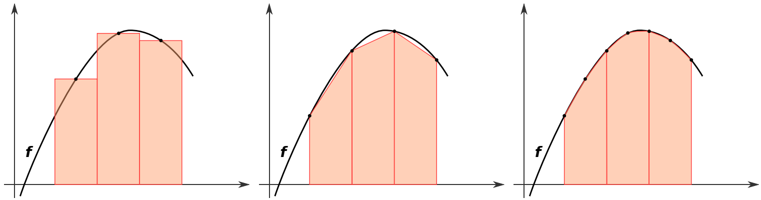
# Théorèmes de convergence pour l'intégration de Monte Carlo

Loïs Paulin

15 Novembre 2019

# Plan

# Integration numérique




(a) Points Milieus (b) Trapèzes (c) Simpson

FIGURE – Méthodes numériques d'integration

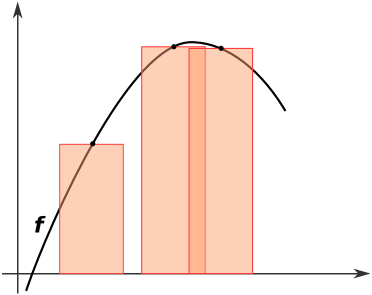
# Intégration de Monte-Carlo

$$\int_{\Omega} f(x)dx = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

avec les $x_i$ choisis stochastiquement

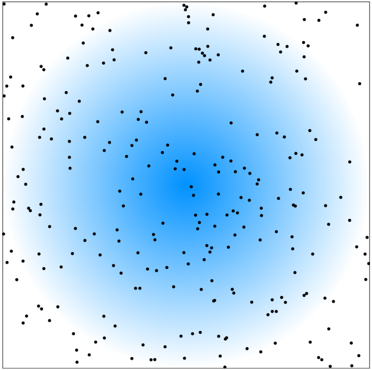
- Indépendant de la dimension
- Indépendant de la complexité du problème

# Intégration de Monte-Carlo




$$\left| \int_\Omega f(x)dx - \frac{1}{n} \sum f(x_i) \right| = O(\frac{1}{\sqrt{n}})$$

# Intégration de Monte-Carlo

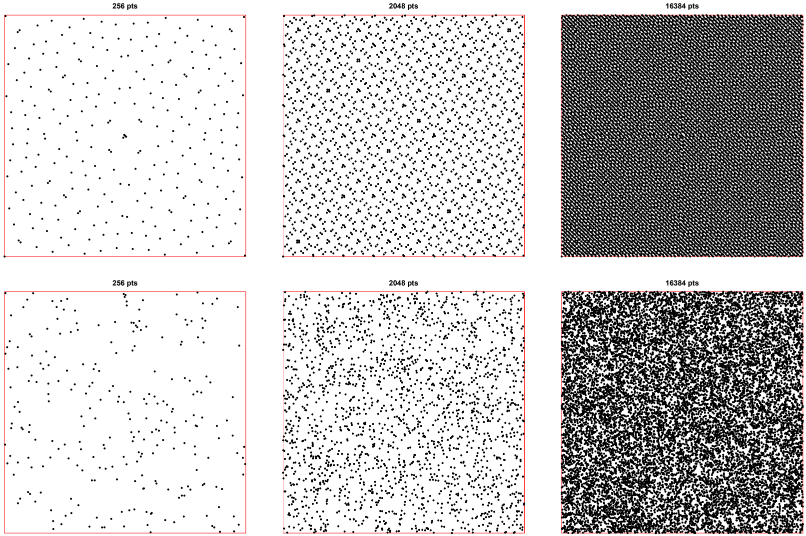


$$\left| \int_\Omega f(x)dx - \frac{1}{n} \sum f(x_i) \right| = O(\frac{1}{\sqrt{n}})$$

# Intégration de Monte-Carlo

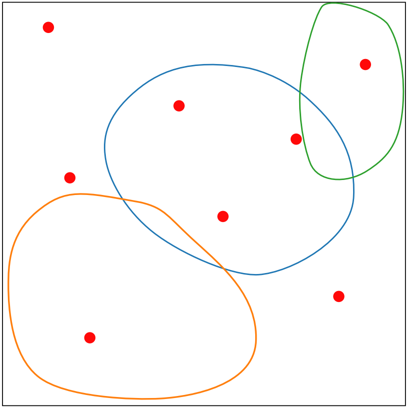
Peut-on mieux faire ?

