DURET Guillaume 5ETI majeure IMI

BRALET Antoine

Compte Rendu du TP d’Estimation

Inférence Bayesienne et Monte-Carlo Markov Chain

Année 2020-2021

I. Introduction

Ce rapport vise à rapporter les résultats des expérimentations menées en Travaux Pratique (TP) le 7 octobre 2020. L’objectif de ce TP était de mettre en place l’algorithme de Metropolis-Hastings à l’aide d’une chaine de Monte-Carlo Markov permettant de simuler une marche aléatoire. En effet, en considérant une loi cible dont un paramètre est inconnu, on cherche ici à trouver une série identiquement distribuée selon cette loi cible. Ne sachant pas mettre en place une telle série, nous utilisons alors une loi dite de proposition dont une série identiquement distribuée est facile à obtenir et telle que pour un donné, . Ainsi puisque la fonction englobe complètement la fonction , si l’on considère la série issue de et identiquement distribuée, n’importe quelle sous-série est également identiquement distribuée dans la zone quelle occupe. C’est pourquoi si l’on ne considère que la sous-série dont les valeurs sont comprises dans la zone occupée par , cette sous-série peut être considérée comme identiquement distribuée. La figure suivante permet d’illustrer ces propos :

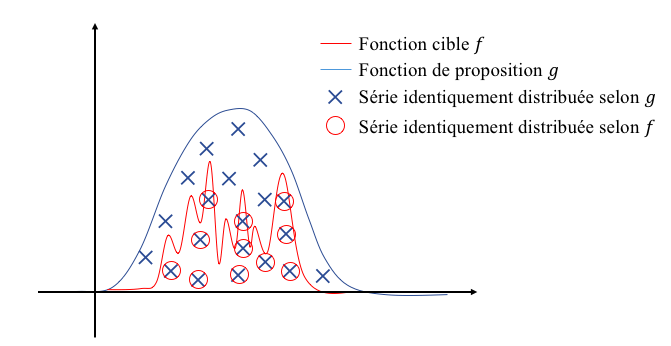


Illustration de la création d’une série identiquement distribuée selon

Afin de créer cette série intelligemment, nous mettons en place une marche aléatoire de type Monte-Carlo Markov Chain (MCMC) qui consiste à prendre en compte la position de la série au temps afin de pouvoir déterminer sa position au temps . Ainsi si notre point se trouve bien dans la zone couverte par , le prochain point créé le sera dans ses environs. Si au contraire, il ne s’y trouve pas, alors on cherchera à faire un plus grand pas de sorte à revenir dans a zone couverte par . De plus, l’algorithme de Metropolis-Hastings utilise un rapport d’acceptation rejet permettant de sélectionner ou non si un point est conservé pour la série identiquement distribuée selon . Ce terme permet de ne pas rester bloquer dans des maxima locaux de la fonction mais de bien couvrir l’ensemble de cette fonction.

II. Commentaires et explications algorithmiques

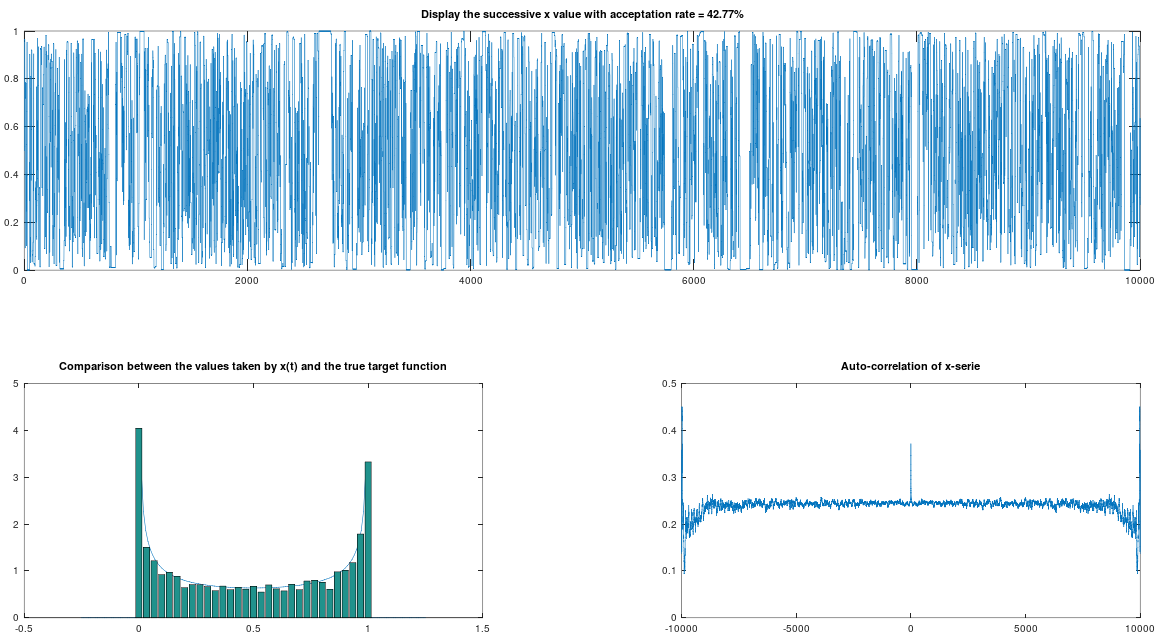
L’algorithme mis en place lors de ce TP se trouve essentiellement dans la fonction MetroHast.m. Néanmoins, pour pouvoir exécuter l’algorithme il faut choisir les paramètres dans le fichier mainEstimation.m et exécuter ce dernier.

