AMPHI 2

1. Estimation de la fonction de répartition, estimation de quantiles
2. Méthodes de Monte Carlo par Chaînes de Markov (MCMC) et Méthode de splitting
3. Echantillonnage d'importance : principe général, et cas Gaussien.
4. (pour en savoir plus) Un exemple d'échantillonnage d'importance adaptatif.

I. Estimation de la fonction de répartition (sur $\mathbb{R}$) et Estimation de quantiles

Dans cette section, $\{X_i, i \geq 1\}$ sont des v.a. i.i.d. de même loi que X, v.a. à valeurs **réelles**.

I-1. Estimation de la fonction de répartition de X

Par définition : $F(x) := \mathbb{P}(X \leq x)$ pour tout $x \in \mathbb{R}$.

Fonction de répartition empirique. Pour tout $x \in \mathbb{R}$, on définit

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \leq x}.$$

On a une autre expression, à l'aide des statistiques d'ordre $X_{(i,n)}$ de l'échantillon $X_1, \cdots, X_n$:

$$X_{(1,n)} \leq X_{(2,n)} \leq \cdots \leq X_{(n-1,n)} \leq X_{(n,n)}$$

pour simplifier la discussion, on suppose que X a une loi à densité par rapport à la mesure de Lebesgue, de sorte que les inégalités précédentes sont strictes avec probabilité 1.

Alors

$$F_n(x) = \begin{cases} 0 & x < X_{(1,n)} \\ \frac{i}{n} & x \in [X_{(i,n)}, X_{(i+1,n)}[\\ 1 & x > X_{(n,n)} \end{cases} \qquad \text{taille des sauts: } 1/n$$

Convergence de la fonction F_n

Convergence ponctuelle (LGN forte) Pour tout $x \in \mathbb{R}$,

$$\lim_n F_n(x) = F(x) \qquad \text{p.s.}$$

Convergence uniforme (Glivenko-Cantelli)[a]

$$\lim_n \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0 \quad \text{p.s.}$$

Vitesse de convergence (TCL) Pour tout $x \in \mathbb{R}$,

$$\sqrt{n}\left(F_n(x) - F(x)\right) \stackrel{loi}{\Longrightarrow} \mathcal{N}\Big(0, F(x)(1-F(x))\Big).$$

et par le lemme de Slutsky

$$\sqrt{n}\frac{\Big(F_n(x) - F(x)\Big)}{\sqrt{F_n(x)(1-F_n(x))}} \stackrel{loi}{\Longrightarrow} \mathcal{N}\Big(0, 1\Big)$$

[a] Theorem 19.1., A.W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press

I-2. Estimation du quantile d'ordre α de X

Notations. X v.a. réelle, de fonction de répartition F.

Définition.

$$Q(\alpha) := \inf\{x \in \mathbb{R} \ : F(x) \geq \alpha\}.$$

Estimateur empirique.

Pour tout $\alpha \in]0,1[$, on pose

$$Q_n(\alpha) := X_{(\lceil n\alpha \rceil, n)} = \inf\{x \in \mathbb{R} : F_n(x) \geq \alpha\}.$$

Exemple. : Sur $n = 1000$ données, le quantile empirique d'ordre 0.99 est $X_{(990,1000)}$ soit la 11-ième plus grande donnée.

Quantiles rares. Si l'on dispose de n données, tous les quantiles empiriques d'ordre supérieur à $(1 - 1/n)$ sont égaux et valent $X_{(n,n)}$; tous ceux d'ordre inférieur à $1/n$ sont égaux et valent $X_{(1,n)}$.

Convergence de $Q_n(\alpha)$, $\alpha \in]0,1[$

Théorème. En tout point de continuité α de Q, on a $Q_n(\alpha) \to Q(\alpha)$, p.s.

Nous avons le résultat plus général[a]: si F est une fonction de répartition sur $\mathbb{R}$ et $\{F_n, n \geq 0\}$ est une suite de fonctions de répartition sur $\mathbb{R}$ t.q. $F_n(t) \to F(t)$ pour tout $t \in \mathbb{R}$ en lequel F est continue, alors on a convergence de leurs fonctions inverses généralisées (notées Q_n et Q) en tout point α en lequel Q est continu.

Preuve : cas général Soit Φ la fonction de répartition d'une v.a. V de loi $\mathcal{N}(0,1)$. Nous avons $|F_n(V) - F(V)| \to 0$ p.s. puisque F a au plus un nombre dénombrable de points de discontinuité. Nous avons les relations (voir Amphi 1)

$$Q_n(\alpha) \leq v \iff \alpha \leq F_n(v) \qquad Q(\alpha) \leq v \iff \alpha \leq F(v)$$

dont nous déduisons que

$$\Phi(Q_n(\alpha)) = \mathbb{P}(F_n(V) < \alpha) \qquad \Phi(Q(\alpha)) = \mathbb{P}(F(V) < \alpha).$$

Par le théorème du Portmanteau, il vient $\lim_n \Phi(Q_n(\alpha)) = \Phi(Q(\alpha))$ en tout point α tel que $\mathbb{P}(F(V) = \alpha) = 0$ ce qui est vérifié en tout point α en lequel Q est continu. On en déduit que $Q_n(\alpha) \to Q(\alpha)$.

[a] Lemma 21.2., A.W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press

Convergence de $Q_n(\alpha)$, $\alpha \in]0,1[$

Théorème. En tout point de continuité α de Q, on a $Q_n(\alpha) \to Q(\alpha), \quad$ p.s.

Remarque. Ce résultat concerne tout estimateur de quantile du moment que l'hypothèse sur la convergence des fonctions de répartition associées est vérifiée. Par suite, on peut aussi prendre $X_{(\lfloor n\alpha \rfloor, n)}$ comme approximation du quantile.

Fluctuations (TCL)

Théorème. Supposons que X possède une densité f strictement positive. Alors

$$\lim_{n\to\infty} X_{(\lceil n\alpha\rceil,n)} = Q(\alpha) \quad \text{p.s.}$$

et l'on a

$$\sqrt{n}(X_{(\lceil n\alpha\rceil,n)} - Q(\alpha)) \overset{loi}{\Longrightarrow} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f^2(Q(\alpha))}\right).$$

Preuve. Voir appendix.

à la variance quand $\alpha \to 0$ et $\alpha \to 1$: la variance est une fonction de $1/f^2(Q(\alpha))$.

II. Méthodes de Monte Carlo par Chaînes de Markov (MCMC)

Méthode de splitting

Références sur la théorie des chaînes de Markov et les méthodes MCMC

- ✓ C.P. Robert and G. Casella, "Monte Carlo Statistical Methods". Springer, 2010.
- ✓ S. Meyn and R.L. Tweedie, "Markov chains and Stochastic Stability". Cambridge, 2009.
- ✓ R. Douc, E. Moulines, P. Priouret and P. Soulier, "Markov Chains". Springer, 2018.

