MACS205 : Méthode de Monte-Carlo

1 Introduction

But du cours : étudier des méthodes aléatoires d'approximation d'intégrales.

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 $I(\varphi) = \int \varphi \, \mathrm{d}\mu$.

Plusieurs cas de figure en pratique :

- φ est une fonction continue avec une expression analytique et on arrive à calculer son intégrale,
- l'intégrale de φ est incalculable. Exemples : Gaussienne ou indicatrice d'ensemble S où l'on ne connaît pas de forme analytique.

Les méthodes considérées sont de la forme suivante :

1. choisir/tirer des points $X_1, \ldots, X_n$ sur S,
2. évaluer $\varphi(X_1), \ldots, \varphi(X_n)$,
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

$$I(\varphi) = \int \varphi \, \mathrm{d}N = \mathbf{E}_\mu(\varphi)$$

D'après la LFGN, si $X_1, \ldots, X_n$ sont i.i.d. de loi μ tel que $\mathbf{E}_\mu |vf| < \infty$ alors $\frac{1}{n} \sum_i \varphi(X_i) \xrightarrow{\text{p.s.}} \mathbf{E}_\mu(\varphi(X_1))$.

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

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

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 (par la LFGN).

...

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

$$\mathbf{P}\left(\left| \sum_{i=1}^n (X_i - \mathbf{E}(X_i)) \right| > \varepsilon \right) \leqslant 2 e^{-\frac{3\varepsilon^2}{n(b-a)^2}} .$$

...

Concentration

...

Déterministe vs aléatoire

On se place dans le cadre de l'approximation de $\int_{[0;1]^d} \varphi(x) \, \mathrm{d}x$ où $\varphi \colon [0;1]^d \longrightarrow \mathbf{R}$.

Méthode déterministe des sommes de Riemann On se donne n^d points équidistants $\left(\frac{i_1}{n}, \ldots, \frac{i_d}{n}\right)$ où $(i_1, \ldots, i_d) \in [\![1 ; n]\!]^d$. La méthode des sommes de Riemann est

$$I_n(\varphi) = \frac{1}{n^d} \sum_{(i_1, \ldots, i_d) \in [\![1;n]\!]^d} \varphi\left(\frac{i_1}{n}, \ldots, \frac{i_d}{n}\right) .$$

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

...

3 Méthode des variables de contrôle

Présentation

Le contexte est maintenant comme Monte-Carlo avec une variable observée en plus :

- on observe $((X_1, \varphi(X_1), Y_1), \dots, (X_n, \varphi(X_n), Y_n))$ i.i.d,
- on cherche $\mathbf{E}[\varphi(X_1)]$,
- on connaît $\mathbf{E}[Y_1]$.

On peut calculer $\frac{1}{n}\sum_{i=1}^n [\varphi(X_i) - (Y_i - \mathbf{E}[Y_i])]$. C'est un estimateur sans biais de variance

$$\frac{1}{n}\operatorname{Var}(\varphi(X_1) - (Y_1 - \mathbf{E}Y_1)) = \frac{1}{n}\mathbf{E}\left[(\varphi(X_1) - (Y_1 - \mathbf{E}Y_1))^2\right] - \frac{1}{n}\mathbf{E}[\varphi(X_1)]^2 .$$

Pour un tel estimateur, le but est de réduire ce risque L^2 au maximum en choisissant bien Y_1. Pour un tel estimateur, on obtient facilement les mêmes résultats que pour Monte-Carlo : forte consistance, normalité asymptotique et estimation consistante de la variance.

Pour faciliter la notation on supposera maintenant $\mathbf{E}Y_1 = 0$.

Rem.
- $Y = 0 \longrightarrow$ Monte-Carlo,
- $Y = \frac{-\varphi\circ L(X) - \varphi(X)}{2} \longrightarrow$ variables antithétiques.

Rem. La méthode des variables de contrôle (VC) est plus performante que MC si $\operatorname{Var}(\varphi(X_1) - Y_1) \leqslant \operatorname{Var}(\varphi(X_1))$ $(\frac{1}{n}\sum \varphi(X_i) + \frac{\varphi\circ L(X_i) - \varphi(X_i)}{2} = \frac{1}{2n}\sum \frac{\varphi(X_i) + \varphi\circ L(X_i)}{2})$.

Afin de prévenir d'une mauvaise variable de contrôle, on définit l'esimateur $\forall \beta \in \mathbf{R}, \hat{\mu}_n(\beta) = \frac{1}{n}\sum_{i=1}^n (\varphi(X_i) - \beta Y_i)$, à utiliser si $\operatorname{Var}(\varphi(X_1) - \beta Y_1) \leqslant \operatorname{Var}(\varphi(X_1))$.

$\longrightarrow \beta^* = \arg\min_\beta \operatorname{Var}(\varphi(X_1) - \beta Y_1)$, $\min_\beta \operatorname{Var}(\varphi - \beta Y) \leqslant \operatorname{Var}(\varphi)$.

Soit $f_1, \dots, f_m$ une collection de fonctions dont on connaît les intégrales. Supposons $\forall L \in [\![1\,;m]\!], \int f_L \,\mathrm{d}\lambda = 0$. Alors VC donne $\frac{1}{n}\sum_{i=1}^n \left[\varphi(u_i) - \sum_{j=1}^m \beta_j f_j(u_i)\right]$.

