

Programmation stochastique

SAA: méthodes adaptatives

Fabian Bastin

IFT-6512 – Hiver 2016

Rappel : nous considérons le problème stochastique

$$\min_{z \in S} g(z) = E_P [G(z, \xi)],$$

où $z \in \mathbb{R}^m$ est un vecteur de variables de décision, S est un sous-ensemble compact de \mathbb{R}^m représentant les solutions réalisables du problème ci-dessus, ξ est un vecteur aléatoire réel défini sur l'espace de probabilité (Ξ, \mathcal{F}, P) et prenant des valeurs dans $(\mathbb{R}^k, \mathcal{B}^k)$ (\mathcal{B}^k est la mesure de Borel), $G : \mathbb{R}^m \times \mathbb{R}^k \rightarrow \mathbb{R}$ est une fonction à valeurs réelles, et $E_P[\cdot]$ est l'espérance par rapport à la mesure P .

Approximation par moyenne échantillonnale :

$$\min_{z \in S} \hat{g}_N(z) = \frac{1}{N} \sum_{i=1}^N G(z, \xi_i),$$

Convergence

On a vu des résultats de consistance pour $N \rightarrow \infty$. De plus, le théorème de la limite centrale nous dit que, si les tirs sont indépendants et identiquement distribués (i.i.d.) (et $g(x)$ fini),

$$\frac{1}{\sqrt{N}}[\hat{g}_N(x) - g(x)] \Rightarrow N(0, \sigma^2(x)),$$

où $\sigma^2(x) = \text{var}(G(x, \xi))$, et \Rightarrow dénote la convergence en probabilité.

Le résultat ci-dessus n'est valide que pour x fixé. Il est nécessaire de poser de plus fortes conditions pour avoir une convergence fonctionnelle.

Remarquons tout d'abord que sous nos hypothèses de travail, $\hat{g}_N(x)$ est continue sur S , et donc peut être considéré comme un point dans l'espace de Banach $C(s)$.

L'espace de Banach $C(S)$

$C(S)$ est l'espace des fonction continues $\psi : S \rightarrow \mathbb{R}$, équipé de la norme sup $\|\psi\| := \sup_{x \in S} |\psi|$

$C(S)$ est un espace de Banach. Un espace de Banach est un espace vectoriel normé complet pour la distance issue de sa norme. un espace métrique M est dit complet ou espace complet si toute suite de Cauchy de M a une limite dans M (c'est-à-dire qu'elle converge dans M).

On va étendre le théorème de la limite centrale (en un point) à un théorème de la limite centrale fonctionnelle), en supposant comme toujours que les tirs sont i.i.d.

Théorème de la limite centrale fonctionnelle

Supposons que les conditions suivantes tiennent :

- 1 Pour tout $x \in S$, la fonction $G(x, \cdot)$ est mesurable (autrement dit, son espérance existe).
- 2 Pour un certain point $\bar{x} \in S$, l'espérance $E_P[G(\bar{x}, \xi)^2]$ est finie.
- 3 La condition de continuité de Lipschitz

$$|G(x_1, \xi) - G(x_2, \xi)| \leq K(\xi) \|x_1 - x_2\|,$$

pour une variable aléatoire à valeurs positives $K(\xi)$ telle que $E[K(\xi)]$ est finie, tient pour tout $x_1, x_2 \in S$ et presque tout ξ . De plus, nous supposons que la variable aléatoire $K(\xi)$ a un moment de second ordre fini.

Théorème de la limite centrale fonctionnelle (2)

Dans ces conditions,

$$N^{1/2}[\hat{g}_n - g] \rightarrow Y \in C(S).$$

Avec la condition i.i.d., Y est tel que pour n'importe quels points $x_1, \dots, x_k \in S$, le vecteur aléatoire $(Y(x_1), \dots, Y(x_k))^T$ a une distribution normale multivariée de matrice de covariance donnée par la matrice de covariance du vecteur $(G(x_1, \xi), \dots, G(x_k, \xi))$.

Dans le cas d'une minimisation globale, nous avons le résultat suivant. Si $N^{1/2}(\hat{g}_N - g) \Rightarrow Y \in C(S)$ (avec $\{\hat{g}_n\}$ et g dans $C(S)$), alors,

$$N^{1/2}(\hat{v}_N - v^*) \Rightarrow \min_{x \in S^*} Y(x).$$

où

$$\hat{v}_N = \min_{x \in S} \hat{g}_N(x) \text{ et } v^* = \min_{x \in S} g(x).$$

Si S^* est un singleton, dans le cas i.i.d. et les conditions précédentes,

$$N^{1/2}(\hat{v}_n - v^*) \Rightarrow N(0, \sigma^2(x^*)).$$

Sous certaines conditions supplémentaires, nous avons aussi la convergence de $E[\hat{v}_N]$ vers v^* .

