Techniques de réduction de la variances : Échantillonnage stratégique, échantillonnage stratifié, hypercube latin

6-601-09 Simulation Monte Carlo

Geneviève Gauthier

HEC Montréal

Rappel
La notation
Propriétés statistiques de l'estimateur
L'échantillonnage stratégique
Échantillonnage stratifié
Carré latin

Estimateur de Monte Carlo

- Soit $X_1, X_2, ..., X_n$ des variables aléatoires indépendantes et identiquement distribuées. Elles sont obtenues à l'aide du générateur de nombres aléatoires de votre logiciel préféré.
- L'*estimateur de Monte Carlo* de la quantité

$$\theta \equiv \mathrm{E}\left[g\left(X\right)\right]$$

est

$$\widehat{\theta}_n \equiv \frac{1}{n}\sum_{i=1}^{n} g\left(X_i\right).$$

- Remarquons que $\widehat{\theta}_n$ est une fonction de l'échantillon $X_1, X_2, ..., X_n$. C'est une variable aléatoire.

Notation

1. Le paramètre devant être estimé : $\theta = \mathrm{E}\left[g\left(X\right)\right]$.

2. La variance : $\sigma^2 = \mathrm{Var}\left[g\left(X\right)\right] < \infty$.

3. L'estimateur : $\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} g\left(X_i\right)$

4. La variance échantillonnale :

$$\left(\widehat{\sigma}^{(n)}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(g\left(X_i\right) - \widehat{\theta}_n\right)^2$$

5. z_α est le quantile d'ordre $1-\alpha$ de la distribution normale centrée et réduite, c'est-à-dire que $\Pr\left[Z \leq z_\alpha\right] = 1-\alpha$.

L'échantillonnage stratégique (Importance sampling)

Mise en contexte

- Rappelons que le but est d'estimer

$$\theta = \mathrm{E}\left[g\left(X\right)\right] = \int_{-\infty}^{\infty} g\left(x\right) f_X\left(x\right) \; dx$$

  où X est une variable aléatoire de distribution quelconque dont les deux premiers moments existent et f_X est sa fonction de densité.
- L'idée de l'échantillonnage stratégique est de modifier la loi de X afin de réduire la variance et, ensuite, de compenser pour le changement de loi.

Changement de mesure

- Soit $\phi : \mathbb{R} \rightarrow [0, \infty)$, une fonction de densité telle que $\phi(x) \neq 0$ si $f_X(x) \neq 0$ (c'est pour éviter de diviser par zéro).
- Notons que

$$\begin{aligned}
\theta &= \mathrm{E}[g(X)] \\
&= \int_{-\infty}^{\infty} g(x) f_X(x) \, dx \\
&= \int_{-\infty}^{\infty} g(x) \frac{f_X(x)}{\phi(x)} \phi(x) \, dx \\
&= \mathrm{E}\left[g(Y) \frac{f_X(Y)}{\phi(Y)}\right]
\end{aligned}$$

où Y est une variable aléatoire possédant la fonction de densité ϕ.

Définition de l'estimateur

- Ainsi, pour estimer θ, nous pouvons générer les variables $X_1, ..., X_n$ selon la loi f_X et utiliser l'estimateur

$$\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} g\left(X_i\right)$$

ou nous pouvons générer les variables aléatoires $Y_1, ..., Y_n$ selon la loi ϕ et utiliser l'estimateur

$$\widehat{\theta}_n^{\phi} \equiv \frac{1}{n}\sum_{i=1}^{n} g\left(Y_i\right)\frac{f_X\left(Y_i\right)}{\phi\left(Y_i\right)}.$$

Changement de mesure

- Notons que l'estimateur de Monte Carlo utilisant l'échantillonnage stratégique est composé d'une somme de variables aléatoires indépendantes et identiquement distribuées. Nous pourrons ainsi :
  - utiliser la loi forte des grands nombres pour obtenir la convergence forte de l'estimateur vers θ
  - et le théorème limite central pour la distribution asymptotique de

$$\sqrt{n}\left(\widehat{\theta}_n^{\phi} - \theta\right).$$

- Quelle est la fonction de densité ϕ qu'il faut prendre?