II-1. Motivations

La loi des grands nombres (faible / forte)

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \mathbb{E}[f(X)] \qquad \text{en proba / p.s.} \tag{1}$$

existe pour d'autres familles de v.a. $\{X_i, i \geq 1\}$ que celles de type "indépendantes de même loi que X". En particulier, certaines *chaînes de Markov* vérifient ce théorème limite.

La méthode de Monte Carlo dite "par Chaînes de Markov" consiste à simuler une chaîne de Markov $\{X_0, X_1, \cdots\}$

- ✓ ayant une unique loi invariante ν spécifiée par l'utilisateur (la loi de X dans (1))
- ✓ et vérifiant des théorèmes limites tels que loi de grands nombre, théorème de la limite centrale, etc

II-2. Chaîne de Markov (rappels)

Définition. Soit une suite de v.a. $\{X_i, i \geq 0\}$ à valeur dans $(E, \mathcal{E})$, définie sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et adaptée à la filtration $\{\mathcal{F}_n, n \geq 0\}$.

C'est une chaîne de Markov (pour la filtration $\{\mathcal{F}_n, n \geq 0\}$) ssi pour tout ensemble $A \in \mathcal{E}$ et pour tout $n \geq 1$,

$$\mathbb{P}(X_n \in A | \mathcal{F}_{n-1}) = \mathbb{P}(X_n \in A | X_{n-1}).$$

Loi du $(n+1)$-uplet $(X_0, \cdots, X_n)$. Soit une chaîne de Markov (homogène) de loi initiale ξ et de noyau de transition P.

Alors la loi de $(X_0, \cdots, X_n)$ est

$$\xi(\mathrm{d}x_0)\, P(x_0, \mathrm{d}x_1)\, P(x_1, \mathrm{d}x_2) \cdots P(x_{n-1}, \mathrm{d}x_n).$$

Exemple. AR(1) gaussien dans $\mathbb{R}^d$

Soit $\rho \in]-1,1[$; et des v.a. indépendantes $Y_1, Y_2, \cdots$ de même loi $\mathcal{N}_d(0, \mathrm{Id})$ et indépendantes de X_0.

La suite aléatoire définie par

$$X_i = \rho\, X_{i-1} + \sqrt{1-\rho^2}\, Y_i \qquad i \geq 1$$

est une Chaîne de Markov homogène, pour la filtration naturelle $\mathcal{F}_n := \sigma(X_0, Y_1, \cdots, Y_n)$.

- Sa loi initiale est celle de X_0.
- Noyau de transition : la loi conditionnelle de X_i sachant X_{i-1} est la loi $\mathcal{N}_d(\rho X_{i-1}, (1-\rho^2)\mathrm{Id})$ dont on déduit le noyau de transition de la chaîne

$$P(x, A) = \frac{1}{(2\pi)^{d/2}(1-\rho^2)^{d/2}} \int_A \exp\left(-0.5(1-\rho^2)^{-1}\|z - \rho x\|^2\right) \mathrm{d}z.$$

Exemple (suite). AR(1) gaussien dans $\mathbb{R}^d$

• Déterminons la loi de X_n:

$$X_n = \rho^n X_0 + \sum_{j=1}^{n} \rho^{n-j} \sqrt{1-\rho^2}\, Y_j$$

$$\sim \rho^n X_0 + \mathcal{N}_d(0, (1-\rho^{2n})\mathrm{Id}).$$

dont nous déduisons que pour tout $n \geq 0$,

$$X_n \sim \mathcal{N}_d(0, \mathrm{Id}) \quad \text{quand } X_0 \sim \mathcal{N}_d(0, \mathrm{Id})$$

Sinon, convergence en loi vers $\mathcal{N}_d(0, \mathrm{Id})$.

Définition: loi invariante. La loi (mesure) ν est invariante pour le noyau de transition P ssi

$$\nu P = \nu \qquad \text{i.e.} \qquad \int \nu(\mathrm{d}x) P(x, \mathrm{d}y) = \nu(\mathrm{d}y).$$

Théorème ergodique. (rappel cours 2A-MAP432) Soit $(X_n)_n$ une chaîne de Markov irréductible et récurrente positive sur un espace dénombrable E, de matrice de transition P, et d'unique loi invariante ν.

Alors, pour toute fonction $g : E \mapsto \mathbb{R}$ vérifiant $\int_E |g(x)| \nu(\mathrm{d}x) < \infty$, on a

$$\mathbf{\frac{1}{n} \sum_{i=1}^{n} g(X_i) \to \int g(x) \nu(\mathrm{d}x)} \quad \text{p.s.}$$

quelle que soit la loi initiale de X_0.

Remarque. moyenne temporelle sur une trajectoire versus moyenne en espace.

Extensions

✓ Il existe des versions valables dans des espaces E plus compliqués ($\mathbb{R}^d$ par exemple).

✓ Il existe un TCL à vitesse $\sqrt{n}$. La variance a une expression non explicite en général - en particulier sous forme d'une serie des auto-corrélations d'ordre k: la variance **n'est pas** la matrice de covariance de $g(X)$ quans $X \sim \nu$.

✓ En Monte Carlo, on se donne la loi cible ν : comment construire une chaîne de Markov $(X_i)_i$ admettant ν pour unique loi invariante ?

- Plusieurs algorithmes (Metropolis-Hastings, Gibbs, ...).
- Deux questions importantes distinctes :
  - S'assurer que la chaîne de Markov X a ν comme loi invariante (facile)
  - Démontrer que le théorème ergodique fonctionne (difficile).

II-3. Le cas des chaînes de Markov réversibles

Définition. La loi ν est réversible pour le noyau de transition P ssi

$$\nu(\mathrm{d}x)\, P(x, \mathrm{d}y) = \nu(\mathrm{d}y)\, P(y, \mathrm{d}x)$$

Dans le cas où E est discret :

$$\nu(x)P(x,y) = \nu(y)P(y,x), \forall (x,y) \in E^2.$$

Proposition. Si ν est réversible pour P alors ν est invariante pour P.

Preuve (cas discret). Montrons que $\sum_x \nu(x)P(x,y) = \nu(y)$ pour tout $y \in E$.

