

Méthodes quasi-Monte Carlo (QMC)

On veut estimer

$$\mu = \int_{[0,1)^s} f(\mathbf{u}) d\mathbf{u}$$

par la moyenne de f sur l'ensemble $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subset [0, 1)^s$:

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n f(\mathbf{u}_i).$$

Méthodes quasi-Monte Carlo (QMC)

On veut estimer

$$\mu = \int_{[0,1)^s} f(\mathbf{u}) d\mathbf{u}$$

par la moyenne de f sur l'ensemble $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subset [0, 1)^s$:

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n f(\mathbf{u}_i).$$

Idée de QMC: Choisir $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ plus uniformément distribué qu'un ensemble de points au hasard.

Ensembles (et suites) de points à faible discrépance, ou hautement uniformes.

Méthodes quasi-Monte Carlo (QMC)

On veut estimer

$$\mu = \int_{[0,1)^s} f(\mathbf{u}) d\mathbf{u}$$

par la moyenne de f sur l'ensemble $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subset [0, 1)^s$:

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n f(\mathbf{u}_i).$$

Idée de QMC: Choisir $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ plus uniformément distribué qu'un ensemble de points au hasard.

Ensembles (et suites) de points à faible discrépance, ou hautement uniformes.

MC: erreur $E_n = \bar{\mu}_n - \mu$ aléatoire.

Méthodes quasi-Monte Carlo (QMC)

On veut estimer

$$\mu = \int_{[0,1)^s} f(\mathbf{u}) d\mathbf{u}$$

par la moyenne de f sur l'ensemble $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subset [0, 1)^s$:

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n f(\mathbf{u}_i).$$

Idée de QMC: Choisir $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ plus uniformément distribué qu'un ensemble de points au hasard.

Ensembles (et suites) de points à **faible discrédance**, ou **hautement uniformes**.

MC: erreur $E_n = \bar{\mu}_n - \mu$ aléatoire.

QMC classique: E_n est déterministe.

Cas simple: une dimension ($s = 1$)

Solutions évidentes:

$$P_n = \mathbb{Z}_n/n = \{0, 1/n, \dots, (n-1)/n\}:$$



qui donne

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=0}^{n-1} f(i/n),$$

Cas simple: une dimension ($s = 1$)

Solutions évidentes:

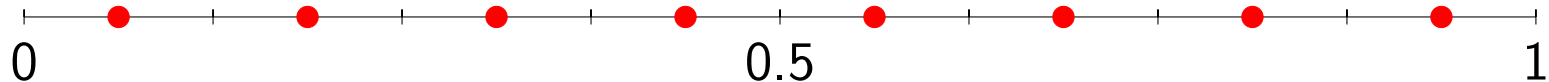
$$P_n = \mathbb{Z}_n/n = \{0, 1/n, \dots, (n-1)/n\}:$$



qui donne

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=0}^{n-1} f(i/n),$$

ou encore $P'_n = \{1/(2n), 3/(2n), \dots, (2n-1)/(2n)\}$:



Si on permet des **poids** différents pour les $f(\mathbf{u}_i)$, on a aussi la **règle du trapèze**:



$$\frac{1}{n} \left[\frac{f(0) + f(1)}{2} + \sum_{i=1}^{n-1} f(i/n) \right],$$

pour laquelle $|E_n| = O(n^{-2})$ si f'' est bornée,

Si on permet des **poids** différents pour les $f(\mathbf{u}_i)$, on a aussi la **règle du trapèze**:



$$\frac{1}{n} \left[\frac{f(0) + f(1)}{2} + \sum_{i=1}^{n-1} f(i/n) \right],$$

pour laquelle $|E_n| = O(n^{-2})$ si f'' est bornée, ou la **règle de Simpson**,

$$\frac{f(0) + 4f(1/n) + 2f(2/n) + \cdots + 2f((n-2)/n) + 4f((n-1)/n) + f(1)}{3n},$$

qui donne $|E_n| = O(n^{-4})$ si $f^{(4)}$ est bornée, etc.

Si on permet des **poids** différents pour les $f(\mathbf{u}_i)$, on a aussi la **règle du trapèze**:



$$\frac{1}{n} \left[\frac{f(0) + f(1)}{2} + \sum_{i=1}^{n-1} f(i/n) \right],$$

pour laquelle $|E_n| = O(n^{-2})$ si f'' est bornée, ou la **règle de Simpson**,

$$\frac{f(0) + 4f(1/n) + 2f(2/n) + \cdots + 2f((n-2)/n) + 4f((n-1)/n) + f(1)}{3n},$$

qui donne $|E_n| = O(n^{-4})$ si $f^{(4)}$ est bornée, etc.

Ici, pour QMC, on se restreint à des **poids égaux**. Plus tard, on va randomiser P_n pour mieux estimer l'erreur et dans ce contexte, les poids égaux sont optimaux.

En $s > 1$ dimensions, solution simpliste: grille rectangulaire

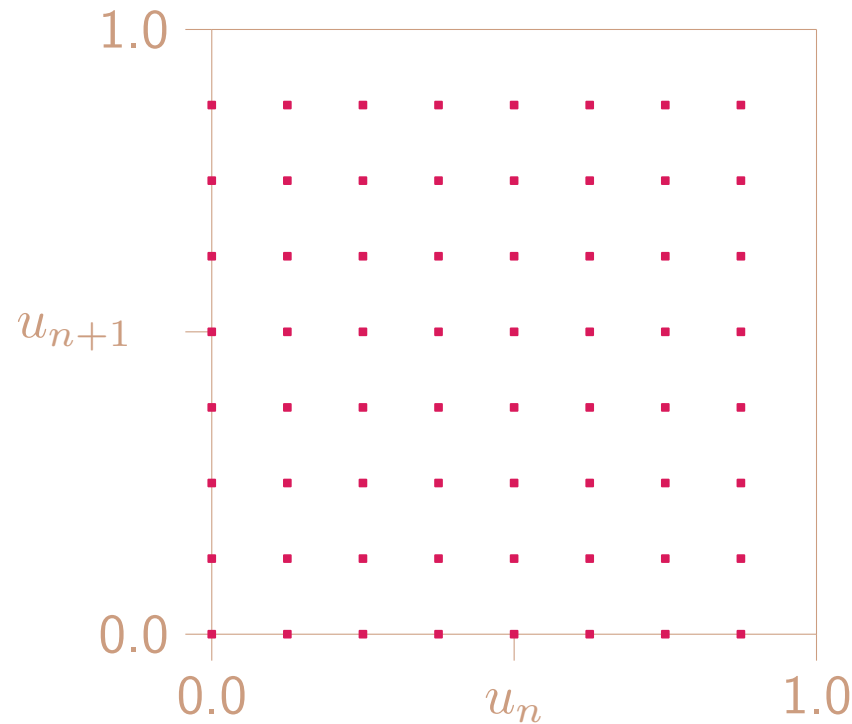
$$P_n = \{(i_1/d, \dots, i_t/d) \text{ tels que } 0 \leq i_j < d \ \forall j\},$$

où $n = d^s$.

En $s > 1$ dimensions, solution simpliste: grille rectangulaire

$$P_n = \{(i_1/d, \dots, i_t/d) \text{ tels que } 0 \leq i_j < d \ \forall j\},$$

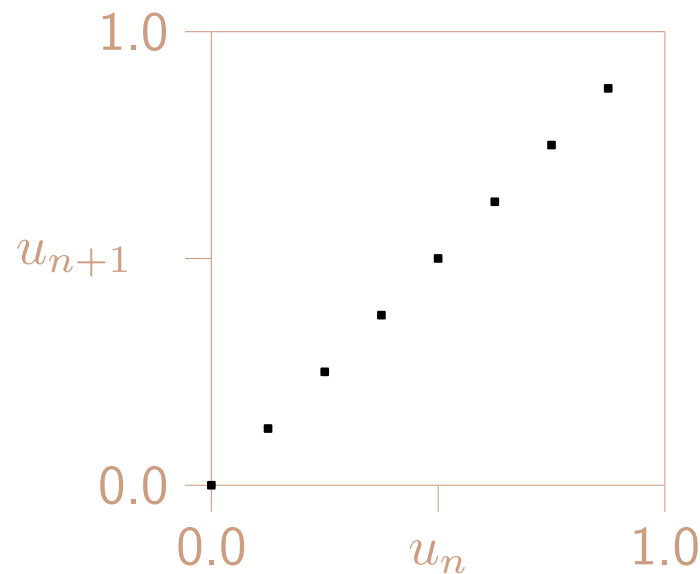
où $n = d^s$. Mais devient vite inutilisable quand s augmente.
Et on perd des points dans les projections.



Idée: on voudrait construire P_n de manière à ce que chaque projection unidimensionnelle soit $\{0, 1/n, \dots, (n-1)/n\}$.

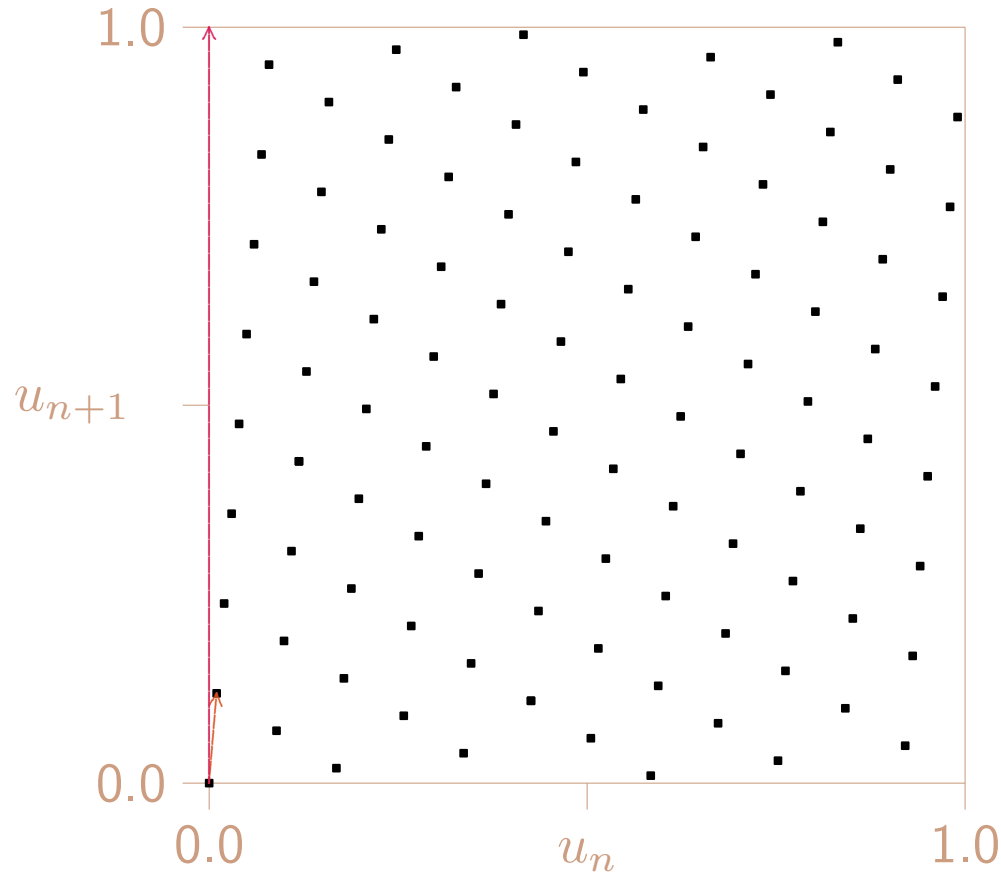
Idée: on voudrait construire P_n de manière à ce que chaque projection unidimensionnelle soit $\{0, 1/n, \dots, (n-1)/n\}$.

Pour cela, on doit énumérer ces n valeurs dans un ordre différent pour chaque coordonnée. (Autrement tous les points seront sur une diagonale.)



I.e., on cherche une **permutation** de \mathbb{Z}_n pour chacune des s coordonnées, et on veut les choisir pour que P_n soit très uniforme sur $[0, 1)^s$.

Exemple: les points produits par un GCL avec $m = 101$ et $a = 12$.
 Les \mathbf{u}_i sont tous les multiples entiers de $\mathbf{v} = (1/101, 12/101)$, modulo 1.



$$\begin{aligned}
 P_n &= \{(x/m, (ax/m) \bmod 1) : x = 0, \dots, m-1\} \\
 &= \{(x/101, (12x/101) \bmod 1) : x = 0, \dots, 100\} \\
 &= \{(0, 0), (1/101, 12/101), (2/101, 43/101), \dots\}.
 \end{aligned}$$

Example mini: Option asiatique avec $d = 2$.

$s = d = 2$, $T = 1$ (année), $t_j = j/2$ pour $j = 0, \dots, 2$,
 $K = 100$, $s_0 = 100$, $r = 0.05$, $\sigma = 0.5$.

$$\mu = e^{-rT} \int_{[0,1)^s} \max \left(0, \frac{1}{s} \sum_{i=1}^s S(0) \cdot \exp \left[(r - \sigma^2/2)t_i + \sigma \sum_{j=1}^i \sqrt{t_j - t_{j-1}} \Phi^{-1}(u_j) \right] - K \right) du_1 \dots du_s.$$

Example mini: Option asiatique avec $d = 2$.

$s = d = 2$, $T = 1$ (année), $t_j = j/2$ pour $j = 0, \dots, 2$,
 $K = 100$, $s_0 = 100$, $r = 0.05$, $\sigma = 0.5$.

$$\mu = e^{-rT} \int_{[0,1)^s} \max \left(0, \frac{1}{s} \sum_{i=1}^s S(0) \cdot \exp \left[(r - \sigma^2/2)t_i + \sigma \sum_{j=1}^i \sqrt{t_j - t_{j-1}} \Phi^{-1}(u_j) \right] - K \right) du_1 \dots du_s.$$

Valeur exacte: $\mu \approx 17.0958$.

Variance MC $\text{Var}[X_i] \approx 934.0$.

Example mini: Option asiatique avec $d = 2$.

$s = d = 2$, $T = 1$ (année), $t_j = j/2$ pour $j = 0, \dots, 2$,
 $K = 100$, $s_0 = 100$, $r = 0.05$, $\sigma = 0.5$.

$$\mu = e^{-rT} \int_{[0,1)^s} \max \left(0, \frac{1}{s} \sum_{i=1}^s S(0) \cdot \exp \left[(r - \sigma^2/2)t_i + \sigma \sum_{j=1}^i \sqrt{t_j - t_{j-1}} \Phi^{-1}(u_j) \right] - K \right) du_1 \dots du_s.$$

Valeur exacte: $\mu \approx 17.0958$.

Variance MC $\text{Var}[X_i] \approx 934.0$.

Approx. QMC avec $n = 100$ points de la fig. (enlève (0,0)): $\bar{\mu}_{100} = 17.6302$.

$\text{MSE}[\bar{\mu}_{100}] = (17.6302 - 17.0958)^2 = 0.2856$.

Pour comparer avec le même nombre de points: $\text{MSE}[\hat{\mu}_{100}] = \text{Var}[\hat{\mu}_{100}] \approx 9.34$.

Ou encore on multiplie le MSE de QMC par 100: 28.56 vs 934.

Les temps de calcul de $\bar{\mu}_{100}$ et $\hat{\mu}_{100}$ sont à peu près les mêmes.

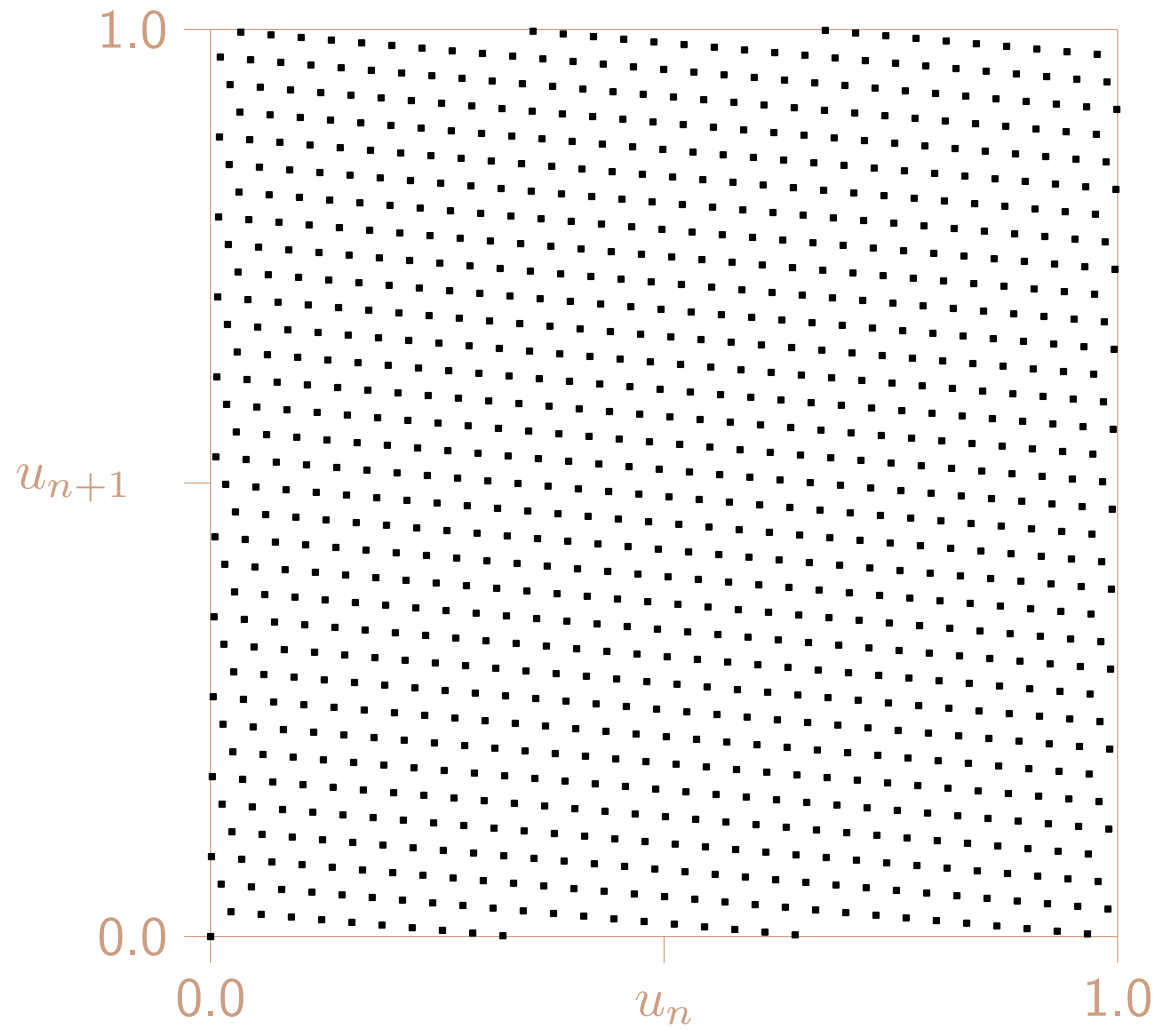
Donc QMC améliore l'efficacité par un facteur d'environ $934/28.56 \approx 32$.

