

# Espérance conditionnelle (EC) ou Monte Carlo conditionnel (CMC)

## Un exemple simple.

Supposons que l'on veut estimer  $\mu = \mathbb{P}[Y_1 + \cdots + Y_t > x]$ .

Estimateur évident:  $X = \mathbb{I}[Y_1 + \cdots + Y_t > x]$ .

# Espérance conditionnelle (EC) ou Monte Carlo conditionnel (CMC)

## Un exemple simple.

Supposons que l'on veut estimer  $\mu = \mathbb{P}[Y_1 + \dots + Y_t > x]$ .

Estimateur évident:  $X = \mathbb{I}[Y_1 + \dots + Y_t > x]$ .

Estimateur CMC:  $X_{e,s} = \mathbb{P}[Y_1 + \dots + Y_t > x \mid Y_1, \dots, Y_s]$ , pour  $s \leq t$ .

# Espérance conditionnelle (EC) ou Monte Carlo conditionnel (CMC)

## Un exemple simple.

Supposons que l'on veut estimer  $\mu = \mathbb{P}[Y_1 + \dots + Y_t > x]$ .

Estimateur évident:  $X = \mathbb{I}[Y_1 + \dots + Y_t > x]$ .

Estimateur CMC:  $X_{e,s} = \mathbb{P}[Y_1 + \dots + Y_t > x \mid Y_1, \dots, Y_s]$ , pour  $s \leq t$ .

Si  $s = t$ : aucun changement.

# Espérance conditionnelle (EC) ou Monte Carlo conditionnel (CMC)

## Un exemple simple.

Supposons que l'on veut estimer  $\mu = \mathbb{P}[Y_1 + \dots + Y_t > x]$ .

Estimateur évident:  $X = \mathbb{I}[Y_1 + \dots + Y_t > x]$ .

Estimateur CMC:  $X_{e,s} = \mathbb{P}[Y_1 + \dots + Y_t > x \mid Y_1, \dots, Y_s]$ , pour  $s \leq t$ .

Si  $s = t$ : aucun changement.

Si  $s = t - 1$  et les  $Y_j$  sont indépendants:

$$X_{e,t-1} = \mathbb{P}[Y_t > x - Y_1 - \dots - Y_{t-1} \mid Y_1, \dots, Y_{t-1}] = 1 - F_t[x - Y_1 - \dots - Y_{t-1}].$$

# Espérance conditionnelle (EC) ou Monte Carlo conditionnel (CMC)

## Un exemple simple.

Supposons que l'on veut estimer  $\mu = \mathbb{P}[Y_1 + \dots + Y_t > x]$ .

Estimateur évident:  $X = \mathbb{I}[Y_1 + \dots + Y_t > x]$ .

Estimateur CMC:  $X_{e,s} = \mathbb{P}[Y_1 + \dots + Y_t > x \mid Y_1, \dots, Y_s]$ , pour  $s \leq t$ .

Si  $s = t$ : aucun changement.

Si  $s = t - 1$  et les  $Y_j$  sont indépendants:

$$X_{e,t-1} = \mathbb{P}[Y_t > x - Y_1 - \dots - Y_{t-1} \mid Y_1, \dots, Y_{t-1}] = 1 - F_t[x - Y_1 - \dots - Y_{t-1}].$$

Si  $s = 0$ :  $X_{e,0} = \mathbb{P}[Y_1 + \dots + Y_t > x] = \mu$  (variance réduite à zéro).

En fait plus  $s$  est petit, plus la variance est réduite.

## Cadre général

Idée: remplacer l'estimateur  $X$  par  $\mathbb{E}[X \mid Z]$  où  $Z$  est une autre v.a.,

## Cadre général

**Idée:** remplacer l'estimateur  $X$  par  $\mathbb{E}[X \mid Z]$  où  $Z$  est une autre v.a.,  
ou plus généralement par  $\mathbb{E}[X \mid \mathcal{G}]$ , où  $\mathcal{G}$  est une  $\sigma$ -algèbre donnant une  
information partielle sur  $X$ .