Nous savons par hypothèse que $\nu(x)P(x,y) = \nu(y)P(y,x)$; on somme à gauche et à droite en x et l'on obtient

$$\sum_{x \in E} \nu(x)P(x,y) = \nu(y) \sum_{x \in E} P(y,x) = \nu(y)P(y,E) = \nu(y).$$

Exemple. AR(1) gaussien dans $\mathbb{R}^d$ (suite)

Montrons que $\nu \equiv \mathcal{N}(0, \mathrm{Id})$ est la loi stationnaire de cette chaîne. Pour ce faire, nous vérifions la condition de réversibilité.

$$\begin{aligned}
\nu(\mathrm{d}x)P(x, \mathrm{d}y) &\propto \exp(-0.5\|x\|^2)\exp(-0.5(1-\rho^2)^{-1}\|y-\rho x\|^2) \\
&= \exp\left(-0.5(1-\rho^2)^{-1}\{(1-\rho^2)\|x\|^2 + \|y\|^2 + \rho^2\|x\|^2 - 2\rho y^\top x\}\right) \\
&= \exp\left(-0.5(1-\rho^2)^{-1}\{\|x\|^2 + \|y\|^2 - 2\rho y^\top x\}\right) \\
&= \text{noter la symétrie} \\
&= \nu(\mathrm{d}y)P(y, \mathrm{d}x).
\end{aligned}$$

Exemple. Chaîne de Metropolis-Hastings

Soit une loi $\pi \, \mathrm{d}\lambda$ sur $\mathbb{R}^d$, de densité π par rapport à une mesure λ sur $\mathbb{R}^d$.

On choisit un mécanisme de proposition de points partant du point courant x: $q(x,y)\lambda(\mathrm{d}y)$ par exemple, dans le cas où λ est la mesure de Lebesgue, on peut prendre $\mathcal{N}(x,\sigma^2)$[une loi gaussienne centrée en x et de variance σ^2.

Le noyau de transition de HM

$$P(x,\mathrm{d}y) := q(x,y)\alpha(x,y)\lambda(\mathrm{d}y) + \delta_x(\mathrm{d}y) \int q(x,z) \left(1 - \alpha(x,z)\right) \lambda(\mathrm{d}z)$$

où

$$\alpha(x,y) := 1 \wedge \frac{\pi(y)}{\pi(x)} \frac{q(y,x)}{q(x,y)}$$

est réversible par rapport à $\pi \, \mathrm{d}\lambda$.

Exemple (suite) : Comment produire une chaîne ayant ce noyau P ?

cible : $\pi(x)\mathrm{d}x$ sur $\mathbb{R}^d$

- Choix du mécanisme de proposition : par exemple

$$q(x,y) \equiv \text{densité de la loi } \mathcal{N}_d(x,C) \text{ évaluée au point } y$$

- Répéter : étant donné X_n

- simuler $X_{n+1/2} \sim \mathcal{N}_d(X_n, C)$
- calculer

$$\alpha(X_n, X_{n+1/2}) := 1 \wedge \frac{\pi(X_{n+1/2})}{\pi(X_n)} \frac{q(X_{n+1/2}, X_n)}{q(X_n, X_{n+1/2})} = 1 \wedge \frac{\pi(X_{n+1/2})}{\pi(X_n)}$$

- tirer $U \sim \mathcal{U}([0,1])$
- Si $U \leq \alpha(X_n, X_{n+1/2})$ alors $X_{n+1} = X_{n+1/2}$; sinon, $X_{n+1} = X_n$.

- Le noyau de cette chaîne est

$$P(x, \mathrm{d}y) := q(x,y)\alpha(x,y)\mathrm{d}y + \delta_x(\mathrm{d}y) \int q(x,z)\left(1 - \alpha(x,z)\right)\mathrm{d}z$$

II-4. Application à la simulation de loi restreinte à A.

Soit ν une loi sur E et $A \subset E$ un ensemble mesurable. On note ν_A la loi ν restreinte à A i.e. (loi conditionnelle) définie par

$$\int f(x)\nu_A(\mathrm{d}x) := \frac{\int_A f(x)\nu(\mathrm{d}x)}{\nu(A)}.$$

Algorithme. Soit P un noyau de transition t.q. $\nu(\mathrm{d}x)P(x,\mathrm{d}y) = \nu(\mathrm{d}y)P(y,\mathrm{d}x)$.

Initialisation : $X_0 \in A$.
Répéter :

- ✓ Tirer $X_{n+1/2} \sim P(X_n, \mathrm{d}x)$
- ✓ Si $X_{n+1/2} \in A$, poser $X_{n+1} = X_{n+1/2}$. Sinon, $X_{n+1} = X_n$.

Remarque. Un exemple de noyau P vérifiant la condition de réversibilité, est un noyau de Metropolis-Hastings admettant ν comme mesure invariante.

Proposition. La loi ν_A est réversible pour la chaîne $\{X_n, n \geq 0\}$.

Preuve. Le noyau de transition de la chaîne est

$$P_A(x, \mathrm{d}y) := P(x, \mathrm{d}y)\mathbf{1}_A(y) + \delta_x(\mathrm{d}y)P(x, A^c) \qquad A^c := E \setminus A.$$

On en déduit, en utilisant $\nu_A(\mathrm{d}x) = \nu(\mathrm{d}x)\mathbf{1}_A(x)/\nu(A)$ et l'hypothèse de réversibilité sur (ν, P)

$$\begin{aligned}
\nu_A(\mathrm{d}x)P_A(x, \mathrm{d}y) &= \nu(A)^{-1}\mathbf{1}_A(x)\mathbf{1}_A(y)\nu(\mathrm{d}x)P(x, \mathrm{d}y) + \delta_y(\mathrm{d}x)\nu_A(\mathrm{d}y)P(y, A^c) \\
&= \nu(A)^{-1}\mathbf{1}_A(x)\mathbf{1}_A(y)\nu(\mathrm{d}y)P(y, \mathrm{d}x) + \delta_y(\mathrm{d}x)\nu_A(\mathrm{d}y)P(y, A^c) \\
&= \nu_A(\mathrm{d}y)P_A(y, \mathrm{d}x).
\end{aligned}$$

Algorithme. Le corollaire est que si la chaîne $(X_n)_n$ vérifie les conditions du théorème ergodique, alors on a pour toute fonction g t.q. $\int |g| \mathrm{d}\nu_A < \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \to \int g(x)\nu_A(\mathrm{d}x)$$

i.e. nous disposons d'un algorithme pour construire des points approchant la loi ν_A à partir d'un mécanisme (le noyau P) qui approche ν.

II-5. Méthode de splitting

Idée du splitting: découper l'évènement rare $\mathbb{P}(g(X) \in A)$ en une suite d'évènements de plus en plus rares:

$$A_0 := \text{tout l'espace} \supset \cdots \supset A_i \supset \cdots \supset A_I := A,$$

et utiliser la relation

$$\mathbb{P}\left(\mathbf{g(X)} \in \mathbf{A}\right) = \prod_{\mathbf{i=1}}^{\mathbf{I}} \mathbb{P}\left(\mathbf{g(X)} \in \mathbf{A_i} | \mathbf{g(X)} \in \mathbf{A_{i-1}}\right).$$

Exemple. $\{|X| \geq 6\} \subset \{|X| \geq 5\} \subset \{|X| \geq 4\} \cdots \subset \{|X| \geq 0\}$.

Gain espéré. On choisit les ensembles A_i de sorte que chaque probabilité conditionnelle n'est pas petite (événement non rare).

En pratique.