Mais tous ses résultats deviennent difficiles à étendre pour le cas d'une optimisation locale.

Convergence des solutions globales (2)

Néanmoins, et c'était prévu, nous voyons que plus N est grand, plus nous devrions avoir un résultat précis. Mais plus N est grand, plus calculer la fonction approximative est coûteux, vu que

$$\hat{g}_N(z) = \frac{1}{N} \sum_{i=1}^N G(z, \xi_i).$$

Qu'est-ce qui nous intéresse ?

$$\min_{z \in S} \hat{g}_N(z) = \frac{1}{N} \sum_{i=1}^N G(z, \xi_i).$$

Autrement dit, pour un niveau donné d'approximation, défini par le nombre de tirs aléatoires, il est possible d'accélérer les premiers itérés au cours de la procédure d'optimisation en considérant des sous-ensembles de l'échantillonnage.

Méthode adaptative externe

Alternativement, on peut commencer avec un faible échantillon et l'étendre des itérations : la procédure d'échantillonnage adaptatif peut être **externe** à l'algorithme, ou **interne**.

Il y a donc **plusieurs stratégies possibles**.

Une approche externe consiste à appliquer l'algorithme d'optimisation de façon répétée avec des échantillons de tailles croissantes, comme exprimé ci-dessous.

Etape 0. Poser $k = 0$, N_{\max} et N_0 , avec $0 < N_0 \leq N_{\max}$. Définir un certain point réalisable \tilde{z} .

Etape 1. Résoudre (approximativement) \hat{g}_{N_k} avec \tilde{z} comme point de départ et soit $z_{N_k}^*$ la solution obtenue.

Etape 2. Si $N_k = N_{\max}$, arrêt. Sinon, poser N_{k+1} tel que $N_k < N_{k+1} < N_{\max}$, et $\tilde{z} = z_{N_k}^*$. Incrémenter k et retour à l'étape Step 1.

Méthode adaptative externe - interne

La difficulté majeure de cette procédure est de quantifier le mot "approximatif" de l'Etape 1. Si aucun soin n'est pris, l'algorithme résultant peut en fait consommer plus de temps que la minimisation directe de $\hat{g}_{N_{\max}}$.

On peut aussi remplacer le critère d'arrêt sur N_{\max} par un test d'hypothèse sur les conditions de criticalité (actuellement, seul le premier ordre a été considéré).

L'**approche interne** est une stratégie non-monotone, dépendante de la méthode d'optimisation sous-jacente. Nous considérerons ici le cas sans contraintes.

Plus précisément, générons un échantillon avant le processus d'optimisation, avec N_{\max} tirs aléatoires i.i.d. A l'itération k , nous utiliserons un sous-ensemble de cette échantillonnage initial, en utilisant N_k des N_{\max} tirs aléatoires.

Détermination de la précision

Pour la simplicité, nous utiliserons les N_k premiers tirs aléatoires. Ceci implique dès lors que \hat{g}_N est une fonction douce bien définie pour chaque choix de N .

Afin de déterminer une taille d'échantillonnage, il convient de mesurer la précision de l'approximation. Soit α_δ le quantile d'une $N(0, 1)$ associé à un certain degré de signification δ , i.e. $P_\xi[-\alpha_\delta \leq X \leq \alpha_\delta] = \delta$, où $X \sim N(0, 1)$.

Nous utiliserons le théorème de la limite centrale,

$$g(z) - \hat{g}_N(z) \Rightarrow N\left(0, \frac{\sigma^2(z)}{N}\right),$$

où $\sigma^2(z)$ est la variance de g en z , pour construire un intervalle de confiance pour $g(z)$ autour de $\hat{g}_N(z)$ est

$$[\hat{g}_N(z) - \epsilon_N^\delta(z), \hat{g}_N(z) + \epsilon_N^\delta(z)],$$

Détermination de la précision (2)

$\epsilon_N^\delta(z)$ est donné par

$$\epsilon_\delta^N(z) = \alpha_\delta \frac{\sigma(z)}{\sqrt{N}}.$$

Typiquement, on choisira $\alpha_{0.9} \approx 1.64$ ou $\alpha_{0.95} \approx 1.96$.

