

Table des matières

1	Notions préliminaires	4
1.1	Introduction	4
1.2	Variables aléatoires	5
1.2.1	Variables aléatoires discrètes	5
1.2.2	Variables aléatoires continues	8
1.3	Estimateur et ses propriétés	12
1.3.1	Définition d'un estimateur	13
1.3.2	Propriétés d'un estimateur	14
1.3.3	Estimateur optimal	17
1.4	La loi forte des grands nombres	18
1.5	Théorème central limite	19
1.6	Génération d'échantillons suivant différentes lois de probabilités	20
1.6.1	Généralité	20
1.6.2	La méthode d'inversion	20
1.6.3	La méthode de rejet	21
1.6.4	La méthode de composition	21
2	Echantillonnage descriptif	22
2.1	Introduction	22
2.2	Description de la méthode	22
2.2.1	La convergence de la méthode	23

2.2.2	Réduction de la variance	26
2.3	Simulation de Monte Carlo	35
2.3.1	Les erreurs d'échantillonnage	36
2.4	Echantillonnage statistique	37
2.5	Echantillonnage aléatoire simple	37
2.6	Echantillonnage descriptif	38
2.6.1	Introduction	38
2.6.2	La procédure d'échantillonnage	38
2.7	Exemple d'utilisation	40
2.8	Conclusion	41

Introduction générale

La simulation est une technique de modélisation du nombre réel, elle permet de représenter le fonctionnement d'un système composé de différents centres d'activités, de mettre en évidence les caractéristiques de ceux-ci et les interactions entre eux, de décrire la circulation des différents objets traités par ces processus, et enfin d'observer le comportement du système dans son ensemble et son évolution dans le temps.

La formulation du phénomène peut s'affectuer aux moyens de modèles mathématiques qui présentent l'avantage de fournir des résultats reflétant une évolution stable du système dans le temps. Les étapes de la simulation sont exécutées dans l'ordre suivant : formulation du problème, modélisation, validation du modèle, programmation, vérification du problème, simulation enfin l'analyse et l'interprétation des résultats.

D'une part, des simulations s'appuient sur un modèle analytique partiel, et d'autre part, bien souvent une étude expérimentale permet de perfectionner un modèle mathématique. Mais aussi la recherche mathématique est vaste et en s'intéresse toujours à développer et améliorer les résultats obtenue.

La simulation de Monte Carlo est une expérience d'échantillonnage qui doit extraire autant d'information possible des variables aléatoires d'entrées qui sont supposées connues. Elle consiste à générer des nombres aléatoires suivant une loi uniforme afin d'obtenir des échantillons des distributions correspondante dans le but d'estimer les paramètres inconnus du système étudié.

Ainsi ce mémoire été réalisé en deux chapitres : dans le premier chapitre nous avons donné des notions préliminaires de la statistique, et qui constituent un bagage pour établir le deuxième chapitre qui aborde les méthodes de Monte Carlo en général et la méthode d'échantillonnage descriptif en particulier avec un exemple d'utilisation.

Chapitre 1

Notions préliminaires

1.1 Introduction

Dans la plupart des phénomènes aléatoires, le résultat d'une épreuve peut se traduire par une grandeur mathématique, très souvent représentée par un nombre entier ou un nombre réel.

La notion mathématique qui représente efficacement ce genre de situation concrète est celle de variable aléatoire.

Ainsi le temps de désintégration d'un atome radioactif, le pourcentage de réponse oui à une question posée dans un sondage ou le nombre d'enfants d'un couple sont des exemples de variables aléatoires.

Définition 1.1 *La population est l'ensemble total d'objets ou de personnes à étudier lors d'une expérience donnée.*

Définition 1.2 *L'échantillon représente la partie de la population sur laquelle porte l'enquête.*

Définition 1.3 *La variable statistique est une caractéristique, définie sur les individus de la population, elle peut être quantitative ou qualitative.*

Définition 1.4 *L'échantillonnage représente l'ensemble des opérations qui ont pour objectif de prélever un certain nombre d'individus dans une population donnée.*

1.2 Variables aléatoires

Soit (Ω, \mathcal{F}, P) un espace probabilisé.

On appelle variable aléatoire sur (Ω, \mathcal{F}, P) toute application :

$$X : \begin{array}{l} \Omega \rightarrow E \\ \omega \mapsto X(\omega) \end{array} . \quad (1.1)$$

Nous distinguons deux types de variables aléatoires : discrètes et continues.

1.2.1 Variables aléatoires discrètes

Une variable aléatoire discrète X est une fonction qui associe à chaque résultat d'une expérience aléatoire un nombre entier naturel K . On note :

$$X : \Omega \rightarrow \mathbb{Z}. \quad (1.2)$$

Loi de probabilité d'une variable aléatoires discrètes

Définition 1.5 *Si X est une variable aléatoire discrète, on donne la loi de probabilité de X en donnant une suite (p_K) où :*

$$p_K = P(X = K). \quad (1.3)$$

Espérance d'une variable aléatoire discrète

En donnant la définition de l'espérance sur les ensembles finis :

Définition 1.6 *L'espérance mathématique d'une variable aléatoire discrète X est la quan-*

tité définit par :

$$E(X) = \sum_{i \in \mathbb{N}} x_i p_i. \quad (1.4)$$

Définition 1.7 1- L'espérance mathématique de X représente ce que X vaut en moyenne, si on répète l'expérience un grand nombre de fois.

2- Il faut remarquer que la somme infinie qui définit l'espérance pourrait fort bien ne pas converger.

Variance d'une variable aléatoire discrète

Définition 1.8 Soit X une variable aléatoire discrète de loi (p_i) , La variance de X se calcule ainsi :

$$Var(X) = \sum_{i \in \mathbb{N}} p_i (x_i - E(X))^2. \quad (1.5)$$

1-La variance de X représente la moyenne des carrés des écarts à l'espérance de X .

2-Elle mesure la tendance de X à la dispersion autour de son espérance, de manière similaire à la variance d'une série statistique.

Théorème 1.1 La variance de X est aussi obtenue par :

$$Var(X) = \sum_{i \in \mathbb{N}} p_i x_i^2 - (E(X))^2. \quad (1.6)$$

Définition 1.9 L'écart-type d'une variable aléatoire X est la racine carrée de sa variance :

$$\sigma(X) = \sqrt{Var(X)}. \quad (1.7)$$

Distributions d'une variable aléatoire discrète

La distribution bernoullide :

Soit $A \in \mathcal{F}$ un événement quelconque (\mathcal{F} tribu de partie de l'ensemble fondamental Ω);

On appelle variable aléatoire indicatrice de A , la variable aléatoire définie par $X = 1_A$:

$$X(w) = 1_A(w) = \begin{cases} 1 & \text{si } w \in A \\ 0 & \text{sin on} \end{cases}$$

Ainsi $X(w) = \{0, 1\}$, il s'agit alors d'une expérience à deux issues seulement (succès, échec)

Avec : $P(X = 1) = p$, $P(X = 0) = 1 - p = q$.

L'espérance mathématique de X est égale à : $E(X) = p$.

La variance mathématique de X est égale à : $Var(X) = pq$.

La distribution binomiale :

Supposons que l'on répète n fois dans les mêmes conditions une expérience aléatoire Bernoulli, dont l'issue se traduit par l'apparition ou la non apparition de A de probabilité p , le résultat d'une expérience étant indépendant des résultats précédent ; soit X le nombre d'apparitions de A parmi ces n expériences, $X(\Omega) = \{0, 1, \dots, n\}$.