# Intuition : Plus Uniforme



# Outline

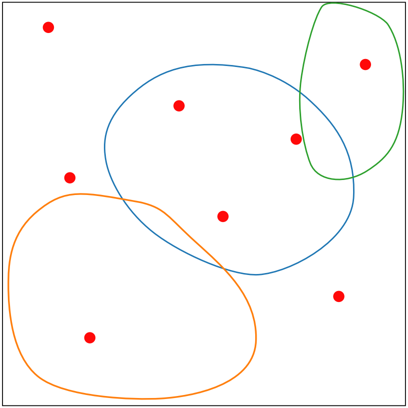
# Discrépance [Nie92, Lem09]



Comparer la proportion de points espérés a la réalisation

- $\mu$ la distribution échantillonnée
- $\mathcal{P}(I)$ la proportion de points tirés dans $I$
- $\mathcal{R}_{\mathcal{P}}(I) = |\mu(I) - \mathcal{P}(I)|$

# Discrépance [Nie92, Lem09]

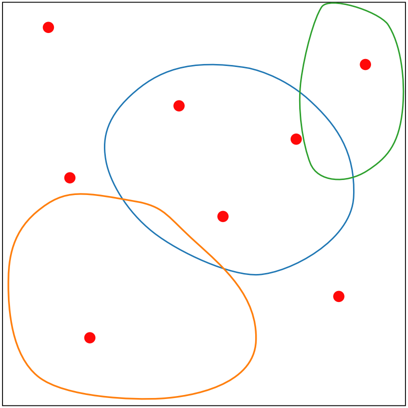


$\mu(I_1) = 1/8, \mathcal{P}(I_1) = 1/8$

Comparer la proportion de points espérés a la réalisation

- $\mu$ la distribution échantillonnée
- $\mathcal{P}(I)$ la proportion de points tirés dans $I$
- $\mathcal{R}_{\mathcal{P}}(I) = |\mu(I) - \mathcal{P}(I)|$

# Discrépance [Nie92, Lem09]

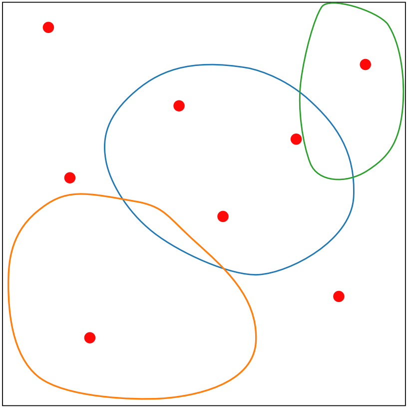


$\mu(I_1) = 1/8, \mathcal{P}(I_1) = 1/8$
$\mu(I_2) = 1/4, \mathcal{P}(I_2) = 3/8$

Comparer la proportion de points espérés a la réalisation

- $\mu$ la distribution échantillonnée
- $\mathcal{P}(I)$ la proportion de points tirés dans $I$
- $\mathcal{R}_{\mathcal{P}}(I) = |\mu(I) - \mathcal{P}(I)|$

# Discrépance [Nie92, Lem09]

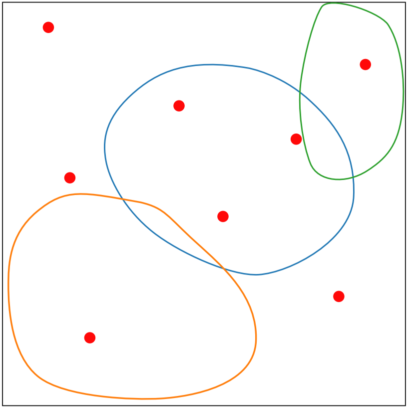


$\mu(I_1) = 1/8, \mathcal{P}(I_1) = 1/8$
$\mu(I_2) = 1/4, \mathcal{P}(I_2) = 3/8$
$\mu(I_3) = 1/4, \mathcal{P}(I_3) = 1/8$

Comparer la proportion de points espérés a la réalisation

- $\mu$ la distribution échantillonnée
- $\mathcal{P}(I)$ la proportion de points tirés dans $I$
- $\mathcal{R}_{\mathcal{P}}(I) = |\mu(I) - \mathcal{P}(I)|$

# Discrépance [Nie92, Lem09]



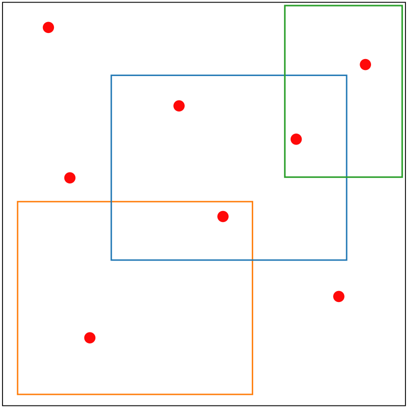
$\mu(I_1) = 1/8, \mathcal{P}(I_1) = 1/8$
$\mu(I_2) = 1/4, \mathcal{P}(I_2) = 3/8$
$\mu(I_3) = 1/4, \mathcal{P}(I_3) = 1/8$