En pratique, nous ne connaissons pas $\sigma^2(z)$, aussi nous l'approximerons par son estimateur

$$\hat{\sigma}_N^2(z) = \frac{1}{N-1} \sum_{i=1}^N (G(z, \xi_i) - \hat{g}_N(z))^2.$$

Nous allons exploiter cette estimation de l'erreur dans le contexte des régions de confiance.

L'idée de base est que si le modèle approxime bien la fonction objectif par rapport à la précision de la fonction objectif elle-même (qui est dépendante de la taille d'échantillonnage), nous présumons que nous pourrions travailler avec une approximation moins précise et dès lors réduire la taille de la taille d'échantillonnage.

D'autre part, si l'adéquation du modèle est pauvre par rapport à la précision de la fonction objectif, nous pouvons augmenter la taille d'échantillonnage dans une tentative de corriger cette déficience.

Nous supposons que les hypothèses exposées lors de l'analyse de consistance tiennent.

Une description formelle de l'algorithme suit.

Algorithme de région de confiance à précision variable.

Etape 0. Initialisation. Un point initial z_0 et un rayon de région de confiance initial Δ_0 sont donnés. Soit des constantes η_1 et η_2 telles que $0 < \eta_1 \leq \eta_2 < 1$ (par exemple, $\eta_1 = 0.01$ et $\eta_2 = 0.75$).

Définissons un nombre minimum de tirs $N_{\min} = N_{\min}^0$ et une taille d'échantillonnage N_0 satisfaisant $\|\nabla_{\theta} \hat{g}_{N_0}(z_0)\| \neq 0$ si $\epsilon_{\delta}^{N_0}(z_{k+1}) \neq 0$, excepté si $N_0 = N_{\max}$. Calculons $\hat{g}_{N_0}(z_0)$ et posons $k = 0$, $t = 0$.

Etape 1. Test d'arrêt. Arrêt si $\|\nabla_{\theta} \hat{g}_{N_k}(z_k)\| = 0$ et soit $N_k = N_{\max}$, soit $\epsilon_{\delta}^{N_k}(z_k) = 0$. Autrement, aller à l'Etape 2.

Etape 2. Définition du modèle. Définissons un modèle $m_k^{N_k}$ de $\hat{g}_{N_k}(\theta)$ dans \mathcal{B}_k . Calculons une nouvelle taille adéquate d'échantillonnage N^+ , et posons $N^- = N_k$.

Etape 3. Calcul du pas. Calculons un pas s_k qui réduit suffisamment le modèle $m_k^{N_k}$ et tel que $z_k + s_k \in \mathcal{B}_k$. Posons

$$\Delta m_k^{N_k} = m_k^{N_k}(z_k) - m_k^{N_k}(z_k + s_k).$$

Etape 4. Comparaison des décroissances. Calculons $\hat{g}_{N^+}(z_k + s_k)$ et définissons

$$\rho_k = \frac{\hat{g}_{N_k}(z_k) - \hat{g}_{N^+}(z_k + s_k)}{\Delta m_k^{N_k}}.$$

Etape 5. Mise à jour de la taille d'échantillonnage. Si $\rho_k < \eta_1$ et $N_k \neq N^+$, modifions N^- ou la taille d'échantillonnage candidate N^+ afin de prendre en compte des différences de variance. Recalculons ρ_k .

Etape 6. Acceptation du point d'essai. Si $\rho_k < \eta_1$, définissons $z_{k+1} = z_k$, $N_{k+1} = N^-$. Sinon, définissons $z_{k+1} = z_k + s_k$ et posons $N_{k+1} = N^+$; incrementons t .
Si $N_{k+1} \neq N^{\max}$, $\|\nabla_{\theta} \hat{g}_{N_{k+1}}(z_{k+1})\| = 0$, et $\epsilon_{\delta}^{N_{k+1}}(z_{k+1}) \neq 0$, augmentons N_{k+1} à une certaine taille inférieure ou égale à N_{\max} telle que $\|\nabla_{\theta} \hat{g}_{N_{k+1}}(z_{k+1})\| \neq 0$ si $N_{k+1} \neq N_{\max}$, et calculons $\hat{g}_{N_{k+1}}(z_{k+1})$.