Cette loi est sous la forme :

$$P(X = k) = C_n^k p^k (1 - p)^{n-k} \text{ avec } p \in [0, 1] \text{ et } n \text{ entier}$$

L'espérance de cette variable aléatoire est :

$$E(X) = np$$

et la variance :

$$Var(X) = npq$$

La distribution de Poisson :

La loi de Poisson peut être considérée comme une approximation de la loi binomiale, mais elle se justifie par elle-même. C'est la loi d'une variable aléatoire X entière positive ou nulle qui vérifie et satisfait :

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \text{ pour } k \in \mathbb{N}.$$

L'espérance et la variance de cette variable aléatoire est :

$$E(X) = Var(X) = \lambda.$$

La distribution géométrique :

On répète un nombre infini d'expérience de Bernoulli de manière indépendante. Il y'a deux résultats possibles à chaque expérience : réussite(de probabilité p) et on arrête, ou échec(probabilité $q = 1 - p$) et on continue, $X(\Omega) = \{1, \dots, \infty\}$ Alors :

$$P(X = n) = pq^{n-1}, \quad E(X) = \frac{1}{p}, \quad Var(X) = \frac{1}{p^2} - \frac{1}{p}.$$

telle que : X est le nombre de répétition d'expérience aléatoire jusqu'au premier succès.

1.2.2 Variables aléatoires continues

Définition 1.10 Soit un univers Ω , on dit que, une variable aléatoire est continue si l'ensemble de valeurs de X est un intervalle de la probabilité que la variable aléatoire X prenne une valeur donnée de l'intervalle est nulle. On ne peut donc plus définir de loi de probabilité comme pour les variables aléatoires discrètes. On va utiliser dans ce cas la fonction de répartition de la variable aléatoire.

Fonction de répartition d'une variable aléatoire continue

Définition 1.11 La fonction de répartition d'une variable aléatoire continue est l'application F définie sur \mathbb{R} par :

$$F(X) = P(X \leq x), \tag{1.8}$$

cette application est à valeur dans l'intervalle $[0, 1]$, puisqu'une probabilité est un nombre compris entre 0 et 1.

Quelques propriétés de la fonction de répartition

1 : $P(X > x) = 1 - F(x)$ pour tout réel x .

2 : $P(a < X \leq b) = F(b) - F(a)$ pour tout réel a et b telle que $a < b$.

3 : La fonction F est croissante et continue sur \mathbb{R} .

4 : $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Densité de probabilité d'une variable aléatoire continue

Définition 1.12 On appelle densité de probabilité de la variable aléatoire continue X toute fonction f définie continue (sauf éventuellement en un nombre fini de points) et positive sur \mathbb{R} et telle que :

$$\int_{\mathbb{R}} f(x) dx = 1. \quad (1.9)$$

Soit X une variable aléatoire continue de fonction de répartition F alors :

1-pour tout réel x , la fonction f définie sur \mathbb{R} par $f(x) = F'(x)$ est une densité de probabilité appelé densité de probabilité de X .

2-on a pour tout réel x :

$$F(x) = \int_{-\infty}^x f(t) dt$$

Espérance mathématique d'une variable aléatoire continue

Définition 1.13 L'espérance mathématique d'une variable aléatoire continue X est le nombre réel(s'il existe), notée $E(X)$ défini par :

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx. \quad (1.10)$$

Variance mathématique d'une variable aléatoire continue

Définition 1.14 La variance mathématique d'une variable aléatoire continue X est le nombre réel (s'il existe), noté $Var(X)$ défini par :

$$Var(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx = \int_{-\infty}^{+\infty} x^2 f(x) dx - (E(X))^2 \quad (1.11)$$

Définition 1.15 L'écart-type d'une variable aléatoire continue X est la racine carrée de sa variance :

$$\sigma(X) = \sqrt{Var(X)}. \quad (1.12)$$

Distributions d'une variable aléatoire continue

La distribution uniforme sur $[a, b]$:

$$F(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x \geq b \end{cases} . \quad (1.13)$$

La loi de la variable aléatoire est alors :

$$f(x) = F'(x) = \begin{cases} 0 & \text{si } x \leq a \\ 1/b - a & \text{si } a \leq x \leq b \\ 0 & \text{si } x \geq b \end{cases} . \quad (1.14)$$

L'espérance et la variance sont :

$$E(X) = \frac{b+a}{2}, \quad Var(X) = \frac{(b-a)^2}{12}. \quad (1.15)$$

Démonstration :

$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{b+a}{2}$$

$$\begin{aligned}
Var(X) &= E(X^2) - (E(X))^2 \\
E(X^2) &= \int_a^b x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{b^3 - a^3}{3(b-a)} \\
\text{Alors : } Var(X) &= \frac{b^3 - a^3}{3(b-a)} - \frac{(b+a)^2}{4} = \frac{(b-a)^2}{12}
\end{aligned}$$

La distribution exponentielle :

La loi exponentielle est généralement utilisée pour modéliser le temps de vie d'un phénomène notée $X \rightsquigarrow \xi(\lambda)$, $\lambda > 0$.

Sa densité est :

$$f(x) = F'(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases} . \quad (1.16)$$

L'espérance mathématique de la variable aléatoire X est égale à :

$$E(X) = 1/\lambda. \quad (1.17)$$

sa variance égale à :

$$V(X) = 1/\lambda^2. \quad (1.18)$$

La distribution normale :

Un grand nombre de variables quantitatives suivent une loi normale (ou loi de Gauss-Laplace).

Les paramètres qui suffisent à définir mathématiquement cette loi sont la moyenne m et l'écart-type σ de cette variable aléatoire.

On note :

$$X \rightsquigarrow N(m, \sigma).$$

La fonction de densité de probabilité est définie par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}. \quad (1.19)$$

La loi normale joue un rôle fondamental en probabilités et statistique mathématique car c'est un modèle fréquemment utilisé dans plusieurs domaines.

L'espérance mathématique de la variable aléatoire X est égale à :

$$E(X) = m, \tag{1.20}$$

et la variance est égale à :

$$Var(X) = \sigma^2. \tag{1.21}$$

1.3 Estimateur et ses propriétés

L'estimation consiste à attribuer des valeurs approchées aux paramètres d'une population (moyenne, variance, etc...) à l'aide d'un échantillon, vérifiant l'hypothèse d'échantillonnage aléatoire simple, de n observations issues de cette population. Autrement dit, dans un échantillon de taille n , on suppose qu'une série statistique x_1, x_2, \dots, x_n correspond à des réalisations de variables aléatoires X_1, X_2, \dots, X_n , il s'agit de trouver une estimation d'un paramètre inconnu de la population totale à partir de ces n réalisations.

Exemple 1.1 Selon la loi des grands nombres \bar{X} et S^2 sont les estimateurs de la moyenne théorique m et la variance théorique σ^2 respectivement et d'après le théorème de la loi forte des grands nombres $\bar{X}_n \xrightarrow{p.s} m$. Et aussi d'après le théorème de loi forte des grands nombres :

$$\frac{1}{n} \left(\sum_{i=1}^n X_i^2 \right) \xrightarrow{p.s} E(X^2).$$

et :

$$\overline{X_n}^2 \xrightarrow{p.s} E^2(X).$$

Donc :

$$\frac{1}{n} \left(\sum_{i=1}^n X_i^2 \right) - \overline{X_n}^2 \xrightarrow{p.s} E(X^2) - E^2(X).$$

Alors :

$$S_n^2 \xrightarrow{p.s} \delta^2.$$

De même la fréquence empirique f d'un évènement est un estimateur de sa probabilité p . Les variables aléatoires \overline{X}_n , S_n^2 et f_n sont appelées alors les estimateurs de m , σ^2 et p .

1.3.1 Définition d'un estimateur

Soit X une variable aléatoire dont la loi dépend d'un paramètre inconnu θ , élément d'un sous ensemble donné Θ de \mathbb{R} appelé espace des paramètres. On cherche à estimer θ à partir d'un échantillon (X_1, X_2, \dots, X_n) de variables aléatoires indépendantes de même loi que X , on notera (x_1, x_2, \dots, x_n) l'échantillon observé. Un estimateur T_n de θ sera une variable aléatoire T_n qui dépend de X telle que :

$$T_n = T_n(X_1, X_2, \dots, X_n),$$

et chaque réalisation $T_n(x_1, x_2, \dots, x_n)$ est un estimateur de θ . Choisir une seule valeur pour estimer θ est un problème d'estimation ponctuelle, choisir un sous ensemble de Θ , dénommée région de confiance, est un problème d'estimation par intervalle dans \mathbb{R} . Ce type de problème sera résolu par l'application :

$$T_n : E^n \rightarrow F$$

qui associera une (ou plusieurs) variable(s) aléatoire(s) à valeur(s) numérique(s) ($F \subset \mathbb{R}$ ou \mathbb{R}^K) à n échantillon (X_1, X_2, \dots, X_n) , application que nous nommerons une statistique. Il s'agit du modèle d'échantillonnage noté $(E, B, (P_\theta; \theta \in \Theta))$ et B est la tribu borélienne associée.