Essayons maintenant avec $m = 65521$ et $a = 944$ ($n = 65520$).

On obtient $\bar{\mu}_{65520} = 17.0963$ et $\text{MSE}[\bar{\mu}_{65520}] \approx 2.5 \times 10^{-7}$.

L'efficacité est augmentée par un facteur d'environ
 $934.0 / (65520 \times 2.5 \times 10^{-7}) \approx 5.7 \times 10^4$.

C'est comme si on faisait 57 000 fois plus de simulations, pour le même prix!



Exemple: Un GCL avec $s = 2$ et $n = 1021$:

Les règles de réseaux (“lattice rules”) se généralisent en s dimensions.

Les règles de réseaux (“lattice rules”) se généralisent en s dimensions.

Un GCL de paramètres (n, a) avec $x_0 = i$ produit les s valeurs successives:

$$i/n, (ai \bmod n)/n, (a^2i \bmod n)/n, \dots, (a^{s-1}i \bmod n)/n.$$

Si on prend ces valeurs pour définir le point \mathbf{u}_i , pour $i = 0, \dots, n-1$, on obtient:

$$P_n = \{\mathbf{u}_i = (i\mathbf{a} \bmod n)/n = i\mathbf{v} \bmod 1 : i = 0, \dots, n-1\},$$

où $\mathbf{a} = (1, a, a^2 \bmod n, \dots, a^{s-1} \bmod n)$ et $\mathbf{v} = \mathbf{a}/n$.

Les règles de réseaux (“lattice rules”) se généralisent en s dimensions.

Un GCL de paramètres (n, a) avec $x_0 = i$ produit les s valeurs successives:

$$i/n, (ai \bmod n)/n, (a^2i \bmod n)/n, \dots, (a^{s-1}i \bmod n)/n.$$

Si on prend ces valeurs pour définir le point \mathbf{u}_i , pour $i = 0, \dots, n-1$, on obtient:

$$P_n = \{\mathbf{u}_i = (i\mathbf{a} \bmod n)/n = i\mathbf{v} \bmod 1 : i = 0, \dots, n-1\},$$

où $\mathbf{a} = (1, a, a^2 \bmod n, \dots, a^{s-1} \bmod n)$ et $\mathbf{v} = \mathbf{a}/n$.

Généralisation:

$$P_n = \{\mathbf{u}_i = i\mathbf{v} \bmod 1, i = 0, \dots, n-1\}$$

pour $\mathbf{a} = (1, a_1, \dots, a_{s-1})$ bien choisi et $\mathbf{v} = \mathbf{a}/n$.

Si $\text{pgcd}(n, a_j) = 1$, alors la coordonnée j parcourt tout \mathbb{Z}_n/n .

Example: Option asiatique avec $s = d = 12$.

Valeur exacte: $\mu \approx 13.122$.

Variance MC $\text{Var}[X_i] \approx 516.3$.

Example: Option asiatique avec $s = d = 12$.

Valeur exacte: $\mu \approx 13.122$.

Variance MC $\text{Var}[X_i] \approx 516.3$.

Approx. QMC avec $n = 100$ points: $\bar{\mu}_{100} = 12.4116$.

$100 \text{MSE}[\bar{\mu}_{100}] \approx 50.5$. Gain par un facteur 10.

Example: Option asiatique avec $s = d = 12$.

Valeur exacte: $\mu \approx 13.122$.

Variance MC $\text{Var}[X_i] \approx 516.3$.

Approx. QMC avec $n = 100$ points: $\bar{\mu}_{100} = 12.4116$.

$100 \text{MSE}[\bar{\mu}_{100}] \approx 50.5$. Gain par un facteur 10.

Approx. QMC avec $n = 65520$ points: $\bar{\mu}_{65520} = 13.1193$.

$65520 \text{MSE}[\bar{\mu}_{65520}] \approx 0.477$. Gain par un facteur 1000.

QMC randomisé (RQMC)

Comment estimer l'erreur avec QMC? Difficile. Pas d'intervalle de confiance.

QMC randomisé (RQMC)

Comment estimer l'erreur avec QMC? Difficile. Pas d'intervalle de confiance.

Mais on peut **randomiser** P_n de manière à:

- (1) préserver la structure et l'uniformité de P_n comme ensemble;
- (2) chaque point de P_n randomisé suit la loi uniforme sur $[0, 1)^s$.

Après la randomisation: $\mathbb{E}[\bar{\mu}_n] = \mathbb{E}[f(\mathbf{U}_i)] = \mu$.

QMC randomisé (RQMC)

Comment estimer l'erreur avec QMC? Difficile. Pas d'intervalle de confiance.

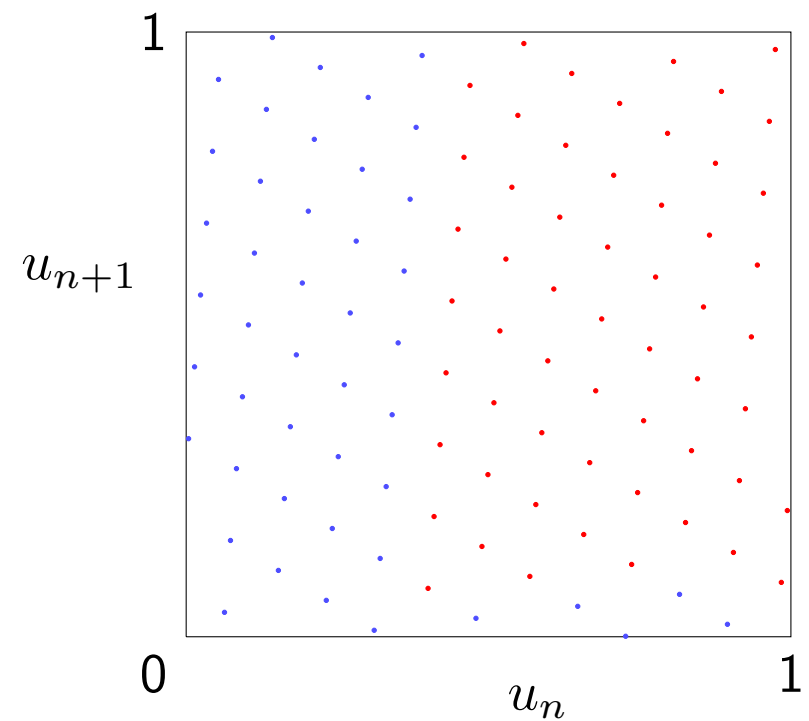
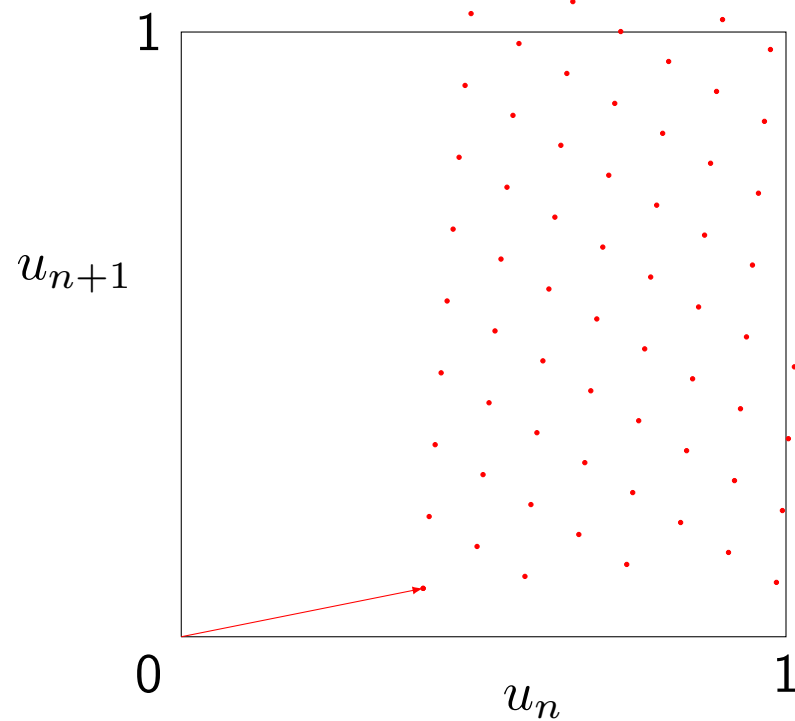
Mais on peut **randomiser** P_n de manière à:

- (1) préserver la structure et l'uniformité de P_n comme ensemble;
- (2) chaque point de P_n randomisé suit la loi uniforme sur $[0, 1)^s$.

Après la randomisation: $\mathbb{E}[\bar{\mu}_n] = \mathbb{E}[f(\mathbf{U}_i)] = \mu$.

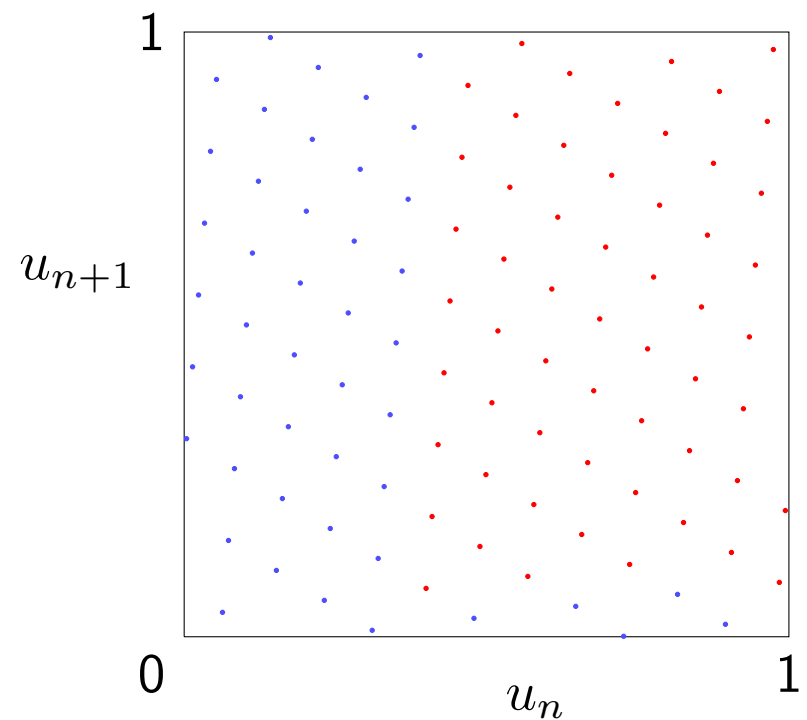
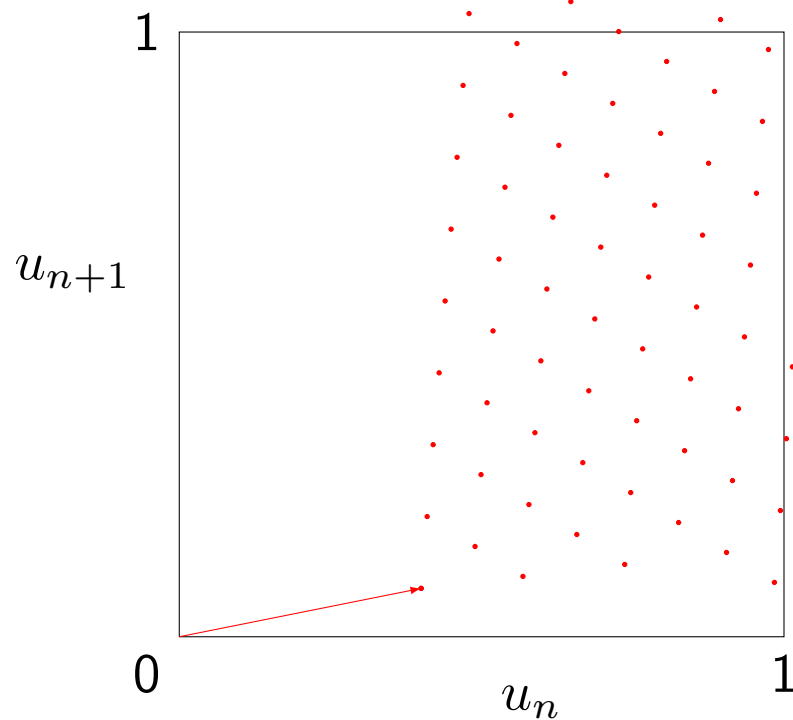
Exemple. Une façon de faire cela est le **décalage aléatoire**: générer un seul point **U** uniformément dans $[0, 1)^s$ et l'ajouter, modulo 1, à **chacun** des points de P_n . Plus besoin d'enlever le point $(0, 0)$.

Exemple avec le LCG (101, 12), avec $\mathbf{U} = (0.40, 0.08)$:



Pour estimer l'erreur, on répète la randomisation r fois, indépendamment, et on calcule la moyenne \bar{X}_r et la variance S_r^2 des r valeurs de $\bar{\mu}_n$.
On a $\mathbb{E}[\bar{X}_r] = \mu$ et $\mathbb{E}[S_r^2] = \text{Var}[\bar{\mu}_n] = r \text{Var}[\bar{X}_r]$.

Exemple avec le LCG (101, 12), avec $\mathbf{U} = (0.40, 0.08)$:



Pour estimer l'erreur, on répète la randomisation r fois, indépendamment, et on calcule la moyenne \bar{X}_r et la variance S_r^2 des r valeurs de $\bar{\mu}_n$.

On a $\mathbb{E}[\bar{X}_r] = \mu$ et $\mathbb{E}[S_r^2] = \text{Var}[\bar{\mu}_n] = r \text{Var}[\bar{X}_r]$.

Permet de calculer un intervalle de confiance pour μ .

Choix de r ? Souvent autour de 10.

Exemple: Option asiatique, $s = 2$.

RQMC (réseau + décalage aléatoire), $n = m$ points, $r = 1000$.

Pour $n = 101$: $\bar{X}_r = 17.076$ et $nS_r^2 = 77.9$.

Pour $n = 65521$: $\bar{X}_r = 17.095$ et $nS_r^2 = 4.03$.

Variance MC: 934.0.

Facteurs de réduction du MSE: 12.0 et 232.

Exemple: Option asiatique, $s = 2$.

RQMC (réseau + décalage aléatoire), $n = m$ points, $r = 1000$.

Pour $n = 101$: $\bar{X}_r = 17.076$ et $nS_r^2 = 77.9$.

Pour $n = 65521$: $\bar{X}_r = 17.095$ et $nS_r^2 = 4.03$.

Variance MC: 934.0.

Facteurs de réduction du MSE: 12.0 et 232.

Exemple: Option asiatique, $s = 12$.

Pour $n = 101$: $\bar{X}_r = 13.089$ et $nS_r^2 = 94.9$.

Pour $n = 65521$: $\bar{X}_r = 13.122$ et $nS_r^2 = 23.0$.

Variance MC: 516.3.

Facteurs de réduction du MSE: 5.4 et 22.