## Cadre général

**Idée:** remplacer l'estimateur  $X$  par  $\mathbb{E}[X \mid Z]$  où  $Z$  est une autre v.a., ou plus généralement par  $\mathbb{E}[X \mid \mathcal{G}]$ , où  $\mathcal{G}$  est une  $\sigma$ -algèbre donnant une **information partielle** sur  $X$ . L'estimateur CMC s'écrit

$$X_e \stackrel{\text{def}}{=} \mathbb{E}[X \mid \mathcal{G}]$$

et on a

$$\mathbb{E}[X_e] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X].$$



## Cadre général

**Idée:** remplacer l'estimateur  $X$  par  $\mathbb{E}[X \mid Z]$  où  $Z$  est une autre v.a., ou plus généralement par  $\mathbb{E}[X \mid \mathcal{G}]$ , où  $\mathcal{G}$  est une  $\sigma$ -algèbre donnant une **information partielle** sur  $X$ . L'estimateur CMC s'écrit

$$X_e \stackrel{\text{def}}{=} \mathbb{E}[X \mid \mathcal{G}]$$

et on a

$$\mathbb{E}[X_e] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X].$$

De plus,

$$\text{Var}[X] = \mathbb{E}\left[ \underbrace{\text{Var}[X \mid \mathcal{G}]}_{\substack{\text{Var. résiduelle} \\ \text{pour } \mathcal{G} \text{ connu} \\ \text{(éliminé par CMC)}}} \right] + \underbrace{\text{Var}[\mathbb{E}[X \mid \mathcal{G}]]}_{\substack{\text{Var. due à la} \\ \text{variation de } \mathcal{G}}} = \mathbb{E}[\text{Var}[X \mid \mathcal{G}]] + \text{Var}[X_e],$$

et donc

$$\text{Var}[X_e] = \text{Var}[X] - \mathbb{E}[\text{Var}[X \mid \mathcal{G}]] \leq \text{Var}[X].$$

(Cas particulier du théorème de Rao-Blackwell.)

Pour minimiser la variance, on doit maximiser  $\mathbb{E}[\text{Var}[X \mid \mathcal{G}]]$ , i.e.,  $\mathcal{G}$  doit contenir le moins d'information possible.

On sait en effet que  $\mathcal{G}_1 \subset \mathcal{G}_2$  implique que  $\mathbb{E}[\text{Var}[X \mid \mathcal{G}_1]] \geq \mathbb{E}[\text{Var}[X \mid \mathcal{G}_2]]$  (exercice).

Pour minimiser la variance, on doit maximiser  $\mathbb{E}[\text{Var}[X \mid \mathcal{G}]]$ , i.e.,  $\mathcal{G}$  doit contenir le moins d'information possible.

On sait en effet que  $\mathcal{G}_1 \subset \mathcal{G}_2$  implique que  $\mathbb{E}[\text{Var}[X \mid \mathcal{G}_1]] \geq \mathbb{E}[\text{Var}[X \mid \mathcal{G}_2]]$  (exercice).

Mais moins  $\mathcal{G}_1$  contient d'information, plus il est difficile de calculer  $X_e$ .

On doit donc faire un compromis.

Dans certains cas,  $X_e$  peut être moins coûteux à calculer que  $X$ .

Pour minimiser la variance, on doit maximiser  $\mathbb{E}[\text{Var}[X \mid \mathcal{G}]]$ , i.e.,  $\mathcal{G}$  doit contenir le moins d'information possible.

On sait en effet que  $\mathcal{G}_1 \subset \mathcal{G}_2$  implique que  $\mathbb{E}[\text{Var}[X \mid \mathcal{G}_1]] \geq \mathbb{E}[\text{Var}[X \mid \mathcal{G}_2]]$  (exercice).

Mais moins  $\mathcal{G}_1$  contient d'information, plus il est difficile de calculer  $X_e$ .

On doit donc faire un compromis.

Dans certains cas,  $X_e$  peut être moins coûteux à calculer que  $X$ .

Cas limites:

Si  $\mathcal{G}$  ne contient aucune information pertinente à  $X$ :  $X_e = \mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X] = \mu$ .

Si  $\mathcal{G}$  permet de calculer  $X$  (i.e.,  $X$  est  $\mathcal{G}$ -mesurable):  $X_e = X$ .

**Exemple.** Réseau d'activités stochastique.

On veut estimer  $\mu = \mathbb{P}[T > x]$ .

