MACS205 : Méthode de Monte-Carlo

1 Introduction

Soit $(S, \mathcal{S}, \mu)$ un espace mesuré où μ est une mesure positive. Soit $\varphi \colon S \to \mathbf{R}$ une fonction intégrable. On cherche à approcher $I(\varphi) = \int \varphi \, \mathrm{d}\mu = \mathbf{E}_\mu(\varphi)$. Deux cas de figure :

- φ est une fonction continue avec une expression analytique et on arrive à calculer son intégrale,
- l'intégrale de φ est incalculable.

Les méthodes de type Monte-Carlo considérées sont de la forme suivante :

1. tirer aléatoirement des points $X_1, \ldots, X_n$ sur S,
2. calculer $\varphi(X_1), \ldots, \varphi(X_n)$,
3. trouver une transformation de $(X_1, \varphi(X_1)), \ldots, (X_n, \varphi(X_n))$ qui approche $I(\varphi)$.

2 La méthode de Monte-Carlo

Algorithme 1 : Monte-Carlo
Générer $X_1, \ldots, X_n$ de façon indépendante sous μ ; Calculer $\varphi(X_1), \ldots, \varphi(X_n)$; **Sorties :** $\hat{I}_n(\varphi) = \hat{I}_n^{(mc)}(\varphi) = \frac{1}{n} \sum_i \varphi(X_i)$

Prop. Si $\int |\varphi| \, \mathrm{d}\mu < \infty$, $\hat{I}_n(\varphi)$ est non-biaisée et fortement consistante. Si de plus $\int |\varphi|^2 \, \mathrm{d}\mu < \infty$ alors $\mathrm{Var}(\hat{I}_n(\varphi)) = \frac{1}{n} \mathrm{Var}(\varphi(X_1)) = \frac{1}{n}\sigma^2$ et $\sqrt{n}\left(\hat{I}_n(\varphi) - I(\varphi)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$.

Estimation de l'erreur

On estime σ^2 par $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n \left(\varphi(X_i) - \hat{I}_n(\varphi)\right)^2$.

Prop. Si $\int |\varphi|^2 \, \mathrm{d}\mu < \infty$ alors $\hat{\sigma}^2$ est sans biais et fortement consistant et $\frac{\sqrt{n}}{\hat{\sigma}}\left(\hat{I}_n(\varphi) - I(\varphi)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

Intervalle de confiance : $\mathbf{P}(I(\varphi) \in \hat{C}(\alpha)) \overset{n \to \infty}{\longrightarrow} 1 - \alpha$ avec $\forall \alpha \in]0;1[, \hat{C}(\alpha) = \left[\hat{I}_n(\varphi) - \frac{\hat{\sigma}^2}{\sqrt{n}}\Phi^-\left(1 - \frac{\alpha}{2}\right), \hat{I}_n(\varphi) - \frac{\hat{\sigma}^2}{\sqrt{n}}\right.$

Inégalités de concentrations

Th (Inégalité de Hoeffding). Soit $X_1, \ldots, X_n$ i.i.d telles que $\forall i \in [\![1;n]\!], a \leqslant X_i \leqslant b$ p.s. Alors

$$\mathbf{P}\left(\left|\sum_{i=1}^n (X_i - \mathbf{E}(X_i))\right| > \varepsilon\right) \leqslant 2 \cdot \exp\left(-\frac{2\varepsilon^2}{n(b-a)^2}\right).$$

Déterministe vs aléatoire en "grande" dimension

Méthode déterministe des sommes de Riemann : soit $\varphi \colon [0;1]^d \longrightarrow \mathbf{R}$, on se donne n^d points équidistants $x_\alpha = \left(\frac{i_1}{n}, \ldots, \frac{i_d}{n}\right)$ où $(i_1, \ldots, i_d) \in [\![1;n]\!]^d$. On calcule $I_n^{(rs)}(\varphi) = \frac{1}{n^d} \sum \varphi(x_\alpha)$.

Prop. Si $\varphi \colon [0;1]^d \longrightarrow \mathbf{R}$ est L-lipschitzienne alors $\left|I_n^{(rs)}(\varphi) - I(\varphi)\right| \leqslant L\frac{\sqrt{d}}{n}$.