Dans un premier nous considérons la loi BETA de paramètre a = 0.5 et b = 0.5 comme étant la loi cible et une loi normale centrée réduite comme loi de proposition. Pour obtenir les résultats de cette expérience, il est important que la variable « loiProp » soit placée à 0.

Il est notable que lors du calcul du rapport d’acceptation-rejet :

en considérant la valeur de la série au temps et la valeur « candidate » pour la série au temps , le calcul de car nous mettons en place une marche aléatoire. De plus sachant que la fonction est symétrique, on a : .

Ainsi, en exécutant l’algorithme avec les paramètres initiaux (« sigq » = 0.5 et « T » = 10000), il est possible d’observer les figures suivantes :

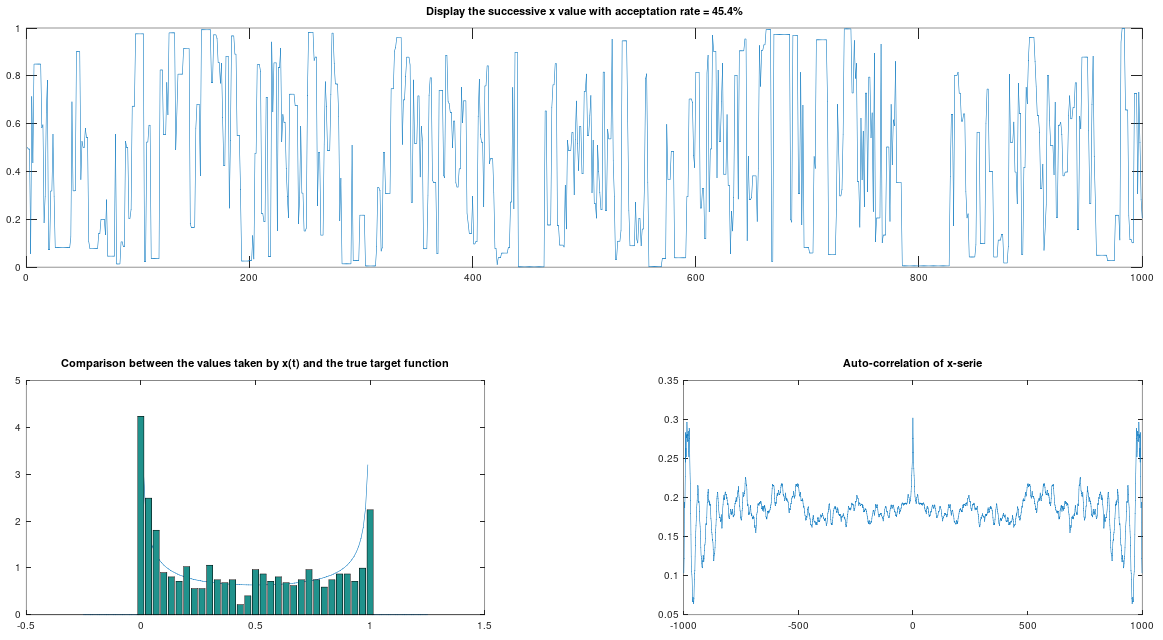


Il est possible ici de vérifier sur la figure en bas à gauche que la série créée permet bien de se rapprocher de la fonction cible. Il est également notable que le taux d’acceptation-rejet est de 42,77% ce qui signifie que l’on conserve autant de valeurs candidates que l’on en rejette. Cette balance est importante car si l’on acceptait trop de valeurs, dans ce cas, on suivrait la loi de proposition de beaucoup trop près et l’on ne se rapprocherait pas suffisamment de la loi cible. Au contraire, si l’on rejetait trop de valeurs, on risquerait d’être bloqué dans un maximum local de la fonction cible et l’on ne pourrait pas étudier l’ensemble de la fonction cible. Cette balance est donc particulièrement importante afin d’avoir une étude raisonnable et fiable de la fonction cible. Ce phénomène peut également être étudié à l’aide de la fonction d’autocorrélation