- ✓ Choix du nombre de niveaux I (pas facile) et des ensembles A_i (pas facile).
- ✓ Approximation des probabilités conditionnelles par moyenne ergodique d'une chaine ν-réversible avec rejet.

Mise en oeuvre pour le calcul de $\mathbb{P}(g(X) \in A)$

1. Se donner $A_0 \subset A_1 \subset \cdots \subset A_I = A$.
2. Simulation indépendante de chaînes de Markov : pour tout $i \in \{1, \cdots, I\}$,
   - ✓ on définit une chaîne $(X_n^i)_n$ construite pour approcher la loi de X sachant $\{g(X) \in A_{i-1}\}$.
   - ✓ On a le théorème ergodique suivant (oubli du point initial)

$$\frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{g(X_n^i) \in A_i} \underset{N \to +\infty}{\longrightarrow} \mathbb{P}\left[g(X) \in A_i | g(X) \in A_{i-1}\right], \qquad \text{p.s.}$$

3. Estimateur final:

$$\mathbb{P}\left[\mathbf{g}(\mathbf{X}) \in \mathbf{A}\right] \approx \prod_{\mathbf{i}=1}^{\mathbf{I}} \left(\frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{n}=1}^{\mathbf{N}} \mathbf{1}_{\mathbf{g}(\mathbf{X}_\mathbf{n}^\mathbf{i}) \in \mathbf{A}_\mathbf{i}} \right); \qquad \mathbf{A} = \mathbf{A}_\mathbf{I}, \ \mathbf{A}_\mathbf{0} = \mathbb{R}.$$

Réglages par l'utilisateur: A_i, I, N...

Mise en oeuvre (suite)

Comment construire une chaîne de Markov dont la mesure invariante est la loi de X sachant $g(X) \in A_{i-1}$, dans le cas gaussien $X \sim \mathcal{N}(0,1)$?

- ✓ choix de X_0 tel que $g(X_0) \in A_{i-1}$
- ✓ itération:

$$X_{n+1/2} = \rho X_n + \sqrt{1-\rho^2}\, Y_n^i \qquad (Y_n^i)_n \overset{i.i.d}{\sim} \mathcal{N}(0,1)$$

$$\text{poser } X_{n+1} = X_{n+1/2} \text{ si } g(X_{n+1/2}) \in A_{i-1}$$

$$\text{et } X_{n+1} = X_n \text{ sinon.}$$

Réglages par l'utilisateur: ρ.

III. Changements de probabilité

Echantillonnage préférentiel (Ech. d'importance)

D'un usage très courant en statistique et probabilité.

✓ Théorie de l'estimation par maximum de vraisemblance: étude des densités (*vraisemblance*) du modèle paramétrique par rapport à une probabilité de référence.

✓ Dans modèles discrets, changements de probabilité fréquents mais très simples.

✓ De manière plus spectaculaire, applications dans la théorie des espaces gaussiens, du processus de Poisson, et plus généralement dans l'étude des martingales.

III-1. Quelques idées simples

Pour l'approximation Monte Carlo de

$$\mu := \mathbb{E}_f\left[h(X)\right] = \int h(x)\, f(x)\lambda(\mathrm{d}x)$$

Stratégie 1 (Monte Carlo naïf).

$$\hat{\mu}_N^{(1)} := \frac{1}{N}\sum_{n=1}^{N} h(X_n) \qquad\qquad X_n \overset{i.i.d.}{\sim} f\,\mathrm{d}\lambda$$

Stratégie 2 (Monte Carlo par Ech. d'Importance).

Pour toute loi $g\,\mathrm{d}\lambda$ t.q. $\{f > 0\} \subset \{g > 0\}$, sur la base de l'observation

$$\int h(x)\, f(x)\lambda(\mathrm{d}x) = \int_{\{f>0\}} h(x)\, f(x)\lambda(\mathrm{d}x) = \int_{\{f>0\}} h(x)\,\frac{f(x)}{g(x)}\,g(x)\lambda(\mathrm{d}x)$$

on propose

$$\hat{\mu}_N^{(2)} := \frac{1}{N}\sum_{n=1}^{N} h(X_n)\frac{f(X_n)}{g(X_n)} \qquad\qquad X_n \overset{i.i.d.}{\sim} g\,\mathrm{d}\lambda$$

Comparaison des stratégies par le biais

Les deux stratégies conduisent à des estimateurs sans biais de μ

$$\mathbb{E}_f\left[\hat{\mu}_N^{(1)}\right] = \mu, \qquad \mathbb{E}_g\left[\hat{\mu}_N^{(2)}\right] = \mu.$$

Comparaison des stratégies par la taille des IC

Des IC asymptotiques de niveau $1-\alpha$ pour l'estimation de μ sont donnés par

$$\hat{\mu}_N^{(1)} \pm \frac{z_{1-\alpha/2}}{\sqrt{N}} \sqrt{\mathbb{V}\mathrm{ar}_f(h(X))}, \qquad \hat{\mu}_N^{(2)} \pm \frac{z_{1-\alpha/2}}{\sqrt{N}} \sqrt{\mathbb{V}\mathrm{ar}_g\left(h(X)\frac{f(X)}{g(X)}\right)}$$

où $z_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ d'une loi $\mathcal{N}(0,1)$.

Meilleure méthode (critère: taille des IC)

On préfèrera la première méthode ssi

$$\mathbb{V}\mathrm{ar}_f(h(X)) << \mathbb{V}\mathrm{ar}_g\left(h(X)\frac{f(X)}{g(X)}\right)$$

ou de façon équivalente (csq de l'égalité des espérances)

$$\int h^2(x)\, f(x)\ \lambda(\mathrm{d}x) << \int \left(h(x)\frac{f(x)}{g(x)}\right)^2 g(x)\ \lambda(\mathrm{d}x).$$

III-2. Le changement de probabilité parfait ... et utopiste

Objectif

$$\mathbb{E}_f\left[h(X)\right] \qquad h > 0$$

Existence du changement de loi parfait. Il existe un changement de loi $g_\star \mathrm{d}\lambda$ qui permet de construire un estimateur par échantillonnage d'importance dont la taille de l'IC asymptotique associé, est nulle :

$$g_\star(x) := \frac{h(x)\, f(x)}{\mathbb{E}_f[h(X)]}$$

la taille de l'IC est proportionnelle à la racine carrée de $\mathbb{V}\mathrm{ar}_{g_\star}(h(X)f(X)/g_\star(X))$. On a

$$\mathbb{E}_{g_\star}\left[\left(h(X)\frac{f(X)}{g_\star(X)}\right)^2\right] = \left(\mathbb{E}_f\left[h(X)\right]\right)^2 \int g_\star(x)\,\lambda(\mathrm{d}x) = \left(\mathbb{E}_f\left[h(X)\right]\right)^2 = \left(\mathbb{E}_{g_\star}\left[h(X)\frac{f(X)}{g_\star(X)}\right]\right)^2$$

Noter : variance nulle, donc estimateur constant et en particulier égal à sa moyenne ... qui est la quantité inconnue !