Définition 1.16 *Un estimateur de θ est une application T_n de E^n dans F qui à un échantillon (X_1, X_2, \dots, X_n) de la loi P_θ associe une variable aléatoire réelle (ou plusieurs dans le cas d'un paramètre multidimensionnel) dont on peut déterminer la loi de probabilité.*

Exemple 1.2 *L'estimateur classique de la moyenne théorique est la moyenne empirique.*

rique :

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Sous la condition d'indépendance, pour un échantillon avec remise :

$$E(\overline{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} n E(X_i) = E(X_i).$$

$$Var(\overline{X}_n) = Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} n Var(X_i) = \frac{\delta^2}{n}.$$

Pour un échantillon sans remise :

$$Var(\overline{X}_n) = \frac{Var(x)}{n} \left(1 - \frac{n-1}{N-1}\right).$$

1.3.2 Propriétés d'un estimateur

Afin de choisir entre plusieurs estimateurs possibles d'un même paramètre, il faut satisfaire certaines propriétés.

Le Biais :

L'estimateur T_n est une variable aléatoire. Lorsque l'on estime θ par T_n on peut commettre une erreur d'estimation qui est la différence entre T_n et θ .

$T_n - \theta$ est donc une variable aléatoire que l'on peut décomposer de façon élémentaire en :

$$T_n - E(T_n) + E(T_n) - \theta.$$

où $E(T_n)$ est l'espérance mathématique de T_n . Le premier terme :

$$T_n - E(T_n),$$

représente les fluctuations aléatoires de T_n autour de sa moyenne et le second terme

$$E(T_n) - \theta,$$

représente une erreur asymétrique impliquée par le fait que T_n varie autour de θ . par conséquent, la quantité

$$E(T_n) - \theta, \tag{1.22}$$

s'appelle le biais. Il est donc souhaitable d'utiliser les estimateurs sans biais.

Définition 1.17 *On dit qu'un estimateur est sans biais si l'espérance mathématique de cet estimateur est égale au paramètre estimé :*

$$E(T_n) = \theta, \forall \theta \in \Theta. \tag{1.23}$$

Définition 1.18 *On dit qu'un estimateur est asymptotiquement sans biais si :*

$$\forall \theta \in \Theta, E(T_n) \rightarrow \theta, \text{ quand } n \rightarrow +\infty. \tag{1.24}$$

Convergence :

La deuxième qualité d'un estimateur est d'être convergent. Il est souhaitable que si $n \rightarrow \infty$, $T_n \rightarrow \theta$. C'est le cas de \overline{X}_n , S_n^2 et F_n . Les estimateurs ne convergent pas nécessairement à la même vitesse, ceci dépend, pour une taille d'échantillon donnée, de la notion de précision d'un estimateur. On dit aussi qu'un estimateur T_n est convergent lorsque sa variance tend vers 0 quand n tend vers l'infini.

Théorème 1.2 *Tout estimateur sans biais dont la variance tend vers 0 est convergent :*

$$E(T_n) = \theta \text{ et } \text{Var}(T_n) \rightarrow 0 \Rightarrow T_n \xrightarrow{p} \theta, n \rightarrow \infty. \tag{1.25}$$

Ce résultat se déduit directement d'inégalité de Bienaymé-Tchebychev :

$$P(|T_n - \theta| > \varepsilon) \leq \frac{\text{Var}(T_n)}{\varepsilon^2} \rightarrow 0, \text{ pour tout } \varepsilon > 0, n \rightarrow \infty. \quad (1.26)$$

Théorème 1.3 *Tout estimateur asymptotiquement sans biais dont la variance tend vers 0 est convergent :*

$$E(T_n) \rightarrow \theta \text{ et } \text{Var}(T_n) \rightarrow 0 \Rightarrow T_n \xrightarrow{p} \theta, n \rightarrow +\infty. \quad (1.27)$$

Exemple 1.3 *Soit (X_n) est une suite de variables aléatoires normales $N(m, \delta)$, mutuellement indépendantes, de même loi :*

1-

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

On a :

$$E(\overline{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} n E(X_i) = E(X_i) = m.$$

\overline{X}_n est donc un estimateur sans biais de m .

d'autre part :

$$\text{Var}(\overline{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} n \text{Var}(X_i) = \frac{\delta^2}{n}.$$

Lorsque $n \rightarrow \infty$, $\text{Var}(\overline{X}_n) \rightarrow 0$, \overline{X}_n est donc un estimateur convergent de m .

2-

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2.$$

est la variance empirique de X dans l'échantillon, cette variance est un estimateur biaisé de δ^2 et le biais vaut $\frac{\delta^2}{n}$.

3-L'estimateur sans biais de δ^2 est :

$$S_n^{*2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{n}{n-1} S_n^2.$$

Il est recommandé d'utiliser S_n^{*2} pour estimer δ^2 .

1.3.3 Estimateur optimal

Précision d'un estimateur où sa qualité

La précision d'un estimateur T_n se mesure par l'erreur quadratique moyenne :

$$EQ(T_n) = E[(T_n - \theta)^2] = Var(T_n) + b_n^2(\theta) \text{ avec } b_n^2(\theta) = E((T_n) - \theta)^2. \quad (1.28)$$

Variance minimal :

Parmi les estimateurs sans biais de θ le plus précis est celui qui a une plus petite variance. Soient deux estimateurs sans biais T_n et T'_n , T_n est meilleur que T'_n si :

$$Var(T_n) \leq Var(T'_n). \quad (1.29)$$

l'orsque on pourrait trouver un troisième estimateur T''_n ayant une variance plus petite $V(T_n)$, il faut poursuivre la recherche mais on ne peut pas améliorer indéfiniment un estimateur !! Nous allons voir comment peut-on régler ce problème.

Inégalité de Frechet-Darmonis-Cramer-Roe(FDCR) :

Si X prend ses valeurs dans un ensemble qui ne dépend pas de θ , si la densité $f(x, \theta)$ est deux fois continûment dérivable par rapport à θ , et sous certaines conditions de régularité, tout estimateur T_n sans biais de θ dont la variance existe vérifié l'inégalité FDCR :

$$\forall \theta \in \Theta, Var(T_n) \geq \frac{1}{I_n(\theta)} .$$

ou $I_n(\theta)$ est la quantité d'information de Fisher définie par :

$$I_n(\theta) = E\left(\frac{\partial \ln L}{\partial \theta}\right)^2 = E\left(-\frac{\partial^2 \ln L}{\partial \theta^2}\right). \quad (1.30)$$

où L : désigne la fonction de vraisemblance.

Les conditions de régularité sont les suivantes : On suppose que l'ensemble des estimateurs Θ est un ensemble ouvert sur lequel la densité $f(x, \theta)$ ne s'annule en aucun point x et est dérivable par rapport à θ . On suppose aussi que l'on peut intervenir dérivation par rapport à θ et intégration, et que la quantité d'information de Fisher est strictement positive.

Efficacité :

La borne inférieure pour la variance des estimateurs sans biais peut être atteinte ou non. Si cette borne est effectivement atteinte par un estimateur, il sera donc le meilleur, selon ce critère parmi l'ensemble des estimateurs sans biais. Cette optimalité est traduite par la définition suivante :

Définition 1.19 *Un estimateur sans biais T_n est efficace si sa variance est égale à la borne inférieure de FDICR :*

$$\text{Var}(T_n) = \frac{1}{I_n(\theta)}. \quad (1.31)$$

1.4 La loi forte des grands nombres

La loi forte des grands nombres permet de justifier la convergence des estimateurs.

