Probabilité d'évènements rares et marches aléatoires conditionnées

Neuvième Colloque Jeunes Probabilistes et Statisticiens, au Mont-Dore

Virgile Caron

Directeur de thèse : Michel Broniatowski

Mai 2010

Importance Sampling

Estimateur naïf

Soit **Z** une v.a. sur $\mathbb{R}$ de loi P et de densité p. Soit $A \subset \mathbb{R}$ tel que $P(A) > 0$. On veut estimer $P(A)$ avec des Z_i copies i.i.d. de **Z** sous p.

$$P_L := \frac{1}{L}\sum_{l=1}^{L} \mathbf{1}_A(Z_l)$$

Par la LGN, cet estimateur converge vers $P(A)$. De plus, il est sans biais.

Estimateur standard

On peut proposer une autre famille d'estimateurs sans biais. Les Y_i sont des copies i.i.d. d'une v.a. **Y** de densité g. Si $supp(p) \subset supp(g)$, Alors

$$P_L^g := \frac{1}{L}\sum_{l=1}^{L} \frac{p(Y_l)}{g(Y_l)}\mathbf{1}_A(Y_l) \to \int \frac{p(x)}{g(x)}\mathbf{1}_A(x)g(x)dx = P(A)$$

Importance Sampling

Choix optimal de g

Le choix optimal de g est $p_{\mathbf{Z}/A}$, la densité de $\mathbf{Z}$ conditionnée par$(\mathbf{Z} \in A)$. Ce choix est impossible sinon $L = 1$ et on obtient un estimateur parfait. Le choix optimal impose de connaitre la réponse à la question posée. On aimerait alors approximer la densité conditionnelle $p_{\mathbf{Z}/A}$.

Cas considéré

On suppose la loi de $\mathbf{X}_1$ de densité $p_{\mathbf{X}}$ sur $\mathbb{R}$ connue tel que les $\mathbf{X}_i$ sont i.i.d., $E(\mathbf{X}_1) = 0$ et $Var\mathbf{X}_1 = 1$. On s'interesse à des conditionnements par $(\mathbf{Z} \in A)$ tel que $\mathbf{Z}$ est une moyenne empirique de n termes avec n fixe :

$$\mathbf{Z} := \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i =: \frac{1}{n} \mathbf{S}_1^n$$

Et on prefere estimer directement à la loi de $((\mathbf{X}_1, ..., \mathbf{X}_n) | \mathbf{Z} \in A)$. On ne peut pas le faire jusqu'à n mais jusqu'à k proche de n.

Importance Sampling

Application

On considère, dans la suite, des ensembles A, éventuellement indexé par n, tel que $\mathbf{Z} \in A$ est un évènement de moyennes(MD) ou de grandes déviations(GD). Ainsi,

$$A := (a_n, \infty)$$

où, en MD, $a_n \to E(\mathbf{X}_1)$ lentement par au dessus tel que

$$a_n \to E(\mathbf{X}_1)$$
$$\sqrt{n}a_n \to \infty$$

et, en GD, $a_n = a$ est une constante.

Importance Sampling

Application

Avec les notations introduites, on veut alors estimer pour n grand fixé :

$$P_n := P\left(\frac{1}{n}\mathbf{S}_1^n \in A\right)$$

Si g une densité d' échantillonnage d'un vecteur Y_1^n tel que g soit le plus proche possible de la densité conditionnelle, alors

$$P_g^{(n)}(\varepsilon_n) := \frac{1}{L}\sum_{l=1}^{L} \frac{p_X\left(Y_1^n(l)\right)}{g\left(Y_1^n(l)\right)} \mathbf{1}_{\varepsilon_n}\left(Y_1^n(l)\right)$$

où

$$\varepsilon_n := \{(x_1, \ldots, x_n) \in \mathbb{R}^n : s_1^n/n > a_n\}$$

avec $s_1^n := x_1 + \ldots + x_n$.

Principe de Gibbs

Kullback-Leibler

On définit la Distance de Kullback-Leibler :

$$K(Q,P) := \int \log \frac{dQ}{dP} dQ$$

$$K(A,P) := \inf K(Q,P); Q \in A$$

Notons :

$$\Pi = \arg\inf K(Q,P); Q \in A$$

où A est un ensemble de mesure.

Th. de Sanov

Théorème Sanov Soit $P_n := \frac{1}{n}\sum \delta_{X_i}$ la mesure empirique où les $\mathbf{X}_i$ sont i.i.d. de loi P.

$$P(P_n \in A) \approx \exp -nK(A,P)$$

Principe de Gibbs

Exemple

Soit $A = \{Q \in \mathcal{P} : \int xQ(dx) \geq a\}$. Alors Π définit par le théorème de Sanov existe et on l'appelle loi **tiltée** définie par

$$\frac{d\Pi^a}{dP}(x) = \frac{\exp tx}{\int \exp tx dP(x)}$$

où $\int v\Pi^a(dv) = a$ et t et a sont en bijection. Dans le cas où il existe une densité, elle s'écrit :

$$\pi^a(x) = \frac{\exp tx}{\Phi(t)} p(x)$$

où Φ est la fonction génératrice des moments.