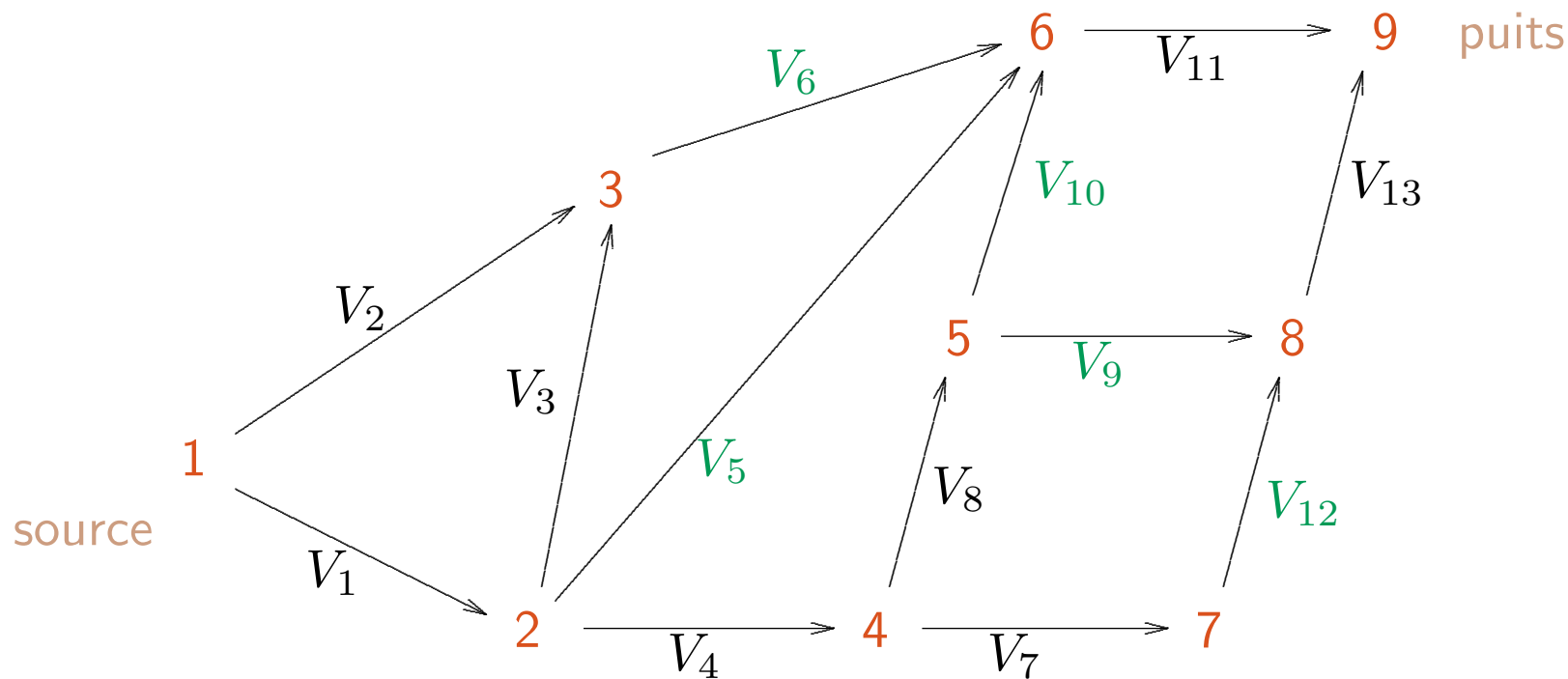
Estimateur naïf:  $X = \mathbb{I}[T > x]$ .