Comparer la proportion de points espérés a la réalisation

- $\mu$ la distribution échantillonnée
- $\mathcal{P}(I)$ la proportion de points tirés dans $I$
- $\mathcal{R}_{\mathcal{P}}(I) = |\mu(I) - \mathcal{P}(I)|$

$$D(\mathcal{P}) = \sup_{I \in \text{Convexes}} R_{\mathcal{P}}(I)$$

# Discrépance Extrème



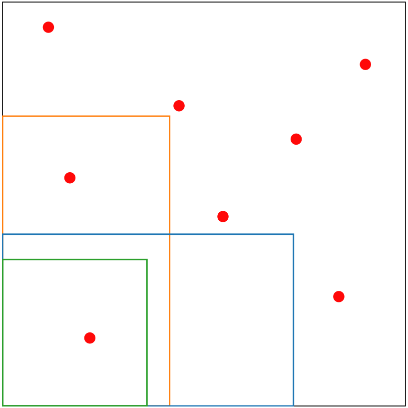
Comparer la proportion de points espérés a la réalisation

- $\mu$ la distribution échantillonnée.
- $\mathcal{P}(I)$ la proportion de points tirés dans $I$.
- $\mathcal{R}_{\mathcal{P}}(I) = |\mu(I) - \mathcal{P}(I)|$
- $J(x, y) = \prod_{i=1}^{d} [x_i, y_i)$.

$\mu(I_1) = 1/8, \mathcal{P}(I_1) = 2/8$
$\mu(I_2) = 1/4, \mathcal{P}(I_2) = 3/8$
$\mu(I_3) = 1/4, \mathcal{P}(I_3) = 2/8$

$$D(\mathcal{P}) = \sup_{x,y \in [0,1)^d} R_{\mathcal{P}}(J(x, y))$$

# Discrépance Star



$\mu(I_1) = 1/8, \mathcal{P}(I_1) = 1/8$
$\mu(I_2) = 1/4, \mathcal{P}(I_2) = 1/8$
$\mu(I_3) = 1/4, \mathcal{P}(I_3) = 2/8$

Comparer la proportion de points espérés a la réalisation

- $\mu$ la distribution échantillonnée
- $\mathcal{P}(I)$ la proportion de points tirés dans $I$
- $\mathcal{R}_{\mathcal{P}}(I) = |\mu(I) - \mathcal{P}(I)|$
- $J(x, y) = \prod_{i=1}^{d} [x_i, y_i)$.

$$D^*(\mathcal{P}) = \sup_{x \in [0,1)^d} R_{\mathcal{P}}(J(0, x))$$

$$D^*(\mathcal{P}) \leq D(\mathcal{P}) \leq 2^{d-1} D^*(\mathcal{P})$$

# Discrépance L2 [Hic98]

## Approximation de la discrépance

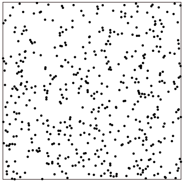
$$T^*(P) = \left( \int_{[0,1)^d} (\mathcal{R}_{\mathcal{P}}(J(0,v)))^2 dv \right)^{\frac{1}{2}}$$

$$T^*(P)^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \prod_{k=1}^{d} (1 - max(x_{i,k}, x_{j,k}))$$

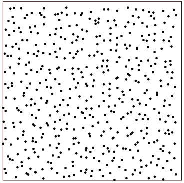
$$- \frac{2^{-s+1}}{n} \sum_{i=1}^{n} \prod_{k=1}^{d} (1 - u_{i,k}^2) + 3^{-d}$$

Algorithme en $O(n \log(n)^d)$ [Hei96]

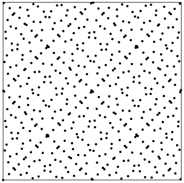
# Exemples de discrépances



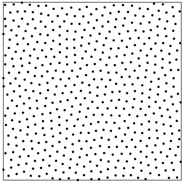
(a) Bruit blanc



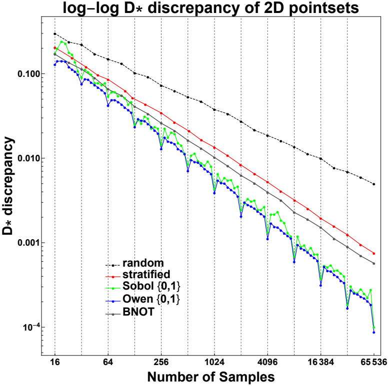
(b) Stratifié



(c) Sobol



(d) Bruit bleu



# Théorème de Koksma-Hlawka [Kok42, Hla61]

### Théorème

*Soit $\mathcal{P} = \{x_1, \ldots, x_n\}$. Si $f$ est une fonction a variation bornée alors*

$$\left| \int_\Omega f(x)dx - \frac{1}{n} \sum_{i=1}^{n} f(x_i) \right| \leq V(f) * D^*(\mathcal{P})$$