Choix de la fonction de densité I

- Commençons par déterminer la variance de $\frac{g(Y) f_X(Y)}{\phi(Y)}$:

$$\begin{aligned}
& \mathrm{Var}\left[g(Y)\frac{f_X(Y)}{\phi(Y)}\right] \\
= \quad & \mathrm{E}\left[\left(g(Y)\frac{f_X(Y)}{\phi(Y)} - \underbrace{\mathrm{E}\left[g(Y)\frac{f_X(Y)}{\phi(Y)}\right]}_{=\theta}\right)^2\right] \\
= \quad & \int_{-\infty}^{\infty}\left(g(x)\frac{f_X(x)}{\phi(x)} - \theta\right)^2 \phi(x)\ dx.
\end{aligned}$$

La fonction de densité optimale $\widetilde{\phi}$ rendra la variance ci-dessus nulle, c'est-à-dire que

$$\left(g(x)\frac{f_X(x)}{\widetilde{\phi}(x)} - \theta\right)\widetilde{\phi}(x) = 0 \text{ pour tout } x \in \mathbb{R}\ .$$

Choix de la fonction de densité II

Rappel
La notation
Propriétés statistiques de l'estimateur
L'échantillonnage stratégique
Échantillonnage stratifié
Carré latin

- Rappelons que

$$\mathrm{Var}\left[g\left(Y\right)\frac{f_X\left(Y\right)}{\phi\left(Y\right)}\right]=\int_{-\infty}^{\infty}\left(g\left(x\right)\frac{f_X\left(x\right)}{\phi\left(x\right)}-\theta\right)^2\phi\left(x\right)\ dx.$$

- La variance sera nulle si

$$\widetilde{\phi}\left(x\right)=g\left(x\right)\frac{f_X\left(x\right)}{\theta}\text{ pour tout }x\in\mathbb{R}.$$

- Or, il n'est pas possible de calculer $\widetilde{\phi}$ car θ est la quantité inconnue que nous cherchons à estimer.

Choix de la fonction de densité III

- Notons que, puisque

$$\mathrm{Var}\left[g(Y)\frac{f_X(Y)}{\phi(Y)}\right] = \mathrm{E}\left[\left(g(Y)\frac{f_X(Y)}{\phi(Y)}\right)^2\right] - \theta^2,$$

minimiser la variance revient à minimiser le deuxième moment de la variable aléatoire $g(Y) f_X(Y) / \phi(Y)$.

- Attention! Un mauvais choix de densité stratégique peut augmenter la variance.

Changement de mesure

- Comme nous venons de le constater, le choix de la fonction de densité stratégique est difficile.
- Par contre, le contexte d'un cas particulier et la possibilité de se limiter à une famille de fonctions de densité permettent une bonne réduction de la variance.
- L'échantillonnage stratégique fonctionne particulièrement bien lorsque des événements rares sont impliqués.
  - C'est le cas lors de la tarification des options "très hors jeu" et lors de l'évaluation de la "valeur à risque".

Échantillonnage stratifié I

Contexte

- Pour expliquer l'échantillonnage stratifié, il nous faut revenir à la base même de la simulation.
- Pour générer une variable aléatoire de fonction de répartition F_X, il suffit de simuler une variable aléatoire U de loi $U(0,1)$ et de la transformer selon la relation suivante $X = F_X^{-1}(U)$.
  - Attention! Ce ne sont pas toutes les fonctions de répartition qui sont inversibles.
  - En effet, rappelons que $\forall 0 \leq u \leq 1$, $F_U(u) = u$, ce qui implique que

$$\mathrm{P}\left[F_X^{-1}(U) \leq x\right] = \mathrm{P}\left[U \leq F_X(x)\right] = F_X(x).$$

- L'échantillonnage stratifié consiste à échantillonner de sorte que certaines probabilités empiriques soient égales à leur correspondante théorique.

Échantillonnage stratifié II

Contexte

- Notons que le quantile d'ordre α, $0 \leq \alpha \leq 1$, de la distribution F_X est $F_X^{-1}(\alpha)$.
  - Attention! F_X doit respecter certaines conditions pour que l'on puisse définir les quantiles de cette façon.
  - En effet, $\mathrm{P}\left[X \leq F_X^{-1}(\alpha)\right] = F_X\left(F_X^{-1}(\alpha)\right) = \alpha$. Ainsi, pour tout $0 \leq \alpha < \beta \leq 1$,

$$\mathrm{P}\left[F_X^{-1}(\alpha) < X \leq F_X^{-1}(\beta)\right] = \beta - \alpha.$$

- D'autre part, pour tout $0 \leq \alpha < \beta \leq 1$, la variable aléatoire

$$X^{\alpha,\beta} \equiv F_X^{-1}\left(U^{\alpha,\beta}\right)$$

sera forcément comprise dans l'intervalle $\left[F_X^{-1}(\alpha), F_X^{-1}(\beta)\right]$ où $U^{\alpha,\beta}$ est une variable aléatoire de loi $U(\alpha, \beta)$.