**Exemple.** Réseau d'activités stochastique.

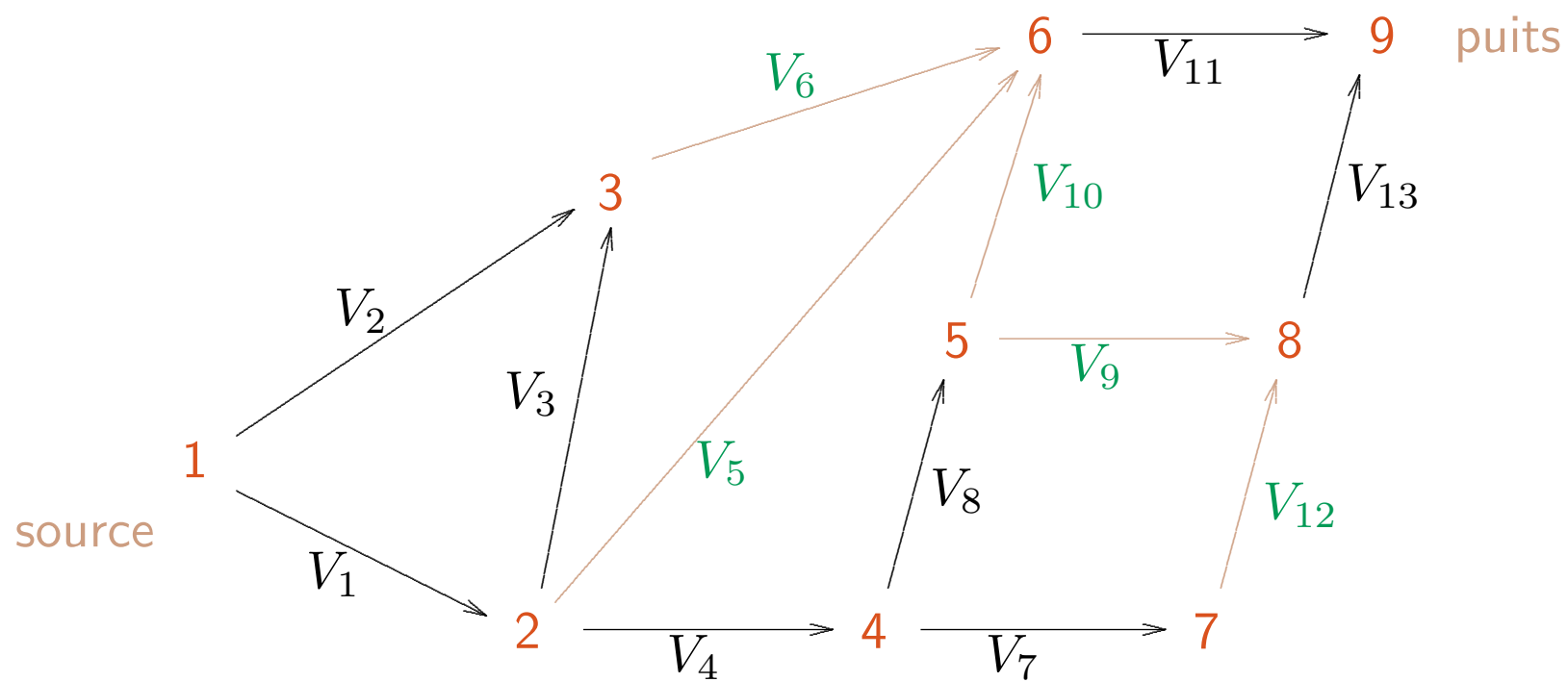
On veut estimer  $\mu = \mathbb{P}[T > x]$ .

Estimateur naïf:  $X = \mathbb{I}[T > x]$ .

Soit  $\mathcal{L} \subseteq \mathcal{A}$  un ensemble d'activités (arcs) tel que chaque chemin de la source au puits contient exactement un arc de  $\mathcal{L}$ . ( $\mathcal{L}$  est une coupe orientée.)









Estimateur CMC:

$$X_e = \mathbb{P}[T > x \mid \{V_j, j \notin \mathcal{L}\}].$$

On le calcule comme suit.

Pour chaque  $l \in \mathcal{L}$ , allant disons de  $a_l$  à  $b_l$ , calculer la longueur  $\alpha_l$  du plus long chemin de la source à  $a_l$ , puis la longueur  $\beta_l$  du plus long chemin de  $b_l$  au puits.

Aucun chemin passant par  $l$  est plus long que  $x$  ssi  $\alpha_l + V_l + \beta_l \leq x$ .

Estimateur CMC:

$$X_e = \mathbb{P}[T > x \mid \{V_j, j \notin \mathcal{L}\}].$$

On le calcule comme suit.

Pour chaque  $l \in \mathcal{L}$ , allant disons de  $a_l$  à  $b_l$ , calculer la longueur  $\alpha_l$  du plus long chemin de la source à  $a_l$ , puis la longueur  $\beta_l$  du plus long chemin de  $b_l$  au puits.