Principe de Gibbs

Conséquence

On déduit de cette loi conditionnelle le principe de Gibbs pour a fixe,

$$\mathcal{L}\left(X_1 / \frac{1}{n} \sum X_i \geq a\right) \rightarrow \Pi^a$$

Ceci correspond à un évènement de GD. On peut en fabriquer une version MD.

Extension

On recherche une extension en densité de $\mathcal{L}\left(X_1^k | \frac{1}{n} \sum X_i \geq a_n\right)$ avec $k = k_n$ le plus grand possible.

Enoncé du Problème

Estimateur

Pour rappel, notre estimateur s'écrit pour des Y_i simulé sous g.

$$P_g^{(n)}(\varepsilon_n) := \frac{1}{L}\sum_{l=1}^{L} \frac{p_X\left(Y_1^n(l)\right)}{g\left(Y_1^n(l)\right)} \mathbf{1}_{\varepsilon_n}\left(Y_1^n(l)\right)$$

Notations

Notons :

$$\mathfrak{P}_n := \mathcal{L}\left(X_1^k / \frac{1}{n}\sum X_i \geq a_n\right) \textit{ et } \ \mathfrak{p}_n \text{ sa densité}$$

On cherche g, une densité sur $\mathbb{R}^k$, proche de $\mathfrak{p}_n$ pour k aussi proche de n que possible, facile à simuler tel que $\mathfrak{p}_n\left(Y_1^k(l)\right) \approx g\left(Y_1^k(l)\right)$ uniquement sur les chemins sous g.

Enoncé du Problème

Etat Actuel

Si k fixe (ou k/n tend vers une constante) et a fixe, il a été démontré (cf. Csiszar[1], Dembo/Zeitouni[2], Diaconis/Freedeman[3]) que $\mathcal{L}\left(X_1^k | \frac{1}{n}\sum X_i \geq a_n\right)$ tends vers la densité produit.

Etat Actuel

Il n'existe pas de résultats en ce sens pour k/n tends vers 1, en GD, en MD, dans $\mathbb{R}^k$ et pas si le conditionnement est $\frac{1}{n}\sum f(X_i) \geq a_n$.

Objectif

Nous voulons maintenant déterminer :

$$p(\mathbf{X}_1^k = Y_1^k | \mathbf{S}_1^n = na_n)$$

Thérorème Principal

Conditions

$$\lim_{n\to\infty} \frac{k}{n} = 1$$

$$\lim_{n\to\infty} n - k = \infty$$

Conditions

On suppose que X_1 est centrée et de variance 1 et satisfait la condition de Cramer, i.e. X_1 a une fonction génératrice des moments finie $\Phi(t) := E \exp tX_1$ dans un voisinage non-vide de 0.

Thérorème Principal

Conditions

Soit ε_n une suite positive qui satisfait

$$\lim_{n\to\infty} \varepsilon_n (\log n)^2 = 0$$

$$\lim_{n\to\infty} \varepsilon_n \sqrt{n-k} = \infty$$

$$\lim_{n\to\infty} \frac{(a_n - E(\mathbf{X}_1))^2}{\varepsilon_n(\log n)} = 0$$

Thérorème Principal

Notations

On définit une densité g_a de façon itérative :

$$g_a(y_1^k) := \prod_{i=0}^{k-1} g_i(y_{i+1}/y_1^i)$$

où chacun des termes approxime $p(\mathbf{X}_{i+1}|\mathbf{X}_1^i, \frac{1}{n}\sum X_i = a_n)$. La densité g_a s'écrit sous la forme

$$g_i(y_{i+1}/x_1^i) = C_i p_{\mathbf{X}}(y_{i+1})\mathfrak{n}(\alpha\beta + a, \alpha, y_{i+1})$$

où $\mathfrak{n}(\mu, \tau, x)$ est la densité normal de moyenne μ et de variance τ en x. Les deux paramètres α et β s'expriment par rapport aux $\mathbf{X}_1^i$ précédent et à la loi de départ. On remarque que g_a est une généralisation de la densité tiltée avec des termes de second ordre.