Algorithme : BTRDA (4)

Etape 6. Acceptation du point d'essai (suite). Si $N_k = N_{k+1}$ ou si une décroissance suffisante a été observée depuis la dernière évaluation de $\hat{g}_{N_{k+1}}$, posons $N_{\min}^{k+1} = N_{\min}^k$. Sinon, définissons $N_{\min}^{k+1} > N_{\min}^k$.

Etape 7. Mise à jour du rayon de la région de confiance.

$$\Delta_{k+1} \in \begin{cases} [\Delta_k, \infty) & \text{si } \rho_k \geq \eta_2, \\ [\gamma_2 \Delta_k, \Delta_k] & \text{si } \rho_k \in [\eta_1, \eta_2), \\ [\gamma_1 \Delta_k, \gamma_2 \Delta_k] & \text{si } \rho_k < \eta_1, \end{cases}$$

Dans cet algorithme, la variable t est utilisé pour comptabiliser le nombre d'itérations réussies.

Remarquons aussi que les algorithmes BTR et BTRDA coïncident si nous fixons N_k à N_{\max} pour tout $k \geq 0$.

Stratégie de taille d'échantillonnage variable

Avant l'optimisation, l'utilisateur choisit une taille d'échantillonnage maximale N_{\max} . Une taille d'échantillonnage minimale N_{\min}^0 est définie pour permettre l'estimation de la précision.

Nous définissons aussi $N_0 = \max\{N_{\min}^0, 0.1 N_{\max}\}$ si $\|\nabla_{\theta} \hat{g}_{N_0}(z_0)\| = 0$ et $\epsilon_{\delta}^{N_0}(z_0) \neq 0$, $N_0 = N_{\max}$ sinon.

Le choix de N^+ dans l'Etape 3 de l'algorithme BTRDA est décrit ci-dessous.

Définissons des constantes ν_1 et χ_1 telles que $\nu_1, \chi_1 \in (0, 1)$. Utilisons $\epsilon_{\delta}^{N_k}(z)$ pour estimer la taille nécessaire pour obtenir une précision égale à la décroissance du modèle, c'est-à-dire

$$N^S = \max \left\{ N_{\min}^k, \left\lceil \frac{\alpha_{\delta}^2 \hat{\sigma}_N(z)}{(\Delta m_k^{N_k})^2} \right\rceil \right\}.$$

Stratégie de taille d'échantillonnage variable (2)

Calculons le rapport entre l'amélioration du modèle et la précision estimée :

$$\tau_1^k = \frac{\Delta m_k^{N_k}}{\epsilon_\delta^{N_k}(z_k)},$$

et le rapport entre la taille d'échantillonnage actuelle et la taille d'échantillonnage suggérée pour la prochaine itération :

$$\tau_2^k = \frac{N_k}{\min\{N_{\max}, N^s\}}.$$

Définissons

$$N' = \begin{cases} \min\{\lceil \chi_1 N_{\max} \rceil, \lceil N^s \rceil\} & \text{si } \tau_1^k \geq 1, \\ \min\{\lceil \chi_1 N_{\max} \rceil, \lceil \tau_1^k N^s \rceil\} & \text{si } \tau_1^k < 1 \text{ et } \tau_1^k \geq \tau_2^k, \\ \lceil \chi_1 N_{\max} \rceil & \text{si } \nu_1 \leq \tau_1^k < 1 \text{ et } \tau_1^k < \tau_2^k, \\ N_{\max} & \text{si } \tau_1^k < \nu_1 \text{ et } \tau_1^k < \tau_2^k. \end{cases}$$

Posons $N^+ = \max\{N', N_{\min}^k\}$.

Stratégie de taille d'échantillonnage variable (3)

Une valeur possible pour χ_1 est 0.5.

Si $\tau_1^k \geq 1$, la décroissance du modèle est plus grande ou égale à la précision estimée, et nous réduisons alors la taille d'échantillonnage au minimum entre N^s et $\lceil \chi_1 N_{\max} \rceil$. L'idée d'utiliser $\lceil \chi_1 N_{\max} \rceil$ vient de l'observation pratique qu'imposer une telle décroissance dans les tailles proposées d'échantillonnage fournit une meilleure performance numérique.

Si $\tau_1^k < 1$, l'amélioration est plus petite que la précision. Cependant, puisque l'échantillonnage a été généré avant le processus d'optimisation, une amélioration suffisante au cours de plusieurs itérations consécutives peut conduire à une amélioration significative en comparaison de la précision de l'approximation, tout en gardant les coûts de calcul plus faibles que si N_{\max} tirs étaient utilisés.