- Meilleure borne possible par la discrepance

### Conjecture

$$D^*(\mathcal{P}) = O(\frac{\log(n)^{d-1}}{n})$$

*(Prouvé pour $d \leq 2$)*

# Variation au sens Hardy-Krause

## Définition

$$V^{(d)}(f) = \sup_{P \in \mathcal{P}} \sum_{J \in P} |\Delta(f; J)|,$$

- $\mathcal{P}$ l'ensemble des partitions P de $[0,1)^d$
- $P$ ensemble des $\{x_1, \dots, x_{n_p}\}$ tels que $\forall i \in \{1, \dots, n_p - 1\}, x_i \in [0,1)^d$ et $x_i \leq x_{i+1}$
- $\Delta(f; J)$ la somme alternée de la fonction aux sommets de $J$

# Variation au sens Hardy-Krause

$$V^{(d)}(f) = \sup_{P \in \mathcal{P}} \sum_{J \in P} |\Delta(f; J)|,$$

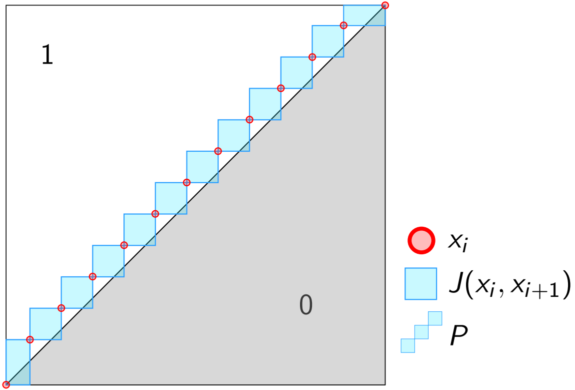



Figure – Exemple d'une discontinuité ayant une variation infinie

# Variation Totale [BCGT13]

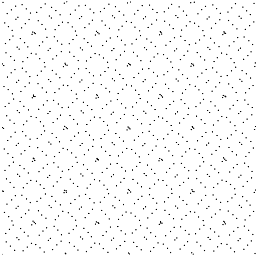
### Définition

$$\mathcal{V}^{(d)}(f) = \sum_{\alpha \in \{0,1\}^d} 2^{d - \|\alpha\|_0} \int_{[0,1]^d} \left| \left( \frac{\partial}{\partial x} \right)^{\alpha} f(x) \right|$$

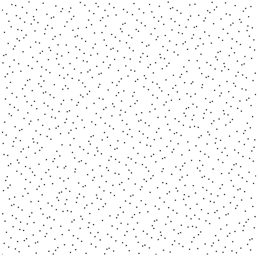
- Considère les discontinuités
- Peut remplacer $V$ dans la borne de Koksma-Hlawka

# Séquences à basse discrépance

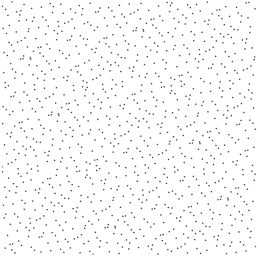
- Séquence = Ensemble "prolongeable" d'échantillons
- $D^*(P) = O(\frac{\log(n)^{d-1}}{n})$
- différentes méthodes :