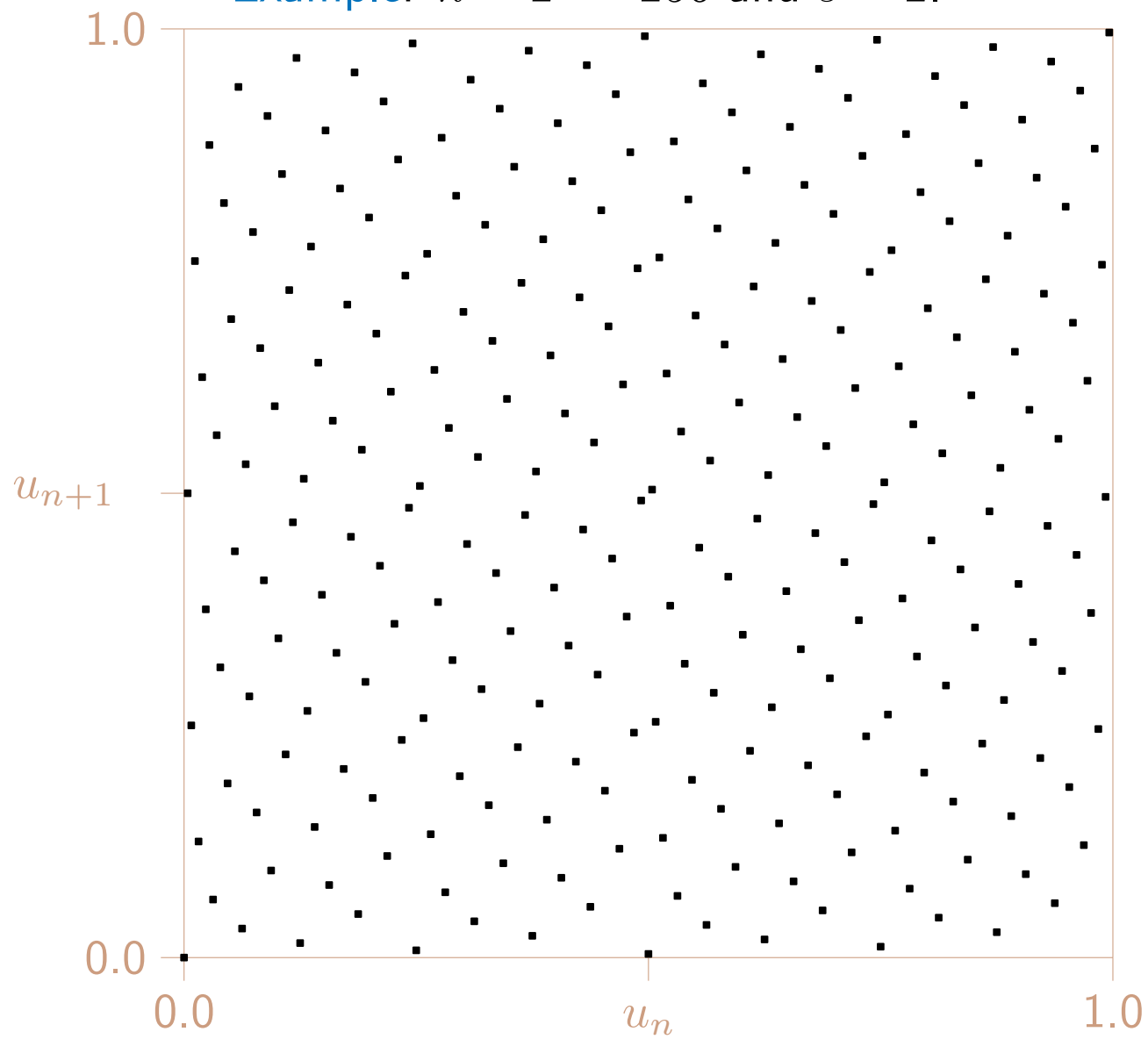
Exemple: Points de Hammersley:

Soient $n = 2^8 = 256$ et $s = 2$. Prenons les points (en binaire):

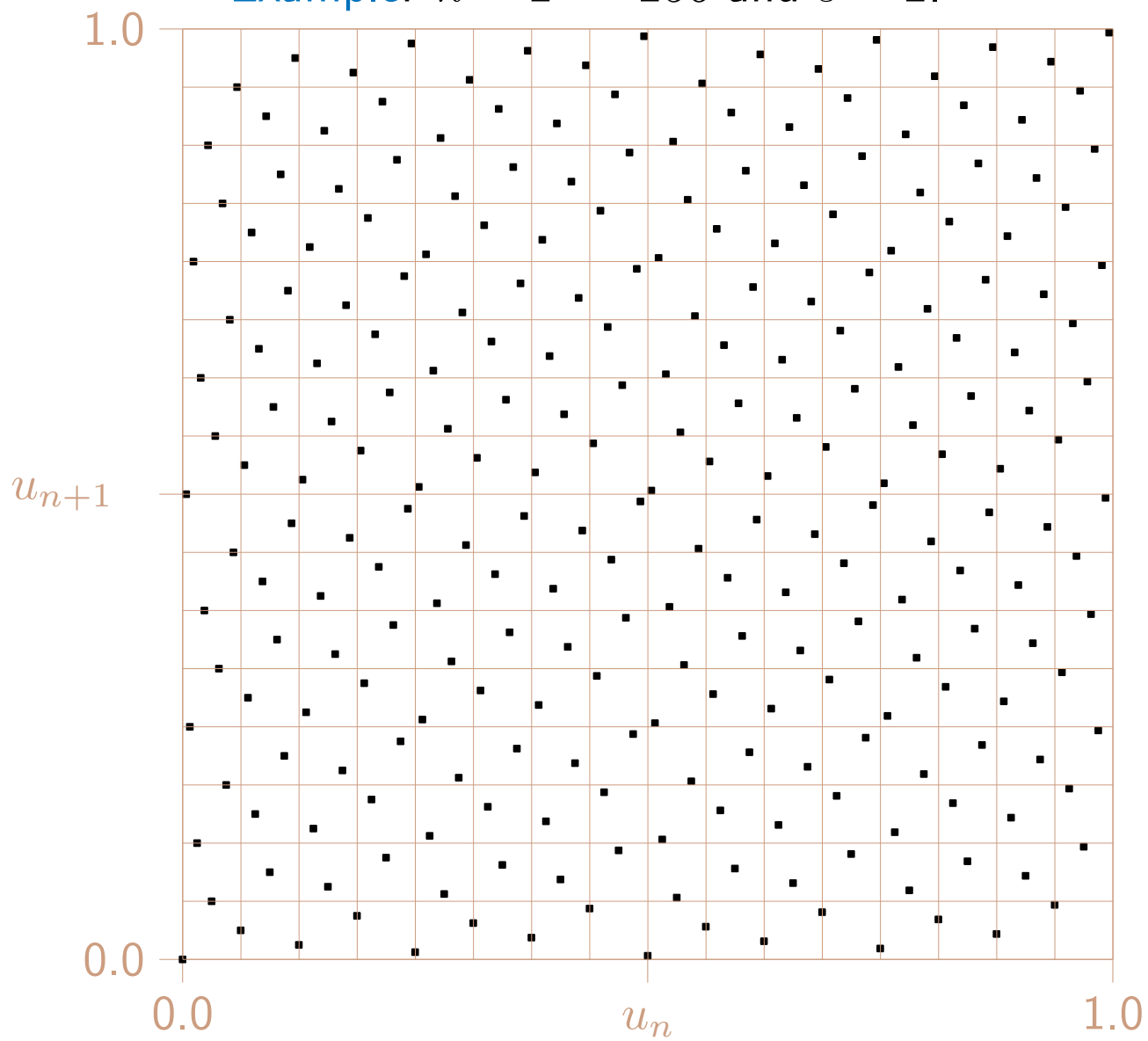
i	$u_{1,i}$	$u_{2,i}$
0	.00000000	.0
1	.00000001	.1
2	.00000010	.01
3	.00000011	.11
4	.00000100	.001
5	.00000101	.101
6	.00000110	.011
\vdots	\vdots	\vdots
254	.11111110	.01111111
255	.11111111	.11111111

À droite: suite de [van der Corput](#) en base 2.

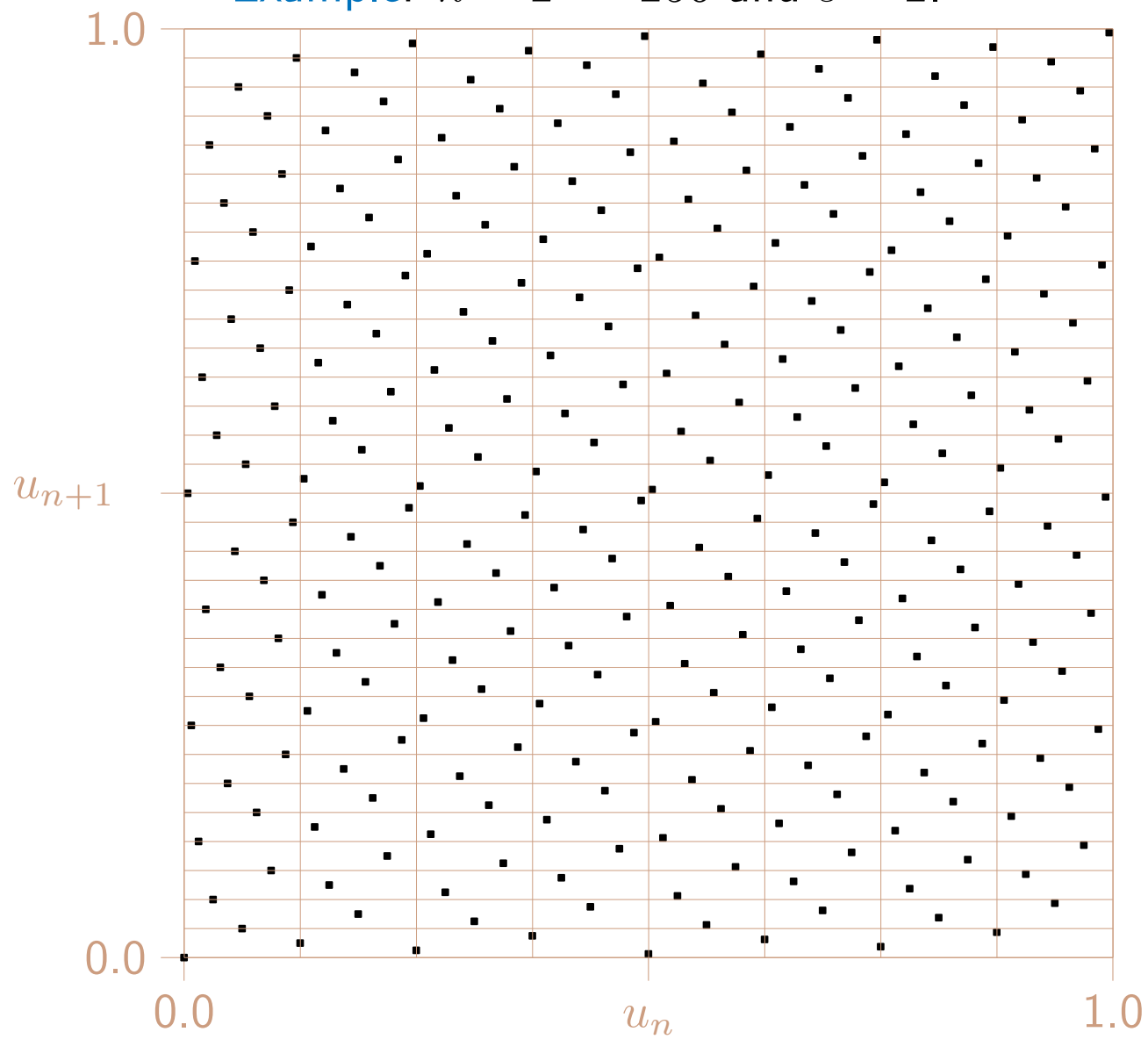
Example: $n = 2^8 = 256$ and $s = 2$.



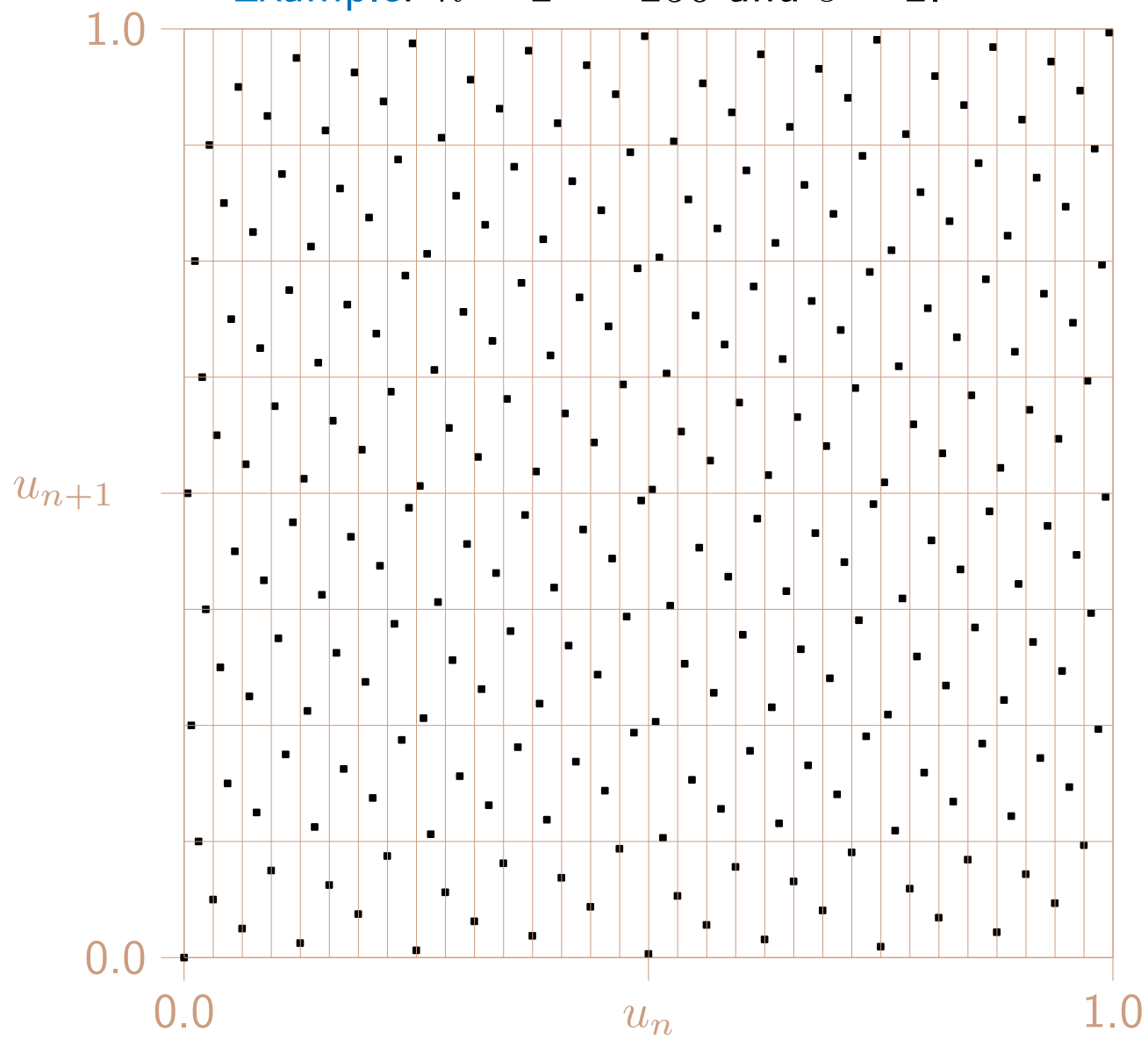
Example: $n = 2^8 = 256$ and $s = 2$.



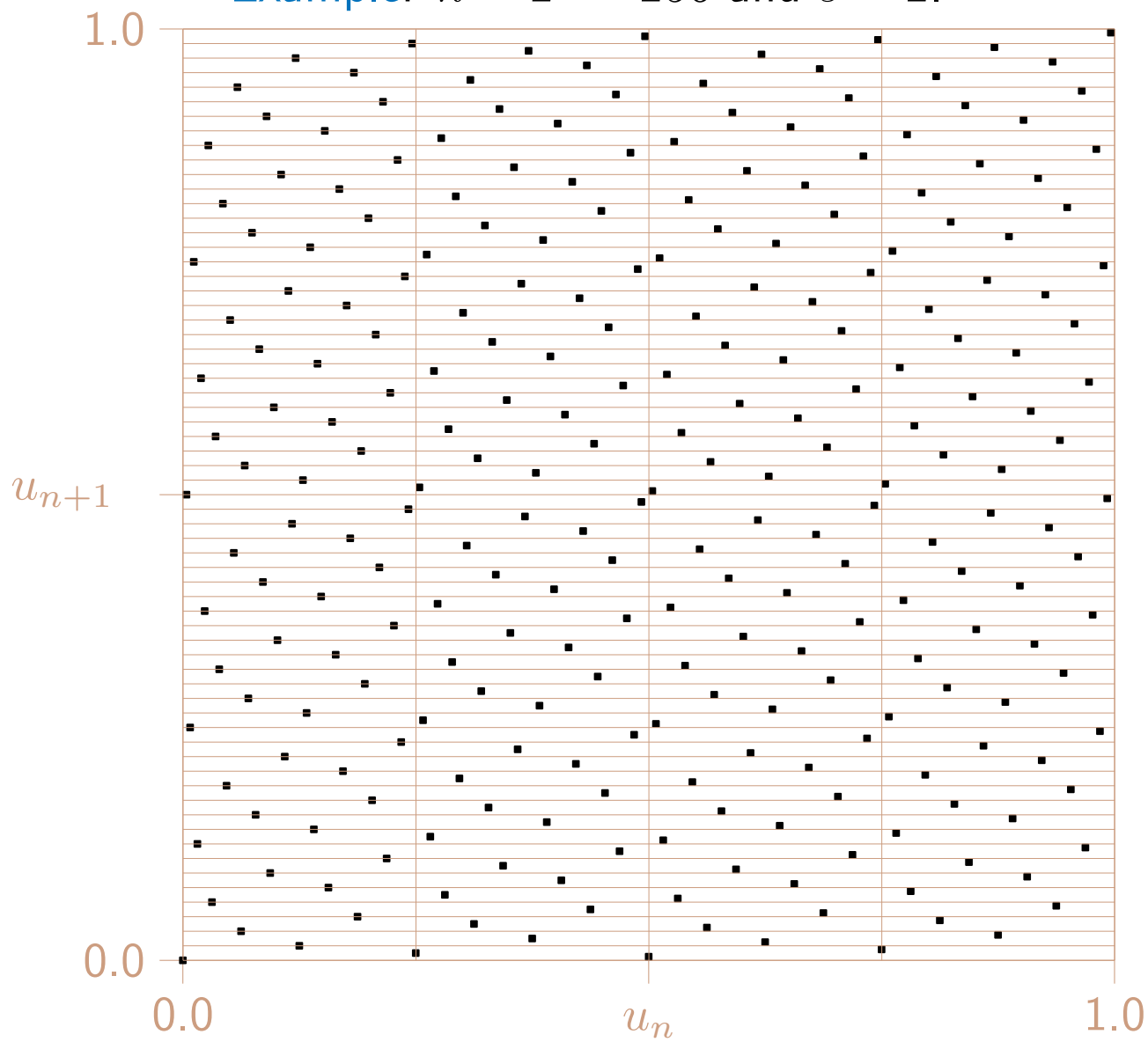
Example: $n = 2^8 = 256$ and $s = 2$.



