

Projet Monte Carlo : ABC pour les Modèles Alpha-Stables

Arthur Neau, Avner El Baz, Mohamed Mahmoud El Ghaouth

ENSAE Paris

25 avril 2025

Plan de la présentation

- 1 Introduction
- 2 Générateur alpha-stable
- 3 Améliorations du générateur
- 4 ABC : Approximate Bayesian Computation
- 5 Méthodes ABC
- 6 Comparaison des méthodes
- 7 Extension multivariée

- Estimer les paramètres de lois alpha-stables sans densité explicite.
- Implémenter et comparer différentes méthodes ABC (Approximate Bayesian Computation).
- Explorer l'utilisation des séquences RQMC (Sobol, Halton).
- Bonus : extension à un cadre multivarié.

- Implémentation du générateur de Chambers-Mallows-Stuck (CMS).
- Permet d'obtenir des lois $S(\alpha, \beta, \gamma, \delta)$ à partir de deux variables :
 - $U \sim \mathcal{U}(-\pi/2, \pi/2)$
 - $W \sim \text{Exp}(1)$
- Formule exacte selon *CMS (1976)*.
- Démonstration mathématique fournie dans le rapport.

- Idée : remplacer les tirages pseudo-aléatoires par des séquences de faible discrédance (Sobol, Halton).
- Sobol et Halton sont des séquences quasi-aléatoires à faible discrédance, conçues pour couvrir uniformément l'espace d'intégration, améliorant ainsi la convergence par rapport au tirage aléatoire classique.
- Transformation des points $[0, 1]^2$:
 - $U = -\pi/2 + \pi u_1$
 - $W = -\log(u_2)$
- Comparaison des estimations $\mathbb{E}[f(X)]$ entre MC et RQMC.

Comparison de Sobol, Halton et PRNG

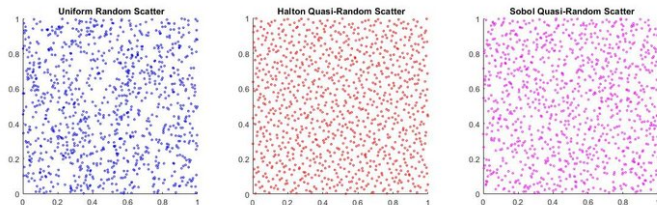


Figure 1 – Comparaison visuelle des distributions de points : Sobol, Halton et pseudo-aléatoire

- Utilisation de la symétrie : $f(X)$ et $f(-X)$ ont même loi.
- Fonction test utilisée : $f(x) = \frac{1}{1+x^2}$.
- Nouvelle estimation : $\hat{\mu}_{\text{ANTI}} = \frac{1}{2N} \sum (f(X_i) + f(-X_i))$.
- Pour évaluer la vitesse de convergence des méthodes :
 - On génère une grande quantité d'observations avec `generate_ABC` pour obtenir une référence précise.
 - Cette estimation est considérée comme la "vraie" valeur de l'espérance.
 - Pour des tailles d'échantillons plus petites, on applique les différentes méthodes (MC, RQMC, RQMC+Antithétique).
 - L'erreur est calculée comme l'écart absolu entre l'estimation et la valeur de référence.

Erreur ABC-MCMC en fonction de ε

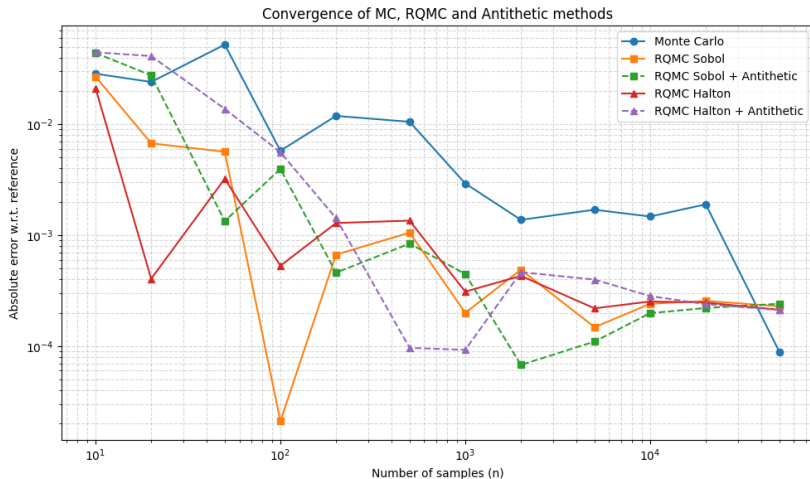


Figure 2 – Comparaison visuelle des distributions de points : Sobol, Halton et pseudo-aléatoire