Stratégie de taille d'échantillonnage variable (4)

Nous considérons alors deux cas.

1. Si $\tau_1^k \geq \tau_2^k$, le rapport entre la taille d'échantillonnage actuelle et la suivante potentielle est plus faible que le rapport entre la décroissance du modèle et l'erreur estimée. Si la taille d'échantillonnage augmente, l'erreur décroît pour un $\Delta m_j^{N_j}$ ($j \geq k$) similaire, et dès lors τ_1^k augmente.

Nous capitalisons sur τ_1^k en calculant une taille d'échantillonnage plus faible que N^s , telle qu'une amélioration de l'ordre de $\epsilon_\delta^{N_k}(z_k)$ serait atteinte en approximativement $\lceil \tau_1^k \rceil$ itérations si τ_1^j est similaire à τ_1^k pour j proche de k .

Nous proposons dès lors d'utiliser le minimum entre $\lceil \chi_1 N_{\max} \rceil$ et $\lceil \tau_1^k N^s \rceil$ comme nouvelle taille d'échantillonnage.

Stratégie de taille d'échantillonnage variable (5)

2. Si $\tau_1^k < \tau_2^k$, il peut néanmoins être plus économique de continuer à travailler avec une plus petite taille d'échantillonnage, définie à nouveau comme $\lceil \chi_1 N_{\max} \rceil$. C'est pourquoi nous choisissons d'utiliser cette plus petite taille d'échantillonnage aussi longtemps que τ_1^k est supérieure à un certain seuil ν_1 (par exemple 0.2). En-deçà de ce seuil, nous considérons que la décroissance est trop petite comparé à la précision, et nous augmentons éventuellement la taille d'échantillonnage.

Si N^+ n'est pas égal à N_k , le calcul de

$$\hat{g}_{N_k}(z_k) - \hat{g}_{N^+}(z_k + s_k)$$

est affecté par le changement dans la variance de l'approximation. Ceci peut conduire à un petit rapport, voire un rapport négatif ρ_k , et ce même quand le modèle $m_k^{N_k}$ donne une bonne prédiction pour la taille d'échantillonnage N^k .

En particulier, $\hat{g}_{N^+}(\theta)$ peut être supérieur à $\hat{g}_{N_k}(z_k)$ pour tout θ dans un voisinage de z_k . Il est dès lors important d'éviter de tels cas, ce qui motive la possible redéfinition de ρ_k , comme décrit ci-après.

Révision de la taille d'échantillonnage

Supposons que $N_k \neq N^+$. Si $\rho_k < \eta_1$, comparons N^+ et N_k . Si $N^+ > N_k$, calculons $\hat{g}_{N^+}(z_k)$, $\Delta m_k^{N^+}$ et $\epsilon_\delta^{N^+}(z_k)$, sinon si $N^+ < N_k$ calculons $\hat{g}_{N^+}(z_k + s_k)$. Posons N^- à $\max\{N_k, N^+\}$, et redéfinissons

$$\rho_k = \frac{\hat{g}_{N^-}(z_k + s_k) - \hat{g}_{N^-}(z_k)}{\Delta m_k^{N^-}}.$$

Bien que nous espérons bénéficier de plus petites tailles d'échantillonnage quand nous sommes loin de la solution, nous devrions être certains que nous utilisons une taille d'échantillonnage égale à N_{\max} au cours des itérations finales, afin de travailler avec la précision voulue. A cette fin, nous augmentons la taille d'échantillonnage minimale quand la stratégie adaptative ne fournit pas des gains numériques suffisants.

Mise à jour de la taille minimale d'échantillonnage

Nous définissons tout d'abord deux vecteurs v et l , de dimension N_{\max} , et, à l'itération $k = 0$, posons $v(N_0) = \hat{g}_{N_0}(z_0)$, $l(N_0) = 0$, tandis que pour $i = 1, \dots, N_{\max}$, $i \neq N_0$, posons $v(i) = +\infty$, $l(i) = -1$.

Au début de l'itération k , $v(i) = \hat{g}_i(z_{h(i)})$, où $h(i)$ correspond à l'index de la dernière itération pour laquelle $N_{h(i)} = i$, et $N_{h(i)-1} \neq N_{h(i)}$ si $h(i) > 0$, ou $+\infty$ si la taille i n'a pas encore été utilisée. $l(i)$ contient le nombre d'itérations réussies jusqu'à l'itération $h(i)$ (incluse), ou -1 si la taille i n'a pas été utilisée.