Échantillonnage stratifié

Définition

1. Soit $0 = a_0 < a_1 < ... < a_{m-1} < a_m = 1$.
2. Définissons les intervalles

$$\begin{aligned} A_i &= \left(F_X^{-1}(a_{i-1}), F_X^{-1}(a_i)\right], \; i \in \{1, ..., m\} \text{ et} \\ A_{m+1} &= \left(F_X^{-1}(a_m), \infty\right). \end{aligned}$$

3. Notons que $\mathrm{P}[X \in A_i] = a_i - a_{i-1}, \; i \in \{1, ..., m+1\}$.
4. Nous simulons un échantillon de taille n.
   - 4.1 Simuler $n(a_i - a_{i-1})$ v. a. de loi $U(a_{i-1}, a_i)$
   - 4.2 Les transformer selon $X^{a_{i-1}, a_i} = F_X^{-1}\left(U^{a_{i-1}, a_i}\right)$
   - 4.3 Nous obtenons $100(a_i - a_{i-1})$ % de notre échantillon dans l'intervalle A_i.
   - 4.4 Les a_i devront être choisis de sorte que $n(a_i - a_{i-1})$ soit entier.
   - 4.5 La probabilité empirique que X soit dans l'intervalle A_i égale sa correspondante théorique.

Échantillonnage stratifié

Interprétation

- L'effet de l'échantillonnage stratifié est de s'assurer que toutes les régions du support de notre variable aléatoire X soit représentées dans notre échantillon.
- En particulier, les valeurs extrêmes qui ont généralement un grand impact sur le calcul de la moyenne, seront bien représentées.
- Cependant, l'échantillon ainsi produit n'est pas composé de variables aléatoires indépendantes et identiquement distribuées, ce qui rend difficile la caractérisation de la distribution de notre estimateur.

Rappel
La notation
Propriétés statistiques de l'estimateur
L'échantillonnage stratégique
Échantillonnage stratifié
Carré latin

Échantillonnage stratifié

Dimensions mulptiples

- En théorie, la stratification peut être utilisée pour simuler des vecteurs aléatoires de toutes dimensions. Il suffit de diviser l'hypercube unitaire $\mathcal{I}^d$ de dimension d, en sous-intervalle de la forme $I = \prod_{j=1}^{d} \left(\alpha_j, \beta_j\right]$ et d'utiliser des variables aléatoires $U\left(\alpha_j, \beta_j\right)$ afin de générer le nombre de points désirés dans I.
- En pratique, lorsque la dimension d est très élevée, il faut des échantillons de très grandes tailles afin que chaque sous-intervalle contienne un nombre raisonnable de points.
- Il existe une solution à ce problème : "*Latin hypercube sampling*". (Boyle, Broadie et Glasserman (1997), p.1280).

Carré latin

Contexte

1. C'est une extension de l'échantillonnage stratifié dans le cas où la dimension du problème est importante.
   1.1 S'il faut simuler d variables et que chacune d'elles est divisée en K strates, alors il faut un échantillon dont la taille minimale est K^d.
   1.2 Utiliser un nombre de strates K petit réduit l'efficacité de la réduction de la variance de la stratification.
   1.3 Si K est grand, alors K^d peut devenir énorme même lorsque la dimension n'est pas très élevée.
      1.3.1 Par exemple, si $K = 20$ et $d = 7$ alors $K^d = 20^7 = 1.28 \times 10^9$.

Carré latin I

Cas uniforme

1. Nous allons décrire la méthode dans le contexte où nous désirons simuler uniformément sur le cube unitaire de dimension d.
2. Fixons K.
3. **Génération de d échantillons**. Pour chaque dimension i, il faut générer de façon indépendante un échantillon stratifié $V_i^{(1)}, ..., V_i^{(K)}$ de l'intervalle $[0, 1]$ en utilisant K strates équiprobables.
   3.1 $V_i^{(j)}$ est donc uniformément distribuée sur l'intervalle $\left[\frac{j-1}{K}, \frac{j}{K}\right)$.

Carré latin II
Cas uniforme

4. Plaçons nos échantillons en colonnes :

$$\begin{matrix} V_1^{(1)} & V_2^{(1)} & \cdots & V_d^{(1)} \\ V_1^{(2)} & V_2^{(2)} & \cdots & V_d^{(2)} \\ \vdots & \vdots & & \vdots \\ V_1^{(K)} & V_2^{(K)} & \cdots & V_d^{(K)} \end{matrix}$$

4.1 Chaque ligne nous donne les coordonnées d'un point dans $[0,1)^d$.

4.1.1 La i ième ligne est identifiable à un point dans le *sous-hypercube* $\left[\frac{i-1}{K}, \frac{i}{K}\right)^d$.

4.1.2 Les K points sont donc à l'intérieur des sous-hypercubes situé le long de la diagonale de $[0,1)^d$.

Carré latin III

Cas uniforme

Rappel
La notation
Propriétés statistiques de l'estimateur
L'échantillonnage stratégique
Échantillonnage stratifié
Carré latin

5. Permutons aléatoirement les éléments d'une colonne.
   - 5.1 $\pi_1, ..., \pi_d$ sont des permutations de $\{1, 2, ..., K\}$ de sorte que les $K!$ permutations possibles ont la même probabilité d'occurence.
     - 5.1.1 Exemple. Si $K = 3$ alors les permutations possibles sont $\Pi = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}$. π_j est choisi uniformément dur l'ensemble Π.