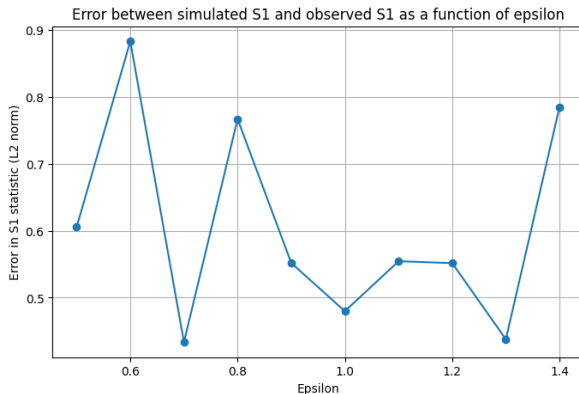
- Basée sur les quantiles (McCulloch, 1986).
- Définition : $S_1(x) = (v_\alpha, v_\beta, \text{IQR}, \text{mean})$, avec :
 - $v_\alpha(x) = \frac{Q_{0.95}(x) - Q_{0.05}(x)}{Q_{0.75}(x) - Q_{0.25}(x)}$ (dispersion)
 - $v_\beta(x) = \frac{(Q_{0.95}(x) + Q_{0.05}(x)) - 2Q_{0.5}(x)}{Q_{0.95}(x) - Q_{0.05}(x)}$ (asymétrie)
 - $\text{IQR}(x) = Q_{0.75}(x) - Q_{0.25}(x)$ (étendue interquartile)
 - $\text{mean}(x) = \frac{1}{n} \sum_{i=1}^n x_i$ (moyenne empirique)
- Utilisée comme statistique résumée commune dans toutes les méthodes ABC.

Approximate Bayesian Computation (ABC) : Processus général

- Objectif : approcher la distribution a posteriori $\pi(\theta \mid y_{\text{obs}})$ sans calcul explicite de vraisemblance.
- Processus général :
 - **Prior** : échantillonner θ depuis une distribution a priori $\pi(\theta)$.
 - **Simulation** : générer des données simulées y_{sim} à partir de θ .
 - **Résumé** : calculer une statistique résumée $S(y_{\text{sim}})$.
 - **Distance** : mesurer la distance entre $S(y_{\text{sim}})$ et $S(y_{\text{obs}})$.
 - **Acceptation** : accepter θ si la distance est inférieure à un seuil ϵ .
- Plus ϵ est petit, plus l'approximation est précise, mais le taux d'acceptation diminue.

ABC Rejection

- Simule des paramètres à partir de la loi a priori.
- Accepte si $\|S(x) - S(x_{\text{obs}})\| < \varepsilon$.
- Simple à mettre en oeuvre mais inefficace pour petit ε .



- Algorithme basé sur Metropolis-Hastings, adapté pour l'ABC (sans calcul de vraisemblance explicite).
- **Proposition** : À chaque itération, on propose un nouveau paramètre θ^* à partir d'une loi de transition $q(\theta^*|\theta)$.
- **Acceptation** : Si la simulation x^* vérifie $\|S(x^*) - S(x_{\text{obs}})\| < \varepsilon$, alors θ^* est accepté avec une probabilité d'acceptation :

$$h = \min \left(1, \frac{q(\theta | \theta^*)}{q(\theta^* | \theta)} \right)$$

sinon on conserve θ .

- Permet d'explorer efficacement l'espace des paramètres, même pour petits ε .