(a) Sobol

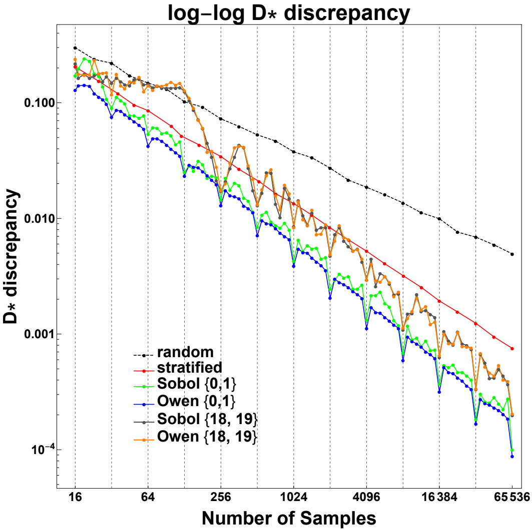


(b) Halton

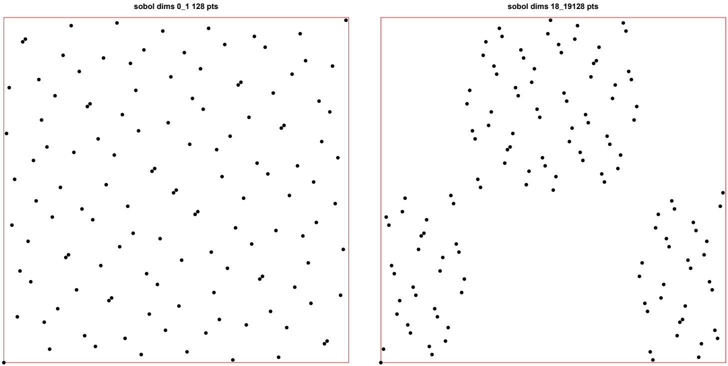


(c) Niederreiter base 3

# Détection de Qualité



# Détection de Qualité




(a) Sobol (b) Sobol dégénéré

# Conclusion

Avantages :

- Calculable
- Minimisable
- Bonnes propriétés liées
- Borne applicable à une large famille d'intégrandes

Désavantages :

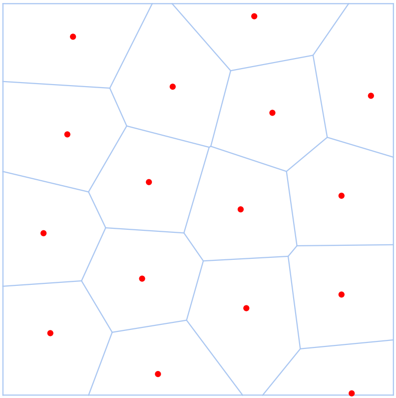
- Non isotrope
- Défauts des séquences sur certaines dimensions

# Outline

# Transport Optimal

## Idée :

Mesurer l'uniformité dans un sens plus intuitif



# Transport Optimal Semi-Discret

## Définition (Monge)

$$W(\mu,\nu) = \inf\left\{ \int_X c(x, T(x)) d\mu(x) \colon T_*(\mu) = \nu \right\},$$

avec $c(x, T(x))$ le coût pour se déplacer de $x$ à $T(x)$

Voir [HHMP19] pour les différents algorithmes pour le calculer

# Transport Optimal Projectif

## Définition [RPDB11]

$$\tilde{W}_p(X, Y) = \left( \int_{\theta \in \Omega} W_p(X_\theta, Y_\theta)^p d\theta \right)^{\frac{1}{p}}$$

where

$X_\theta = \{\langle X_i, \theta \rangle\}$,
$\Omega = \left\{\theta \in \mathbb{R}^d : \ \|\theta\| = 1\right\}$,
$W_p(X_\theta, Y_\theta)$ la distance de Wasserstein 1D entre $X_\theta$ et $Y_\theta$

Mesure équivalente au transport optimal [Bon13]

# Dualité de Kantorovich-Rubinstein

Soit

$$\|f\|_L = \sup\left\{\frac{f(x)-f(y)}{\|x-y\|} : x, y \in \Omega; x \neq y\right\},$$

On a

$$W(\mu,\nu) = \sup\left\{\int_\Omega f d\mu - \int_\Omega f d\nu : \ \|f\|_L \leq 1\right\},$$

donc

$$\left|\int_\Omega f d\mu - \int_\Omega f d\nu\right| \leq \|f\|_L W(\mu,\nu),$$

et

$$\left|\int_\Omega f d\mu - \frac{1}{n}\sum_{i=1}^n f(x_i)\right| \leq \|f\|_L W(\mu,\nu).$$

# Dualité de Kantorovich-Rubinstein

Soit

$$\|f\|_L = \sup\left\{\frac{f(x)-f(y)}{\|x-y\|} : x, y \in \Omega; x \neq y\right\},$$

On a

$$W(\mu,\nu) = \sup\left\{\int_\Omega f d\mu - \int_\Omega f d\nu : \ \|f\|_L \leq 1\right\},$$

donc

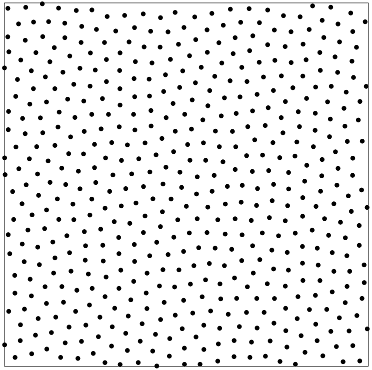
$$\left|\int_\Omega f d\mu - \int_\Omega f d\nu\right| \leq \|f\|_L W(\mu,\nu),$$

et

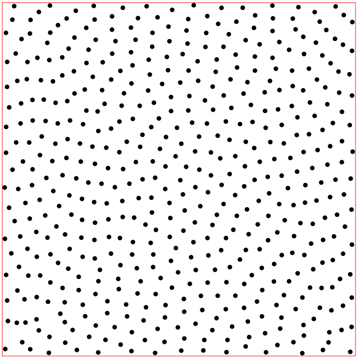
$$\left|\int_\Omega f d\mu - \frac{1}{n}\sum_{i=1}^n f(x_i)\right| \leq \|f\|_L W(\mu,\nu).$$