Aucun chemin passant par  $l$  est plus long que  $x$  ssi  $\alpha_l + V_l + \beta_l \leq x$ .

Conditionnellement à  $\{V_j, j \in \mathcal{B}\}$ , on a cette condition avec probabilité  $\mathbb{P}[V_l \leq x - \alpha_l - \beta_l] = F_l[x - \alpha_l - \beta_l]$ .

Estimateur CMC:

$$X_e = \mathbb{P}[T > x \mid \{V_j, j \notin \mathcal{L}\}].$$

On le calcule comme suit.

Pour chaque  $l \in \mathcal{L}$ , allant disons de  $a_l$  à  $b_l$ , calculer la longueur  $\alpha_l$  du plus long chemin de la source à  $a_l$ , puis la longueur  $\beta_l$  du plus long chemin de  $b_l$  au puits.

Aucun chemin passant par  $l$  est plus long que  $x$  ssi  $\alpha_l + V_l + \beta_l \leq x$ .

Conditionnellement à  $\{V_j, j \in \mathcal{B}\}$ , on a cette condition avec probabilité  $\mathbb{P}[V_l \leq x - \alpha_l - \beta_l] = F_l[x - \alpha_l - \beta_l]$ .

Puisque les  $V_l$  sont indépendants, on obtient

$$X_e = 1 - \prod_{l \in \mathcal{L}} F_l[x - \alpha_l - \beta_l].$$

Cet estimateur peut être moins coûteux à calculer que  $X$ .

Estimateur CMC:

$$X_e = \mathbb{P}[T > x \mid \{V_j, j \notin \mathcal{L}\}].$$

On le calcule comme suit.

Pour chaque  $l \in \mathcal{L}$ , allant disons de  $a_l$  à  $b_l$ , calculer la longueur  $\alpha_l$  du plus long chemin de la source à  $a_l$ , puis la longueur  $\beta_l$  du plus long chemin de  $b_l$  au puits.

Aucun chemin passant par  $l$  est plus long que  $x$  ssi  $\alpha_l + V_l + \beta_l \leq x$ .

Conditionnellement à  $\{V_j, j \in \mathcal{B}\}$ , on a cette condition avec probabilité  $\mathbb{P}[V_l \leq x - \alpha_l - \beta_l] = F_l[x - \alpha_l - \beta_l]$ .

Puisque les  $V_l$  sont indépendants, on obtient

$$X_e = 1 - \prod_{l \in \mathcal{L}} F_l[x - \alpha_l - \beta_l].$$

Cet estimateur peut être moins coûteux à calculer que  $X$ .

Avec l'exemple tiré de Avramidis et Wilson (1998), cette méthode réduit la variance environ par un facteur 4.

Estimateur CMC:

$$X_e = \mathbb{P}[T > x \mid \{V_j, j \notin \mathcal{L}\}].$$

On le calcule comme suit.

Pour chaque  $l \in \mathcal{L}$ , allant disons de  $a_l$  à  $b_l$ , calculer la longueur  $\alpha_l$  du plus long chemin de la source à  $a_l$ , puis la longueur  $\beta_l$  du plus long chemin de  $b_l$  au puits.

Aucun chemin passant par  $l$  est plus long que  $x$  ssi  $\alpha_l + V_l + \beta_l \leq x$ .

Conditionnellement à  $\{V_j, j \in \mathcal{B}\}$ , on a cette condition avec probabilité  $\mathbb{P}[V_l \leq x - \alpha_l - \beta_l] = F_l[x - \alpha_l - \beta_l]$ .

Puisque les  $V_l$  sont indépendants, on obtient

$$X_e = 1 - \prod_{l \in \mathcal{L}} F_l[x - \alpha_l - \beta_l].$$

Cet estimateur peut être moins coûteux à calculer que  $X$ .

Avec l'exemple tiré de Avramidis et Wilson (1998), cette méthode réduit la variance environ par un facteur 4.

Dans cet exemple, on peut aussi utiliser les  $V_j$ , et peut-être les  $V_j^2$ , comme variables de contrôle. On sait que ces variables sont corrélées avec  $X_e$ .