Mais un tel choix signifie que $\mathbb{E}_f[h(X)]$ est connue : non.

Ce calcul aide néanmoins à choisir un changement de loi $g\,\mathrm{d}\lambda$.

III-3. Application à l'estimation de quantile $(X \in \mathbb{R})$

Objectif : pour $\alpha \in]0,1[$,

$$Q(\alpha) := \inf\{x \in \mathbb{R}, \mathbb{P}_f(X \leq x) \geq \alpha\}; \qquad \text{sous } \mathbb{P}_f,\ X \sim f \mathrm{d}\lambda$$

Stratégie 1. $\{X_i, i \geq 1\}$ i.i.d. de loi $f\,\mathrm{d}\lambda$

$$Q_n^{(1)}(\alpha) := \inf\{x \in \mathbb{R}, \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{X_i \leq x} \geq \alpha\} = X_{(\lceil n\alpha \rceil, n)}$$

Stratégie 2. $\{X_i, i \geq 1\}$ i.i.d. de loi $g\,\mathrm{d}\lambda$

$$Q_n^{(2)}(\alpha) := \inf\{x \in \mathbb{R}, \frac{1}{n}\sum_{i=1}^n \frac{f(X_i)}{g(X_i)} \mathbf{1}_{X_i \leq x} \geq \alpha\}$$

basée sur la relation

$$\mathbb{P}_f(X \leq x) = \mathbb{E}_g\left[1_{X \leq x} \frac{f(X)}{g(X)}\right]$$

Il faut faire un tri ordonné des X_i et observer que cette seconde fonction de répartition empirique saute de $n^{-1} f(X_i)/g(X_i)$ au point X_i.

III-4. Changement de loi gaussien, cas $\mathbb{R}$

Notations. Sous $\mathbb{P}_{(\mu,\sigma^2)}$, $X \sim \mathcal{N}(\mu,\sigma^2)$.

Formules de changement de loi.

$$\mathbb{E}_{(\mathbf{0},\mathbf{1})}\left[\mathbf{h}(\mathbf{X})\right] = \sigma\,\mathbb{E}_{(\mu,\sigma^{\mathbf{2}})}\left[\mathbf{h}(\mathbf{X})\exp\left(-\frac{\mathbf{1}}{\mathbf{2}\sigma^{\mathbf{2}}}\{(\sigma^{\mathbf{2}}-\mathbf{1})\mathbf{X}^{\mathbf{2}}+\mathbf{2}\mu\mathbf{X}-\mu^{\mathbf{2}}\}\right)\right] \tag{2}$$

dont on déduit

$$\mathbb{E}_{(\mu,\sigma^{\mathbf{2}})}\left[\mathbf{h}(\mathbf{X})\right] = \frac{\tilde{\sigma}}{\sigma}\,\mathbb{E}_{(\tilde{\mu},\tilde{\sigma}^{\mathbf{2}})}\left[\mathbf{h}(\mathbf{X})\exp\left(\frac{\mathbf{1}}{\mathbf{2}}\mathcal{Q}(\mathbf{X})\right)\right] \tag{3}$$

avec

$$\mathcal{Q}(X) := \left(\frac{1}{\tilde{\sigma}^2}-\frac{1}{\sigma^2}\right)X^2 + 2\left(\frac{\mu}{\sigma^2}-\frac{\tilde{\mu}}{\tilde{\sigma}^2}\right)X - \left(\frac{\mu^2}{\sigma^2}-\frac{\tilde{\mu}^2}{\tilde{\sigma}^2}\right)$$

Cas particulier: $\tilde{\sigma}^2 = \sigma^2$ (changement de moyenne)

$$\mathbb{E}_{(\mu,\sigma^2)}\left[\mathbf{h}(\mathbf{X})\right] = \mathbb{E}_{(\tilde{\mu},\sigma^2)}\left[\mathbf{h}(\mathbf{X}) \exp\left(\frac{\mu-\tilde{\mu}}{\sigma^2}\mathbf{X} - \frac{1}{2\sigma^2}\{\mu^2 - \tilde{\mu}^2\}\right)\right] \qquad (4)$$

En écrivant $\tilde{\mu} = \mu + \tau\sigma^2$, il vient

$$\mathbb{E}_{(\mu,\sigma^2)}\left[h(X)\right] = \mathbb{E}_{(\mu+\tau\sigma^2,\sigma^2)}\left[h(X) \exp\left(-\tau(X-\mu) + \frac{\tau^2\sigma^2}{2}\right)\right] \qquad (5)$$

Exemple: Probabilités de queue. Objectif

$$p_\star := \mathbb{P}_{(0,1)}(X > c)$$

La taille de l'IC asymptotique de niveau $(1-\alpha)$ basé sur l'estimateur

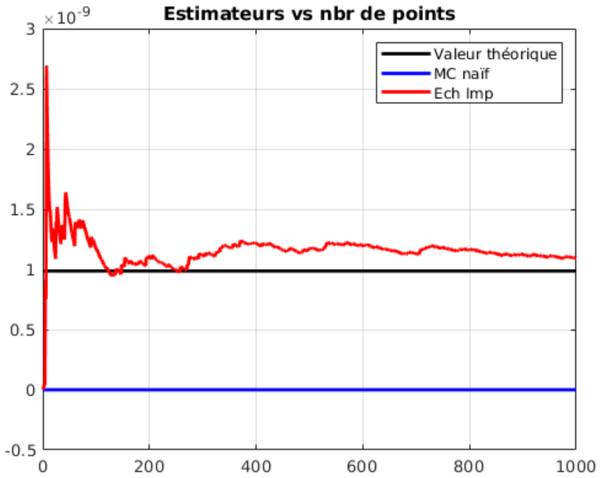
$$\hat{\mu}_N^{(2)} := \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{X_n > c} \exp\left(0.5\tau^2 - \tau X_n\right) \qquad X_n \overset{i.i.d.}{\sim} \mathcal{N}(\tau, 1)$$

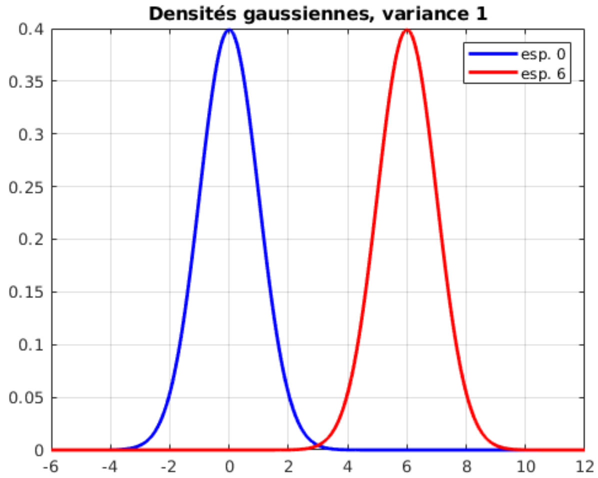
est proportionnelle à la racine carrée de

$$\mathbb{E}_{(\tau,1)}\left[\mathbf{1}_{X>c} \exp\left(\tau^2 - 2\tau X\right)\right] - p_\star^2 = \mathbb{E}_{(0,1)}\left[\mathbf{1}_{X>c} \exp\left(0.5\tau^2 - \tau X\right)\right] - p_\star^2.$$