Avec Monte-Carlo la méthode de même ordre se fait avec évaluation en n^d v.a tirées selon $\mathcal{U}([0;1]^d)$ et l'on a $\mathrm{Var}\left(\hat{I}_{n^d}(\varphi)\right) = \frac{1}{n^d}\sigma^2$ et $\mathbf{E}\left[\left|\hat{I}_{n^d}(\varphi) - I(\varphi)\right|\right] \leqslant \frac{\sigma}{n^{d/2}}$.

Méthode des variables antithétiques

Soit $Z \sim \mu$ v.a telle que $\mathbf{E}[\varphi(Z)^2] < \infty$ et $\{Z_k, k \geqslant 0\}$ i.i.d selon μ. On a $\hat{I}_n^{(av)}(\varphi) = \frac{1}{2n}\sum_{i=1}^n (\varphi(Z_i) + \varphi(L(Z_i)))$.

Ex. $U_1, \dots, U_n \sim \mathcal{U}([a;b])$ et $L(u) := a+b-u$, ou si $Z \sim \mathcal{N}(\mu, 1)$ alors $2\mu - Z \sim \mathcal{N}(\mu, 1)$.

Prop. Si $\mathbf{E}|\varphi(Z)|^2 < \infty$ alors :

- $\mathrm{Var}\left(\hat{I}_{2n}(\varphi)\right) \geqslant \mathrm{Var}\left(\hat{I}_n^{(av)}\right) \iff \mathrm{Cov}(\varphi(Z), \varphi(L(Z))) \leqslant 0$,
- si φ est réelle croissante et $\varphi \circ L$ décroissante (ou inversement) alors $\mathrm{Cov}(\varphi(Z), \varphi(L(Z))) \leqslant 0$.

Lem. Soit Z une v.a réelle, $g\colon \mathbf{R} \to \mathbf{R}$ croissante avec $\mathbf{E}[g(Z)^2] < \infty$ et $\tilde{g}\colon \mathbf{R} \to \mathbf{R}$ décroissante avec $\mathbf{E}[\tilde{g}(Z)^2] < \infty$. Alors $\mathrm{Cov}(g(Z), \tilde{g}(Z)) \leqslant 0$.

3 Méthode des variables de contrôle

Le contexte est comme Monte-Carlo avec une variable observée en plus : $((X_1, Z_1), \dots, (X_n, Z_n))$ i.i.d dans $S \times \mathbf{R}$, $X_1 \sim \mu$ et $\mathbf{E}[Z_1]$ est connu. Soit $\varphi\colon S \to \mathbf{R}$ tel que $\mathbf{E}|\varphi(X_1)| < \infty$, on cherche $I_\mu = \mathbf{E}[\varphi(X_1)]$.

On peut se ramener à $\mathbf{E}Z_1 = 0$, et on pose $\hat{I}_n^{(cv)} = \frac{1}{n}\sum_{i=1}^n (\varphi(X_i) - Z_i)$.

Prop. Si $\mathbf{E}|\varphi(X_1)| < \infty$ et $\mathbf{E}|Z_1| < \infty$, $\hat{I}_n^{(cv)}$ est sans biais et fortement consistant. Si de plus $\mathbf{E}\left[|\varphi(X_1)|^2\right] < \infty$ et $\mathbf{E}\left[|Z_1|^2\right] < \infty$ alors :

- $\mathrm{Var}\left(\hat{I}_n^{(cv)}\right) = \frac{1}{n}\mathrm{Var}(\varphi(X_1) - Z_1)$ et $\hat{I}_n^{(cv)}$ est asymptotiquement normal avec variance $\sigma^2 = \mathrm{Var}(\varphi(X_1) - Z_1)$, i.e $\sqrt{n}\left(\hat{I}_n^{(cv)} - I\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$,
- un estimateur consistant de σ^2 est $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n \left((\varphi(X_i) - Z_i) - \hat{I}_n^{(cv)}\right)^2$.

Rem. Cela comprend Monte-Carlo : $Z_1 = 0$, et les variables antithétiques : $Z_1 = \frac{1}{2}(\varphi(X_1) - (\varphi \circ L)(X_1))$.

Rem. VC est plus performante que MC si $\mathrm{Var}(\varphi(X_1) - Z_1) \leqslant \mathrm{Var}(\varphi(X_1))$.