### Exemple.

Prix d'un sous-jacent modélisé par un mouvement brownien géométrique. Supposons que l'on remplace le paramètre de volatilité  $\sigma$  (constant) par un processus de volatilité stochastique  $\{\sigma(t), t \geq 0\}$ .

On obtient

$$dS(t) = rS(t)dt + \sigma(t)S(t)dB(t)$$

où  $B(\cdot)$  est un mouvement Brownien standard, en supposant que  $\sigma$  est indépendant de  $B$ .

### Exemple.

Prix d'un sous-jacent modélisé par un mouvement brownien géométrique. Supposons que l'on remplace le paramètre de volatilité  $\sigma$  (constant) par un processus de **volatilité stochastique**  $\{\sigma(t), t \geq 0\}$ .

On obtient

$$dS(t) = rS(t)dt + \sigma(t)S(t)dB(t)$$

où  $B(\cdot)$  est un **mouvement Brownien** standard, en supposant que  $\sigma$  est indépendant de  $B$ .

Pour évaluer une option sous un tel modèle, on peut simuler le processus de volatilité  $\sigma$ , puis (dans certains cas) calculer l'espérance conditionnelle à la trajectoire de  $\sigma$ .

### Exemple.

Prix d'un sous-jacent modélisé par un mouvement brownien géométrique. Supposons que l'on remplace le paramètre de volatilité  $\sigma$  (constant) par un processus de **volatilité stochastique**  $\{\sigma(t), t \geq 0\}$ .

On obtient

$$dS(t) = rS(t)dt + \sigma(t)S(t)dB(t)$$

où  $B(\cdot)$  est un **mouvement Brownien** standard, en supposant que  $\sigma$  est indépendant de  $B$ .

Pour évaluer une option sous un tel modèle, on peut simuler le processus de volatilité  $\sigma$ , puis (dans certains cas) calculer l'espérance conditionnelle à la trajectoire de  $\sigma$ .

On pourrait aussi utiliser comme variables de contrôle des v.a.'s qui dépendent de la trajectoire de  $\sigma$  et dont on connaît l'espérance.



**Exemple.** Évaluation d'une option "down-and-in" à barrière discrète.

La valeur de l'actif  $S(t)$  est observée aux instants  $0 = t_0 < t_1 < \dots < t_d = T$ .

Le revenu de l'option est

$$X = e^{-rT} \max(0, S(T) - K) \cdot \mathbb{I}[T_\ell \leq T]$$

où  $\ell$  est une constante et  $T_\ell = \inf\{t_j : S(t_j) < \ell\}$ .

**Exemple.** Évaluation d'une option "down-and-in" à barrière discrète.

La valeur de l'actif  $S(t)$  est observée aux instants  $0 = t_0 < t_1 < \dots < t_d = T$ .

Le revenu de l'option est

$$X = e^{-rT} \max(0, S(T) - K) \cdot \mathbb{I}[T_\ell \leq T]$$

où  $\ell$  est une constante et  $T_\ell = \inf\{t_j : S(t_j) < \ell\}$ .

**Estimateur CMC:**  $\mathcal{G}$  contient la trajectoire du processus  $S$  jusqu'au temps  $\min(T_\ell, T)$ , i.e.,

$$X_e = \begin{cases} \mathbb{E}[X \mid T_\ell, S(T_\ell)] & \text{si } T_\ell \leq T, \\ 0 & \text{sinon.} \end{cases}$$

**Exemple.** Évaluation d'une option "down-and-in" à barrière discrète.

La valeur de l'actif  $S(t)$  est observée aux instants  $0 = t_0 < t_1 < \dots < t_d = T$ .

Le revenu de l'option est

$$X = e^{-rT} \max(0, S(T) - K) \cdot \mathbb{I}[T_\ell \leq T]$$

où  $\ell$  est une constante et  $T_\ell = \inf\{t_j : S(t_j) < \ell\}$ .

**Estimateur CMC:**  $\mathcal{G}$  contient la trajectoire du processus  $S$  jusqu'au temps  $\min(T_\ell, T)$ , i.e.,

$$X_e = \begin{cases} \mathbb{E}[X \mid T_\ell, S(T_\ell)] & \text{si } T_\ell \leq T, \\ 0 & \text{sinon.} \end{cases}$$

Si  $S$  est un mouvement brownien géométrique, on peut calculer  $X_e$  via Black-Scholes pour un modèle sur horizon  $T - T_\ell$  avec valeur initiale  $S(T_\ell)$  (voir exercices).