Elle est minimale en $\tau_\star$ tel que $\tau_\star > c$. **En pratique, $\tau_\star$ et l'espérance sont non explicites.**

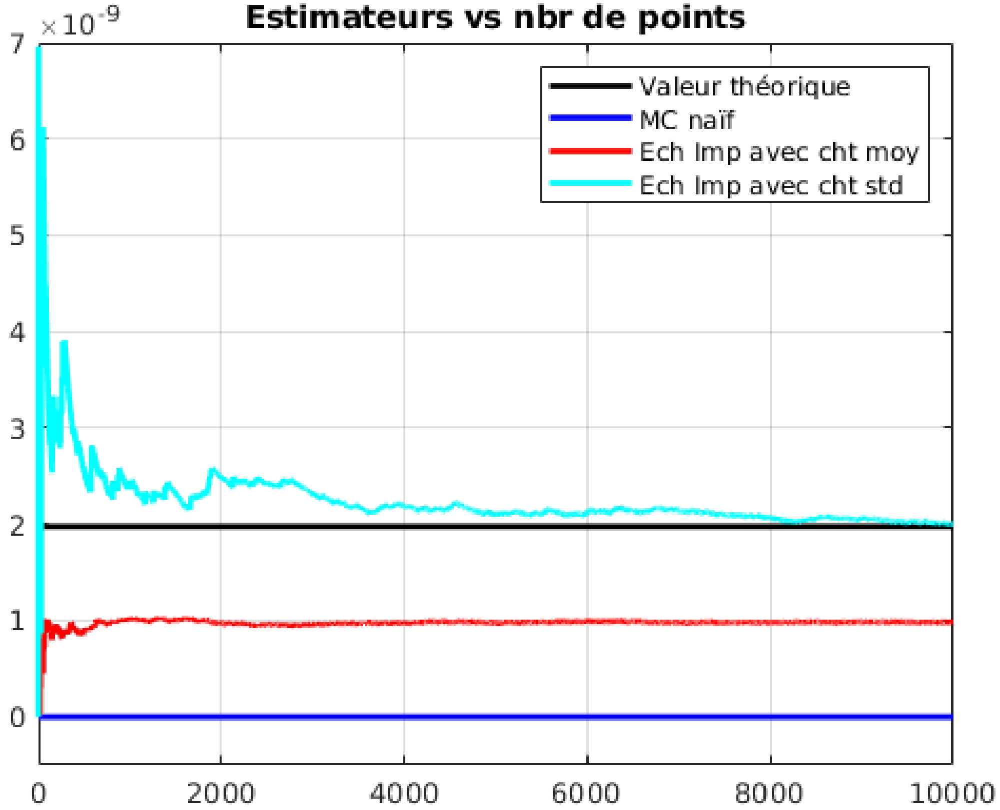
Une illustration numérique $\mathbb{P}_{(0,1)}(X > c)$ par Monte Carlo naïf (bleu), puis avec changement de moyenne (rouge, $0 \to 6$). Cas $c = 6$.





c	Valeur exacte de $\mathbb{P}_{(0,1)}(X > c)$	Moyenne empirique (MC naïf)	Demi-largeur Int. Confiance (à 95%)	Moyenne empirique (Ech. préf.)	Demi-largeur Int. Confiance (à 95%)	Réduction de variance
1	1,59E-001	1,48E-001	6,96E-003	1,58E-001	3,74E-003	3.46
2	2,28E-002	2,12E-002	2,82E-003	2,24E-002	6,77E-004	17.4
3	1,35E-003	1,70E-003	8,07E-004	1,34E-003	4,84E-005	1/279
4	3,17E-005	0,00E+000	0,00E+000	3,24E-005	1,34E-006	∞
5	2,87E-007	0,00E+000	0,00E+000	2,90E-007	1,35E-008	∞

Illustration numérique 2. $\mathbb{P}_{(0,1)}(|X| > c)$ par Monte Carlo naïf (bleu), puis avec changement de moyenne (rouge, $(0 \to 6)$), puis avec seul changement de variance (cyan, écart-type $1 \to 9$). Cas $c = 6$.



Attention au choix aveugle de changement de probabilités.

Corollaire (du changement de lois (3))

Pour tout $\tilde{\mu} \in \mathbb{R}$ et $\tilde{\sigma} > 0$,

$$\begin{aligned} \mathbb{E}_{(\mu,\sigma^2)}\left[\mathbf{h}(\mathbf{X})\right] &= \mathbb{E}_{(\tilde{\mu},\tilde{\sigma}^2)}\left[h\left(\frac{\sigma}{\tilde{\sigma}}(X-\tilde{\mu})+\mu\right)\right] \\ &= \frac{\sigma}{\tilde{\sigma}}\mathbb{E}_{(\mu,\sigma^2)}\left[\mathbf{h}\left(\frac{\sigma}{\tilde{\sigma}}(\mathbf{X}-\tilde{\mu})+\mu\right)\exp\left(\frac{\mathbf{1}}{\mathbf{2}}\mathcal{Q}(\mathbf{X})\right)\right] \end{aligned} \tag{6}$$

avec

$$\mathcal{Q}(X) := \left(\frac{1}{\sigma^2}-\frac{1}{\tilde{\sigma}^2}\right)X^2 + 2\left(\frac{\tilde{\mu}}{\tilde{\sigma}^2}-\frac{\mu}{\sigma^2}\right)X - \left(\frac{\tilde{\mu}^2}{\tilde{\sigma}^2}-\frac{\mu^2}{\sigma^2}\right).$$

Ainsi, sans changer la v.a. X mais en compensant par un **poids d'importance**, on peut construire une autre v.a. de même espérance mais pas nécessairement de même variance.

➥ importantes applications en simulations Monte Carlo et échantillonnage préférentiel (voir section IV).

III-5. Changement de loi gaussien, cas $\mathbb{R}^d$

Notations. Sous $\mathbb{P}_{(\mu,\Gamma)}$, $\quad X \sim \mathcal{N}_d(\mu, \Gamma)$.