Théorème 1.4 *(loi forte des grands nombres)*

Soit $(X_n)_{n \geq 0}$ une suite des variables aléatoires iid à valeurs dans \mathbb{R}^d ($d \in \mathbb{N}^$)*

On suppose que $E(|X_1|) < +\infty$ (ie finie), alors :

$$\frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{p.s} E(X) \quad (1.32)$$

Rappel : "ps" signifie "presque sûrement", ce théorème nous dit pour quoi l'approximation de MC est valide (et sous quel hypothèse).

1.5 Théorème central limite

Le théorème central limite permet de construire des intervalles de confiance asymptotiques.

Théorème 1.5 Soit $\{Y_n\}_{n \geq 1}$ une suite de v.a réelles iid de carré intégrable, on suppose que :

$$\sigma^2 = \text{Var}(Y_i) > 0, \quad (1.33)$$

et on pose :

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i \text{ et } \hat{\sigma}_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2}. \quad (1.34)$$

Alors :

$$\sqrt{n} \left(\frac{\bar{Y}_n - E(Y_i)}{\hat{\sigma}_n} \right) 1_{\{\hat{\sigma}_n > 0\}} \rightsquigarrow N(0, 1), \quad (1.35)$$

converge en loi, en particulier, pour tout réel :

$$c > 0, P \left[\sqrt{n} \left| \frac{\bar{Y}_n - E[Y_i]}{\hat{\sigma}_n} \right| \leq c \right] \rightarrow 1 - \alpha_c \text{ quand } n \rightarrow \infty, \quad (1.36)$$

où

$$\alpha_c = P[|X| > c], \text{ pour, } X \rightsquigarrow N(0, 1) \quad (1.37)$$

Pour N suffisamment grand, la probabilité d'avoir :

$$E(Y_i) \in [\bar{Y}_n \pm c * \hat{\sigma}_n / \sqrt{n}], \quad (1.38)$$

et proche de $1 - \alpha_c$.

1.6 Génération d'échantillons suivant différentes lois de probabilités

1.6.1 Généralité

La simulation de valeur d'une variable aléatoire réelle suivant une loi de probabilité autre que la loi uniforme standard repose sur une propriété mathématique connue. Nous rappelons dans ce qui suit cette propriété et nous montrons comment elle est utilisée pour simuler des lois de probabilité non uniforme. D'autre part il y'a trois méthodes principales pour engendrer des variables aléatoires qui obéissent à une distribution donnée, à savoir la méthode d'inversion, la méthode de rejection et la méthode de composition

b-Rappel d'une propriété mathématique :

Toute variable aléatoire X a valeur dans \mathbb{R}^d peut être simulée sous la forme :

$$X = f(U)$$

où $U = (U_1, U_2, \dots, U_q)$ est uniformément répartie sur $[0, 1]$.

Remarque 1.1 1- Il s'agit d'une égalité en loi.

2- La fonction a une expression explicite et n'est pas nécessairement unique.

1.6.2 La méthode d'inversion

Théorème 1.6 *Théorème d'inversion*

Soit F une fonction de répartition sur \mathbb{R} qui continue est strictement croissante.

On note :

$$F^{-1}(y) = \inf\{x \in \mathbb{R} / F(x) \geq y\}, \quad (1.39)$$

l'inverse généralisé de F .

Soit U de loi uniforme sur $[0, 1]$.

Alors :

- 1- $X = F^{-1}(u)$ a pour fonction de répartition F .
- 2- F est continue sur \mathbb{R} .

1.6.3 La méthode de rejet

On veut simuler une variable aléatoire de loi de densité f , et on suppose qu'il existe une loi de densité g simulable facilement telle que :

$$\forall x \in \mathbb{R}, f(x) \leq kg(x). \quad (1.40)$$

où k est une constante réelle, on pose :

$$\alpha(x) = \frac{f(x)}{kg(x)}. \quad (1.41)$$

Proposition 1.1 *Générons un couple (X_1, U_1) de variable aléatoire indépendantes telle que Y_1 a une loi de densité g et U_1 suite une loi uniforme sur $[0, 1]$.*

- Si $U_1 \leq \alpha(X_1)$, posons $X = X_1$.
- Sinon rejetons X_1 et recommençons en générant une suite $(X_n, U_n)_{n \geq 2}$ de variable indépendant de même loi que (X_1, U_1) jusqu'à l'instant p où $U_p \leq \alpha(X_p)$. Posons alors $X = X_p$. la v.a X ainsi simulé et de loi de densité f .

1.6.4 La méthode de composition

Cette méthode consiste à remplacer une densité de fonction $f(x)$ d'une variable aléatoire X par un mélange probabiliste de fonctions de densités $g_n(x)$ judicieusement choisies.

Autrement dit, elle exploite une relation du type :

$$f(x) = \sum g_n(x) \times p_n \quad (1.42)$$

Chapitre 2

Echantillonnage descriptif

2.1 Introduction

Nous allons commencer ce chapitre par le principe de la méthode de Monte-Carlo, ensuite nous allons aborder la méthode d'échantillonnage descriptif avec un exemple d'illustration.

2.2 Description de la méthode

Le terme méthode Monte-Carlo désigne toute méthodes visant à calculer une valeur numérique en utilisant des procédés aléatoires, c'est à dire des techniques probabilistes. le nom de ces méthodes fait allusion aux jeux de hasard pratique à M-C et a été inventé en 1947 par N.Metropolis, dans un article publié pour la première fois en 1949 pour utiliser une méthode MC, on doit tout d'abord mettre sous la forme d'une espérance la quantité que l'on cherche à calculer c'est souvent simple par exemple, calculer l'intégrale :

$$I = \int_0^1 g(x)dx. \quad (2.1)$$

La méthode MC consiste à écrire I sous la forme :

$$I = E(g(u)), \quad (2.2)$$

telle que $U \rightsquigarrow U_{[0,1]}$ de la loi uniforme sur l'intervalle $[0, 1]$ mais peut être plus compliqué (comme les équations aux dérivées partielles). A l'issue de cette étape, il reste à calculer une quantité de la forme $E(X)$ c'est à dire l'espérance de la variable aléatoire X , on dispose alors d'une suite $(X_i)_{1 \leq i \leq N}$ de N réalisations de la variable aléatoire X . On approxime alors $E(X)$ par :

$$E(X) = \frac{1}{N}(X_1 + \dots + X_N). \quad (2.3)$$

On cherche à calculer :

$$I = \int_{[0,1]^d} f(u_1, u_2, \dots, u_d) du_1, du_2, \dots, du_d. \quad (2.4)$$

Mise sous forme d'espérance; Application au cas du calcul d'une intégrale :

On pose :

$$X = f(u_1, u_2, \dots, u_d),$$

où u_1, u_2, \dots, u_d sont des réalisations de la loi uniforme sur l'intervalle $[0, 1]$ alors :

$$E(X) = E(f(u_1, u_2, \dots, u_d)) = I.$$

2.2.1 La convergence de la méthode

D'après la loi forte des grands nombres que nous avons présentée dans le premier chapitre, la méthode de Monte-Carlo est convergente.

a-L'estimateur de Monte carlo :

Les méthodes de Monte-Carlo sont basées sur la loi forte des grands nombres. Soit X une variable aléatoire, on simule un grand nombre de variable aléatoire $(X_n)_{n \geq 0}$ indé-

pendantes et de même loi que X . On prend en suite la moyenne des valeurs prises :

$$\frac{1}{n} \sum_{i=1}^n X_i. \quad (2.5)$$

Soient Y_1, Y_2, \dots, Y_n , n réalisations iid de la variable Y l'estimateur de Monte-Carlo est définie comme la moyenne empirique :

$$\hat{\theta}_n = g\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) = g(\bar{Y}), \quad (2.6)$$

par la loi forte des grands nombres, \bar{Y} est une suite d'estimateurs de $E(Y)$.