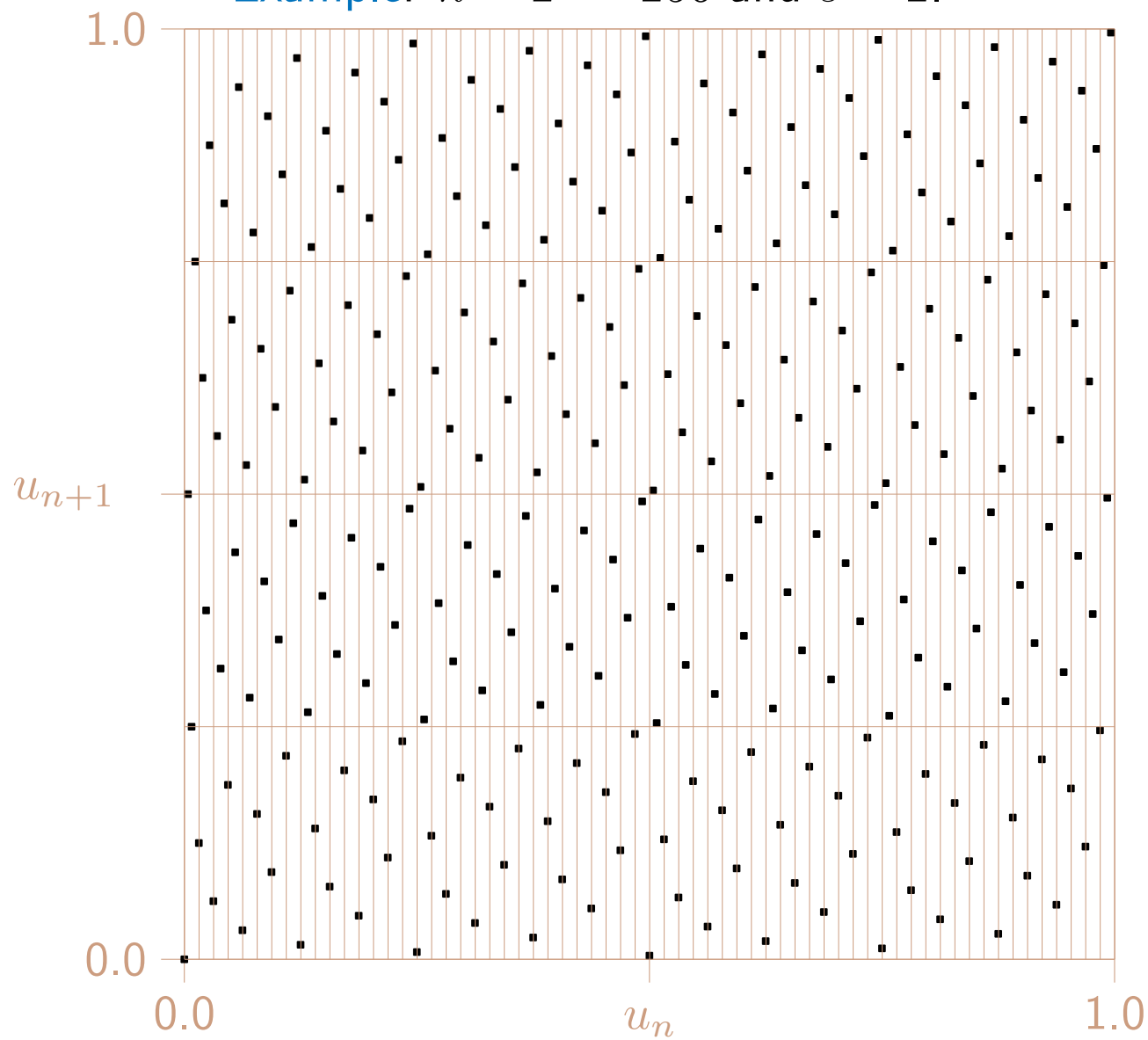
Example: $n = 2^8 = 256$ and $s = 2$.



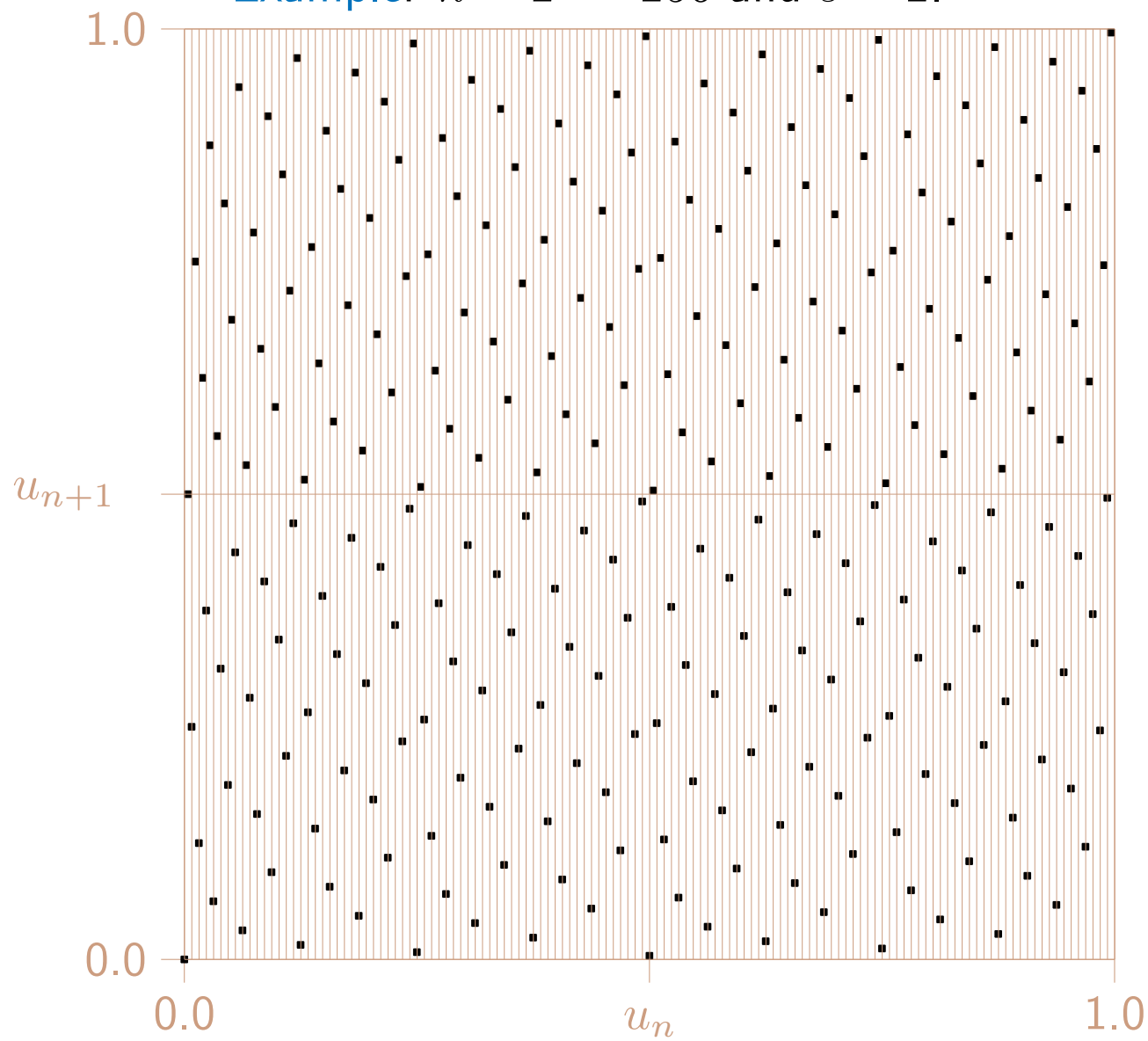
Example: $n = 2^8 = 256$ and $s = 2$.



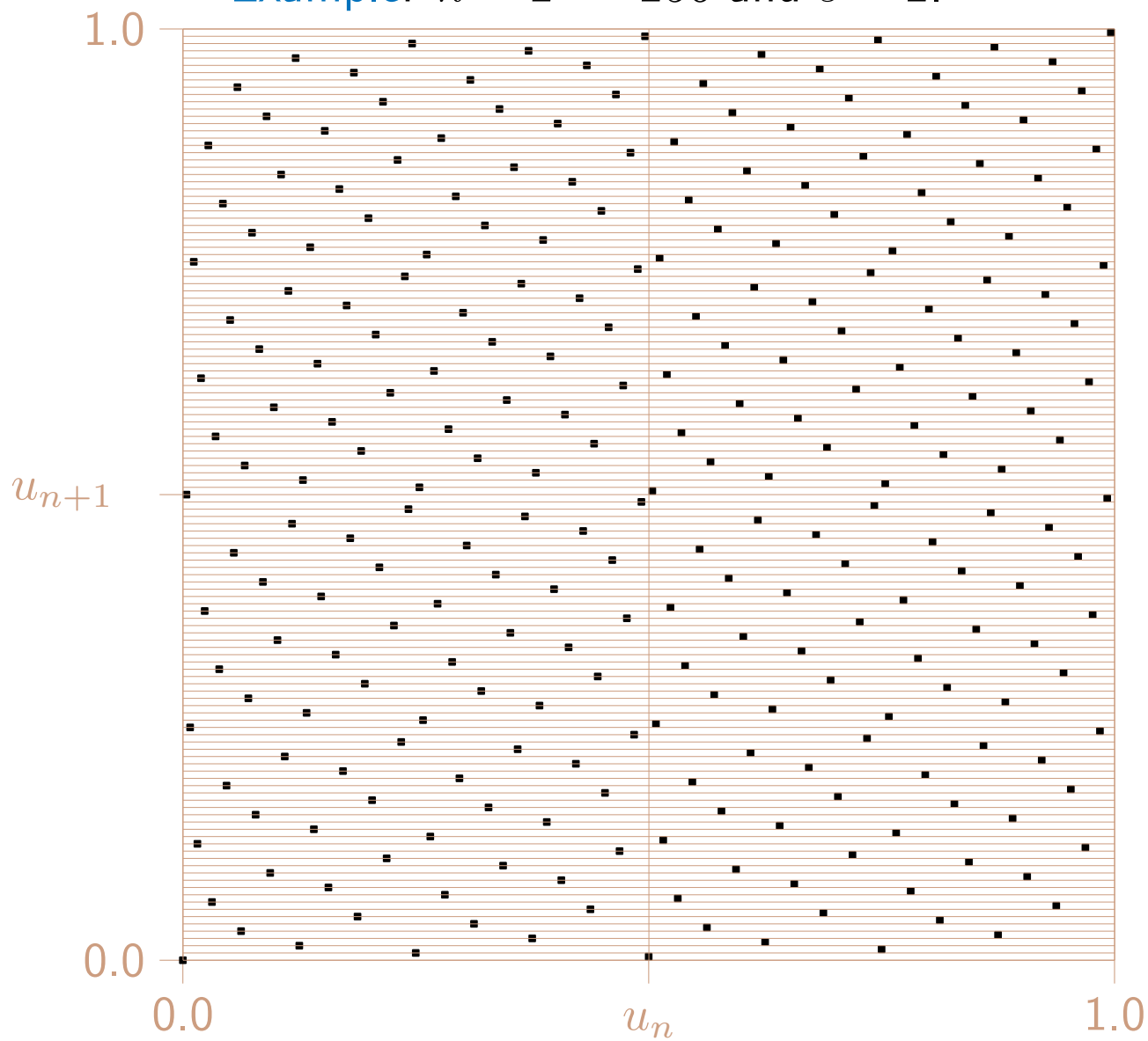
Example: $n = 2^8 = 256$ and $s = 2$.



Example: $n = 2^8 = 256$ and $s = 2$.



Example: $n = 2^8 = 256$ and $s = 2$.



En général, on peut prendre $n = 2^k$ points.

Si on partitionne $[0, 1)^2$ en rectangles de tailles 2^{-k_1} par 2^{-k_2} où $k_1 + k_2 \leq k$, chaque rectangle contiendra exactement le même nombre de points. On dit que les points sont **équidistribués** pour cette partition.

En général, on peut prendre $n = 2^k$ points.

Si on partitionne $[0, 1)^2$ en rectangles de tailles 2^{-k_1} par 2^{-k_2} où $k_1 + k_2 \leq k$, chaque rectangle contiendra exactement le même nombre de points. On dit que les points sont **équidistribués** pour cette partition.

Ce type de P_n est un cas particulier d'un **réseau digital** en base 2.

En général, on peut prendre $n = 2^k$ points.

Si on partitionne $[0, 1)^2$ en rectangles de tailles 2^{-k_1} par 2^{-k_2} où $k_1 + k_2 \leq k$, chaque rectangle contiendra exactement le même nombre de points. On dit que les points sont **équidistribués** pour cette partition.

Ce type de P_n est un cas particulier d'un **réseau digital** en base 2.

Pour un réseau digital en base 2, il suffit de choisir s permutations de $\{0, 1, \dots, 2^k - 1\}$, une pour chaque coordonnée, puis diviser par 2^k .

En général, on peut prendre $n = 2^k$ points.

Si on partitionne $[0, 1)^2$ en rectangles de tailles 2^{-k_1} par 2^{-k_2} où $k_1 + k_2 \leq k$, chaque rectangle contiendra exactement le même nombre de points. On dit que les points sont **équidistribués** pour cette partition.

Ce type de P_n est un cas particulier d'un **réseau digital** en base 2.

Pour un réseau digital en base 2, il suffit de choisir s permutations de $\{0, 1, \dots, 2^k - 1\}$, une pour chaque coordonnée, puis diviser par 2^k .

Se généralise en base $b > 2$.

On peut aussi avoir $s = \infty$ et/ou $n = \infty$ (**suite infinie** de points).

Randomisation.

Avec un décalage aléatoire modulo 1, l'équidistribution dans les boîtes digitales ne tient plus (il faudrait aussi décaler les boîtes).

Mais l'équidistribution tient si on fait un **décalage aléatoire digital** en base 2: On génère un point $\mathbf{U} \sim U(0, 1)^s$ et on fait un XOR bit à bit avec les points.

Exemple pour $s = 2$:

$$\begin{aligned}\mathbf{u}_i &= (0.01100100\dots, 0.10011000\dots) \\ \mathbf{U} &= (0.01001010\dots, 0.11101001\dots) \\ \mathbf{u}_i \oplus \mathbf{U} &= (0.00101110\dots, 0.01110001\dots).\end{aligned}$$

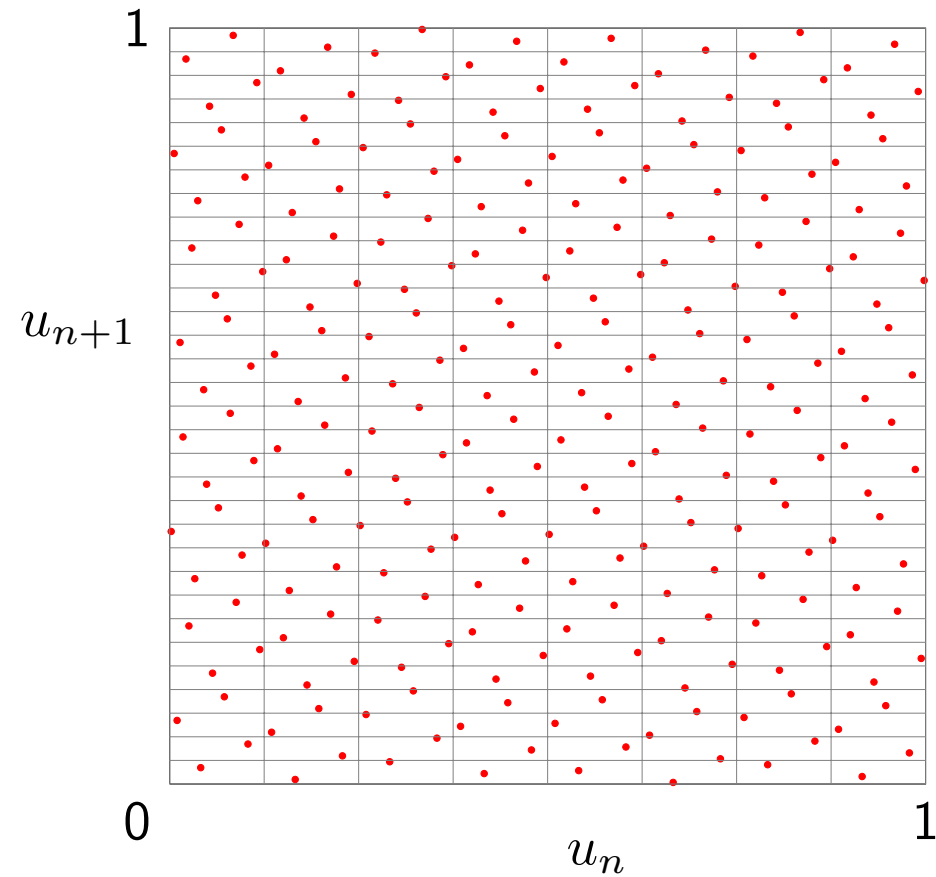
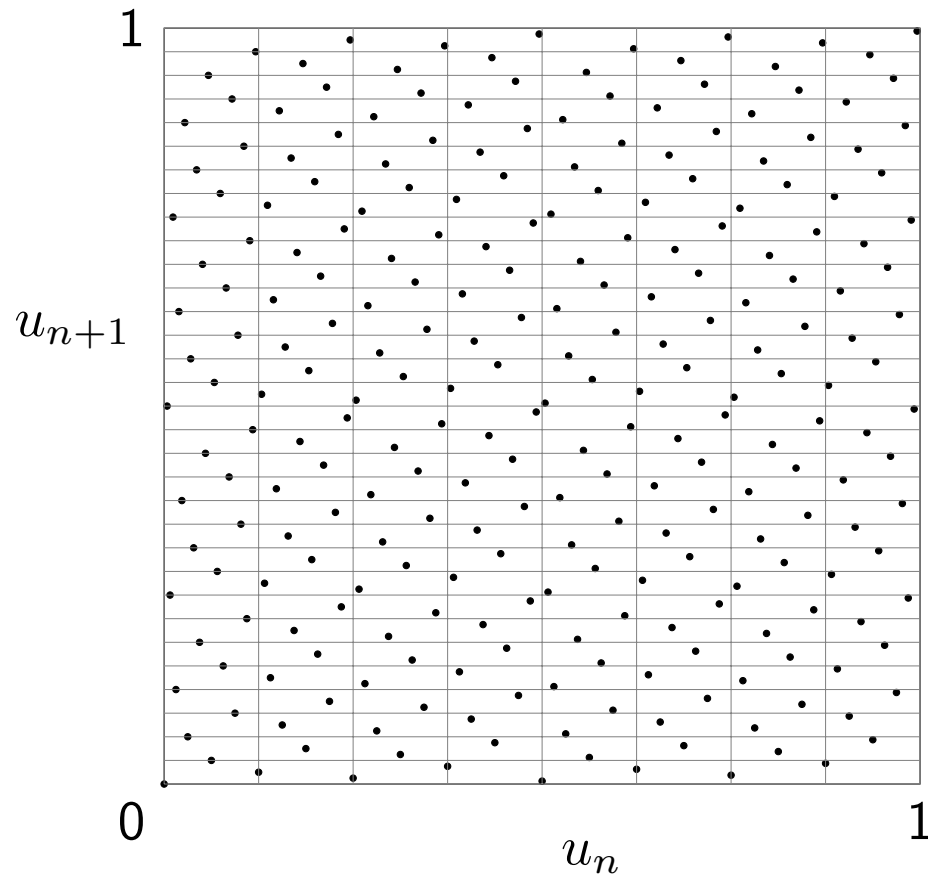
Expliquer pourquoi cela préserve l'équidistribution.

$$\begin{aligned}\mathbf{u}_i &= (0.***, 0.*****) \\ \mathbf{U} &= (0.010, 0.11101) \\ \mathbf{u}_i \oplus \mathbf{U} &= (0.***, 0.*****)\end{aligned}$$

Exemple avec

$$\begin{aligned}\mathbf{U} &= (0.1270111220, 0.3185275653) \\ &= (0.00100000100000111100, 0.01010001100010110000)_2.\end{aligned}$$

Change les bits 3, 9, 15, 16, 17, 18 de $u_{i,1}$ et les bits 2, 4, 8, 9, 13, 15, 16 de $u_{i,2}$.