Thérorème Principal

Theorem

On suppose satisfaite les conditions ci-dessus. Soit Y_1^n un échantillon de loi $\mathfrak{P}_n$. Alors

$$\mathfrak{p}_n\left(Y_1^k\right) := p(\mathbf{X}_1^k = Y_1^k / \mathbf{S}_1^n = na) = g_a(Y_1^k)(1 + O_{\mathfrak{P}_n}(\varepsilon_n (\log n)^2)).$$

Lemme d'inversion

Lemme d'inversion

Soient $\mathfrak{R}_n$ et $\mathfrak{S}_n$ deux mesures de probabilités sur $\mathbb{R}^n$ de densités respectives $\mathfrak{r}_n$ et $\mathfrak{s}_n$.

Lemma

On suppose que pour une suite ε_n qui tend vers 0 quand n tends vers l'infini, on a :

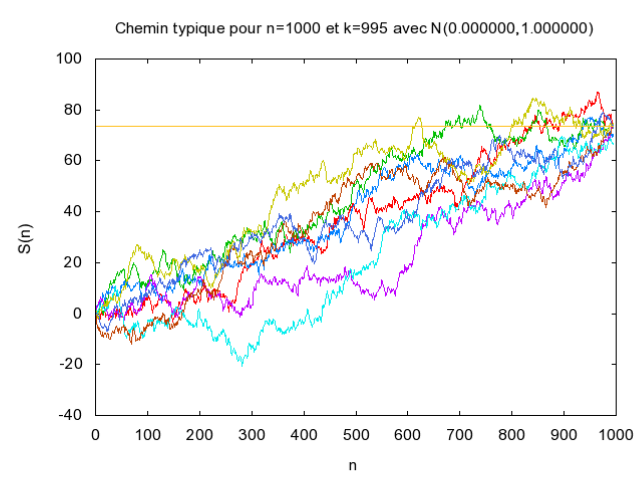
$$\mathfrak{r}_n\left(Y_1^n\right) = \mathfrak{s}_n\left(Y_1^n\right)\left(1 + o_{\mathfrak{R}_n}(\varepsilon_n)\right)$$

quand n tends vers ∞. Alors

$$\mathfrak{s}_n\left(Y_1^n\right) = \mathfrak{r}_n\left(Y_1^n\right)\left(1 + o_{\mathfrak{S}_n}(\varepsilon_n)\right)$$

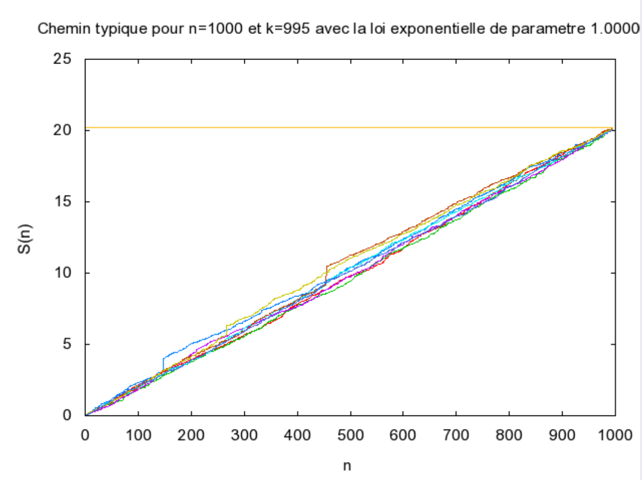
Graphes

Trajectoires



Graphes

Trajectoires



Choix de k

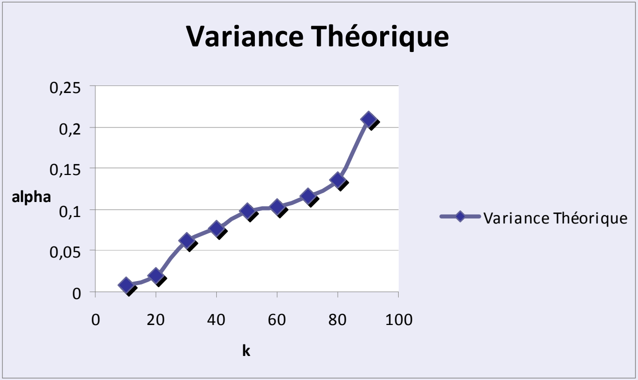
Choix de k

Le paramètre clé dans cette approximation est k qui représente la longueur maximale de la sous-trajectoire pour laquelle l'approximation est bonne à $\alpha\%$ près. Afin de déterminer la valeur maximal de k possible, on calcule :

$$RE(k) := \sqrt{Var_{g_a} \frac{\mathfrak{p}_n\left(Y_1^k\right) - g_a\left(Y_1^k\right)}{g_a\left(Y_1^k\right)}}.$$