au sens où :

$$\bar{Y}_n \xrightarrow{p.s} E(Y), \quad (2.7)$$

ceci est démontré dans le théorème suivant :

Théorème 2.1 *Soit $\{Y_n\}_{n>0}$ une suite de variable aléatoire iid telle que $E(\|Y_i\|) < \infty$, D'après le théorème de la loi des grands nombres :*

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{p.s} E(Y) \quad (2.8)$$

Si g est une fonction continue au point θ

alors :

$$\hat{\theta}_n = g(\bar{Y}) \xrightarrow{p.s} g(E(Y)). \quad (2.9)$$

Remarque 2.1 *La méthode MC ne peut donc s'utiliser que pour des variables aléatoires intégrables. Pour avoir une idée de l'intérêt de la méthode, il faut pouvoir évaluer l'erreur comise.*

Définition 2.1 L'erreur comise ε_n est définie par :

$$\varepsilon_n = E(X) - \frac{1}{n} \sum_{i=1}^n X_i. \quad (2.10)$$

Le théorème de la limite central donne un asymptotique de l'erreur ε_n , mais de nature aléatoire. Il dit que la loi de l'erreur finit par ressembler à une loi gaussienne centrée.

Exemple 2.1 Soit f une fonction mesurable sur $[0, 1]$ on cherche à calculer :

$$p = \int_{f(x) \geq \lambda} dx.$$

où : λ une constante donnée.

Soit : X la variable aléatoire telle que :

$$X = 1_{f(u) \geq \lambda},$$

où :

u est une réalisation d'une variable aléatoire $U \rightsquigarrow U_{[0,1]}$.

Alors :

$$p = E(X) \text{ et } \sigma^2 = \text{Var}(X) = p(1 - p).$$

A l'issus de n tirage indépendants selon la loi de X , on a :

$$p_n = \frac{X_1 + X_2 + \dots + X_n}{n} \simeq p + \frac{\sigma}{\sqrt{n}} G.$$

pour faire une erreur moyenne $\frac{\sigma}{\sqrt{n}}$ de l'ordre de 0.01, comme $p(1 - p) \leq \frac{1}{4}$, il faut prendre n de l'ordre 2500. Aussi l'intervalle de confiance à 95% est alors :

$$]p_n - 1.96 * 0.01, p_n + 1.96 * 0.01[$$

si la valeur de p à estimer est de l'ordre de 0.5, l'erreur est acceptable.

Par contre, si p est proche de 0 (ou de 1). Comme $\sigma \simeq \sqrt{p}$ (ou bien $\sqrt{1-p}$), l'erreur relative est :

$$\frac{\sigma}{\sqrt{n}E(X)} = \frac{1}{\sqrt{np}},$$

donc il faut beaucoup de tirage aléatoire (n grand), pour estimer un p convenable.

2.2.2 Réduction de la variance

Nous venons de voir que la vitesse de convergence de la méthode de Monte-Carlo est de l'ordre $\frac{\sigma}{\sqrt{n}}$. Pour améliorer cette méthode il existe de nombreuses techniques dites de réduction de variance, qui cherchent à minimiser (diminuer) la valeur de la variance σ^2 . L'idée générale est de donner une autre représentation sous forme d'espérance de la quantité à calculer :

$$E(X) = E(Y) \text{ avec } Var(Y) < Var(X). \quad (2.11)$$

Nous cherchons à diminuer la variance. Nous allons passer en revue quelques-unes de ces méthodes qui sont applicables..

a-Echantillonnage préférenciel (fct d'importance) :

Supposons que l'on cherche à calculer $E(g(X))$ et que la loi de X soit $f(x)dx$ (sur \mathbb{R}), la quantité que l'on cherche à évaluer vaut donc :

$$E(g(X)) = \int_{\mathbb{R}} g(x)f(x)dx. \quad (2.12)$$

Soit maintenant, h la densité d'une autre loi telle que :

$$h > 0, \text{ et } \int_{\mathbb{R}} h(x)dx = 1$$

(ie h et une densité de probabilité), il est clair que $E(g(X))$ peut aussi s'écrire comme suit

$$E(g(X)) = \int_{\mathbb{R}} \frac{g(x)f(x)}{h(x)} h(x) dx. \quad (2.13)$$

Cela signifie que ;

$$E(g(X)) = E\left(\frac{g(Y)f(Y)}{h(Y)}\right), \quad (2.14)$$

Si Y suit la loi $h(x)dx$ sous p . On a donc une autre méthode de calcul de $E(g(X))$ en utilisant n tirages de Y , (Y_1, \dots, Y_n) , et en approxinant $E(g(X))$ par :

$$\frac{1}{n} \left(\frac{g(Y_1)f(Y_1)}{h(Y_1)} + \dots + \frac{g(Y_n)f(Y_n)}{h(Y_n)} \right).$$

Si l'on pose :

$$Z = \frac{g(Y)f(Y)}{h(Y)},$$

on aura amélioré l'algorithme si :

$$Var(Z) < Var(g(X))$$

Il est facile de calculer la variance de Z :

$$Var(Z) = E(Z^2) - (E(Z))^2 = \int_{\mathbb{R}} \frac{g^2(x)f^2(x)}{h(x)} dx - (E(g(X)))^2 \quad (2.15)$$

Si $g(X) > 0$, on peut vérifier qu' en prenant :

$$h(x) = \frac{g(x)f(x)}{E(g(X))}$$

on annule $Var(Z)$.

Il ne faut pas trop donner d'importance à ce résultats car il repose sur le fait que l'on connaît $E(g(X))$, et c'est justement la quantité que l'on cherche à calculer. Cela permet cependant de justifier l'heuristique suivante : prendre $h(x)$ aussi proche que possible de

$|g(x)f(x)|$ puis la normaliser (diviser par $\int h(x)dx$ de façon à obtenir une densité dont la loi est facilement simulable.

b-Variable de contrôle :

Dans sa version la plus simple, il s'agit d'écrire $E(f(X))$ sous la forme :

$$E(f(X)) = E(f(X) - h(X)) + E(h(X)) \quad (2.16)$$

avec $E(h(X))$ qui peut se calculer explicitement et $Var(f(X) - h(X))$ sensiblement plus petite que $Var(f(X))$. On utilise alors une méthode de Monte-Carlo pour évaluer ;

$$E(f(X) - h(X)),$$

et le calcul discrète pour $E(h(X))$.

Commençons pour donner un exemple simple.

Exemple 2.2 *Supposons que l'on veuille calculer :*

$$\int_0^1 \exp(x) dx$$

comme au voisinage de 0 ;

$$\exp(x) \approx 1 + x$$

on peut écrire : $h(x) = 1 + x$ et $E(h(x)) = E(1 + x) = 1 + E(x) = 1 + \frac{1}{2} = \frac{3}{2}$

$$\int_0^1 \exp(x) dx = \int_0^1 (\exp(x) - 1 - x) dx + \frac{3}{2}$$

Il est facile de vérifier que la variance de la méthode diminue alors sensiblement.

c-Variable antithétique :

Supposons que l'on cherche à calculer :

$$I = \int_0^1 f(x)dx, \quad (2.17)$$

comme $x \rightarrow 1 - x$ laisse invariante la mesure dx , on a aussi :

$$I = \frac{1}{2} \int_0^1 (f(x) + f(1 - x))dx. \quad (2.18)$$

On peut donc calculer I de la façon suivante. on tire n variables aléatoire U_1, \dots, U_n suivants une loi uniforme sur $[0, 1]$ et indépendant et on approxime I par :

$$\begin{aligned} I_{2n} &= \frac{1}{n} \left[\frac{1}{2} (f(U_1) + f(1 - U_1)) + \dots + \frac{1}{2} (f(U_n) + f(1 - U_n)) \right] \\ &= \frac{1}{2n} (f(U_1) + f(1 - U_1) + \dots + f(U_n) + f(1 - U_n)), \end{aligned} \quad (2.19)$$

lorsque l'on compare cette méthode à une méthode de M-C directe, à l'issue de $2n$ tirages on peut montrer que si la fonction f est continue monotone la qualité de l'approximation s'améliore. On peut généraliser ce genre d'idées en dimension supérieure et à d'autre transformations préservant la loi de la variable aléatoire.