Pour prévenir d'une mauvaise variable de contrôle, on définit l'estimateur $\frac{1}{n}\sum_{i=1}^n (\varphi(X_i) - \beta Z_i)$, à utiliser si $\mathrm{Var}(\varphi(X_1) - \beta Z_1) \leqslant \mathrm{Var}(\varphi(X_1))$. C'est vérifié avec $\beta^* = \arg\min_\beta \mathrm{Var}\,(\varphi(X_1) - \beta Y_1) = \mathbf{E}[\varphi(X_1)Z_1]/\mathbf{E}[Z_1^2]$.

Propriétés asymptotiques : $Z_1 \in \mathbf{R}^m$

On estime $I = \mathbf{E}[\varphi(X_1)]$ par $\hat{I}_n^{(cv)}(\beta) = \frac{1}{n}\sum_{i=1}^n \left(\varphi(X_i) - \beta^\top Z_i\right)$ où $\beta \in \mathbf{R}^m$, avec encore $\mathbf{E}Z_1 = 0$.

La valeur théorique pour minimiser la variance, si $\mathbf{E}\left[Z_1 Z_1^\top\right]$ est inversible, est $\beta^* = \mathbf{E}\left[Z_1 Z_1^\top\right]^{-1} \mathbf{E}[Z_1\varphi($

En pratique on l'estime. Notons $Z_{n,m} = \begin{pmatrix} Z_1 & \cdots & Z_n \end{pmatrix}^\top \in \mathbf{R}^{n\times m}$, et $\varphi_i = (\varphi(X_1), \dots, \varphi(X_n))^\top \in \mathbf{R}^n$.

Alors $\hat{\beta}_n = (Z_{n,m}^\top Z_{n,m})^+ Z_{n,m}^\top \varphi_n = \left(\frac{1}{n}\sum Z_i Z_i^\top\right)^{-1} \frac{1}{n}\sum Z_i \varphi(X_i)$ (en notant A^+ l'inverse généralisé de A).

Prop. Supposons $\mathbf{E}|\varphi(X_1)| < \infty$, $\forall k \in [\![1;m]\!], \mathbf{E}\left|\varphi(X_1)Z_{k,1}\right| < \infty$ et $\mathbf{E}[Z_1 Z_1^\top]$ existe et est inversible. Alors $\hat{I}_n^{(cv)}(\hat{\beta}_n) \xrightarrow{\text{p.s.}} I$ (fortement consistant). Si de plus $\mathbf{E}|\varphi(X_1)|^2 < \infty$, alors $\sqrt{n}\left(\hat{I}_n^{(cv)}(\hat{\beta}_n) - I\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_m^2)$ avec $\sigma_m^2 = \mathrm{Var}\left(\varphi(X_1) - \beta^{*\top} Z_1\right)$.

Rem. L'estimation de $\hat{\beta}_n$ n'a pas d'effet en l'asymptotique (c'est comme si on connnaissait β^*).

Rem. D'autres estimateurs de β^* peuvent être légitimes sous condition d'inversibilité et de consistance.

Rem. On a $\forall m \geqslant 0, \sigma_{m+1} \leqslant \sigma_m$ et σ_0^2 correspond à la variance de Monte-Carlo. Un estimateur de la variance de $\hat{I}_n^{(cv)}(\hat{\beta}_n)$ est donné par $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \left(\varphi(X_i) - \hat{\beta}_n^\top Z_i - \hat{I}_n^{(cv)}(\hat{\beta}_n)\right)^2$.

Prop. Supposons $\mathbf{E}|\varphi(X_1)|^2 < \infty$, $\forall k \in [\![1\,;m]\!], \mathbf{E}\left|\varphi(X_1)Z_{k,1}\right| < \infty$ et $\mathbf{E}[Z_1 Z_1^\top]$ est inversible. Alors $\hat{\sigma}^2 \underset{n\to\infty}{\longrightarrow} \sigma_m^2$.

Complexité du calcul

Règles du temps de calcul : générer X_1, générer $Z_{1,k}$ pour un k et évaluer $\varphi(X_1)$ comptent chacun pour une opération élémentaire.