Les formules de changement de loi s'obtiennent de façon analogue au cas réel. En particulier

$$\mathbb{E}_{(\mathbf{O},\mathrm{Id})}\left[\mathbf{h}(\mathbf{X})\right] = \sqrt{\det(\boldsymbol{\Gamma})}\,\mathbb{E}_{(\mu,\boldsymbol{\Gamma})}\left[\mathbf{h}(\mathbf{X})\exp\left(-\frac{\mathbf{1}}{\mathbf{2}}\{\|\mathbf{X}\|^{\mathbf{2}} - (\mathbf{X}-\mu)^\top\boldsymbol{\Gamma}^{-\mathbf{1}}(\mathbf{X}-\mu)\}\right)\right]$$

III-6. D'une façon générale

$$\mathbb{E}_\nu\left[h(X)\right] = \int h(x)\mathrm{d}\nu(x) = \int h(x)\frac{\mathrm{d}\nu(\mathbf{x})}{\mathrm{d}\tilde{\nu}(\mathbf{x})}\,\mathrm{d}\tilde{\nu}(x) = \mathbb{E}_{\tilde{\nu}}\left[h(X)\,\mathbf{Z^{-1}}\right]$$

Définition. De manière générale un changement de probabilité de ν à $\tilde{\nu}$ est défini par une v.a. $Z \geq 0$ (**vraisemblance** $\frac{\mathrm{d}\tilde{\nu}}{\mathrm{d}\nu}$ ou dérivée de Radon-Nikodym) telle que $\mathbb{E}_\nu(Z) = 1$.

Propriétés. Formules de passage de ν à $\tilde{\nu}$ ou inversement (quand $Z > 0$) :

$$\mathbb{E}_{\tilde{\nu}}(\mathbf{Y}) = \mathbb{E}_\nu(\mathbf{YZ}), \qquad \mathbb{E}_\nu(\mathbf{Y}) = \mathbb{E}_{\tilde{\nu}}(\mathbf{YZ^{-1}}).$$

Application à la simulation Monte Carlo. Il s'agit de simuler $W := YZ^{-1}$ sous $\tilde{\nu}$ au lieu de Y sous ν:

$$\mathbb{E}_\nu(\mathbf{Y}) = \mathbb{E}_{\tilde{\nu}}(\mathbf{YZ^{-1}}) \approx \frac{\mathbf{1}}{\mathbf{N}}\sum_{\mathbf{n=1}}^{\mathbf{N}} \mathbf{W_n}$$

La méthode est meilleure si $\mathbb{V}\mathrm{ar}_{\tilde{\nu}}\left(\mathbf{W}\right) \ll \mathbb{V}\mathrm{ar}_\nu\left(\mathbf{Y}\right)$.
Au lieu d'être équipondérées avec poids $\omega_n := N^{-1}$ (stratégie 1), les simulations sont pondérées avec poids $\tilde{\omega}_\mathbf{n} := \mathbf{N^{-1}(Z^{-1})_n}$ (stratégie 2).

Conclusion

Grands principes pour mettre en œuvre les chgts de proba $\mathbb{P} \to \mathbb{Q}$

1. **Approche 1.** Décrire la loi $\mathbb{Q}$ puis déduire le ratio d'importance Z^{-1}:
   - ✓ on impose la distribution après changement de probabilités (facile, intuition du problème)
   - ✓ dans le cas de densités explicites, Z explicite (ratio des densités)
   - dans les autres cas, Z pas facilement explicitable (pbm de simulation).
2. **Approche 2.** Se donner une variable aléatoire Z (pour définir la vraisemblance), puis caractériser la loi $\mathbb{Q}$ induite:
   - ✓ facile de générer des variables Z positives d'espérance 1 sous $\mathbb{P}$
   - loi $\mathbb{Q}$ le plus souvent non explicite (pas dans le répertoire classique)
   - trouver des formes de Z manipulables: pour simuler, pour interpréter la distribution après changement de probabilités...

Il existe un certain nombre de changements de probabilités bien connus.

IV. Pour en savoir plus: Ech. d'importance adaptatif

Réf: B. Jourdain and J. Lelong, "Robust Adaptive Importance Sampling for Normal Random Vectors". Ann Appl Prob, 2009.

Objectifs

Objectif:

$$\mathbb{E}(h(X)) \qquad X \sim \mathcal{N}_d(0, \mathrm{Id}).$$

Notations. $\mathbb{E}$ designe l'espérance sous la loi "d'origine" i.e. la loi avec laquelle le problème est formulé; pour alléger les notations, on n'indique pas ici $(0, \mathrm{Id})$ en indice.

Changement de loi. Gaussien, avec drift $\theta \in \mathbb{R}^d$ sur la moyenne uniquement $(0 \to \theta)$. En mimant la preuve de (6) appliquée avec $\tilde{\mu} = -\theta$

$$\mathbb{E}[h(X)] = \mathbb{E}\left[h(X+\theta)\exp(-\theta^\top X - \frac{\|\theta\|^2}{2})\right]. \tag{7}$$

Quel choix optimal de θ? Crtière d'optimalité basé sur la minimisation de la variance (minimisation de la taille de l'IC asymptotique déduit du TCL). Minimiser la variance est équivalent à miniser le moment d'ordre 2 (voir slides précédents, section III).

Expressions du moment d'ordre 2, noté ν

1ère expression. Du terme de droite dans (7)

$$\nu(\theta) := \mathbb{E}\left[h^2(X+\theta)\exp(-2\theta^\top X - \|\theta\|^2)\right]$$

On cherche à minimiser ν sur $\mathbb{R}^d$.

✓ Calcul d'une approximation $\hat{\nu}_N$ de la fonction $\nu(\theta)$ par Monte Carlo :

$$\hat{\nu}_N(\theta) := \frac{1}{N}\sum_{n=1}^{N} h^2(X_n+\theta)\exp(-2\theta^\top X_n - \|\theta\|^2) \qquad X_n \overset{i.i.d.}{\sim} \mathcal{N}_d(0, \mathrm{Id}).$$

✓ Minimisation de la fonction $\hat{\nu}_N$ par un algorithme d'optimisation $\to$ Convexité de la fonction $\hat{\nu}_N$?

Si oui, algorithme de minimisation de Newton...

Mais la fonction $\theta \mapsto h^2(x+\theta)\exp(-2\theta^\top x - \|\theta\|^2)$ n'a pas de propriété de convexité particulière

2nde expression. En changeant une nouvelle fois de probabilité (i.e. mimant la preuve de (6) appliquée avec $\tilde{\mu} = +\theta$), on a

$$\nu(\theta) = \mathbb{E}\left[h^2(X) \exp(-\theta^\top X + \frac{\|\theta\|^2}{2})\right].$$

Approximation Monte Carlo

Seule une approximation Monte Carlo de ce critère est possible, par exemple

$$\hat{\nu}_N(\theta) := \frac{1}{N}\sum_{n=1}^{N} h^2(X_n) \exp(-\theta^\top X_n + \frac{\|\theta\|^2}{2}) \qquad X_n \overset{i.i.d.}{\sim} \mathcal{N}_d(0, \mathrm{Id}).$$

Cette fois, on a la convexité de $\hat{\nu}_N$ puisque pour tout $x \in \mathbb{R}^d$, la fonction $\theta \mapsto \upsilon(\theta) := \exp(-\theta^\top x + \frac{\|\theta\|^2}{2})$ est convexe.