d-Méthode de stratification :

C'est une méthode bien connue des statisticiens et souvent utilisée dans les sondages
Supposons que l'on cherche à calculer I , avec :

$$I = E(g(X)) = \int_{\mathbb{R}^d} g(x)f(x)dx. \quad (2.20)$$

où X est une variable aléatoire à valeur dans \mathbb{R}^d suivant la loi $f(x)dx$. On se donne une partition $(D_i, 1 \leq i \leq m)$ de \mathbb{R}^d on décompose alors I de la façon suivante :

$$I = \sum_{i=1}^m E(1_{\{X \in D_i\}}g(X)) = \sum_{i=1}^m E(g(X))/X \in D_i \quad (2.21)$$

lorsque que l'on connaît les nombres $p_i = P(X \in D_i)$, on peut utiliser une méthode de Monte-Carlo pour estimer les intégrales :

$$I_i = E(g(X)/X \in D_i). \quad (2.22)$$

Supposons que l'on approxime l'intégrale I_i par \tilde{I}_i à l'aide de n_i tirages indépendants, la variance de l'erreur d'approximation est donnée par :

$$\frac{\sigma_i^2}{n_i}, \quad (2.23)$$

si l'on note :

$$\sigma_i^2 = Var(g(X)/X \in D_i). \quad (2.24)$$

On approxime ensuite I par \tilde{I} avec :

$$\tilde{I} = \sum_{i=1}^m p_i \tilde{I}_i, \quad (2.25)$$

les échantillons servant à obtenir les estimateur \tilde{I}_i étant supposés indépendants on montre facilement que la variance de l'estimateur \tilde{I} vaut :

$$\sum_{i=1}^m p_i^2 \frac{\sigma_i^2}{n_i}, \quad (2.26)$$

il est alors naturel de minimiser cette erreur pour un nombre total de tirages fixé.

$$\sum_{i=1}^m n_i = n \quad (2.27)$$

On peut vérifier que les n_i qui minimise la variance de \tilde{I} sont données par :

$$n_i = n \frac{p_i \sigma_i}{\sum_{i=1}^m p_i \sigma_i} \quad (2.28)$$

le minimum de la variance de \tilde{I} vaut alors :

$$\frac{1}{n} \left(\sum_{i=1}^m p_i \sigma_i \right)^2 \quad (2.29)$$

Il est inférieur à la variance que l'on obtiendrait avec n tirages aléatoires par la méthode de Monte-Carlo classique. En effet, cette variance vaut :

$$\text{var}(g(X)) = E(g(X)^2) - E(g(X))^2 = \sum_{i=1}^m p_i E(g(X)^2 / X \in D_i) - \left(\sum_{i=1}^m E(g(X) / X \in D_i) \right)^2 \quad (2.30)$$

D'où en faisant intervenir les variances conditionnelle σ_i :

$$\text{var}(g(X)) = \sum_{i=1}^m p_i \text{var}(g(X) / X \in D_i) + \sum_{i=1}^m p_i E(g(X) / X \in D_i)^2 - \left\{ \sum_{i=1}^m p_i E(g(X) / X \in D_i) \right\}^2 \quad (2.31)$$

On utilise alors, deux fois l'inégalité de convexité pour x^2 :

$$\left(\sum_{i=1}^m p_i a_i \right)^2 \leq \sum_{i=1}^m p_i a_i^2 \sum_{i=1}^m p_i = 1 \quad (2.32)$$

pour montrer que :

$$\text{var}(g(X)) \geq p_i \text{var}(g(X) / X \in D_i) \geq \left(\sum_{i=1}^m p_i \sigma_i \right)^2 \quad (2.33)$$

Ceci prouve que, sous réserve que l'on fasse une affectation optimale des tirages, on peut obtenir par stratification un estimateur de variance moindre. Notons cependant que l'on ne peut que rarement calculer les σ_i , ce qui limite la portée de cette technique (mais on peut toujours les estimer à l'aide d'un premier tirage de Monte-Carlo). Notons aussi

qu'il est possible d'obtenir un estimateur de variance supérieur à l'estimateur initial si l'affectation des points aux domaines est quelconque. Il existe malgré tout d'autre stratégie d'affectation des points par domaines qui réduisent forcément la variance. Par exemple la stratégie qui affecte un nombre de points proportionnel à la probabilité du domaine :

$$n_i = np \tag{2.34}$$

on obtient alors un estimateur de variance égal à :

$$\frac{1}{n} \sum_{i=1}^m p_i \sigma_i^2 \tag{2.35}$$

or nous venons de voir que :

$$\sum_{i=1}^m p_i \sigma_i^2 \tag{2.36}$$

est un majurant de $var(g(X))$. Cette stratégie d'allocation est parfois utilisé lorsque l'on sait expliciter les probabilités p_i , pour des considérations approfondies sur ces techniques on pourra consulter.

e-Valeur moyenne où conditionnement

Supposons que l'on cherche à calculer :

$$E(g(X, Y)) = \int g(x, y) f(x, y) dx dy, \tag{2.37}$$

où $f(x, y) dx dy$ est la loi du couple (X, Y) .

Si l'on pose :

$$h(x) = \frac{1}{m(x)} \int g(x, y) f(x, y) dy, \tag{2.38}$$

avec :

$$m(x) = \int f(x, y) dy, \tag{2.39}$$

il est clair que $E(g(X, Y)) = E(h(X))$. En effet la loi de X est $m(x)dx$, et donc

$$E(h(X)) = \int m(x)h(x)dx = \int dx \int g(x, y)f(x, y)dy = E(g(X, Y)), \quad (2.40)$$

on peut retrouver ce résultat en notant que :

$$E(g(X, Y)/X) = h(X), \quad (2.41)$$

cette interprétation comme une espérance conditionnelle permet de prouver que :

$$\text{var}(h(X)) \leq \text{var}(g(X, Y)). \quad (2.42)$$

Si l'on peut calculer explicitement la fonction $h(X)$, il est préférable d'utiliser une méthode de Monte-Carlo pour $h(X)$.

f-Suites à discrédance faible :

Une autre façon d'améliorer les méthodes de type M-C est de renoncer au caractère aléatoire des tirages et de tirer les points de façon "plus ordonnée". On cherche à trouver des suites $(x_i)_{i \geq 0}$ déterministe permettant d'approximer les intégrales par une formule de la forme :

$$\int_{[0,1]^d} f(x)dx \approx \lim_{n \rightarrow +\infty} \frac{1}{n} (f(x_1) + \dots + f(x_n)) \quad (2.43)$$

On parle dans ce cas de méthode de **quasi MC**. On peut trouver des suites telles que la vitesse de la convergence de l'approximation soit de l'ordre de $K \frac{\log(n)^4}{n}$, mais à condition que la fonction f possède une certaine régularité, ce qui est sensiblement meilleur qu'une méthode de M-C, c'est ce genre de suite que l'on appelle une suite à discrédance faible. Commençons par donner la définition d'une suite équirépartie.

Définition 2.2 *On dit que $(x_n)_{n \geq 1}$ est une suite équirépartie sur $[0, 1]^d$ si l'une des propriétés suivantes (équivalentes) est vérifiée. (si x et y sont deux points de $[0, 1]^d$, $x \leq y$ si*

et seulement si par définition $x_i \leq y_i$ pour tout $1 \leq i \leq d$)

1- pour tout $y = (y^1, \dots, y^d) \in [0, 1]^d$:

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n 1_{\{x_k \in [0, y]\}} = \prod_{i=1}^d y^i = \text{volume}([0, y]) \quad (2.44)$$

où

$$[0, y] = \{z \in [0, 1]^d, z \leq y\} \quad (2.45)$$

2- $D_n^*(x) = \sup \left| \frac{1}{n} \sum_{i=1}^n 1_{\{x_k \in [0, y]\}} - \text{volume}([0, y]) \right| \rightarrow 0$.