Estimation de ε qui permet d'avoir la plus petite erreur

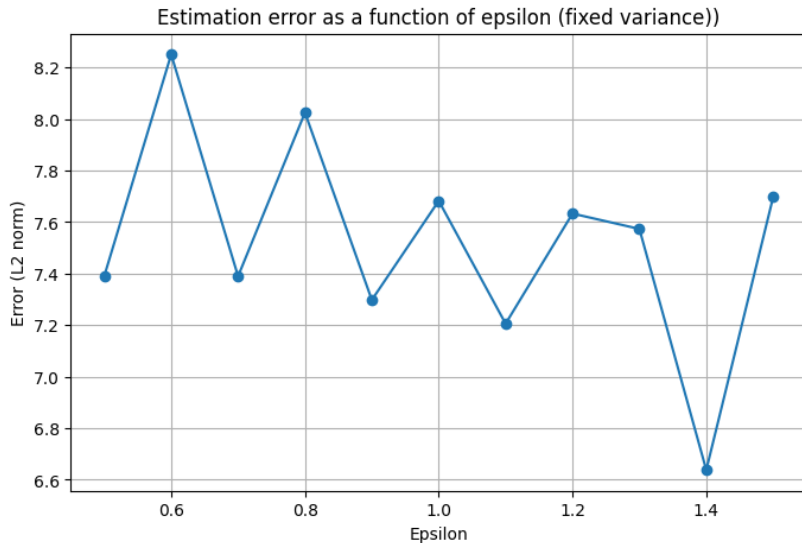
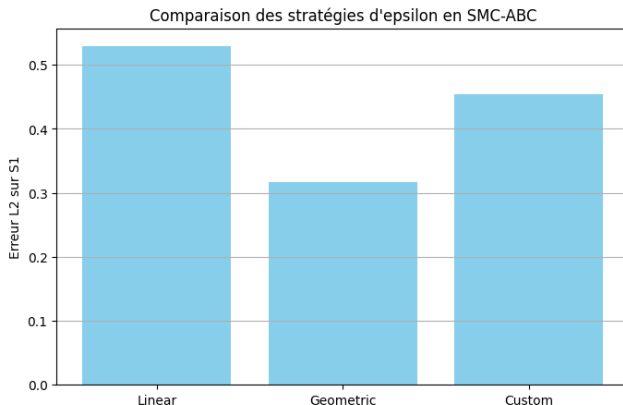


Figure 3 – Erreur d'estimation en fonction de

- Utilise une **population de particules** pour approximer la loi postérieure.
- À chaque itération :
 - On **réduit progressivement** le seuil ε (sélection plus stricte).
 - Les particules sont **mutées** (perturbées) pour explorer l'espace des paramètres.
 - Chaque particule reçoit un **poids adaptatif** selon sa distance aux données observées.
- Si l'**ESS** (*Effective Sample Size*) devient trop faible, on **rééchantillonne** les particules pour éviter la dégénérescence.
- Permet un **compromis efficace** entre précision (petit ε) et coût de calcul.

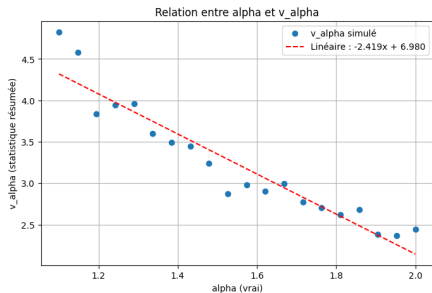
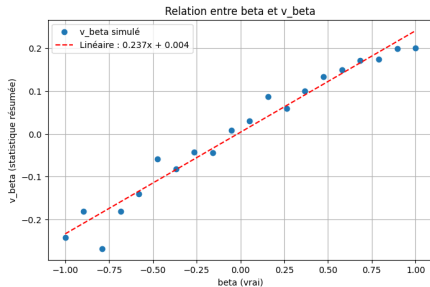
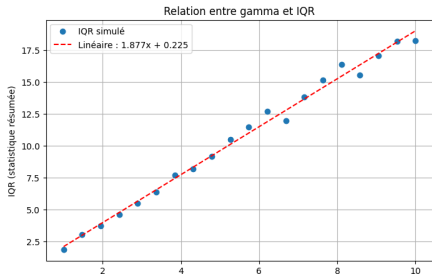
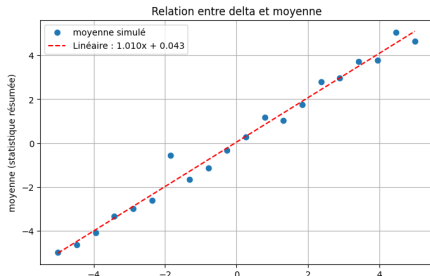
Choix des ε dans SMC-ABC

- L'algorithme SMC-ABC dépend du **choix de la séquence des seuils** ε .
- Ici, on compare plusieurs stratégies pour faire décroître ε (linéaire, géométrique, personnalisée).