Carré latin IV

Cas uniforme

5.2 Les lignes de

$$\begin{array}{cccc} V_1^{\pi_1(1)} & V_2^{\pi_2(1)} & \cdots & V_d^{\pi_d(1)} \\ V_1^{\pi_1(2)} & V_2^{\pi_2(2)} & \cdots & V_d^{\pi_d(2)} \\ \vdots & \vdots & & \vdots \\ V_1^{\pi_1(K)} & V_2^{\pi_2(K)} & \cdots & V_d^{\pi_d(K)} \end{array}$$

correspondent à des points dans $[0,1)^d$ mais ils n'appartiennent plus aux sous-hypercubes $\left[\frac{i-1}{K}, \frac{i}{K}\right)^d$ formant la diagonale de $[0,1)^d$.

5.2.1 Chaque ligne correspond à un point uniformément distribué sur $[0,1)^d$.

5.2.2 Les points formés par chacune des lignes ne sont pas indépendants.

5.2.3 Si on projette les points sur la i ème dimension, nous aurons une stratification de l'intervalle $[0,1)$.

Carré latin V

Cas uniforme

Rappel
La notation
Propriétés statistiques de l'estimateur
L'échantillonnage stratégique
Échantillonnage stratifié
Carré latin

Exemple. $d = 2$, $K = 5$.

•				
	○			
		$*$		
			$\times$	
				$+$

$$\pi_1 = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 4 \\ 5 \end{pmatrix} \longrightarrow \pi_2 = \begin{pmatrix} 1 \\ 5 \\ 4 \\ 3 \\ 2 \end{pmatrix}$$

			$*$	
				○
•				
		$\times$		
	$+$			

Dans cet exemple, un échantillonnage stratifié aurait demandé un échantillon de taille $5^2 = 25$.

Carré latin

Algorithme - cas uniforme

Algorithme. Pour générer un carré latin de taille K en dimension d, il faut

1. Générer des v.a. $U_i^{(j)}$ indépendantes de loi uniforme sur l'intervalle $[0,1)$ où $i = 1, 2, \ldots, d$ et $j = 1, 2, \ldots, K$.
2. $\pi_1, \ldots, \pi_d$ sont des permutations aléatoires et indépendantes de $\{1, 2, \ldots, K\}$.
   - 2.1 Pour simuler une permutation aléatoire, il suffit de simuler uniformément sur l'ensemble $\{1, 2, \ldots, K\}$, puis de simuler uniformément sur les $K-1$ points restants, et ainsi de suite jusqu'à ce qu'il ne reste qu'une seule valeur.
   - 2.2 Il est possible de forcer une des permutations à être l'identité sans modifier la distribution de l'échantillon.
3. $V_i^{(j)} = \frac{\pi_i(j) - 1 + U_i^{(j)}}{K}$, $i = 1, 2, \ldots, d$ et $j = 1, 2, \ldots, K$.

Carré latin

Algorithme - méthode d'inversion

1. En utilisant la méthode d'inversion, il est possible d'appliquer le carré latin à d'autres distributions que l'uniforme.
2. **Exemple**. Pour engendrer un hypercube latin de taille K d'une $N(\mathbf{0}, \mathbf{I})$ dans $\Re^d$ il suffit de poser
   2.1 $Z_i^{(j)} = \Phi^{-1}\left(V_i^{(j)}\right)$, $i = 1, 2, ..., d$ et $j = 1, 2, ..., K$ où Φ est la fonction de répartition d'un v.a. de loi normale centrée et réduite.
   2.2 L'échantillon consiste des vecteurs $\mathbf{Z}^{(j)} = \left(Z_1^{(j)}, ..., Z_d^{(j)}\right)$, $j = 1, 2, ..., K$.
   2.3 La projection de ces k points sur n'importe laquelle des dimensions donnera un échantillon stratifié de taille K de la loi normale univariée.
   2.4 Cette méthodologie est basée sur l'indépendance des d distributions marginales. On peut obtenir la dépendance en posant $\mathbf{X}^{(j)} = \mathbf{AZ}^{(j)}$ mais les marginales de $\mathbf{X}^{(j)}$ ne seront vraisemblablement plus stratifiées.