Méthode	Nombre d'opérations élémentaires
Monte Carlo	$O(n)$
Calcul de $\hat{\beta}_n$	$O(m^2 n + m^3)$
Variables de contrôle (avec $\hat{\beta}_n$ donné)	$O(mn)$

Donc la méthode avec variable de contrôle est mieux que Monte Carlo lorsque $\frac{\sigma_m}{\sigma_0} \leqslant \frac{1}{m}$.

4 Échantillonage préférentiel

Contexte : estimation de $I_\lambda = \int \varphi \,\mathrm{d}\lambda$ où $\varphi\colon S \to \mathbf{R}$ est intégrable et $S \subseteq \mathbf{R}^d$.

Def. Si $g\colon \mathbf{R}^d \to \mathbf{R}$, on définit son support comme l'ensemble fermé $S_g = \overline{\{x \in \mathbf{R}^d \mid g(x) \neq 0\}}$.

L'échantillonage d'importance se base sur la formule suivante : pour toute densité f telle que $S_f \supset S_\varphi$,

$$I_\lambda = \int_{S_\varphi} \varphi(x)\,\mathrm{d}x = \int_{S_f} \varphi(x)\,\mathrm{d}x = \int_{S_f} \frac{\varphi(x)}{f(x)} f(x)\,\mathrm{d}x = \mathbf{E}_{X\sim f}\left[\frac{\varphi(X)}{f(X)}\right]$$

L'échantillonage d'importance "naïf" consiste à générer $X_1, \dots, X_n \sim f$ i.i.d puis appliquer Monte-Carlo :

$$\hat{I}_n^{(is)} = \frac{1}{n}\sum_{i=1}^n \frac{\varphi(X_i)}{f(X_i)}\,.$$

La distribution associée à f est appelée **distribution d'échantillonage**, ou bien l'**échantilloneur**.

Prop. Si $\int |\varphi| < \infty$ et $S_\varphi \subseteq S_f$, $\hat{I}_n^{(is)} \overset{\text{p.s.}}{\longrightarrow} I_\lambda$. Si de plus $\int \frac{\varphi^2(x)}{f(x)}\,\mathrm{d}x < \infty$ alors :

- $\sqrt{n}\left(\hat{I}_n^{(is)} - I_\lambda\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, r_f^2(\varphi)\right)$ avec $r_f^2(\varphi) = \operatorname{Var}\left(\frac{\varphi}{f}\right) = \int \left(\frac{\varphi(x)}{f(x)} - I_\lambda\right)^2 f(x)\,\mathrm{d}x$
- $\frac{\sqrt{n}}{\hat{r}_n}\left(\hat{I}_n^{(is)} - I_\lambda\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ avec $\hat{r}_n^2 = \frac{1}{n}\sum\left(\frac{\varphi(X_i)}{f(X_i)} - \hat{I}_n^{(is)}\right)^2$ l'estimateur de la variance.

Cette méthode est naïve car f n'est pas choisie par rapport à φ.

Rem. Si on cherche $\mathbf{E}[\varphi(Z)]$ où $Z \sim g$, alors prendre $g \cdot \varphi$ à la place de φ.

Par ailleurs il existe deux méthodes de réduction de la variance où l'on s'adapte à φ :

- variable de contrôle : approcher φ dans une certaine base $\longrightarrow$ pas de choix d'échantilloneur,
- changer la mesure d'échantillonage.

Réduction de la variance

On remarque que $r_f^2(\varphi) = 0 \iff \varphi/f \overset{\text{p.p}}{=} I_\lambda$.

Th. Si $\int |\varphi| < \infty$, alors parmi les densités f telle que $\int \frac{\varphi^2}{f}\,\mathrm{d}\lambda < \infty$, i.e. $S_\varphi \subseteq S_f$, le minimiseur de $r_f^2(\varphi)$ est unique et donné par $f^* = \frac{|\varphi|}{\int |\varphi|\,\mathrm{d}\lambda}$. La variance associée est $r_{f^*}^2(\varphi) = \left(\int |\varphi|\,\mathrm{d}\lambda\right)^2 - \left(\int \varphi\,\mathrm{d}\lambda\right)^2$.