Example: Même expérience qu’avec le réseau et décalage aléatoire, mais ici avec les points de Hammersley en base 2, randomisés avec un “scramble matriciel” (à voir plus tard) + décalage aléatoire digital.

Example: Même expérience qu’avec le réseau et décalage aléatoire, mais ici avec les points de Hammersley en base 2, randomisés avec un “scramble matriciel” (à voir plus tard) + décalage aléatoire digital.

Exemple: Option asiatique, $s = 2$.

Pour $n = 2^{10} = 1024$: $\bar{X}_r = 17.096$ et $nS_r^2 = 1.815$.

Pour $n = 2^{16} = 65536$: $\bar{X}_r = 17.096$ et $nS_r^2 = 0.034$.

Variance MC: 934.0.

Facteurs de réduction du MSE: 515 et 27,120.

Example: Même expérience qu'avec le réseau et décalage aléatoire, mais ici avec les points de Hammersley en base 2, randomisés avec un “scramble matriciel” (à voir plus tard) + décalage aléatoire digital.

Exemple: Option asiatique, $s = 2$.

Pour $n = 2^{10} = 1024$: $\bar{X}_r = 17.096$ et $nS_r^2 = 1.815$.

Pour $n = 2^{16} = 65536$: $\bar{X}_r = 17.096$ et $nS_r^2 = 0.034$.

Variance MC: 934.0.

Facteurs de réduction du MSE: 515 et 27,120.

Exemple: Option asiatique, $s = 12$. Points de Sobol' en 12 dimensions.

Pour $n = 2^{10} = 1024$: $\bar{X}_r = 13.122$ et $nS_r^2 = 6.2$.

Pour $n = 2^{16} = 65536$: $\bar{X}_r = 13.122$ et $nS_r^2 = 1.7$.

Variance MC: 516.3.

Facteurs de réduction du MSE: 84 et 304.

Discrédance, convergence et borne de Koksma-Hlawka

Comment mesurer l'uniformité d'un ensemble de points?

Idée: pour B un sous-ensemble quelconque de $[0, 1)^s$, la fraction de P_n qui se trouve dans B devrait être à peu près égale au volume de B .

Possible pour tout B ?

Discrépance, convergence et borne de Koksma-Hlawka

Comment mesurer l'uniformité d'un ensemble de points?

Idée: pour B un sous-ensemble quelconque de $[0, 1)^s$, la fraction de P_n qui se trouve dans B devrait être à peu près égale au volume de B .

Possible pour tout B ? Non.

Discrépance, convergence et borne de Koksma-Hlawka

Comment mesurer l'uniformité d'un ensemble de points?

Idée: pour B un sous-ensemble quelconque de $[0, 1)^s$, la fraction de P_n qui se trouve dans B devrait être à peu près égale au volume de B .

Possible pour tout B ? **Non.**

On peut se limiter à une famille \mathcal{I} de boîtes B , par exemple la famille \mathcal{I}_s^* de toutes les boîtes rectangulaires alignées avec les axes et ayant un coin à l'origine (les intervalles de forme $[0, \mathbf{u})$ en s dimensions).

Discr pance, convergence et borne de Koksma-Hlawka

Comment mesurer l'uniformit  d'un ensemble de points?

Id e: pour B un sous-ensemble quelconque de $[0, 1)^s$, la fraction de P_n qui se trouve dans B devrait  tre   peu pr s  gale au volume de B .

Possible pour tout B ? **Non.**

On peut se limiter   une famille \mathcal{I} de bo tes B , par exemple la famille \mathcal{I}_s^* de toutes les bo tes rectangulaires align es avec les axes et ayant un coin   l'origine (les intervalles de forme $[0, \mathbf{u})$ en s dimensions).

On d finit $\Delta(B) = |(fraction\ de\ P_n\ qui\ est\ dans\ B) - volume(B)|$.

Discr pance, convergence et borne de Koksma-Hlawka

Comment mesurer l'uniformit  d'un ensemble de points?

Id e: pour B un sous-ensemble quelconque de $[0, 1)^s$, la fraction de P_n qui se trouve dans B devrait  tre   peu pr s  gale au volume de B .

Possible pour tout B ? **Non.**

On peut se limiter   une famille \mathcal{I} de bo tes B , par exemple la famille \mathcal{I}_s^* de toutes les bo tes rectangulaires align es avec les axes et ayant un coin   l'origine (les intervalles de forme $[0, \mathbf{u})$ en s dimensions).

On d finit $\Delta(B) = |(\text{fraction de } P_n \text{ qui est dans } B) - \text{volume}(B)|$.

Mesure de **discr pance** pour P_n :

$$D_n^*(P_n) = \sup_{B=[0, \mathbf{u}) \in \mathcal{I}_s^*} \Delta(B).$$

Discr pance, convergence et borne de Koksma-Hlawka

Comment mesurer l'uniformit  d'un ensemble de points?

Id e: pour B un sous-ensemble quelconque de $[0, 1)^s$, la fraction de P_n qui se trouve dans B devrait  tre   peu pr s  gale au volume de B .

Possible pour tout B ? **Non.**

On peut se limiter   une famille \mathcal{I} de bo tes B , par exemple la famille \mathcal{I}_s^* de toutes les bo tes rectangulaires align es avec les axes et ayant un coin   l'origine (les intervalles de forme $[0, \mathbf{u})$ en s dimensions).

On d finit $\Delta(B) = |(\text{fraction de } P_n \text{ qui est dans } B) - \text{volume}(B)|$.

Mesure de **discr pance** pour P_n :

$$D_n^*(P_n) = \sup_{B=[0, \mathbf{u}) \in \mathcal{I}_s^*} \Delta(B).$$

Il y a d'autres possibilit s: moyenne \mathcal{L}_p au lieu du sup, autre choix de \mathcal{I} , etc.

Inégalité de **Koksma-Hlawka** (erreur pire cas):

$$|\bar{\mu}_n - \mu| \leq \|f - \mu\| D_n^*(P_n)$$

où $\|f - \mu\|$ mesure la **variation de f** au sens de Hardy et Krause.

Inégalité de **Koksma-Hlawka** (erreur pire cas):

$$|\bar{\mu}_n - \mu| \leq \|f - \mu\| D_n^*(P_n)$$

où $\|f - \mu\|$ mesure la **variation de f** au sens de Hardy et Krause.

On connaît des suites infinies de points $P_\infty = \{\mathbf{u}_1, \mathbf{u}_2, \dots\}$ telle que pour $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$, $D_n^*(P_n) = O(n^{-1}(\ln n)^s)$.

Inégalité de **Koksma-Hlawka** (erreur pire cas):

$$|\bar{\mu}_n - \mu| \leq \|f - \mu\| D_n^*(P_n)$$

où $\|f - \mu\|$ mesure la **variation de f** au sens de Hardy et Krause.

On connaît des suites infinies de points $P_\infty = \{\mathbf{u}_1, \mathbf{u}_2, \dots\}$ telle que pour $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$, $D_n^*(P_n) = O(n^{-1}(\ln n)^s)$.

Asymptotiquement, c'est mieux que $O(n^{-1/2})$.

Inégalité de **Koksma-Hlawka** (erreur pire cas):

$$|\bar{\mu}_n - \mu| \leq \|f - \mu\| D_n^*(P_n)$$

où $\|f - \mu\|$ mesure la **variation de f** au sens de Hardy et Krause.

On connaît des suites infinies de points $P_\infty = \{\mathbf{u}_1, \mathbf{u}_2, \dots\}$ telle que pour $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$, $D_n^*(P_n) = O(n^{-1}(\ln n)^s)$.

Asymptotiquement, c'est mieux que $O(n^{-1/2})$.

Mais: la borne est difficile à calculer et ne peut être pratique que lorsque s ne dépasse pas 8 ou 10.

Questions:

Est-ce que $E_n \rightarrow 0$ plus vite avec QMC que MC quand $n \rightarrow \infty$?

Questions:

Est-ce que $E_n \rightarrow 0$ plus vite avec QMC que MC quand $n \rightarrow \infty$?

Comment construire P_n et mesurer son uniformité?

Questions:

Est-ce que $E_n \rightarrow 0$ plus vite avec QMC que MC quand $n \rightarrow \infty$?

Comment construire P_n et mesurer son uniformité?

Facile en 1 dimension. Mais en plusieurs dimensions?

Pour s grand, il faut trop de points pour vraiment bien remplir l'espace.

Est-ce que QMC peut quand même fonctionner? Pourquoi?

Questions:

Est-ce que $E_n \rightarrow 0$ plus vite avec QMC que MC quand $n \rightarrow \infty$?

Comment construire P_n et mesurer son uniformité?

Facile en 1 dimension. Mais en plusieurs dimensions?

Pour s grand, il faut trop de points pour vraiment bien remplir l'espace.

Est-ce que QMC peut quand même fonctionner? Pourquoi?

Dans plusieurs cas, f peut s'approximer assez bien par une somme de fonctions f_I dont chacune ne dépend que de peu de coordonnées de \mathbf{u} :

$$\begin{aligned} f(\mathbf{u}) &= f(u_1, \dots, u_s) \\ &= \sum_{I \subseteq \{1, \dots, s\}} f_I(\mathbf{u}) = \mu + \sum_{i=1}^s \dot{f}_{\{i\}}(u_i) + \sum_{i,j=1}^s \dot{f}_{\{i,j\}}(u_i, u_j) + \dots \end{aligned}$$

où f_I ne dépend que de $\{u_i, i \in I\}$.

Questions:

Est-ce que $E_n \rightarrow 0$ plus vite avec QMC que MC quand $n \rightarrow \infty$?

Comment construire P_n et mesurer son uniformité?

Facile en 1 dimension. Mais en plusieurs dimensions?

Pour s grand, il faut trop de points pour vraiment bien remplir l'espace.

Est-ce que QMC peut quand même fonctionner? Pourquoi?

Dans plusieurs cas, f peut s'approximer assez bien par une somme de fonctions f_I dont chacune ne dépend que de peu de coordonnées de \mathbf{u} :

$$\begin{aligned} f(\mathbf{u}) &= f(u_1, \dots, u_s) \\ &= \sum_{I \subseteq \{1, \dots, s\}} f_I(\mathbf{u}) = \mu + \sum_{i=1}^s \dot{f}_{\{i\}}(u_i) + \sum_{i,j=1}^s \dot{f}_{\{i,j\}}(u_i, u_j) + \dots \end{aligned}$$

où f_I ne dépend que de $\{u_i, i \in I\}$.

Il suffit alors que les **projections** correspondantes de P_n soient bien uniformes.

Choix de la loi d'échantillonnage

Exemple: Intégrale sur \mathbb{R} . Supposons que l'on veut estimer

$$\mu = \int_{-\infty}^{\infty} f(x) dx < \infty,$$

où $f : \mathbb{R} \rightarrow [0, \infty)$.

Soit g la densité d'une v.a. Y telle que $g(x) > 0$ quand $f(x) > 0$. On a

$$\mu = \int_{-\infty}^{\infty} [f(y)/g(y)]g(y)dy = \mathbb{E}[f(Y)/g(Y)].$$

Choix de la loi d'échantillonnage

Exemple: Intégrale sur \mathbb{R} . Supposons que l'on veut estimer

$$\mu = \int_{-\infty}^{\infty} f(x) dx < \infty,$$

où $f : \mathbb{R} \rightarrow [0, \infty)$.

Soit g la densité d'une v.a. Y telle que $g(x) > 0$ quand $f(x) > 0$. On a

$$\mu = \int_{-\infty}^{\infty} [f(y)/g(y)]g(y)dy = \mathbb{E}[f(Y)/g(Y)].$$

Donc pour estimer μ (sans biais), il suffit de générer n copies i.i.d. de Y , disons Y_1, \dots, Y_n , et prendre la moyenne:

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n f(Y_i)/g(Y_i).$$

Par exemple, g pourrait être une densité normale.

On a $\text{Var}[\hat{\mu}_n] = \text{Var}[f(Y)/g(Y)]/n$, où

$$\text{Var} \left[\frac{f(Y)}{g(Y)} \right] = \mathbb{E} \left[\frac{f^2(Y)}{g^2(Y)} \right] - \mu^2 = \int_{-\infty}^{\infty} \frac{f^2(y)}{g^2(y)} g(y) dy - \mu^2 = \int_{-\infty}^{\infty} \frac{f^2(y)}{g(y)} dy - \mu^2.$$

Attention: si $g(y) \ll f(y)$ quelque part, ou si $f^2(y)/g(y)$ converge vers 0 trop lentement lorsque $y \rightarrow \pm\infty$, par exemple, alors la variance peut être très grande, voire infinie. Donc le choix de g est crucial.

Exemple (simple): estimer une probabilité.

On veut estimer $p = \mathbb{P}[Y > y_0] = \mathbb{E}[\mathbb{I}[Y > y_0]]$, où Y est exponentielle de paramètre (taux) λ , i.e., $\pi(y) = \lambda e^{-\lambda y}$ pour $y \geq 0$.

Exemple (simple): estimer une probabilité.

On veut estimer $p = \mathbb{P}[Y > y_0] = \mathbb{E}[\mathbb{I}[Y > y_0]]$, où Y est exponentielle de paramètre (taux) λ , i.e., $\pi(y) = \lambda e^{-\lambda y}$ pour $y \geq 0$.

(C'est une illustration purement académique; on sait que $\mathbb{P}[Y > y] = e^{-\lambda y}$.)

Exemple (simple): estimer une probabilité.

On veut estimer $p = \mathbb{P}[Y > y_0] = \mathbb{E}[\mathbb{I}[Y > y_0]]$, où Y est exponentielle de paramètre (taux) λ , i.e., $\pi(y) = \lambda e^{-\lambda y}$ pour $y \geq 0$.

(C'est une illustration purement académique; on sait que $\mathbb{P}[Y > y] = e^{-\lambda y}$.)

Monte Carlo: générer Y selon la bonne densité exponentielle π , soit $Y = -\ln(1 - U)/\lambda$, et calculer $X = \mathbb{I}[Y > y_0]$.

Répéter n fois et calculer la moyenne des n réalisations de X .

Exemple (simple): estimer une probabilité.

On veut estimer $p = \mathbb{P}[Y > y_0] = \mathbb{E}[\mathbb{I}[Y > y_0]]$, où Y est exponentielle de paramètre (taux) λ , i.e., $\pi(y) = \lambda e^{-\lambda y}$ pour $y \geq 0$.

(C'est une illustration purement académique; on sait que $\mathbb{P}[Y > y] = e^{-\lambda y}$.)

Monte Carlo: générer Y selon la bonne densité exponentielle π , soit $Y = -\ln(1 - U)/\lambda$, et calculer $X = \mathbb{I}[Y > y_0]$.

Répéter n fois et calculer la moyenne des n réalisations de X .

Soit π_0 une autre densité exponentielle, de paramètre $\lambda_0 \neq \lambda$. On a

$$p = \int_0^\infty \mathbb{I}[y \geq y_0] \pi(y) dy = \int_0^\infty \mathbb{I}[y \geq y_0] \frac{\pi(y)}{\pi_0(y)} \pi_0(y) dy = \mathbb{E}[X_{\text{is}}],$$

où

$$X_{\text{is}} = \mathbb{I}[Y_0 \geq y_0] \frac{\pi(Y_0)}{\pi_0(Y_0)} = \mathbb{I}[Y_0 \geq y_0] \frac{\lambda}{\lambda_0} \exp[-(\lambda - \lambda_0)Y_0]$$

et $Y_0 = -\ln(1 - U)/\lambda_0$ est une v.a. exponentielle de taux λ_0 .

Exemple (simple): estimer une probabilité.

On veut estimer $p = \mathbb{P}[Y > y_0] = \mathbb{E}[\mathbb{I}[Y > y_0]]$, où Y est exponentielle de paramètre (taux) λ , i.e., $\pi(y) = \lambda e^{-\lambda y}$ pour $y \geq 0$.

(C'est une illustration purement académique; on sait que $\mathbb{P}[Y > y] = e^{-\lambda y}$.)

Monte Carlo: générer Y selon la bonne densité exponentielle π , soit $Y = -\ln(1 - U)/\lambda$, et calculer $X = \mathbb{I}[Y > y_0]$.

Répéter n fois et calculer la moyenne des n réalisations de X .

Soit π_0 une autre densité exponentielle, de paramètre $\lambda_0 \neq \lambda$. On a

$$p = \int_0^\infty \mathbb{I}[y \geq y_0] \pi(y) dy = \int_0^\infty \mathbb{I}[y \geq y_0] \frac{\pi(y)}{\pi_0(y)} \pi_0(y) dy = \mathbb{E}[X_{\text{is}}],$$

où

$$X_{\text{is}} = \mathbb{I}[Y_0 \geq y_0] \frac{\pi(Y_0)}{\pi_0(Y_0)} = \mathbb{I}[Y_0 \geq y_0] \frac{\lambda}{\lambda_0} \exp[-(\lambda - \lambda_0)Y_0]$$

et $Y_0 = -\ln(1 - U)/\lambda_0$ est une v.a. exponentielle de taux λ_0 .

