}^{'}$.

la densité optimale qui minimise la constante asymptotique pour $0 < \alpha < 1$ *est :*

$$h^*(t) = \left\{ \varphi^{'}(t) \left(\boldsymbol{S}^{-1} \boldsymbol{A} \boldsymbol{S}^{-1} \right) \varphi(t) \right\}^{1/(\alpha+2)}$$
$$/ \int_0^1 \left\{ \varphi^{'}(u) \left(\boldsymbol{S}^{-1} \boldsymbol{A} \boldsymbol{S}^{-1} \right) \varphi(u) \right\}^{1/(\alpha+2)} \mathrm{d}u.$$

Corollaire

L'estimateur $L^{trap}_{\boldsymbol{T}^*_n}(g)$*, construit à partir de* $\boldsymbol{T}^*_n$ *et engendré par la densité* $h^*(t)$ *pour* $0 < \alpha < 1$ *vérifie :*

$$\lim_{n\longrightarrow\infty} n^{1+\alpha}\left(\psi\left(Var(L^{trap}_{\boldsymbol{T}^*_n}(g))\right) - \psi\left(Var(L(g))\right)\right)$$

$$= -\zeta(-\alpha)\lambda_\alpha tr(\boldsymbol{M})\left(\int_0^1\left\{\varphi'(t)\left(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}^{-1}\right)\varphi(t)\right\}^{1/(\alpha+2)}\mathrm{d}t\right)^{\alpha+2}.$$

Plans d'échantillonnage robustes selon le critère du Minimax

Soit

$$h^*(t) = \{v(t)\}^{1/(\alpha+2)} \Big/ \int_0^1 \{v(u)\}^{1/(\alpha+2)}\, \mathrm{d}u, \tag{2}$$

où $0 < \alpha \leq 1$ et

$$v(t) = \varphi^{'}(t) \left(\boldsymbol{S}^{-1}\boldsymbol{M}\boldsymbol{S}^{-1}\right) \varphi(t).$$

Le plan d'échantillonnage asymptotiquement optimal $\boldsymbol{T}_n^*$ engendré par la densité optimale $h^*(t)$ peut ne pas être robuste par rapport à une mauvaise spécification de la fonction d'autocovariance R et de la fonction φ. Pour cette raison, on propose le critère du minimax pour la détermination du plan d'échantillonnage optimal et d'appeler une densité d'échantillonnage h^* est optimale selon le critère du Minimax pour l'estimation de β ou de l'$\mathrm{AUC}(g)$ si elle minimise

$$\max \left\{\Psi_v(h)\, | v \in \Lambda\right\},$$

où

$$\Psi_v(h) = \int_0^1 \frac{v(t)}{h^{1+\alpha}(t)}\, \mathrm{d}t,$$

et Λ est une classe appropriée des fonctions définie par :

$$\Lambda = \left\{ v \in C\,[0,1] \ \ \text{tel que} \int_0^1 v(t)\mathrm{d}t \leq \epsilon \right\}.$$

Théorème

Le plan d'échantillonnage régulier $\boldsymbol{T_n^*}$ *engendré par la densité* $h^*(t)$ *définie par (2) est optimal selon le critère du minimax.*

1. **Définition du modèle et plans d'échantillonnages**

2. **Estimation de la fonction de concentration et de l'AUC**
   - Modèle de régression linéaire multiple
   - Estimateur BLUE et plan d'échantillonnage optimal
   - Estimateur simple et plan d'échantillonnage optimal

3. **Perspectives**

Références :

[1] BELOUNI, M. ET BENHENNI, K. (2012). Plans d'expérience pour l'estimation de la courbe de concentration et de l'AUC. *Comptes Rendus de l'Académie des Sciences de Paris (CRAS)*, 350, Issue 13, 707–710.

[2] SACKS, J. AND YLVISAKER, D (1966). Designs for regression problems with correlated errors, *Ann. Math, 37, 66–89.*

[3] CAMBANIS, S. AND SU, Y (1993). Sampling designs for regression coefficient estimation with correlated errors, *Annals of the Institute of Statistical Mathematics, 46,707–722.*

Perspectives

Plans d'échantillonnage optimaux pour l'estimation non-paramétrique de la fonction de concentration et de l'AUC :

- Résultats asymptotiques (*n* grand).
- Taille fixée et petite : application à la pharmacocinétique (estimation + algorithmes d'optimalité).