Échantillonage préférentiel paramétrique

On se donne une famille paramétrique $\mathcal{F} = \{f_\theta, \theta \in \Theta\}$ de densités par rapport à la mesure de Lebesgue pour lesquelles on sait générer des v.a, avec $\Theta \subset \mathbf{R}^q, q \geqslant 1$. On suppose $\forall \theta \in \Theta, S_\varphi \subseteq S_{f_\theta}$.

Soit $\theta^* \in \arg\min_{\theta\in\Theta} r^2_{f_\theta}(\varphi) = \arg\min_{\theta\in\Theta} \int \frac{\varphi^2}{f_\theta} \mathrm{d}\lambda$. On cherche à estimer par simulation cette variance. On utilise l'algorithme suivant, avec en entrée $n \in \mathbf{N}^*$, $\mathcal{F}$ et f_0 l'échantillon initial.

(i) Soit $n_1 < n$ et $n_2 = n - n_1$. Générer $X_1, \ldots, X_{n_1} \sim f_0$ i.i.d.

(ii) Calculer $\hat{\theta}^{(1)}_{n_1} = \arg\min_{\theta\in\Theta} \hat{\psi}n_1{}^{(1)}(\theta)$ où $\hat{\psi}^{(1)}_{n_1}(\theta) := \frac{1}{n_1}\sum_{i=1}^{n_1} \frac{\varphi(X_i)^2}{f_\theta(X_i) f_0(X_i)}$.

(iii) Générer $Z_1, \ldots, Z_{n_2} \sim f_{\hat{\theta}_{n_1}}$.

(iv) Calculer $\frac{1}{n_2}\sum_{i=1}^{n_2} \frac{\varphi(Z_i)}{f_{\hat{\theta}^{(1)}_{n_1}}(Z_i)}$ ou $\frac{1}{n}\left(\sum_{i=1}^{n_1} \frac{\varphi(X_i)}{f_0(X_i)} + \sum_{i=1}^{n_2} \frac{\varphi(Z_i)}{f_{\hat{\theta}^{(1)}_{n_1}}(Z_i)}\right)$.

On peut aussi utiliser la méthode par vraisemblance où l'on remplace $\hat{\psi}^{(1)}_{n_1}$ par $\hat{\psi}^{(2)}_{n_1} : \theta \mapsto \frac{1}{n_1}\sum_{i=1}^{n_1} \ln\left(\frac{f_\theta(X_i)}{|\varphi(X_i)|}\right)\frac{|\varphi(X_i)|}{f_0(X_i)}$. On appelle $\hat{I}^{(is)}_{n,1}$ et $\hat{I}^{(is)}_{n,2}$ les deux estimateurs obtenus.

Lem. Supposons Θ compact, ψ continue sur Θ, qu'il existe un unique $\theta^* \in \arg\min_{\theta\in\Theta}\psi(\theta)$, que $\sup_{\theta\in\Theta}\left|\hat{\psi}(\theta) - \psi(\theta)\right| \xrightarrow[n]{\mathbf{P}} 0$ et que $\hat{\theta}_{n_1}$ minimise $\hat{\psi}$. Alors $\left|\hat{\theta}_{n_1} - \theta^*\right| \xrightarrow[n_1]{\mathbf{P}} 0$.

Prop. $\hat{I}^{(is)}_{n,1}$ et $\hat{I}^{(is)}_{n,2}$ sont des estimateurs sans biais de I_λ.

Prop. Soit $v_0 = \mathrm{Var}\left(\frac{\varphi(X_1)}{f_0(X_1)}\right)$. Supposons $v_0 < \infty$ et $\sup_{\theta\in\Theta}\int \frac{\varphi(x)^2}{f_\theta(x)}\mathrm{d}x < \infty$. Alors on a $\mathrm{Var}\left(\hat{I}^{(is)}_{n,1}\right) = \frac{1}{n_2}\left(\mathbf{E}[\psi^{(1)}(\hat{\theta}_{n_1})] - I^2_\lambda\right)$ et $\mathrm{Var}\left(\hat{I}^{(is)}_{n,2}\right) = \frac{1}{n^2}\left[n_1 v_0 + n_2\left(\mathbf{E}[\psi^{(1)}(\hat{\theta}_{n_1})] - I^2_\lambda\right)\right]$.

La complexité est en $O(n + n_1^2 + n_2) = O(n)$.