Ainsi, on a un estimateur sans biais peu importe $\lambda_0 > 0$.

On a aussi

$$\begin{aligned}
 \text{Var}[X_{\text{is}}] &= \mathbb{E}[X_{\text{is}}^2] - p^2 \\
 &= \int_{y_0}^{\infty} \frac{\pi^2(y)}{\pi_0^2(y)} \pi_0(y) dy - p^2 \\
 &= \int_{y_0}^{\infty} (\lambda/\lambda_0)^2 \exp[-2(\lambda - \lambda_0)y] \lambda_0 \exp[-\lambda_0 y] dy - p^2 \\
 &= \begin{cases} \frac{\lambda^2}{\lambda_0(2\lambda - \lambda_0)} \exp[-(2\lambda - \lambda_0)y_0] - p^2 & \text{si } 0 < \lambda_0 < 2\lambda, \\ \infty & \text{sinon.} \end{cases}
 \end{aligned}$$

Pour $\lambda_0 = \lambda$, la variance est $p(1 - p)$. Lorsque λ_0 s'approche de 0 ou de 2λ , la variance tend vers l'infini.

On a aussi

$$\begin{aligned}
 \text{Var}[X_{\text{is}}] &= \mathbb{E}[X_{\text{is}}^2] - p^2 \\
 &= \int_{y_0}^{\infty} \frac{\pi^2(y)}{\pi_0^2(y)} \pi_0(y) dy - p^2 \\
 &= \int_{y_0}^{\infty} (\lambda/\lambda_0)^2 \exp[-2(\lambda - \lambda_0)y] \lambda_0 \exp[-\lambda_0 y] dy - p^2 \\
 &= \begin{cases} \frac{\lambda^2}{\lambda_0(2\lambda - \lambda_0)} \exp[-(2\lambda - \lambda_0)y_0] - p^2 & \text{si } 0 < \lambda_0 < 2\lambda, \\ \infty & \text{sinon.} \end{cases}
 \end{aligned}$$

Pour $\lambda_0 = \lambda$, la variance est $p(1 - p)$. Lorsque λ_0 s'approche de 0 ou de 2λ , la variance tend vers l'infini. Exercice: La variance est **minimisée** pour $\lambda_0 = \lambda + 1/y_0 - (\lambda^2 + 1/y_0^2)^{1/2} < \lambda$.

On a aussi

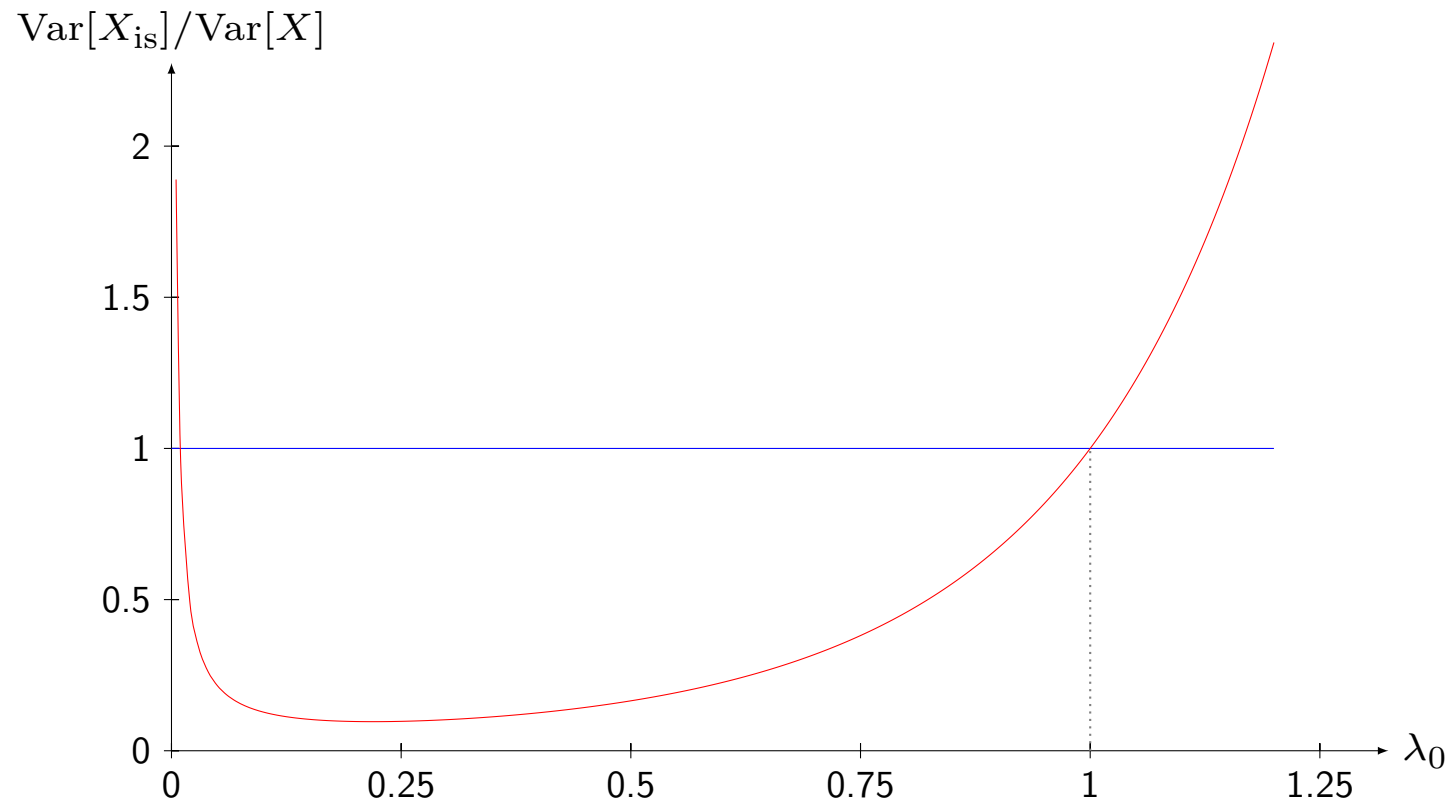
$$\begin{aligned}
 \text{Var}[X_{\text{is}}] &= \mathbb{E}[X_{\text{is}}^2] - p^2 \\
 &= \int_{y_0}^{\infty} \frac{\pi^2(y)}{\pi_0^2(y)} \pi_0(y) dy - p^2 \\
 &= \int_{y_0}^{\infty} (\lambda/\lambda_0)^2 \exp[-2(\lambda - \lambda_0)y] \lambda_0 \exp[-\lambda_0 y] dy - p^2 \\
 &= \begin{cases} \frac{\lambda^2}{\lambda_0(2\lambda - \lambda_0)} \exp[-(2\lambda - \lambda_0)y_0] - p^2 & \text{si } 0 < \lambda_0 < 2\lambda, \\ \infty & \text{sinon.} \end{cases}
 \end{aligned}$$

Pour $\lambda_0 = \lambda$, la variance est $p(1 - p)$. Lorsque λ_0 s'approche de 0 ou de 2λ , la variance tend vers l'infini. Exercice: La variance est **minimisée** pour $\lambda_0 = \lambda + 1/y_0 - (\lambda^2 + 1/y_0^2)^{1/2} < \lambda$.

Cet exemple montre que remplacer π par une autre densité g peut réduire la variance, mais peut aussi l'augmenter, et même la rendre infinie. La variance est souvent très sensible au choix de g .

Exemple numérique: $\lambda = 1$ et $y_0 = 4$.

$\text{Var}[X_{\text{is}}]/\text{Var}[X]$ en fonction de λ_0 :



Variance minimale pour $\lambda_0 \approx 0.2192$ et $\text{Var}[X_{\text{is}}]/\text{Var}[X] \approx 0.0962$ pour ce λ_0 .

Ainsi, pour estimer $p = \mathbb{P}[Y > 4]$, on change la moyenne de Y pour $1/\lambda_0 \approx 1/.2192 \approx 4.56$, qui est proche de 4.

Autre choix de densité g pour cet exemple: une exponentielle tronquée à $[y_0, \infty)$:

$$g(y) = \frac{\pi(y)}{\mathbb{P}[Y > y_0]} = \frac{\lambda e^{-\lambda y}}{e^{-\lambda y_0}} = \lambda e^{-\lambda(y-y_0)}$$

pour $y > y_0$, et 0 ailleurs. C'est la densité de $y_0 + X$ où X est une v.a. exponentielle de taux λ . On peut donc générer $Y = y_0 - \ln(1 - U)/\lambda$.

Autre choix de densité g pour cet exemple: une exponentielle tronquée à $[y_0, \infty)$:

$$g(y) = \frac{\pi(y)}{\mathbb{P}[Y > y_0]} = \frac{\lambda e^{-\lambda y}}{e^{-\lambda y_0}} = \lambda e^{-\lambda(y-y_0)}$$

pour $y > y_0$, et 0 ailleurs. C'est la densité de $y_0 + X$ où X est une v.a. exponentielle de taux λ . On peut donc générer $Y = y_0 - \ln(1 - U)/\lambda$.

On peut écrire

$$p = \int_{y_0}^{\infty} \lambda e^{-\lambda y} dy = \int_{y_0}^{\infty} \frac{\lambda e^{-\lambda y}}{g(y)} g(y) dy = \int_{y_0}^{\infty} e^{-\lambda y_0} g(y) dy = \mathbb{E}[X_{\text{is}}],$$

où

$$X_{\text{is}} = \mathbb{I}[Y \geq y_0] \exp[-\lambda y_0] = \exp[-\lambda y_0] = p,$$

car on a toujours $Y \geq y_0$.

Ce X_{is} est donc un estimateur sans biais de variance zero!

Autre choix de densité g pour cet exemple: une exponentielle tronquée à $[y_0, \infty)$:

$$g(y) = \frac{\pi(y)}{\mathbb{P}[Y > y_0]} = \frac{\lambda e^{-\lambda y}}{e^{-\lambda y_0}} = \lambda e^{-\lambda(y-y_0)}$$

pour $y > y_0$, et 0 ailleurs. C'est la densité de $y_0 + X$ où X est une v.a. exponentielle de taux λ . On peut donc générer $Y = y_0 - \ln(1 - U)/\lambda$.

On peut écrire

$$p = \int_{y_0}^{\infty} \lambda e^{-\lambda y} dy = \int_{y_0}^{\infty} \frac{\lambda e^{-\lambda y}}{g(y)} g(y) dy = \int_{y_0}^{\infty} e^{-\lambda y_0} g(y) dy = \mathbb{E}[X_{\text{is}}],$$

où

$$X_{\text{is}} = \mathbb{I}[Y \geq y_0] \exp[-\lambda y_0] = \exp[-\lambda y_0] = p,$$

car on a toujours $Y \geq y_0$.

Ce X_{is} est donc un estimateur sans biais de variance zero!

En théorie, de tels estimateurs “magiques” existent dans presque tous les cas. Mais en pratique, il sont très difficiles à trouver et à implanter. Par contre, on peut souvent les approximer et construire ainsi des estimateurs plus performants.

Example. Soient Y_1 et Y_2 des v.a. indép. de densités π_1 et π_2 , sur \mathbb{R} . On gagne

$$X = \begin{cases} Y_1 + Y_2 - K & \text{si } Y_1 \leq a \text{ and } Y_1 + Y_2 \geq b, \\ 0 & \text{sinon,} \end{cases}$$

où $K > 0$, et a et b sont des constantes. On veut estimer $\mu = \mathbb{E}[X]$.

Example. Soient Y_1 et Y_2 des v.a. indép. de densités π_1 et π_2 , sur \mathbb{R} . On gagne

$$X = \begin{cases} Y_1 + Y_2 - K & \text{si } Y_1 \leq a \text{ and } Y_1 + Y_2 \geq b, \\ 0 & \text{sinon,} \end{cases}$$

où $K > 0$, et a et b sont des constantes. On veut estimer $\mu = \mathbb{E}[X]$.

MC standard: générer Y_1 et Y_2 selon π_1 et π_2 , et calculer X .

Exemple. Soient Y_1 et Y_2 des v.a. indép. de densités π_1 et π_2 , sur \mathbb{R} . On gagne

$$X = \begin{cases} Y_1 + Y_2 - K & \text{si } Y_1 \leq a \text{ and } Y_1 + Y_2 \geq b, \\ 0 & \text{sinon,} \end{cases}$$

où $K > 0$, et a et b sont des constantes. On veut estimer $\mu = \mathbb{E}[X]$.

MC standard: générer Y_1 et Y_2 selon π_1 et π_2 , et calculer X .

Stratégie IS: On veut éviter de gaspiller des échantillons dans la région où $X = 0$. Générer Y_1 selon sa densité conditionnelle à $Y_1 \leq a$, puis générer Y_2 selon sa densité conditionnelle à $Y_1 + Y_2 \geq b$, i.e., tronquée à l'intervalle $[b - Y_1, \infty)$.

Example. Soient Y_1 et Y_2 des v.a. indép. de densités π_1 et π_2 , sur \mathbb{R} . On gagne

$$X = \begin{cases} Y_1 + Y_2 - K & \text{si } Y_1 \leq a \text{ and } Y_1 + Y_2 \geq b, \\ 0 & \text{sinon,} \end{cases}$$

où $K > 0$, et a et b sont des constantes. On veut estimer $\mu = \mathbb{E}[X]$.

MC standard: générer Y_1 et Y_2 selon π_1 et π_2 , et calculer X .

Stratégie IS: On veut éviter de gaspiller des échantillons dans la région où $X = 0$.

Générer Y_1 selon sa densité conditionnelle à $Y_1 \leq a$,

puis générer Y_2 selon sa densité conditionnelle à $Y_1 + Y_2 \geq b$,

i.e., tronquée à l'intervalle $[b - Y_1, \infty)$.

La nouvelle densité de Y_1 est

$$g_1(y) = \frac{\pi_1(y)}{\mathbb{P}[Y_1 \leq a]} = \frac{\pi_1(y)}{F_1(a)}$$

pour $y \leq a$, et celle de Y_2 conditionnelle à $Y_1 = y_1$ est

$$g_2(y \mid y_1) = \frac{\pi_2(y)}{\mathbb{P}[Y_2 \geq b - y_1]} = \frac{\pi_2(y)}{(1 - F_2(b - y_1))}$$

pour $y \geq b - y_1$, où F_1 et F_2 sont les fonctions de répartition de Y_1 et Y_2 .

$$g_2(y \mid y_1) = \frac{\pi_2(y)}{\mathbb{P}[Y_2 \geq b - y_1]} = \frac{\pi_2(y)}{(1 - F_2(b - y_1))}$$

pour $y \geq b - y_1$, où F_1 et F_2 sont les fonctions de répartition de Y_1 et Y_2 .

On a

$$\begin{aligned} \mu &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X \pi_2(y_2) \pi_1(y_1) dy_2 dy_1 \\ &= \int_{-\infty}^a \int_{b-y_1}^{\infty} X \frac{\pi_2(y_2) \pi_1(y_1)}{g_2(y_2 \mid y_1) g_1(y_1)} g_2(y_2 \mid y_1) g_1(y_1) dy_2 dy_1 \\ &= \int_{-\infty}^a \int_{b-y_1}^{\infty} X F_1(a) (1 - F_2(b - y_1)) g_2(y_2 \mid y_1) g_1(y_1) dy_2 dy_1 \\ &= \mathbb{E}_0[X_{\text{is}}], \end{aligned}$$

où

$$X_{\text{is}} = X F_1(a) (1 - F_2(b - Y_1))$$

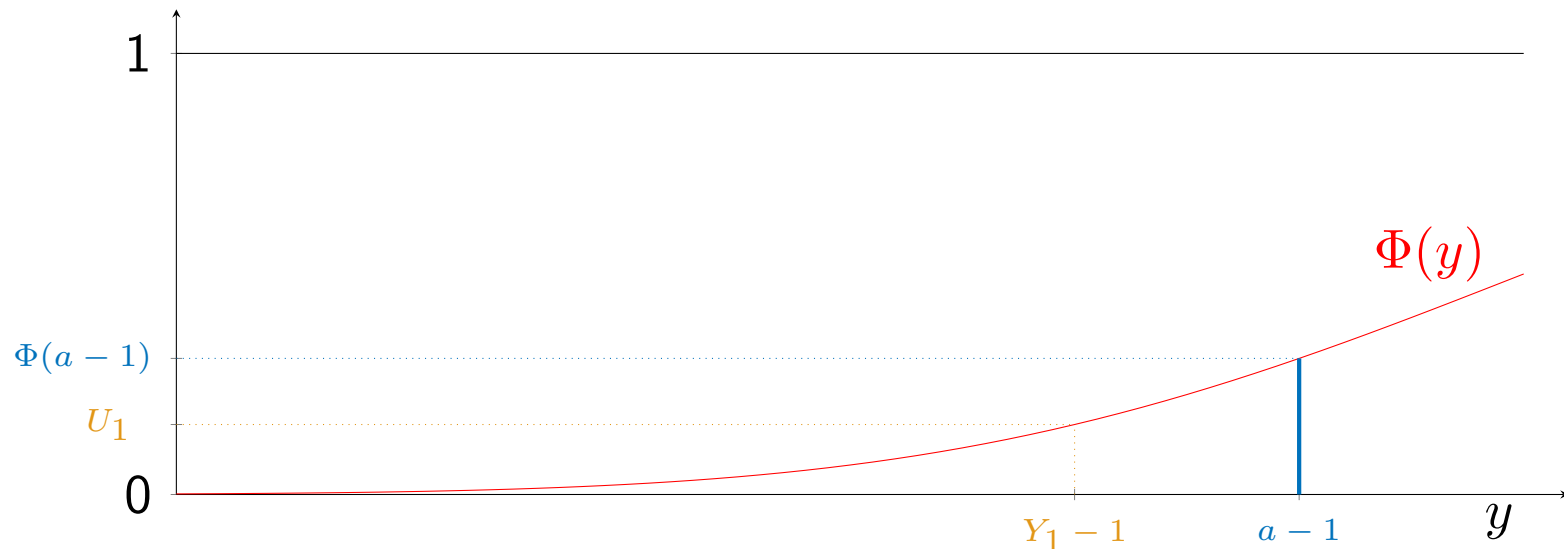
et \mathbb{E}_0 désigne l'espérance sous g_1 et g_2 .