Ex. (f_L) polynômes, (f_L) base de Fourier ou (f_L) indicatrices.

Propriétés asymptotiques

Soient $((X_i, Y_i))_i$ une suite de v.a i.i.d à valeurs dans $S \times \mathbf{R}^m$. On définit l'estimateur de $\mathbf{E}[\varphi(X_1)]$ par $\forall \beta \in \mathbf{R}^m, \hat{\mu}_n(\beta) = \frac{1}{n}\sum_{i=1}^n \left(\varphi(X_i) - \beta^\top Y_i\right)$.

Comme dans l'intro, on suppose $\mathbf{E}Y_1 = \begin{pmatrix} \mathbf{E}[Y_{1,1}] \\ \vdots \\ \mathbf{E}[Y_{1,m}] \end{pmatrix} = 0$.

$\{\mu_n(\beta), \beta \in \mathbf{R}^m\}$ est une collection d'estimateurs sans biais. Trouvons l'élément de variance minimale :

$$\begin{aligned}
\beta^* &= \arg\min_\beta \frac{1}{n}\operatorname{Var}\left(\varphi - \beta^\top T\right) \\
&= \arg\min_\beta \operatorname{Var}\left(\varphi - \beta^\top T\right) \\
&= \arg\min_\beta \mathbf{E}\left[(\varphi - \beta^\top Y)^2\right] - \mathbf{E}[\varphi]^2 \\
&= \arg\min_\beta \mathbf{E}\left[(\varphi - \beta^\top Y)^2\right]
\end{aligned}$$

Si $\mathbf{E}[Y_1 Y_1^\top]$ est inversible, les équations normales / du premier ordre admettent une unique solution :

$$\beta^* = \mathbf{E}[Y_1 Y_1^\top]^{-1}\mathbf{E}[Y_1 \varphi(X_1)]$$

Il faut utiliser $\hat{\mu}_n(\beta^*)$, mais β^* est inconnue.

Idée : estimer β^* sur les données $\longrightarrow \hat{\beta}$, et utiliser $\hat{\mu}_n(\hat{\beta})$, qui a la même variance asymptotique que $\hat{\mu}_n(\beta^*)$.

Si $\frac{1}{n}\sum_{i=1}^n Y_i Y_i^\top$ est inversible :

$$\hat{\beta} = \arg\min_{\beta\in\mathbf{R}^m} \underbrace{\frac{1}{n}\sum_{i=1}^n \left([\varphi(X_i) - \beta^\top Y_i] - \hat{\mu}_n(\beta)\right)^2}_{\text{estimateur classique de la covariance}}$$

Ce choix ne va pas entrainer de changement à l'asymptotique mais pratique il procure de meilleurs performance. Donc $\hat{\beta} = \arg\min_\beta \frac{1}{n}\sum_{i=1}^n \left((\varphi(X_i) - \bar{\varphi}) - \beta^\top (Y_i - \bar{Y})\right)^2$.

Notons

$$Z_{n,m} = \begin{pmatrix} Y_{11} - \bar{Y}_1 & \cdots & Y_{1m} - \bar{Y}_m \\ \vdots & & \vdots \\ Y_{n1} - \bar{Y}_1 & \cdots & Y_{nm} - \bar{Y}_m \end{pmatrix} \in \mathbf{R}^{n\times m}, \qquad Y_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{im} \end{pmatrix} \in \mathbf{R}^m$$

(Y_i est la covariable du problème de régression). On a $\bar{Y}_k = \frac{1}{n}\sum_{i=1}^n Y_{ik}$.

Notons également $\Psi_n = \begin{pmatrix} \varphi(X_1) - \bar{\varphi} \\ \vdots \\ \varphi(X_n) - \bar{\varphi} \end{pmatrix}$. Alors $\hat{\beta} = \arg\min_\beta \left\| \Psi_n - Z_{n,m}\beta \right\|^2$.

Le théorème de projection nous donne une unique solution qui, si $Z_{n,m}^\top Z_{n,m}$ est inversible, vérifie :

$$(Z_{n,m}^\top Z_{n,m})\beta = Z_{n,m}^\top \Psi_n$$

$$\hat{\beta} = (Z_{n,m}^\top Z_{n,m})^{-1} Z_{n,m}^\top \Psi_n$$

Prop (asymptotique de $\hat{\mu}_n(\beta)$)**.** Supposons que $\mathbf{E}|\varphi(X_1)| < \infty$, $\forall k \in [[1\,;m]]$, $\mathbf{E}|\varphi(X_1)Y_{1k}| < \infty$ et $\mathbf{E}[Y_1 Y_1^\top]$ existe et est inversible. Alors $\hat{\mu}_n(\hat{\beta}) \xrightarrow{\text{p.s.}} \mathbf{E}[\varphi(X_1)]$. Si de plus $\mathbf{E}|\varphi(X_1)|^2 < \infty$, alors $\sqrt{n}\left(\hat{\mu}_n(\hat{\beta} - \mathbf{E}[\varphi(X_1)]\right) \longrightarrow \mathcal{N}(0, \sigma_m^2)$ avec $\sigma_m^2 = \mathrm{Var}\left(\varphi(X_1) - \beta^{*\top} Y_1\right)$.