# Minimiser le Transport Optimal

Descente de gradient



(a) BNOT[DGBOD12] (Transport Optimal)



(b) SOTS (Transport Optimal Projectif)

# Conclusion

## Avantages

- Mesure intuitive
- Échantillonnage préferentiel

## Désavantages

- Plus coûteux
- Classe de fonctions plus restreinte

## À faire

- Trouver la valeur minimale théorique en fonction de $n$

# Outline

# $(\mathcal{M}, \mu)$-Uniformité [Nie03]

## Intuition

Critère binaire d'uniformité proche de la discrépance.

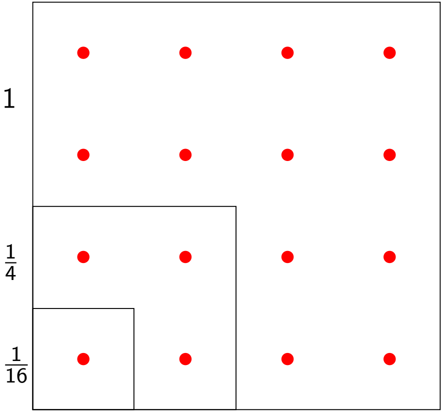



FIGURE – Exemple d'une réalisation $(\mathcal{M}, \mu)$-uniforme

# $(\mathcal{M}, \mu)$-Uniformité [Nie03]

## Intuition

Critère binaire d'uniformité proche de la discrépance.

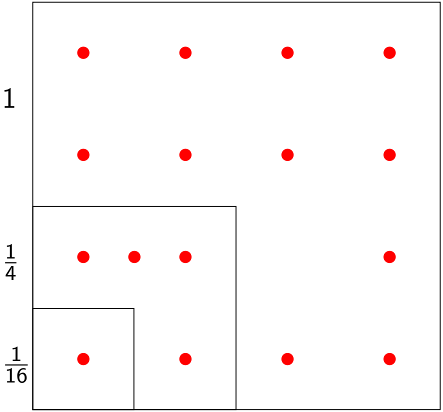



Figure – Exemple d'une réalisation non $(\mathcal{M}, \mu)$-uniforme

# $(\mathcal{M}, \mu)$-Uniformité [Nie03]

Soit

- $(\Omega, \mathcal{B}, \mu)$ un espace de probabilité
- $\Omega$ notre domaine d'intégration
- $\mathcal{B}$ une $\sigma$-algèbre de sous ensembles de $\Omega$
- $\mu$ une mesure de probabilité definie sur $\mathcal{B}$
- $\mathcal{M}$ un sous ensemble de $\mathcal{B}$
- $\mathcal{P} = \{x_1, \dots, x_n\}$ un ensemble de $n$ points
- $\mathcal{P}(M)$ la proportion de points de $\mathcal{P}$ dans $M$

## Définition

$\mathcal{P}$ est $(\mathcal{M}, \mu)$-uniforme si $\forall M \in \mathcal{M}, \mathcal{P}(M) = \mu(M)$

# Borne d'intégration

### Théorème

*Si $\mathcal{M}$ est une partition de $\Omega$ et que $\mathcal{P}$ est $(\mathcal{M}, \mu)$-uniforme alors*

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(\mathbf{x}_n) - \int_X f d\mu \right| \leq \sum_{j=1}^{|\mathcal{M}|} \mu(M_j)(\sup_{M_j}(f) - \inf_{M_j}(f))$$

On défini $\delta(M) = \sup_{x,y \in M} d(x,y)$ et $\delta(\mathcal{M}) = \sup_{M \in \mathcal{M}} \delta(M)$

### Théorème

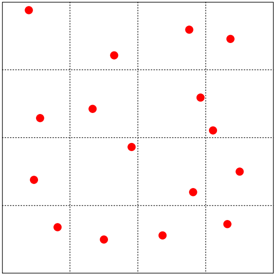
*Si $\mathcal{M}$ est une partition de $\Omega$ et que $\mathcal{P}$ est $(\mathcal{M}, \mu)$-uniforme alors*

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(\mathbf{x}_n) - \int_X f d\mu \right| \leq \sup_{\substack{x,y \in X \\ d(x,y) \leq \delta(\mathcal{M})}} |f(x) - f(y)|$$

# Exemple

## Stratifié

Par définition $(\mathcal{M}, \mu)$-uniforme avec $\mathcal{M}$ l'ensemble des strates et $\mu$ la loi uniforme.