de la série . On peut noter ici (en bas à gauche de la figure) que cette fonction d’autocorrélation connait un pic significatif en 0, ce qui signifie que la série n’est pas du tout corrélée et donc que ses valeurs sont indépendantes les unes des autres. Cette indépendance est particulièrement importante car cela signifie que notre série est bien identiquement distribuée et que son évolution ne dépend en aucun cas de ses valeurs précédentes, phénomène qui permet d’étudier l’ensemble de la fonction cible et non pas seulement une partie de celle-ci. Il est important de rappeler ici que seule la partie centrale de la fonction d’autocorrélation doit être prise en compte car, l’autocorrélation étant un calcul d’intégrale de la superposition entre la série et sa translation temporelle, les premières valeurs ne contiennent que très peu d’échantillons et sont particulièrement bruitées.

III. Expérimentations et variation de paramètres

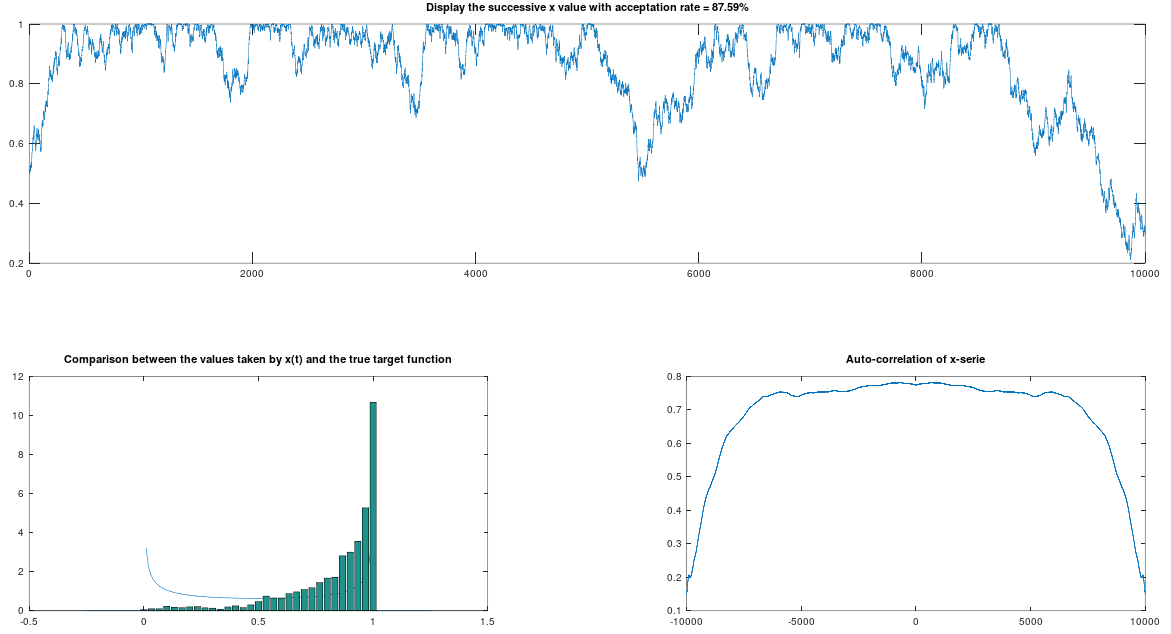
1. Modification du nombre d’échantillons

La modification de nombre d’échantillons dans la série influence directement les résultats obtenus ainsi que la précision de l’histogramme final des valeurs de la série comme en témoigne la figure ci-dessous où la variable « T » vaut 1000. On peut y constater aisément que la série a passé du temps sur le pic de gauche et n’a eu que très peu de « temps » pour étudier le pic de droite de la fonction cible. En contrepartie, plus le nombre d’échantillons est grand, plus l’algorithme sera précis mais sera lent. Le compromis temps de calcul contre précision est toujours important à rechercher.