Graphe

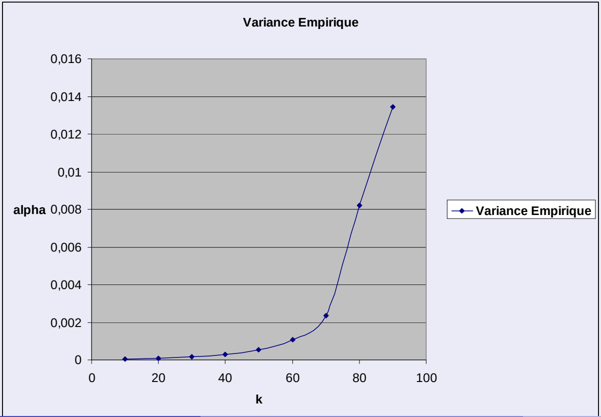
Variance Théorique



Exemple

Variance Empirique

Pour la loi exponentielle de paramètre 1, le calcul de loi conditionnelle est exacte. On peut donc calculer la même quantité que précedemment.



Integration

Integration

Soit $\mathbf{T} := \mathbf{S}_1^n / n$ distribué sous $\mathcal{E}_n$. Ainsi, pour tout ensemble A,

$$P(\mathbf{T} \in A) = \mathfrak{P}_n\left(\frac{\mathbf{S}_1^n}{n}\right)$$

La masse de la loi de $\mathbf{T}$ est concentrée en un petit voisinage de a_n.

Integration

Si Y_1^k est un vecteur aléatoire généré sous $\mathfrak{p}_n$ et sous des hypothèses générales, alors on obtient :

$$\mathfrak{p}_n(Y_1^k) = \overline{g_n}(Y_1^k)\left(1 + o_{\mathfrak{P}_n}(a_n(\log n)^{2+\delta})\right)$$

Propriété de l'IS

Calcul de la Variance

La variance d'un estimateur d'IS de P_n où l'échantillonage est effectué sous la densité g s'écrit

$$VarP_g^{(n)}(\mathcal{E}) = \frac{1}{L}\left(E_g\left(P_g(l)\right)^2 - P_n^2\right)$$

avec

$$P_g(l) := \frac{p\left(Y_1^n(l)\right)}{g\left(Y_1^n(l)\right)} \mathbf{1}_{\mathcal{E}_n}\left(\Sigma_1^n(l)\right)$$

Réduction de la Variance IS Classique

L'IS classique est défini par L simulations d'un échantillon de taille n i.i.d. $X_1^n(j)$, $1 \leq j \leq L$, sous la densité tiltée π^{a_n} de façon non adaptative. L'erreur relative de l'estimateur P_n^{IS} est donnée par

$$RE(\overline{P_n}) := \frac{Var\overline{P_n}}{P_n^2} = \frac{\sqrt{2\pi}\sqrt{n}}{L} a_n(1 + o(1))$$

Propriété de l'IS

Réduction de la Variance

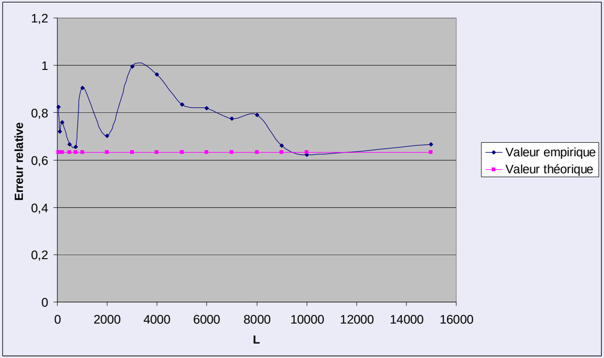
On prouve que notre estimateur IS fourni une nette amélioration par rapport au schéma classique. En effet, dans notre cas, nous obtenons pour l'erreur relative de notre estimateur :

$$RE(\widehat{P_n}) = \frac{\sqrt{2\pi}\sqrt{n-k-1}}{L} a_n(1+o(1))$$

Propriété de l'IS

Réduction de la Variance : Empirique

Sur ce graphe, on a tracé le rapport des deux variances empiriques en fonction de k. On retrouve bien $\sqrt{n-k}/\sqrt{n}$, ce qui montre un net gain. En termes pratiques, c'est une réduction du nombre d'itérations.



Difficultés rencontrées

Inversion de m

On est amené à résoudre un grand nombre de fois :

$$\frac{d}{dt}\log\Phi(t) = c$$

Simulation

La simulation selon N(a,b,x)p(x) peut être compliquer.

Développements Actuels et Futurs

Développements Actuels

Conditionnement par $\frac{1}{n}\sum f(X_i) \geq a_n$, $\frac{1}{n}\sum \alpha_i X_i \geq a_n$.
Etude avec des X_i dans $\mathbb{R}^d$.

Développements Futurs

Recherche des maximas d'une fonction par des méthodes qui généralisent le recuit simulé.

Biblio

Biblio