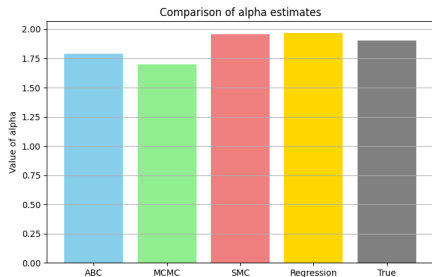
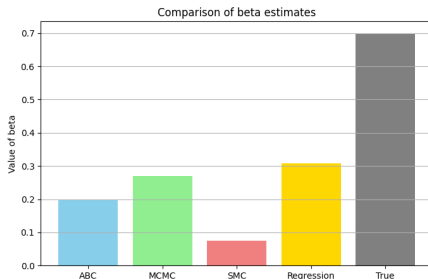
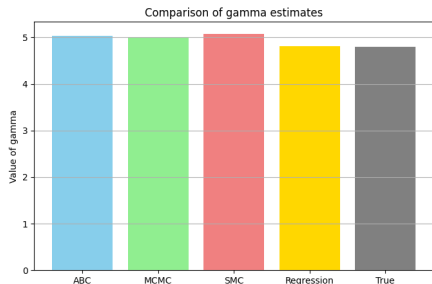
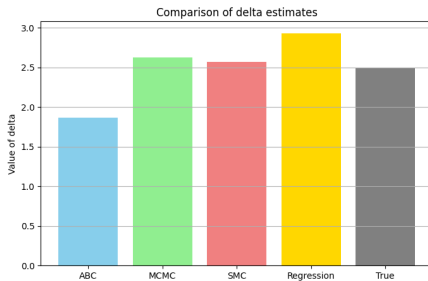
- **Objectif** : corriger les biais dus à un grand ε en ajustant les paramètres simulés.
- **Méthode** :
 - Simuler des couples $(\theta, S(\theta))$.
 - Effectuer une régression locale pondérée de θ sur $S(\theta)$.
 - Correction : $\theta^* = \theta - \hat{\beta}(S(\theta) - S(y_{\text{obs}}))$.
- **Hypothèse clé** : il faut que la relation entre θ et $S(\theta)$ soit **approximativement linéaire**.
- **Avantages** : meilleure estimation, réduction du biais, meilleure utilisation des données même avec un ε plus grand.

Vérification de la linéarité entre θ et $S(\theta)$



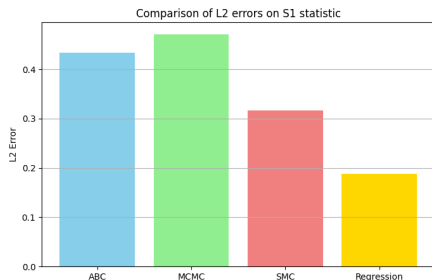
- Analyse des erreurs sur S_1 .
- Efficacité (temps de calcul, nombre de simulations).
- Valeur finale de ε utilisée.
- Plots des différents $\theta = (\alpha, \beta, \gamma, \delta)$.

Plot des paramètres finaux en comparaison aux réels

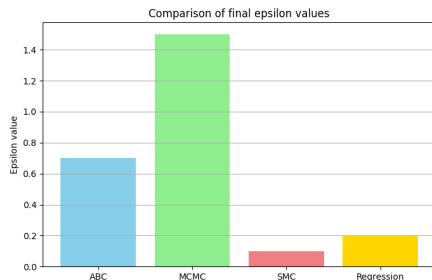


Comparaison des erreurs et des ε finaux

- Le paramètre ε joue un rôle crucial dans les méthodes ABC : il contrôle la précision de l'approximation.
- Un ε plus petit signifie que seules les simulations très proches des données observées sont acceptées.
- Cela améliore la précision, mais augmente fortement le **coût computationnel**, car beaucoup de tirages sont rejetés.



Erreurs finales (L2 sur S_1)



Valeurs finales de ε

- Génération de lois alpha-stables en dimension d via projection sphérique.
- Représentation spectrale : combinaison linéaire de vecteurs unitaires.
- Résultat : Y suit une loi alpha-stable multivariée exacte.

- Adaptation du schéma SMC-ABC au cas multivarié.
- Statistique résumée S_{multi} calculée par variable.
- Visualisation des directions S estimées vs. vraies via PCA.

- SMC-ABC (avec epsilon géométrique) est le meilleur compromis.
- RQMC (Sobol, Halton) réduit la variance à faible coût.
- Extension multivariée valide les méthodes dans des contextes plus complexes.

Merci pour votre attention !