Expérience: Supposons π_1 et $\pi_2 \sim N(1, 1)$, $K = 1$, $b = 2$, et $a = 1/2$.
Essayer MC vs IS, avec $n = 10^5$, et comparer les variances.

Expérience: Supposons π_1 et $\pi_2 \sim N(1, 1)$, $K = 1$, $b = 2$, et $a = 1/2$.
Essayer MC vs IS, avec $n = 10^5$, et comparer les variances.

On a $F_1(a) = \mathbb{P}[Y_1 < a] = \mathbb{P}[Y_1 - 1 < a - 1] = \Phi(a - 1)$.

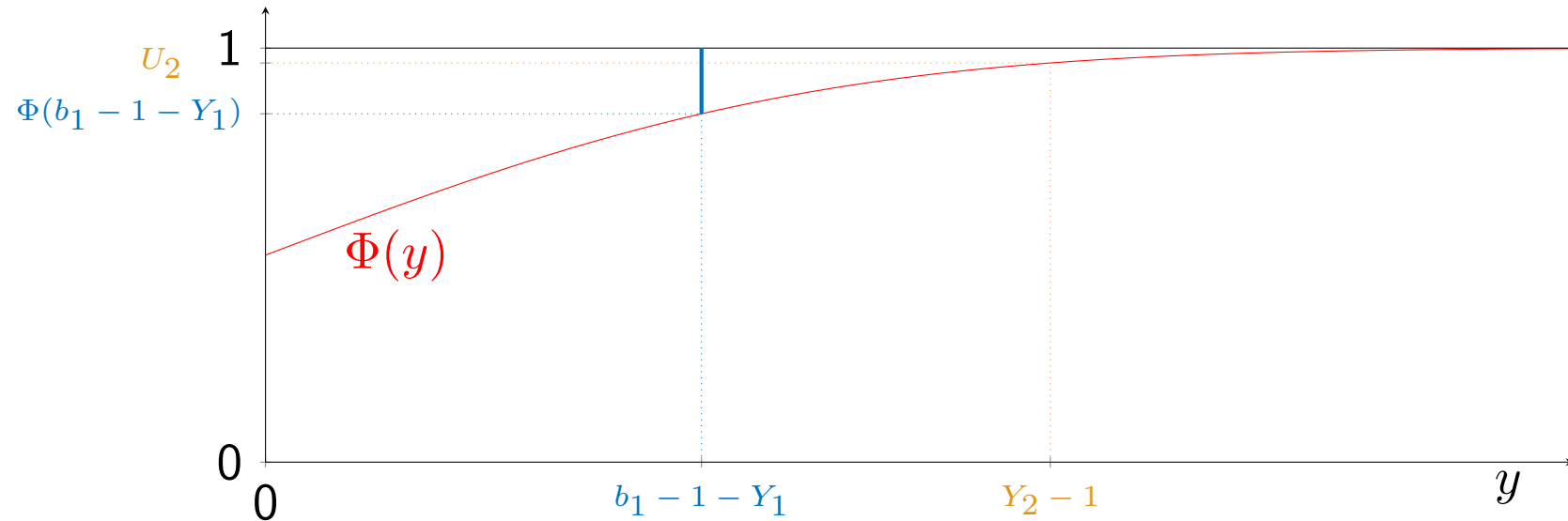
On pose $U_1 \sim \text{Uniforme}(0, \Phi(a - 1))$ et $Y_1 = 1 + \Phi^{-1}(U_1)$.



On a ensuite

$$1 - F_2(b - Y_1) = \mathbb{P}[Y_2 > b - Y_1] = \mathbb{P}[Y_2 - 1 > b - 1 - Y_1] = 1 - \Phi(b - 1 - Y_1).$$

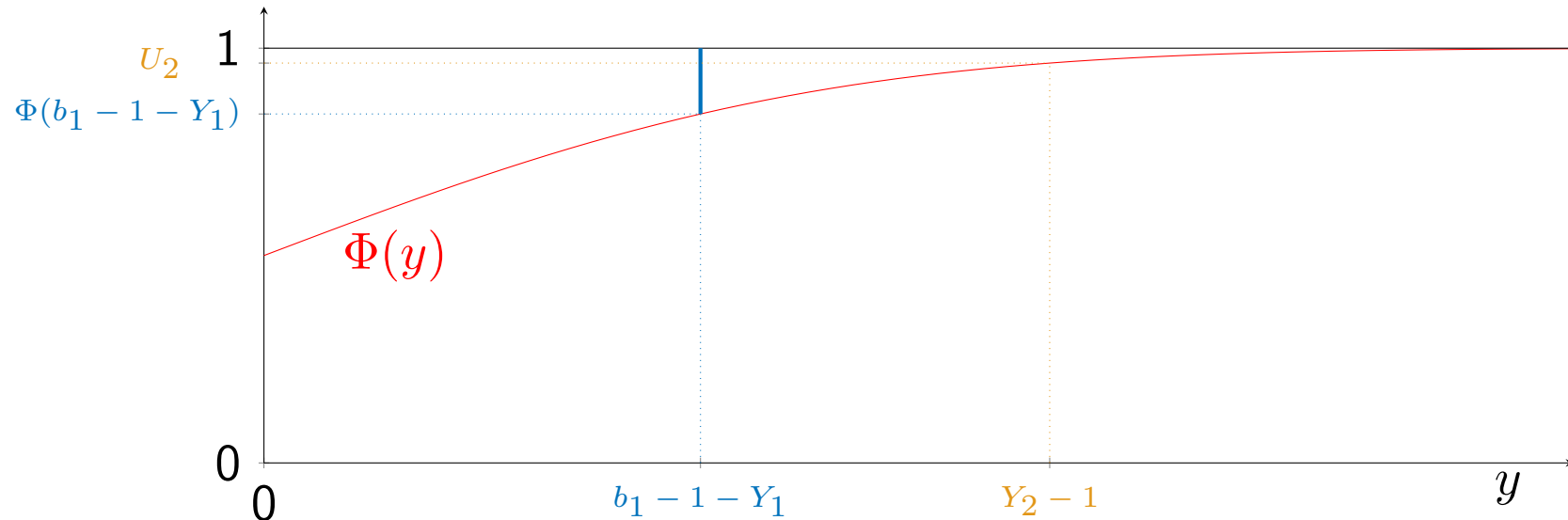
On pose $U_2 \sim \text{Uniforme}(\Phi(b - 1 - Y_1), 1)$ et $Y_2 = 1 + \Phi^{-1}(U_2)$.



On a ensuite

$$1 - F_2(b - Y_1) = \mathbb{P}[Y_2 > b - Y_1] = \mathbb{P}[Y_2 - 1 > b - 1 - Y_1] = 1 - \Phi(b - 1 - Y_1).$$

On pose $U_2 \sim \text{Uniforme}(\Phi(b - 1 - Y_1), 1)$ et $Y_2 = 1 + \Phi^{-1}(U_2)$.

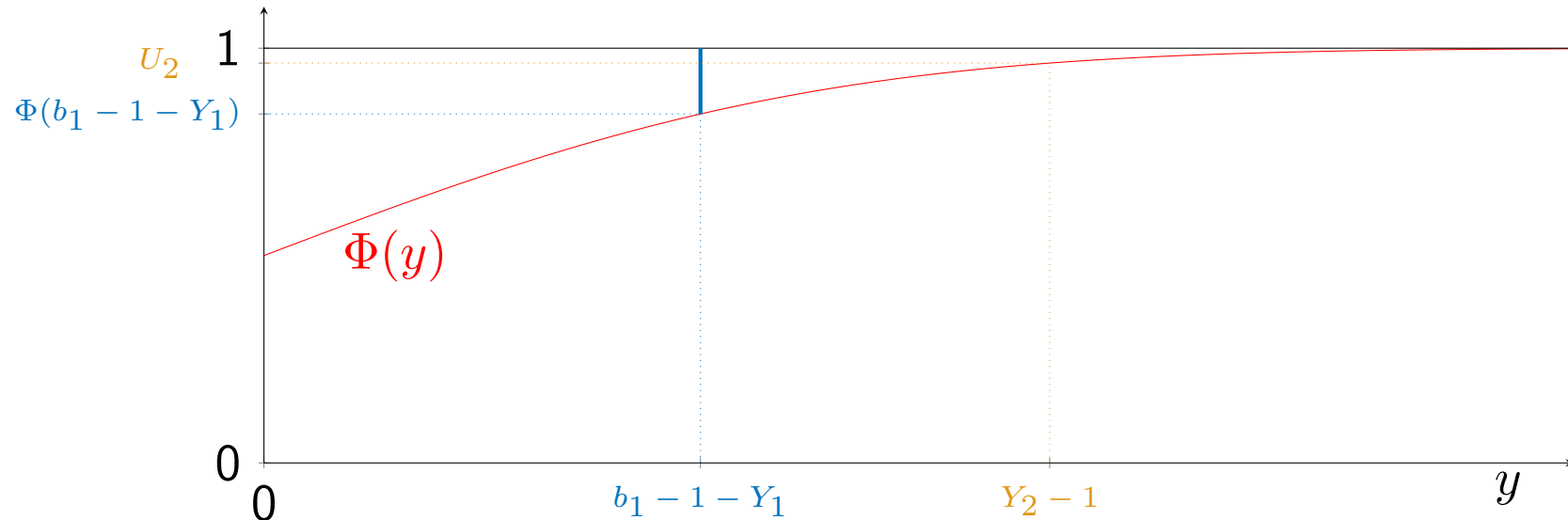


On calcule l'estimateur $X_{is} = X\Phi(a - 1)(1 - \Phi(b - 1 - Y_1))$.

On a ensuite

$$1 - F_2(b - Y_1) = \mathbb{P}[Y_2 > b - Y_1] = \mathbb{P}[Y_2 - 1 > b - 1 - Y_1] = 1 - \Phi(b - 1 - Y_1).$$

On pose $U_2 \sim \text{Uniforme}(\Phi(b - 1 - Y_1), 1)$ et $Y_2 = 1 + \Phi^{-1}(U_2)$.



On calcule l'estimateur $X_{\text{is}} = X\Phi(a - 1)(1 - \Phi(b - 1 - Y_1))$.

La variance empirique S_n^2 est environ 40 fois plus petite avec X_{is} qu'avec X .

Estimator	$\hat{\mu}_n$	S_n^2	95% confidence interval
X	0.0733	0.1188	(0.071, 0.075)
X_{is}	0.0742	0.0027	(0.074, 0.075)

Importance sampling: cas général

Rappel: On veut estimer $\mu = \int_{[0,1)^s} f(\mathbf{u}) d\mathbf{u} = \mathbb{E}[f(\mathbf{U})]$ où $\mathbf{U} \sim U(0, 1)^s$.

Importance sampling: cas général

Rappel: On veut estimer $\mu = \int_{[0,1)^s} f(\mathbf{u}) d\mathbf{u} = \mathbb{E}[f(\mathbf{U})]$ où $\mathbf{U} \sim U(0, 1)^s$.

Soit $f(\mathbf{U}) = h(\mathbf{Y})$ où $\mathbf{Y} = \tau(\mathbf{U})$ est un vecteur aléatoire de loi connue.

Importance sampling: cas général

Rappel: On veut estimer $\mu = \int_{[0,1]^s} f(\mathbf{u}) d\mathbf{u} = \mathbb{E}[f(\mathbf{U})]$ où $\mathbf{U} \sim U(0, 1)^s$.

Soit $f(\mathbf{U}) = h(\mathbf{Y})$ où $\mathbf{Y} = \tau(\mathbf{U})$ est un vecteur aléatoire de loi connue.

Exemple: pour le réseau d'activités stochastique $\mathbf{Y} = (V_1, \dots, V_{13})$.

Importance sampling: cas général

Rappel: On veut estimer $\mu = \int_{[0,1]^s} f(\mathbf{u}) d\mathbf{u} = \mathbb{E}[f(\mathbf{U})]$ où $\mathbf{U} \sim U(0, 1)^s$.

Soit $f(\mathbf{U}) = h(\mathbf{Y})$ où $\mathbf{Y} = \tau(\mathbf{U})$ est un vecteur aléatoire de loi connue.

Exemple: pour le réseau d'activités stochastique $\mathbf{Y} = (V_1, \dots, V_{13})$.

La façon directe d'estimer μ est:

Générer \mathbf{Y} selon la bonne loi et calculer $h(\mathbf{Y})$.

Importance sampling: cas général

Rappel: On veut estimer $\mu = \int_{[0,1]^s} f(\mathbf{u}) d\mathbf{u} = \mathbb{E}[f(\mathbf{U})]$ où $\mathbf{U} \sim U(0, 1)^s$.

Soit $f(\mathbf{U}) = h(\mathbf{Y})$ où $\mathbf{Y} = \tau(\mathbf{U})$ est un vecteur aléatoire de loi connue.

Exemple: pour le réseau d'activités stochastique $\mathbf{Y} = (V_1, \dots, V_{13})$.

La façon directe d'estimer μ est:

Générer \mathbf{Y} selon la bonne loi et calculer $h(\mathbf{Y})$.

Supposons que \mathbf{Y} a une densité $\pi(\mathbf{y})$ sur \mathbb{R}^d et $h : \mathbb{R}^d \rightarrow \mathbb{R}$. On a

$$\mu = E_{\pi}[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y}) \pi(\mathbf{y}) d\mathbf{y}.$$

Échantillonnage stratégique (“importance sampling” (IS)):

Soit g une autre densité, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. Alors,

$$\mu = E_{\pi}[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y}$$

Échantillonnage stratégique (“importance sampling” (IS)):

Soit g une autre densité, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. Alors,

$$\mu = E_{\pi}[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^d} [h(\mathbf{y})\pi(\mathbf{y})/g(\mathbf{y})]g(\mathbf{y})d\mathbf{y}$$

Échantillonnage stratégique (“importance sampling” (IS)):

Soit g une autre densité, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. Alors,

$$\begin{aligned}\mu &= E_{\pi}[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^d} [h(\mathbf{y})\pi(\mathbf{y})/g(\mathbf{y})]g(\mathbf{y})d\mathbf{y} \\ &= E_g[h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})].\end{aligned}$$

Échantillonnage stratégique (“importance sampling” (IS)):

Soit g une autre densité, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. Alors,

$$\begin{aligned}\mu &= E_{\pi}[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^d} [h(\mathbf{y})\pi(\mathbf{y})/g(\mathbf{y})]g(\mathbf{y})d\mathbf{y} \\ &= E_g[h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})].\end{aligned}$$

Ainsi, si \mathbf{Y} est généré selon la densité g , $X_{\text{is}} = h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})$ est un estimateur sans biais de μ . Pour tenir compte du changement de densité de π à g , on multiplie $X = h(\mathbf{Y})$ par le rapport de vraisemblance $L(\mathbf{Y}) = \pi(\mathbf{Y})/g(\mathbf{Y})$.

Échantillonnage stratégique (“importance sampling” (IS)):

Soit g une autre densité, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. Alors,

$$\begin{aligned}\mu &= E_{\pi}[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^d} [h(\mathbf{y})\pi(\mathbf{y})/g(\mathbf{y})]g(\mathbf{y})d\mathbf{y} \\ &= E_g[h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})].\end{aligned}$$

Ainsi, si \mathbf{Y} est généré selon la densité g , $X_{\text{is}} = h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})$ est un estimateur sans biais de μ . Pour tenir compte du changement de densité de π à g , on multiplie $X = h(\mathbf{Y})$ par le rapport de vraisemblance $L(\mathbf{Y}) = \pi(\mathbf{Y})/g(\mathbf{Y})$. Pour $g \equiv \pi$, on a $L(\mathbf{Y}) = 1$.

Échantillonnage stratégique (“importance sampling” (IS)):

Soit g une autre densité, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. Alors,

$$\begin{aligned}\mu &= E_{\pi}[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^d} [h(\mathbf{y})\pi(\mathbf{y})/g(\mathbf{y})]g(\mathbf{y})d\mathbf{y} \\ &= E_g[h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})].\end{aligned}$$

Ainsi, si \mathbf{Y} est généré selon la densité g , $X_{\text{is}} = h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})$ est un estimateur sans biais de μ . Pour tenir compte du changement de densité de π à g , on multiplie $X = h(\mathbf{Y})$ par le rapport de vraisemblance $L(\mathbf{Y}) = \pi(\mathbf{Y})/g(\mathbf{Y})$. Pour $g \equiv \pi$, on a $L(\mathbf{Y}) = 1$.

Même principe pour les lois discrètes.

Échantillonnage stratégique (“importance sampling” (IS)):

Soit g une autre densité, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. Alors,

$$\begin{aligned}\mu &= E_{\pi}[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^d} [h(\mathbf{y})\pi(\mathbf{y})/g(\mathbf{y})]g(\mathbf{y})d\mathbf{y} \\ &= E_g[h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})].\end{aligned}$$

Ainsi, si \mathbf{Y} est généré selon la densité g , $X_{\text{is}} = h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})$ est un estimateur sans biais de μ . Pour tenir compte du changement de densité de π à g , on multiplie $X = h(\mathbf{Y})$ par le rapport de vraisemblance $L(\mathbf{Y}) = \pi(\mathbf{Y})/g(\mathbf{Y})$. Pour $g \equiv \pi$, on a $L(\mathbf{Y}) = 1$.

Même principe pour les lois discrètes.

Pourquoi changer la densité de π à g ?

- (1) il est parfois trop difficile de générer des v.a. selon π ;
- (2) un bon choix de g peut réduire la variance comparativement à π .

Choix optimal de g .

Si $h \geq 0$ et si on prend $g(\mathbf{y})$ proportionnel à $h(\mathbf{y})\pi(\mathbf{y})$, alors

$X_{\text{is}} = h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})$ devient une constante: sa variance est zero!

Choix optimal de g .