Rappel : t contient le nombre total d'itérations réussies rencontrées jusqu'à l'itération k (incluse).

Mise à jour de la taille minimale d'échantillonnage (2)

Supposons que $N_k \neq N_{k+1}$. Soit une constante $\gamma_3 \in (0, 1]$. Si

$$v(N_{k+1}) - \hat{g}_{N_{k+1}}(z_{k+1}) \geq \gamma_3 \nu_1(t - l(N_{k+1})) \epsilon_\delta^{N_{k+1}}(z_{k+1}),$$

posons $N_{\min}^{k+1} = N_{\min}^k$. Sinon, augmentons la taille d'échantillonnage minimale : posons

$$N_{\min}^{k+1} \in \{N_{k+1} + 1, \dots, N_{\max}\}.$$

Posons $l(N_{k+1}) = t$ et $v(N_{k+1}) = \hat{g}_{N_{k+1}}(z_{k+1})$.

Une valeur pratique pour γ_3 est 0.5. Notons que $N_{\min}^{k+1} > N^k$ si le test ci-dessus n'est pas satisfait.

De plus, nous avons que si $N_k \neq N_{k+1}$, $t - l(N_{k+1}) \geq 1$. Ceci est clairement vérifié si $l(N_{k+1}) = -1$, aussi sans perte de généralité, nous supposons que $l(N_{k+1}) \geq 0$. Au début de l'itération k , nous avons $l(N_i) \leq t$, $i = 1, \dots, N_{\max}$.

Mise à jour de la taille minimale d'échantillonnage (3)

Si $\rho_k \geq \eta_1$, t est incrémenté de 1 au cours de l'Etape 6 de l'algorithme de région de confiance, aussi $l(N_{k+1}) < t$ dans l'algorithme ci-dessus.

Si $\rho_k < \eta_1$, de l'algorithme de mise à jour de la taille d'échantillonnage, $N_k < N_{k+1}$ puisque des réductions de tailles d'échantillonnage peuvent seulement se produire lors d'itérations réussies. Ceci implique aussi que $l(N_{k+1}) < l(N_k) \leq t$.

Finalement, notons que si $N_k \neq N_{\max}$, nous ne pouvons pas exclure le cas pathologique dans lequel z_k est un point critique au premier ordre pour \hat{g}_{N_k} .

Si $\epsilon_{\delta}^{N_k}(z_k) \neq 0$, l'algorithme ne s'arrête pas, mais puisque le modèle est quadratique, aucune décroissance n'est atteinte si H_k est défini positif.

Précautions additionnelles

Afin d'éviter cette situation, nous forçons dès lors une croissance de N_{k+1} quand cette situation se produit.

En pratique cependant, la norme du gradient change habituellement lentement dans un voisinage d'un tel point critique, et un petit gradient conduit typiquement à une petite décroissance du modèle, ce qui mène à l'augmentation de la taille d'échantillonnage et N_{\max} est atteinte avant que cette précaution ne soit mise en oeuvre.

Théorème

Sous certaines hypothèses de régularité, si

$$\exists \kappa > 0 \text{ such that } \epsilon_{\delta}^{N_k}(z_k) \geq \kappa,$$

pour tout k suffisamment grand, alors, presque sûrement, l'algorithme converge en un nombre fini d'itération avec un nombre final de tirs aléatoires égal à N_{\max} , ou le nombre d'itérations est infini et il existe un certain j tel que pour toutes les itérations i , $i \geq j$, N_i est égal à N_{\max} .

Preuve : voir Bastin, Cirillo et Toint, *An adaptive Monte Carlo algorithm for computing mixed logit estimators*, Computational Management Science 3(1), pp. 55–79, 2006.

Nous pouvons alors prouver la convergence au premier et second ordre, et ce pour le SAA avec N_{\max} tirs.

Exemple

Modèle logit mélangés : maximisation de vraisemblance stochastique

$$\max_{\theta} LL(\theta) = \max_{\theta} \frac{1}{N} \sum_{n=1}^N \ln E[P_{ij_i}(x, \theta, \xi)].$$

Modèle de choix de mode : données [Mobidrive](#) (Axhausen and al., 2002)

$N = 5799$ observations, $R_{\max} = 2000$ tirs par individu, 14 paramètres (dimension d'intégration : 3 variables normales).