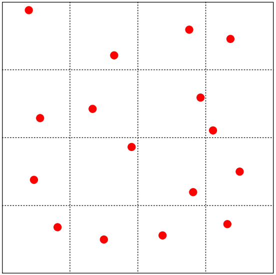
- $\delta(\mathcal{M}) = O(\frac{1}{\sqrt{n}})$
- Pour $f$ Lipschitz,

$$\sup_{\substack{x,y \in X \\ d(x,y) \leq \delta(\mathcal{M})}} |f(x) - f(y)| = O(\frac{1}{\sqrt{n}})$$

# Exemple

## Stratifié

Par définition $(\mathcal{M}, \mu)$-uniforme avec $\mathcal{M}$ l'ensemble des strates et $\mu$ la loi uniforme.



- $\delta(\mathcal{M}) = O(\frac{1}{\sqrt[d]{n}})$
- Pour $f$ Lipschitz,

$$\sup_{\substack{x,y \in X \\ d(x,y) \leq \delta(\mathcal{M})}} |f(x) - f(y)| = O(\frac{1}{\sqrt[d]{n}})$$

# Conclusion

## Avantages

- Critère élégant
- Souplesse d'application

## Désavantages

- Preuve à refaire pour chaque ensemble de point
- Borne très faible ?

## À faire

- Étudier de la bibliographie plus récente pour trouver l'intérêt [Kel06, Kel13]

# Le mot de la fin

Les échantillons sont construits en respectant des propriétés qui sont le fondement de leur éfficacité pour l'intégration.
Ces propriétés peuvent être détruite par les manipulations faites sur les échantillons.

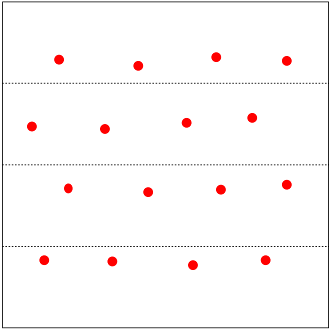


FIGURE – Exemple de mauvaise manipulation d'échantillons

# Le mot de la fin

Les échantillons sont construits en respectant des propriétés qui sont le fondement de leur éfficacité pour l'intégration.
Ces propriétés peuvent être détruite par les manipulations faites sur les échantillons.

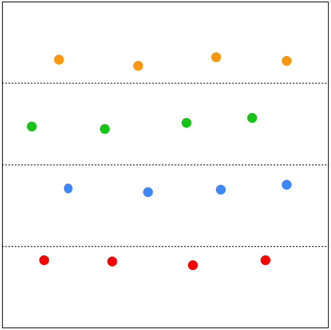


FIGURE – Exemple de mauvaise manipulation d'échantillons

# Le mot de la fin

Les échantillons sont construits en respectant des propriétés qui sont le fondement de leur éfficacité pour l'intégration.
Ces propriétés peuvent être détruite par les manipulations faites sur les échantillons.

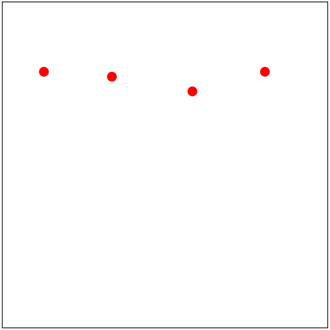


FIGURE – Exemple de mauvaise manipulation d'échantillons

# Le mot de la fin

Les échantillons sont construits en respectant des propriétés qui sont le fondement de leur éfficacité pour l'intégration.
Ces propriétés peuvent être détruite par les manipulations faites sur les échantillons.

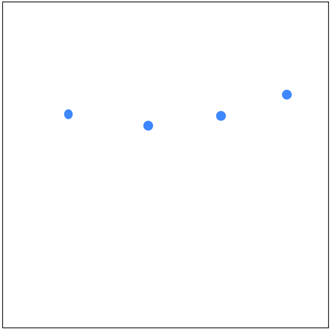


FIGURE – Exemple de mauvaise manipulation d'échantillons

# Le mot de la fin

Les échantillons sont construits en respectant des propriétés qui sont le fondement de leur éfficacité pour l'intégration.
Ces propriétés peuvent être détruite par les manipulations faites sur les échantillons.