Si $h \geq 0$ et si on prend $g(\mathbf{y})$ proportionnel à $h(\mathbf{y})\pi(\mathbf{y})$, alors

$X_{\text{is}} = h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})$ devient une constante: sa variance est zero!

Ce choix optimal est habituellement trop difficile à réaliser en pratique, mais cette propriété nous indique quand même dans quel sens modifier π pour obtenir g .

Choix optimal de g .

Si $h \geq 0$ et si on prend $g(\mathbf{y})$ proportionnel à $h(\mathbf{y})\pi(\mathbf{y})$, alors

$X_{\text{is}} = h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})$ devient une constante: sa variance est zero!

Ce choix optimal est habituellement trop difficile à réaliser en pratique, mais cette propriété nous indique quand même dans quel sens modifier π pour obtenir g .

Exemple. On veut estimer $p = \mathbb{P}\{A\} = \mathbb{E}[\mathbb{I}[A]]$ où p est petit et $\mathbb{I}[A] = h(\mathbf{Y})$ où \mathbf{Y} a une densité π .

Choix optimal de g .

Si $h \geq 0$ et si on prend $g(\mathbf{y})$ proportionnel à $h(\mathbf{y})\pi(\mathbf{y})$, alors

$X_{\text{is}} = h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})$ devient une constante: sa variance est zero!

Ce choix optimal est habituellement trop difficile à réaliser en pratique, mais cette propriété nous indique quand même dans quel sens modifier π pour obtenir g .

Exemple. On veut estimer $p = \mathbb{P}\{A\} = \mathbb{E}[\mathbb{I}[A]]$ où p est petit et $\mathbb{I}[A] = h(\mathbf{Y})$ où \mathbf{Y} a une densité π .

La densité IS optimale pour \mathbf{Y} est dans ce cas $g(\mathbf{y}) = h(\mathbf{y})\pi(\mathbf{y})/p = \mathbb{I}[A]\pi(\mathbf{y})/p$.

Choix optimal de g .

Si $h \geq 0$ et si on prend $g(\mathbf{y})$ proportionnel à $h(\mathbf{y})\pi(\mathbf{y})$, alors

$X_{\text{is}} = h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})$ devient une constante: sa variance est zero!

Ce choix optimal est habituellement trop difficile à réaliser en pratique, mais cette propriété nous indique quand même dans quel sens modifier π pour obtenir g .

Exemple. On veut estimer $p = \mathbb{P}\{A\} = \mathbb{E}[\mathbb{I}[A]]$ où p est petit et $\mathbb{I}[A] = h(\mathbf{Y})$ où \mathbf{Y} a une densité π .

La densité IS optimale pour \mathbf{Y} est dans ce cas $g(\mathbf{y}) = h(\mathbf{y})\pi(\mathbf{y})/p = \mathbb{I}[A]\pi(\mathbf{y})/p$.

C'est la densité de \mathbf{Y} conditionnelle à A .

Cette densité réduit la variance à zero.

Exemple: Probabilité de ruine d'une compagnie d'assurance.

Exemple: Probabilité de ruine d'une compagnie d'assurance.

La compagnie encaisse des primes au taux $c > 0$.

Les réclamations arrivent selon un processus de Poisson $\{N(t), t \geq 0\}$ de taux λ et leurs tailles sont des v.a. i.i.d. C_1, C_2, \dots de densité h .

Exemple: Probabilité de ruine d'une compagnie d'assurance.

La compagnie encaisse des primes au taux $c > 0$.

Les réclamations arrivent selon un processus de Poisson $\{N(t), t \geq 0\}$ de taux λ et leurs tailles sont des v.a. i.i.d. C_1, C_2, \dots de densité h . La réserve au temps t est

$$R(t) = R(0) + ct - \sum_{j=1}^{N(t)} C_j.$$

Exemple: Probabilité de ruine d'une compagnie d'assurance.

La compagnie encaisse des primes au taux $c > 0$.

Les réclamations arrivent selon un processus de Poisson $\{N(t), t \geq 0\}$ de taux λ et leurs tailles sont des v.a. i.i.d. C_1, C_2, \dots de densité h . La réserve au temps t est

$$R(t) = R(0) + ct - \sum_{j=1}^{N(t)} C_j.$$

On veut estimer $\mu = \mathbb{P}[\inf_{t>0} R(t) < 0]$, la probabilité de ruine.

Exemple: Probabilité de ruine d'une compagnie d'assurance.

La compagnie encaisse des primes au taux $c > 0$.

Les réclamations arrivent selon un processus de Poisson $\{N(t), t \geq 0\}$ de taux λ et leurs tailles sont des v.a. i.i.d. C_1, C_2, \dots de densité h . La réserve au temps t est

$$R(t) = R(0) + ct - \sum_{j=1}^{N(t)} C_j.$$

On veut estimer $\mu = \mathbb{P}[\inf_{t>0} R(t) < 0]$, la probabilité de ruine.

L'instant où $R(t)$ devient négatif (s'il a lieu) doit se produire à l'instant d'une réclamation.

Exemple: Probabilité de ruine d'une compagnie d'assurance.

La compagnie encaisse des primes au taux $c > 0$.

Les réclamations arrivent selon un processus de Poisson $\{N(t), t \geq 0\}$ de taux λ et leurs tailles sont des v.a. i.i.d. C_1, C_2, \dots de densité h . La réserve au temps t est

$$R(t) = R(0) + ct - \sum_{j=1}^{N(t)} C_j.$$

On veut estimer $\mu = \mathbb{P}[\inf_{t>0} R(t) < 0]$, la probabilité de ruine.

L'instant où $R(t)$ devient négatif (s'il a lieu) doit se produire à l'instant d'une réclamation. Si R_j est la réserve juste après la j -ième réclamation, alors

$$R_j = R_{j-1} + A_j c - C_j$$

pour $j \geq 1$, où $R_0 = R(0)$ et les A_j sont les durées entre les réclamations.

Le processus $\{R_j, j \geq 1\}$ est une **marche aléatoire** sur \mathbb{R} .
On a $\mu = \mathbb{P}[T < \infty]$ où $T = \inf\{j \geq 1 : R_j < 0\}$.

Le processus $\{R_j, j \geq 1\}$ est une **marche aléatoire** sur \mathbb{R} .

On a $\mu = \mathbb{P}[T < \infty]$ où $T = \inf\{j \geq 1 : R_j < 0\}$.

Pour estimer μ par MC, il faudrait générer n copies i.i.d. de T , disons, T_1, \dots, T_n , poser $X_i = \mathbb{I}[T_i < \infty]$, et estimer μ par \bar{X}_n , la moyenne des X_i .

Le processus $\{R_j, j \geq 1\}$ est une **marche aléatoire** sur \mathbb{R} .

On a $\mu = \mathbb{P}[T < \infty]$ où $T = \inf\{j \geq 1 : R_j < 0\}$.

Pour estimer μ par MC, il faudrait générer n copies i.i.d. de T , disons, T_1, \dots, T_n , poser $X_i = \mathbb{I}[T_i < \infty]$, et estimer μ par \bar{X}_n , la moyenne des X_i .

Mais calculer T peut prendre un temps infini!

Le processus $\{R_j, j \geq 1\}$ est une **marche aléatoire** sur \mathbb{R} .

On a $\mu = \mathbb{P}[T < \infty]$ où $T = \inf\{j \geq 1 : R_j < 0\}$.

Pour estimer μ par MC, il faudrait générer n copies i.i.d. de T , disons, T_1, \dots, T_n , poser $X_i = \mathbb{I}[T_i < \infty]$, et estimer μ par \bar{X}_n , la moyenne des X_i .

Mais calculer T peut prendre un temps infini!

Si $T = \infty$, on en sera jamais certain si on ne simule que pour un temps fini.

Le processus $\{R_j, j \geq 1\}$ est une **marche aléatoire** sur \mathbb{R} .

On a $\mu = \mathbb{P}[T < \infty]$ où $T = \inf\{j \geq 1 : R_j < 0\}$.

Pour estimer μ par MC, il faudrait générer n copies i.i.d. de T , disons, T_1, \dots, T_n , poser $X_i = \mathbb{I}[T_i < \infty]$, et estimer μ par \bar{X}_n , la moyenne des X_i .

Mais calculer T peut prendre un temps infini!

Si $T = \infty$, on en sera jamais certain si on ne simule que pour un temps fini.

Et même si on le pouvait, si μ est très petit (e.g., 10^{-9} ou moins) il faudrait un n énorme pour l'estimer avec une précision relative raisonnable. Pour estimer $\mu = 10^{-9}$ avec moins de 10% d'erreur, il faut $n \approx 10^{11}$.

Le processus $\{R_j, j \geq 1\}$ est une **marche aléatoire** sur \mathbb{R} .

On a $\mu = \mathbb{P}[T < \infty]$ où $T = \inf\{j \geq 1 : R_j < 0\}$.

Pour estimer μ par MC, il faudrait générer n copies i.i.d. de T , disons, T_1, \dots, T_n , poser $X_i = \mathbb{I}[T_i < \infty]$, et estimer μ par \bar{X}_n , la moyenne des X_i .

Mais calculer T peut prendre un temps infini!

Si $T = \infty$, on en sera jamais certain si on ne simule que pour un temps fini.

Et même si on le pouvait, si μ est très petit (e.g., 10^{-9} ou moins) il faudrait un n énorme pour l'estimer avec une précision relative raisonnable. Pour estimer $\mu = 10^{-9}$ avec moins de 10% d'erreur, il faut $n \approx 10^{11}$.

Solution: IS.

Le processus $\{R_j, j \geq 1\}$ est une **marche aléatoire** sur \mathbb{R} .

On a $\mu = \mathbb{P}[T < \infty]$ où $T = \inf\{j \geq 1 : R_j < 0\}$.

Pour estimer μ par MC, il faudrait générer n copies i.i.d. de T , disons, T_1, \dots, T_n , poser $X_i = \mathbb{I}[T_i < \infty]$, et estimer μ par \bar{X}_n , la moyenne des X_i .

Mais calculer T peut prendre un temps infini!

Si $T = \infty$, on en sera jamais certain si on ne simule que pour un temps fini.

Et même si on le pouvait, si μ est très petit (e.g., 10^{-9} ou moins) il faudrait un n énorme pour l'estimer avec une précision relative raisonnable. Pour estimer $\mu = 10^{-9}$ avec moins de 10% d'erreur, il faut $n \approx 10^{11}$.

Solution: IS.

On change les densités des A_j et C_j pour favoriser la ruine, de manière à ce qu'elle se produise avec probabilité 1, et on multiplie l'estimateur par le rapport de vraisemblance approprié.

On verra plus tard que le changement de mesure suivant est approprié dans ce cas:
remplacer la densité $h(x)$ des C_j par

$$h_\theta(x) = h(x)e^{\theta x}$$

On verra plus tard que le changement de mesure suivant est approprié dans ce cas:
remplacer la densité $h(x)$ des C_j par

$$h_\theta(x) = h(x)e^{\theta x}/M_h(\theta),$$

où

$$M_h(\theta) = \int_{-\infty}^{\infty} h(x)e^{\theta x} dx$$

est la constante de normalisation pour ramener h_θ à une densité,

On verra plus tard que le changement de mesure suivant est approprié dans ce cas: remplacer la densité $h(x)$ des C_j par

$$h_\theta(x) = h(x)e^{\theta x}/M_h(\theta),$$

où

$$M_h(\theta) = \int_{-\infty}^{\infty} h(x)e^{\theta x} dx$$

est la constante de normalisation pour ramener h_θ à une densité, et augmenter le taux λ à

$$\lambda_\theta = \lambda + \theta c,$$

On verra plus tard que le changement de mesure suivant est approprié dans ce cas: remplacer la densité $h(x)$ des C_j par

$$h_\theta(x) = h(x)e^{\theta x}/M_h(\theta),$$

où

$$M_h(\theta) = \int_{-\infty}^{\infty} h(x)e^{\theta x} dx$$

est la constante de normalisation pour ramener h_θ à une densité, et augmenter le taux λ à

$$\lambda_\theta = \lambda + \theta c,$$

où θ est la plus grande racine de l'équation $M_h(\theta) = (\lambda + \theta c)/\lambda$.

On verra plus tard que le changement de mesure suivant est approprié dans ce cas: remplacer la densité $h(x)$ des C_j par

$$h_\theta(x) = h(x)e^{\theta x}/M_h(\theta),$$

où

$$M_h(\theta) = \int_{-\infty}^{\infty} h(x)e^{\theta x} dx$$

est la constante de normalisation pour ramener h_θ à une densité, et augmenter le taux λ à

$$\lambda_\theta = \lambda + \theta c,$$

où θ est la plus grande racine de l'équation $M_h(\theta) = (\lambda + \theta c)/\lambda$.

On suppose ici que $M_h(\theta) < \infty$.

On verra plus tard que le changement de mesure suivant est approprié dans ce cas: remplacer la densité $h(x)$ des C_j par

$$h_\theta(x) = h(x)e^{\theta x}/M_h(\theta),$$

où

$$M_h(\theta) = \int_{-\infty}^{\infty} h(x)e^{\theta x} dx$$

est la constante de normalisation pour ramener h_θ à une densité, et augmenter le taux λ à

$$\lambda_\theta = \lambda + \theta c,$$

où θ est la plus grande racine de l'équation $M_h(\theta) = (\lambda + \theta c)/\lambda$.

On suppose ici que $M_h(\theta) < \infty$.

Sous cette nouvelle loi de probabilité, on a $\mathbb{P}[T < \infty] = 1$ et

$$L(A_1, C_1, A_2, C_2, \dots) = e^{\theta(R_T - R_0)} < e^{-\theta R_0}$$

(le rapport de vraisemblance correspondant) est un estimateur sans biais de μ .

Exemple numérique: prenons $\lambda = 1$ et supposons que les tailles des demandes sont exponentielles de taux $\beta = 1/2$ (moyenne de 2), et que $R(0) = 200$.

Exemple numérique: prenons $\lambda = 1$ et supposons que les tailles des demandes sont exponentielles de taux $\beta = 1/2$ (moyenne de 2), et que $R(0) = 200$.

La densité modifiée dans ce cas est exponentielle de taux $\beta_\theta = \beta - \theta$. On obtient:

c	θ	λ_θ	β_θ	$\hat{\mu}_n$	S_n^2
3	0.1667	1.5	0.3333	2.2×10^{-15}	6.3×10^{-31}
5	0.3000	2.5	0.2000	3.5×10^{-27}	6.9×10^{-54}
10	0.4000	5.0	0.1000	3.6×10^{-36}	2.3×10^{-71}

L'erreur relative est $< 1\%$ dans tous les cas.

Avec MC ordinaire, on a $\sigma^2 \approx \mu$.

IS réduit la variance par un facteur $\approx \hat{\mu}_n / S_n^2$ dans le tableau.

Pour $c = 10$, par exemple, IS divise la variance par 1.5×10^{35} .

Exemple numérique: prenons $\lambda = 1$ et supposons que les tailles des demandes sont exponentielles de taux $\beta = 1/2$ (moyenne de 2), et que $R(0) = 200$.

La densité modifiée dans ce cas est exponentielle de taux $\beta_\theta = \beta - \theta$. On obtient:

c	θ	λ_θ	β_θ	$\hat{\mu}_n$	S_n^2
3	0.1667	1.5	0.3333	2.2×10^{-15}	6.3×10^{-31}
5	0.3000	2.5	0.2000	3.5×10^{-27}	6.9×10^{-54}
10	0.4000	5.0	0.1000	3.6×10^{-36}	2.3×10^{-71}

L'erreur relative est $< 1\%$ dans tous les cas.

Avec MC ordinaire, on a $\sigma^2 \approx \mu$.

IS réduit la variance par un facteur $\approx \hat{\mu}_n / S_n^2$ dans le tableau.

Pour $c = 10$, par exemple, IS divise la variance par 1.5×10^{35} .

Pour une erreur relative de 10%, il faut $n \approx 183$ avec IS $n \approx 2.8 \times 10^{37}$ sans IS.

Dans certaines situations, on **connait** π **seulement à un facteur près**.

Si on génère $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ i.i.d. selon g , un estimateur sans biais de μ serait:

$$X_{\text{is},n} = \frac{1}{n} \sum_{i=1}^n h(\mathbf{Y}_i) L(\mathbf{Y}_i).$$

Si on peut seulement calculer $\tilde{L}(\mathbf{Y}_i) = \kappa L(\mathbf{Y}_i)$ pour une constante κ inconnue, on peut prendre

$$\frac{\sum_{i=1}^n h(\mathbf{Y}_i) \tilde{L}(\mathbf{Y}_i)}{\sum_{i=1}^n \tilde{L}(\mathbf{Y}_i)} = \frac{\sum_{i=1}^n h(\mathbf{Y}_i) L(\mathbf{Y}_i)}{\sum_{i=1}^n L(\mathbf{Y}_i)} \xrightarrow{\text{p.s.}} \frac{E_g[h(\mathbf{Y}) L(\mathbf{Y})]}{E_g[L(\mathbf{Y})]} = \mu$$

quand $n \rightarrow \infty$.

Dans certaines situations, on **connait** π **seulement à un facteur près**.

Si on génère $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ i.i.d. selon g , un estimateur sans biais de μ serait:

$$X_{\text{is},n} = \frac{1}{n} \sum_{i=1}^n h(\mathbf{Y}_i) L(\mathbf{Y}_i).$$

Si on peut seulement calculer $\tilde{L}(\mathbf{Y}_i) = \kappa L(\mathbf{Y}_i)$ pour une constante κ inconnue, on peut prendre

$$\frac{\sum_{i=1}^n h(\mathbf{Y}_i) \tilde{L}(\mathbf{Y}_i)}{\sum_{i=1}^n \tilde{L}(\mathbf{Y}_i)} = \frac{\sum_{i=1}^n h(\mathbf{Y}_i) L(\mathbf{Y}_i)}{\sum_{i=1}^n L(\mathbf{Y}_i)} \xrightarrow{\text{p.s.}} \frac{E_g[h(\mathbf{Y}) L(\mathbf{Y})]}{E_g[L(\mathbf{Y})]} = \mu$$

quand $n \rightarrow \infty$.

Cet estimateur est **biaisé** pour n fini, mais souvent plus efficace que $X_{\text{is},n}$.