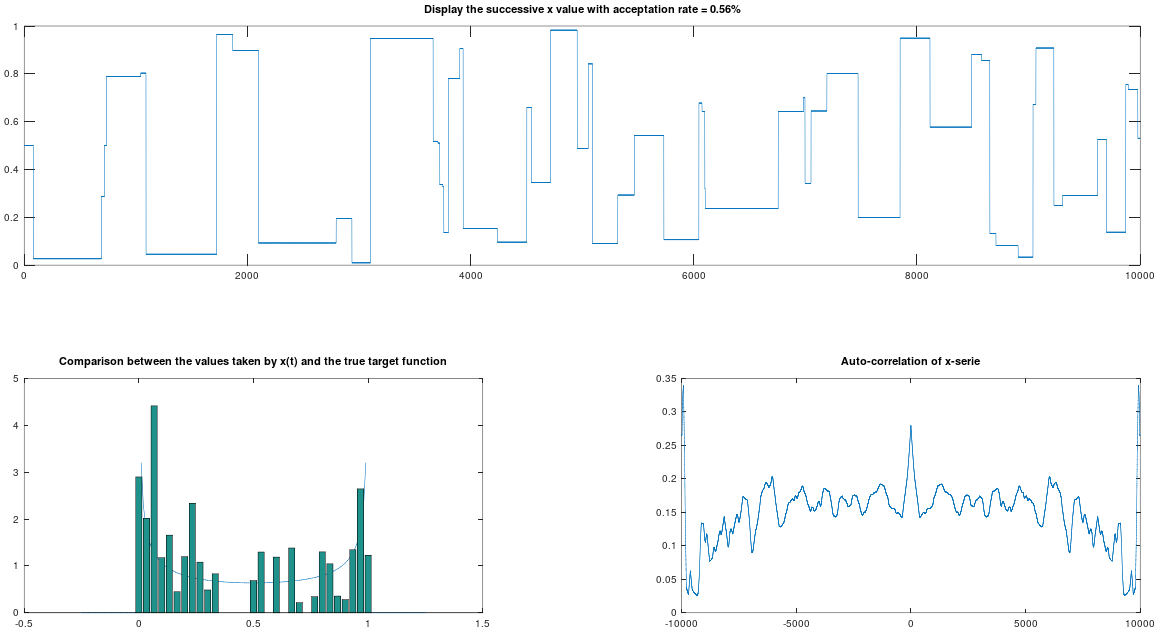


2. Modification du pas

Afin d’accélérer le processus tout en gardant une bonne précision, il est également possible de modifier un autre paramètre : le pas d’avancée. Mais une fois de plus des précautions sont à prendre avec ce paramètre « sigq » dans l’algorithme. En effet, si celui-ci est trop petit (0.001 sur la figure ci-dessous), alors on peut observer que l’algorithme aura tendance à rester toujours très proche d’un temps à un temps ce qui ne permettra pas d’étudier l’ensemble de la fonction cible même avec un nombre d’échantillons très grand, ici 10000. On peut alors de plus constater que la fonction d’autocorrélation est beaucoup plus stable et statue donc sur une forte corrélation et donc dépendance d’un échantillon à l’autre de la série. En témoigne également le taux d’acceptation-rejet qui est particulièrement fort : 87.59%, ce qui signifie que la majeure partie des échantillons candidats sont acceptés et sont très proches les uns des autres (à cause du pas de petite valeur).



Au contraire si l’on choisit un pas trop grand (50 sur la figure ci-dessous), l’on observe un tout autre phénomène : cette fois ci le taux d’acceptation rejet est de 0.56% ce qui signifie qu’une très grande majorité des échantillons candidats ont été rejetés et donc on observe une série très crénelée à cause de ces rejets. Par conséquent, l’autocorrélation est elle aussi beaucoup moins prononcée en 0 ce qui s’explique par cette succession d’échantillons identiques. Aussi, l’histogramme final est particulièrement perturbé car les valeurs les plus répétées ne sont pas nécessairement issues du bon fonctionnement de l’algorithme mais de la forte répétition des échantillons à cause d’un pas beaucoup trop grand.



IV. Modification de loi cible et de proposition

V. Conclusion