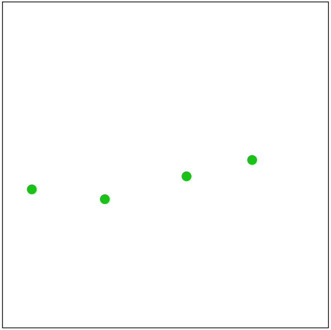


FIGURE – Exemple de mauvaise manipulation d'échantillons

# Le mot de la fin

Les échantillons sont construits en respectant des propriétés qui sont le fondement de leur éfficacité pour l'intégration.
Ces propriétés peuvent être détruite par les manipulations faites sur les échantillons.

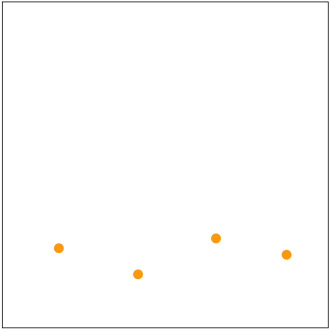


FIGURE – Exemple de mauvaise manipulation d'échantillons

# Le mot de la fin

Les échantillons sont construits en respectant des propriétés qui sont le fondement de leur éfficacité pour l'intégration.
Ces propriétés peuvent être détruite par les manipulations faites sur les échantillons.

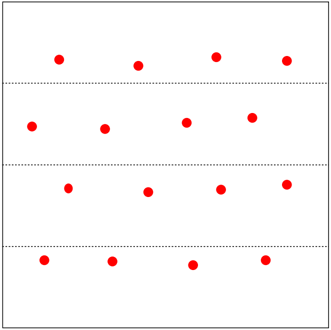


FIGURE – Exemple de mauvaise manipulation d'échantillons

# Le mot de la fin

Les échantillons sont construits en respectant des propriétés qui sont le fondement de leur éfficacité pour l'intégration.
Ces propriétés peuvent être détruite par les manipulations faites sur les échantillons.

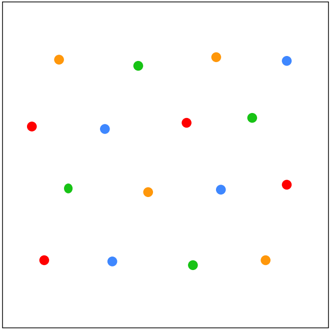


FIGURE – Exemple de mauvaise manipulation d'échantillons

# Références I


Luca Brandolini, Leonardo Colzani, Giacomo Gigante, and Giancarlo Travaglini, *On the koksma–hlawka inequality*, Journal of Complexity **29** (2013), no. 2, 158–172.

Nicolas Bonnotte, *Unidimensional and evolution methods for optimal transportation*, Ph.D. thesis, Paris 11, 2013.

Fernando De Goes, Katherine Breeden, Victor Ostromoukhov, and Mathieu Desbrun, *Blue noise through optimal transport*, ACM Transactions on Graphics (TOG) **31** (2012), no. 6, 171.

Stefan Heinrich, *Efficient algorithms for computing the $L_2$-discrepancy*, Mathematics of Computation of the American Mathematical Society **65** (1996), no. 216, 1621–1633.

Agathe Herrou, Mathieu Heitz, Jocelyn Meyron, and Loïs Paulin, *(almost)-optimal transport*, https ://projet.liris.cnrs.fr/origami/math/, 2019.

Fred Hickernell, *A generalized discrepancy and quadrature error bound*, Mathematics of Computation of the American Mathematical Society **67** (1998), no. 221, 299–322.

# Références II


- Edmund Hlawka, *Funktionen von beschränkter variatiou in der theorie der gleichverteilung*, Annali di Matematica Pura ed Applicata **54** (1961), no. 1, 325–333.
- Alexander Keller, *Myths of computer graphics*, Monte Carlo and Quasi-Monte Carlo Methods 2004, Springer, 2006, pp. 217–243.
- ______, *Quasi-monte carlo image synthesis in a nutshell*, Monte Carlo and Quasi-Monte Carlo Methods 2012, Springer, 2013, pp. 213–249.
- JF Koksma, *Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1*, Mathematica B (Zutphen) **11** (1942), no. 7-11, 43.
- Christian Lemieux, *Monte carlo and quasi-monte carlo sampling*, Springer, New York, NY, USA, 2009.
- Harald Niederreiter, *Random number generation and quasi-monte carlo methods*, vol. 63, Siam, 1992.
- ______, *Error bounds for quasi-monte carlo integration with uniform point sets*, Journal of computational and applied mathematics **150** (2003), no. 2, 283–292.

# Références III

Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot, *Wasserstein barycenter and its application to texture mixing*, International Conference on Scale Space and Variational Methods in Computer Vision, Springer, 2011, pp. 435–446.