Le gradient et le hessien sont donnés par

$$\nabla \upsilon(\theta) = (\theta - x)\upsilon(\theta) \qquad \nabla^2 \upsilon(\theta) = \left(\mathrm{Id} + (\theta - x)(\theta - x)^\top\right) \upsilon(\theta).$$

Convexité mais pas stricte: la borne inférieure sur $\nabla^2 \hat{\nu}_N(\theta)$ peut être petite...

Astuce ! l'algorithme d'optimisation recherche les zeros d'une fonction, alors ...

Tout zéro $\theta_\star$ de $\nabla \hat{\nu}_N(\theta)$ vérifie

$$\theta_\star \sum_{n=1}^N h^2(X_n) \exp(-\theta_\star^\top X_n + \frac{\|\theta_\star\|^2}{2}) = \sum_{n=1}^N h^2(X_n) \exp(-\theta_\star^\top X_n + \frac{\|\theta_\star\|^2}{2})\, X_n$$

et ce sont aussi les zeros de

$$\theta \mapsto \theta - \frac{\sum_{n=1}^N X_n h^2(X_n) \exp(-\theta^\top X_n)}{\sum_{n=1}^N h^2(X_n) \exp(-\theta^\top X_n)}$$

qui est le gradient de la fonction **fortement convexe**

$$\theta \mapsto u_N(\theta) := \frac{\|\theta\|^2}{2} + \log\left(\sum_{n=1}^N h^2(X_n) \exp(-\theta^\top X_n)\right).$$

la matrice Hessienne vaut

$$\nabla^2 u_N(\theta) = \mathrm{Id} + \frac{\sum_{n=1}^N \omega_n(\theta) X_n X_n^\top}{\sum_{n=1}^N \omega_n(\theta)} - \frac{\{\sum_{n=1}^N \omega_n(\theta) X_n\}\{\sum_{n=1}^N \omega_n(\theta) X_n\}^\top}{(\sum_{n=1}^N \omega_n(\theta))^2} \geq \mathrm{Id}$$

en ayant posé $\omega_n(\theta) := h^2(X_n) \exp(-\theta^\top X_n)$.

Résultats de convergence

Théorème. Soit $\theta_{N,\star}$ le zero de u_N.

1. $\theta_{N,*}$ converge vers un zero de ν (p.s. + TCL à vitesse $\sqrt{N}$).
2. $\frac{1}{N}\sum_{n=1}^{N} h(X_n + \theta_{N,*}) \exp(-\theta_{N,*}^\top X_n - \frac{\|\theta_{N,*}\|^2}{2})$ converge vers $\mathbb{E}(h(X))$ p.s. et avec un TCL de variance minimale.

Remarque. Au fil des modifications des θ, on utilise seulement X_n et $h(X_n)$ (on ne resimule pas les $h(X_n)$ ➠ gain en temps calcul).

En pratique:

- Méthode plus longue que Monte Carlo simple, mais gain sur la variance.
- Reste assez générique et robuste.

Des extensions et des variantes: voir réf. B. Jourdain and J. Lelong.

ANNEXES AMPHI 2

Preuve - Convergence quantile empirique - TCL

On souhaite démontrer la limite

$$\mathcal{P}_n(t) = \mathbb{P}\left(X_{(\lceil n\alpha \rceil, n)} < Q(\alpha) + t \frac{\sqrt{\alpha(1-\alpha)}}{f(Q(\alpha))} \frac{1}{\sqrt{n}} \right) \to \int_{-\infty}^{t} \frac{\exp(-u^2/2)}{\sqrt{2\pi}} \, du, \quad \forall t.$$

On remarque que $\left\{ \mathbf{X}_{(\lceil \mathbf{n}\alpha \rceil, \mathbf{n})} < \mathbf{Q}(\alpha) + \mathbf{t} \frac{\sqrt{\alpha(\mathbf{1}-\alpha)}}{\mathbf{f}(\mathbf{Q}(\alpha))} \frac{\mathbf{1}}{\sqrt{\mathbf{n}}} \right\} = \left\{ \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{n}} \mathbf{Y}_{\mathbf{j},\mathbf{n}} \geq \lceil \mathbf{n}\alpha \rceil \right\}$

avec $\mathbf{Y}_{\mathbf{j},\mathbf{n}} := \mathbf{1}_{\mathbf{X}_\mathbf{j} < \mathbf{Q}(\alpha) + \mathbf{t} \frac{\sqrt{\alpha(\mathbf{1}-\alpha)}}{\mathbf{f}(\mathbf{Q}(\alpha))} \frac{\mathbf{1}}{\sqrt{\mathbf{n}}}}$. Les $(Y_{j,n})_{1 \leq j \leq n}$ sont des v.a. de Bernoulli de paramètre

$$p_n = F\left(Q(\alpha) + t \frac{\sqrt{\alpha(1-\alpha)}}{f(Q(\alpha))} \frac{1}{\sqrt{n}} \right) = F(Q(\alpha)) + f(Q(\alpha)) t \frac{\sqrt{\alpha(1-\alpha)}}{f(Q(\alpha))} \frac{1}{\sqrt{n}} + o(n^{-1/2})$$

$$= \alpha + t\sqrt{\alpha(1-\alpha)} \frac{1}{\sqrt{n}} + o(n^{-1/2}).$$

$$\Longrightarrow \mathcal{P}_n(t) = \mathbb{P}\left(\frac{\sum_{j=1}^n (Y_{j,n} - p_n)}{\sqrt{np_n(1-p_n)}} \geq \frac{\lceil n\alpha \rceil - np_n}{\sqrt{np_n(1-p_n)}} \right) \text{ avec } \frac{\lceil n\alpha \rceil - np_n}{\sqrt{np_n(1-p_n)}} \to -t.$$

Théorème. **(TCL de Lindeberg-Lévy)** À n fixe, considérons des variables aléatoires $(Z_{n,j})_{1 \leq j \leq n}$ indépendantes, bornées uniformément en j et n, chacune étant de variance $\sigma^2_{n,j}$. Alors, si $s_n^2 = \sum_{j=1}^{n} \sigma^2_{n,j} \to \infty$ et si $t_n \to t \in \mathbb{R}$, on a:

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{\sum_{j=1}^{n} (Z_{n,j} - \mathbb{E}(Z_{n,j}))}{s_n} < t_n \right) = \int_{-\infty}^{t} \frac{\exp(-u^2/2)}{\sqrt{2\pi}} \, du.$$

Application. On déduit

$$\mathcal{P}_n \to \int_{-t}^{\infty} \frac{\exp(-u^2/2)}{\sqrt{2\pi}} \, du = \int_{-\infty}^{t} \frac{\exp(-u^2/2)}{\sqrt{2\pi}} \, du.$$