3- Pour tout fonction f **Riman** intégrable (ie bornée et dx presque sure continue) définie sur $[0, 1]^d$:

$$\lim \frac{1}{n} \sum_{i=1}^n f(x_k) = \int_{[0, 1]^d} f(x) dx \quad (2.46)$$

$D_n^*(x)$ est appelée la discrèpance à l'origine de la suite x

Remarque 2.2 Si $(U_n)_{n \geq 1}$ désigne une suite de variables aléatoires indépendantes et de loi uniforme sur $[0, 1]$, les suites aléatoires $(U_n(\omega))_{n \geq 1}$ seront presque sûrement équirépartie, de plus une loi du logarithme itéré pour la discrèpance :

$$p.s \overline{\lim} \sqrt{\frac{2n}{\log(\log(n))}} D_n^*(u) = 1. \quad (2.47)$$

on dit qu'une suite est à discrèpance faible si sa discipance est asymptotiquement meilleure que celle d'une suite aléatoire. On peut prouver que la discrèpance d'une suite infinie vérifie forcément :

$$D_n^* > c_d \frac{\log(n)^{\max(\frac{d}{2}, 1)}}{n}, \quad (2.48)$$

pour un nombre infini de valeur de n où c_d est une constante ne dépendant que de d .

On connaît de nombreuses suites à discrèpance faible **d-dimensionnelles**. Les meilleures discrèpances asymptotiques connues sont de l'ordre de $((\log(n))^d/n)$ ces suites ont une discrèpance **quasi optimale** vu la remarque précédente. Ces suites sont asymptotiquement meilleures qu'une suite de nombres aléatoires. Cependant, dans la pratique ie pour

des valeurs de n entre 10^3 et 10^6 , les discrédances des meilleures suites connues ne sont pas aussi bonne que les résultats asymptotiques pourraient le laisser espérer particulièrement pour des dimensions supérieures à la dizaine. Une autre intérêt des suites à discrédance faible est de donner une estimation à priori de l'erreur commise lors de l'intégration numérique, pour des fonction à variation finie, par l'intermédiaire de la formule de **Koksma-Hlawka**. Contrairement aux suites aléatoires qui fournissent des intervalles de confiance pour une probabilité donnée, cette majoration est effective et déterministe. Il faut cependant relativiser l'intérêt de cette majoration en notant qu'elle est presque toujours très éloignée de la valeur réelle de l'erreur et que la variation d'une fonction est une quantité très difficile à évaluer.

La proposition suivante explicite cette majoration :

Proposition 2.1 (*Inégalité de **koksma-Hlawka***)

Si g est une fonction à variation finie au sens de **Hardy** et **Krouse** de variation $V(g)$, alors :

$$\forall n \geq 1, \left| \frac{1}{N} \sum_{k=1}^N g(x_k) - E(X) \right| \leq V(g) D_N^*(x). \quad (2.49)$$

2.3 Simulation de Monte Carlo

Les méthodes de simulation reposent toutes sur la capacité à produire une suite des variables aléatoires indépendantes suivant une distribution f donnée. La production d'une telle suite dépend elle même de la capacité que l'on a à produire une séquence des variables aléatoires uniformes et indépendantes. En pratique, la production par des moyennes informatiques d'une suite de variables uniformes iid est impossible : toute suite de nombre résulte de la mise en oeuvre d'un algorithme, et par conséquent est déterministe. L'objectif que l'on s'assigne n'est donc pas la production de véritables variables aléatoires mais la production de séquences déterministes imitant le mieux possible le comportement d'une suite de variables aléatoires uniformes iid.

2.3.1 Les erreurs d'échantillonnage

L'incertitude associée aux résultats de simulation nait du besoin de représenter, dans une histoire nie, toutes les informations contenues dans une variable aléatoire d'entrée. En simulation, l'utilisation de l'échantillonnage aléatoire génère toujours des erreurs d'échantillonnage car les estimateurs obtenues sont des fonctions des valeurs de la variable d'entrée. Il est évident que la variabilité causée par les erreurs d'échantillonnage est indésirable.

En simulation stochastique, ils existent deux types d'erreurs d'échantillonnage, appelés, effet ensembliste et effet séquentiel a montré l'existence d'un autre type d'erreur, appelée effet d'interaction ensemble-séquence et a obtenu des résultats empiriques sur quelques problèmes.

Définissons maintenant les différents types d'erreurs :

Erreur de type 1 :

L'erreur de type 1 appelée effet ensembliste, est liée à l'ensemble des valeurs produites par la loi uniforme dans la méthode de MC. Cet effet est caractérisé par la différence entre la distribution empirique et la distribution théorique ou par la différence entre le graphe de la série de données (histogramme) et le graphe de la fonction de probabilité. L'effet ensembliste est contrôlé si la procédure d'échantillonnage est parfaitement maîtrisée.

Erreur de type 2 :

L'erreur de type 2 appelée effet séquentiel, est induite par la suite de l'ensemble des valeurs produites par la loi uniforme dans la méthode de MC. Cet effet est caractérisé par l'ordre dans lequel les valeurs de la loi uniforme se présente. Contrairement à l'effet ensembliste, l'effet séquentiel est difficile à contrôler en simulation.

Erreur de type 3 :

Ce type d'erreur est défini par les erreurs qui ne sont expliqués ni par l'effet ensemble ni par l'effet séquentiel mais par leurs interactions. De la même manière que l'effet séquentiel, il est difficile à contrôler dans certains cas.

Parmi les méthodes qui réduisent les erreurs d'échantillonnage, nous citons les méthodes de réduction de variance.

2.4 Echantillonnage statistique

pour recueillir des informations sur une population statistique, on dispose de deux méthodes :

1- La méthode exhaustive ou recensement où chaque individu de la population est étudié selon les caractères étudiés.

2- La méthode de sondage ou d'échantillonnage qui conduit à n'examiner qu'une fraction de la population, un échantillon.

Pour que les résultats observés lors d'une étude soient généralisables à la population statistique, l'échantillon doit être représentatif de cette dernière, c'est à dire qu'il doit refléter fidèlement sa composition et sa complexité. Seul l'échantillonnage aléatoire assure la représentativité de l'échantillon.

2.5 Echantillonnage aléatoire simple

Prélevement au hasard, et de façon indépendante, d'un certain nombre n d'éléments de la population statistique de N éléments ou unité de l'échantillonnage d'une population à N individus, chaque individu possède ainsi la même probabilité de faire partie d'un échantillon de n individus et chacun des échantillons possibles de taille n possède la même probabilité d'être constitué.

L'échantillonnage aléatoire simple assure l'indépendance des erreurs (ε_i) ie l'absence

d'autocorrélation parmi les données relatives à un même caractère. Cette indépendance est indispensable à la validité de plusieurs tests statistiques, est un fragment d'un ensemble prélevé pour juger de cet ensemble, fraction de la population statistique sur laquelle des mesures sont faites pour connaître les propriétés de cette population.

2.6 Echantillonnage descriptif

2.6.1 Introduction

La méthode d'échantillonnage descriptif est l'une des méthodes de simulation de Monte-Carlo qui sont connues aussi par les méthodes d'échantillonnages aléatoires, et qui sont basées sur la production des échantillons en générant des nombres aléatoires suivant une loi uniforme à partir des variables d'entrées du modèle.

Dans la pratique il est souvent coûteux de prendre des échantillons suffisamment grands en raison de l'espace mémoire et du temps machine nécessaire pour son utilisation. Par conséquent, les estimateurs de l'échantillonnage aléatoire varient grâce à l'effet ensembliste et séquentiel de cette méthode.

La méthode d'échantillonnage descriptif évite l'effet ensembliste et garde l'effet séquentiel, mais de ce fait elle a ses inconvénients à savoir :

Les estimateurs produits sont biaisés et encore en pratique son fonctionnement strict exige une connaissance préalable de la taille d'échantillon requise.

