

Master professionnel II : Ingénierie mathématique : Option Statistique

Statistique Bayésienne.

Anne PHILIPPE
Université de Nantes
Laboratoire de Mathématiques Jean Leray

Fiche 7. Méthodes de Monte Carlo

EXERCICE 1. CONTRÔLE ET VISUALISATION DE LA CONVERGENCE DES ESTIMATEURS DE MONTE CARLO

On observe $k = 10$ variables aléatoires iid X_1, \dots, X_k sur $\{0, 1\}$. On note $\sum X_i = 7$. On veut estimer

$$\theta = \log(p/(1-p))$$

où $p = P(X_1 = 1)$. On choisit comme loi a priori sur la paramètre p la loi beta $(1/2, 1/2)$.

1) Ecrire l'estimateur de Bayes sous la forme

$$\hat{\theta} = \int_0^1 h(\theta)f(\theta) d\theta$$

avec $h \in L^2(f)$ et f une densité de probabilité.

- 2) Décrire la mise en oeuvre de l'estimateur de Monte Carlo de $\hat{\theta}$. On le note I_n où n est le nombre de variables simulées.
- 3) Représenter l'estimateur de Monte Carlo en fonction de n .
- 4) Sur la trajectoire simulée à la question précédente, évaluer la variance de l'estimateur de Monte Carlo.
- 5) En déduire un intervalle de confiance de niveau asymptotique $1 - \alpha = 95\%$
- 6) Superposer l'estimateur et ses régions de confiance.

Méthode alternative : on estime la variance et les régions de confiance par une méthode non asymptotique basée sur la simulation. Pour tout n , on simule un échantillon de taille N de même loi que I_n . La variance, les quantiles, etc de la loi de I_n sont alors estimés à partir de cet échantillon.

Commande R .**La démarche**

- On construit une matrice n lignes et N colonnes de nombres aléatoires distribués suivant la loi de densité f . On note A cette matrice
- On calcule la suite des estimateurs $\{I_k, k = 1, \dots, n\}$ sur les N colonnes

```
h = fonction(x) ... à définir ...
cummean = fonction(x) cumsum(x)/(1:length(x))
B = apply(h(A),2,cummean)
```

La j ème ligne de la matrice B contient un échantillon suivant la loi de l'estimateur de Monte Carlo I_j .

Evaluation de la variance de I_n

- Calculer la variance de chacune des lignes


```
V = apply(B,1,var)
```

 pour obtenir une estimation de la variance de $I_k, k = 1 \dots, n$

Régions de confiance

- Estimer les quantiles d'ordre $\alpha/2$ et $1 - \alpha/2$ de la loi de I_n . Utiliser la fonction **quantile** et **apply** sur les lignes de la matrice B .
- Représenter sur un même graphique l'estimateur de Monte Carlo et sa région de confiance en fonction de la taille de l'échantillon

- 7) Représenter l'estimation de la variance en fonction de la taille de l'échantillon
- 8) Représenter sur un même graphique l'estimateur I_n et les intervalles de confiance en fonction de n
- 9) Comparer les deux méthodes mises en oeuvre.

EXERCICE 2. ALGORITHME SIR : CHOIX DE m EN FONCTION DE n

On cherche à simuler par la méthode SIR un échantillon de taille n suivant la loi Gamma $\Gamma(1.5, 1)$ à partir d'un échantillon de taille m suivant la loi exponentielle de paramètre 1

1. Programmer l'algorithme SIR.

Indication : utiliser la fonction **sample** pour l'étape de rééchantillonnage.

On cherche à calibrer m en fonction de n .

2. Pour évaluer les performances de cet algorithme, on teste l'ajustement à la loi Gamma $\Gamma(1.5, 1)$ par le test de Komogorov de niveau 5% pour $N = 500$ échantillons simulés de façon indépendante.

Comparer la proportion d'échantillons rejetés avec l'erreur de première espèce 5% dans les situations suivantes

n	50	500	1000
-----	----	-----	------

et

m/n	5	10	50	100
-------	---	----	----	-----

3. Pourrait on remplacer l'algorithme SIR par l'algorithme d'acceptation rejet ?

EXERCICE 3.

Soit X_1, \dots, X_k iid suivant une loi gaussienne $\mathcal{N}(\theta, 1)$. Le choix de la loi a priori sur θ est la loi de Cauchy

$$\pi(\theta) = \frac{1}{\pi} \frac{1}{1 + \theta^2}$$

- 1) Tester l'hypothèse : "la distribution des observations est gaussienne" sur les données suivantes
<http://www.math.sciences.univ-nantes.fr/~philippe/lecture/data-gauss>
puis tester l'hypothèse
 H_0 La variance est égale à 1
- 2) Calculer la loi a posteriori.
- 3) Peut on calculer explicitement l'estimateur de Bayes de θ ?
- 4) Proposer et implémenter un algorithme d'acceptation-rejet pour simuler suivant la loi a posteriori. Indication : prendre la loi gaussienne de moyenne \bar{X}_k et de variance $1/k$ comme loi instrumentale
- 5) A partir d'échantillons simulés via l'algorithme d'acceptation-rejet
 - Construire une approximation de la densité de la loi a posteriori
 - Donner une approximation de l'estimateur de Bayes. Estimer la précision de votre approximation.
 - Donner une approximation de la région de confiance bayésienne de niveau 95%.
- 6) En utilisant la même loi instrumentale que pour l'algorithme d'acceptation-rejet, mettre en oeuvre un algorithme SIR pour générer des nombres aléatoires suivant la loi a posteriori
- 7) A partir d'échantillons simulés via l'algorithme SIR
 - Construire une approximation de la densité de la loi a posteriori
 - Donner une approximation de l'estimateur de Bayes. Estimer la précision de votre approximation.
 - Donner une approximation de la région de confiance bayésienne de niveau 95%.
- 8) Comparer les résultats obtenus et conclure.