2.6.2 La procédure d'échantillonnage

Supposons que le problème de simulation en étude contient une variable aléatoire d'entrée X . Soit H sa fonction de répartition et H^{-1} son inverse telle que :

$$X = H^{-1}(R) \tag{2.50}$$

où : R suit la loi uniforme entre 0 et 1.

La procédure de génération d'échantillons descriptifs est la suivante :

1- On subdivise l'intervalle $[0, 1]$ en n sous-intervalles équiprobables.

Soit :

$$\{r_i, i = 1, 2, \dots, n\}, \quad (2.51)$$

l'ensemble de n points réguliers où :

$$r_i = \frac{i - 0.5}{n}, i = 1, 2, \dots, n. \quad (2.52)$$

2- On génère les valeurs de l'échantillonnage sans remise

$$x_{d_i} = H^{-1}(r_i), i = 1, 2, \dots, n. \quad (2.53)$$

3- On garde en mémoire les valeurs de l'échantillon afin de les utiliser à la demande de la simulation .

Dans une histoire, les estimateurs obtenus par la simulation, en utilisant l'échantillonnage descriptif sont données par :

$$(Y_r)_j = F_j(r_1, r_2, \dots, r_n), j = \overline{1, k}. \quad (2.54)$$

où :

-(Y_r)_j est l'estimateur du paramètre inconnu $\theta_j, j = \overline{1, k}$.

- r_1, r_2, \dots, r_n sont des nombres réguliers dépendants uniformément distribués entre 0 et 1.

Etant donné que l'échantillonnage descriptif utilise des nombres réguliers, on parle alors de la surface des réponses régulière. Sa définition est identique à celle de la surface des réponses sauf qu'au lieu de considérer l'ensemble

$$\{u_1, u_2, \dots, u_n\}, \quad (2.55)$$

nous utilisons l'ensemble $\{r_1, r_2, \dots, r_n\}$.

2.7 Exemple d'utilisation

Dans cet exemple nous allons appliquer la méthode ED sur le cas d'une variable d'entrée qui suit la loi exponentielle de paramètre $\lambda > 0$. Nous avons utilisé le langage *R* pour effectuer les calculs.

Algorithme 2.2 :

```
n <- 31
x <- c()
r <- c()
h <- function(r){h = -(1/lambda) * log(1 - r)}
for(i in 1 : n)
  {r[i] <- -(i - 0.5)/n
    x [i] <- -h(r[i])
    r
    x
  }
mean(x)
```

les résultats obtenue sont sous-dissus :

i	$r_i = (i - 0.5)/n$	$x_i = -(1/\lambda)\log(1 - r_i)$
1	0.05	0.1025866
2	0.15	0.3250379
3	0.25	0.5753641
4	0.35	0.8615658
5	0.45	1.1956740
6	0.55	1.5970154
7	0.65	2.0996442
8	0.75	2.7725887
9	0.85	3.7942400
10	0.95	5.9914645

Comme nous l'avons indiqué précédemment, le théorème de la loi des grands nombres assure la convergence de l'estimateur de Monte-Carlo vers l'espérance mathématique :

$$\text{mean}(X) = 1.931518.$$

2.8 Conclusion

La simulation et la modélisation sont très liées, mais la simulation peut prendre des formes différentes de celles de la modélisation.

La simulation n'est pas destinée à remplacer la modélisation mais bien plutôt à la compléter, un modèle est toujours une image simplifiée de la réalité mais une bonne utilisation de la simulation requiert une bonne connaissance des limites de validité du modèle.

l'échantillonnage descriptif est l'une des méthodes de simulation qui peut être utilisée pour produire des valeurs d'entrée pour l'estimation de l'espérance de fonctions de variables de sortie.

Bibliographie

- [1] Aissani A. Modélisation et simulation. OPU, 2007.
- [2] Bontempi, G. Modélisation et simulation. Tech. rep., Département d'Informatique, 2008.
- [3] Baranger, C., and Mathiaud, J. Méthode de monte-carlo, 2012/2013.
- [4] Cours statistique mathématique Master1, 2015-2016.
- [5] Dimov, I.T. : Monte Carlo methods for applied scientists. World Scientific Publishing Co. Pte, Ltd Singapore (2008).
- [6] Fort, G., and Moulines, E. Méthodes de monte carlo et applications à la finance. Tech. rep., 29 juin 2009.
- [7] Fishman, G.S. : Monte-Carlo Concepts algorithms and applications. Springer-Verlag, Berlin (1997).
- [8] Grigoriu, M. : A spectral-based Monte Carlo algorithm for generating samples of nonstationary Gaussian processes. Monte Carlo methods and Application. 16, 143167 (2010) 77.
- [9] The mersenne twister generator. [http : en.wikipedia.org/wiki/Mersennetwister](http://en.wikipedia.org/wiki/Mersennetwister), Mai (2013).
- [10] Keller, A., Heinrich, S., Niederreiter, H. : Monte Carlo and Quasi Monte Carlo methods 2006. Springer, Berlin (2008).
- [11] M.G.Kendall and S.Babington, Random ness and random sampling numbers, J.Roy. Stat. Soc. 101 (1938) 147-166.

- [12] Lehmann, E. L. Some concepts of dependence. *Ann. Math. Statist.*, 35, 1137-1153 (1966).
- [13] Loh, W.L. : On Latin hypercube sampling. *The annals of statistics*. 24, 2058-2080 (1996).
- [14] Millet, A. *Méthodes de Monte-Carlo*. Université Paris VII et I, 2011.
- [15] MacKay, M.D, Beckman, R. J, Conover, W. J. :A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 21, 239-245 (1979).
- [16] Niederreiter, H. : *Random number generation and quasi monte Carlo methods*. Society for Industrial and Applied Mathematics, Philadelphia (1992).
- [17] Ourbih-Tari, M., Aloui, A. : Sampling methods and parallelism into Monte Carlo simulation, *Journal of Statistics, Advances in Theory and Applications*. 1, 169-192 (2009).
- [18] Ourbih-Tari M, Aloui A, Alioui A. : A software component which generates regular numbers from nested scriptive sampling. *The Proceedings of the European Simulation Modelling conference*. Edited by Marwan Al-Akaidi. Leicester : United Kingdom, 23-25 (2009). 78.
- [19] Ourbih, M. *Cour de simulation à évènements discrets*, 2010/2011.
- [20] Pidd, M. : *Computer simulation in management science* (5ed). John Wiley and Sons : Chichester (2004).
- [21] Ramberg, J.S., Schmeiser, B.W. : An Approximate Method for generating Symmetric Random Variables. *Communications of the ACM*. 15, 987- 990 (1972).
- [22] Saliby E. : Descriptive Sampling : A better approach to Monte Carlo simulation *Journal of the Operational Research Society*. 41(12) : 1133-1142 (1990).
- [23] Schmitz, T.L., Kim, H.S. : Monte Carlo evaluation of periodic error uncertainty. *Precision Engineering*. 31, 251-259 (2007).
- [24] Sobol, I.M. : *A primer for the Monte Carlo method*. CRS Press, Florida (1994).

- [25] Tari, M., Dahmani, A. : Flowshop simulator using different sampling methods. *Operational Research An International Journal*. 5, 261-272 (2005).
- [26] Tari, M., Dahmani, A. :The three phase discreteevent simulation using some sampling methods. *International Journal of Applied Mathematics & Statistics*. 3, 37-48 (2005).
- [27] Tari, M., Dahmani, A. : Refined Descriptive Sampling A better approach to Monte Carlo simulation. *Simulation Modelling Practice and Theory*. 14, 143-160.
- [28] Trignan, J. Probabilités et statistiques et leurs applications : Brevet de technicien supérieur, bréal ed. instituts universitaires de technologie, octobre 1990.
- [29] Laouj Farida. site d'intérieur. Cours statistique mathématique Master 1 2015-2016.