**Exemple.** Le centre d'appels.

Soit  $L$  le nombre d'abandons dans une journée.

On peut estimer  $\mathbb{E}[L]$  par  $L$ .

**Exemple.** Le centre d'appels.

Soit  $L$  le nombre d'abandons dans une journée.

On peut estimer  $\mathbb{E}[L]$  par  $L$ .

Supposons maintenant que l'on rend les abandons invisibles.

$\mathcal{G}$  représente tout le reste de l'information.

L'Estimateur CMC est  $L_e = \mathbb{E}[L \mid \mathcal{G}]$ .

### Exemple. Le centre d'appels.

Soit  $L$  le nombre d'abandons dans une journée.

On peut estimer  $\mathbb{E}[L]$  par  $L$ .

Supposons maintenant que l'on rend les abandons invisibles.

$\mathcal{G}$  représente tout le reste de l'information.

L'Estimateur CMC est  $L_e = \mathbb{E}[L \mid \mathcal{G}]$ .

Supposons pour simplifier que les clients ne peuvent abandonner qu'en arrivant.

Un client qui doit attendre abandonne avec probabilité  $p$ .

Soit  $\lambda(t)$  le taux d'arrivée au temps  $t$ , et

$I(t)$  un indicateur que tous les agents sont occupés au temps  $t$ .

Le processus d'arrivée conditionnel (invisible) des clients qui abandonnent est un processus de Poisson de taux  $\{p\lambda(t)I(t), t \geq 0\}$ .

### Exemple. Le centre d'appels.

Soit  $L$  le nombre d'abandons dans une journée.

On peut estimer  $\mathbb{E}[L]$  par  $L$ .

Supposons maintenant que l'on rend les abandons invisibles.

$\mathcal{G}$  représente tout le reste de l'information.

L'Estimateur CMC est  $L_e = \mathbb{E}[L \mid \mathcal{G}]$ .

Supposons pour simplifier que les clients ne peuvent abandonner qu'en arrivant.

Un client qui doit attendre abandonne avec probabilité  $p$ .

Soit  $\lambda(t)$  le taux d'arrivée au temps  $t$ , et

$I(t)$  un indicateur que tous les agents sont occupés au temps  $t$ .

Le processus d'arrivée conditionnel (invisible) des clients qui abandonnent est un processus de Poisson de taux  $\{p\lambda(t)I(t), t \geq 0\}$ .

L'espérance conditionnelle du nombre d'abandons est donc

$$L_e = \int_{9.75}^{15} p\lambda(t)I(t)dt = 30p \int_{9.75}^{11} I(t)dt + 60p \int_{11}^{14} I(t)dt + 30p \int_{14}^{15} I(t)dt.$$

Facile à calculer et on sait que  $\text{Var}[L_e] < \text{Var}[L]$ .

## CMC étendu et MC filtré

Soit

$$X = C_1 + \cdots + C_T$$

où  $T$  est aléatoire. L'estimateur CMC est  $X_e = \mathbb{E}[C_1 + \cdots + C_T \mid \mathcal{G}]$ .



## CMC étendu et MC filtré

Soit

$$X = C_1 + \cdots + C_T$$

où  $T$  est aléatoire. L'estimateur CMC est  $X_e = \mathbb{E}[C_1 + \cdots + C_T \mid \mathcal{G}]$ .

L'estimateur CMC étendu (ECMC) a la forme

$$X_{ee} = \mathbb{E}[C_1 \mid \mathcal{G}_1] + \cdots + \mathbb{E}[C_T \mid \mathcal{G}_T].$$

Souvent plus facile à calculer que  $X_e$ .

## CMC étendu et MC filtré

Soit

$$X = C_1 + \cdots + C_T$$

où  $T$  est aléatoire. L'estimateur CMC est  $X_e = \mathbb{E}[C_1 + \cdots + C_T \mid \mathcal{G}]$ .

L'estimateur CMC étendu (ECMC) a la forme

$$X_{ee} = \mathbb{E}[C_1 \mid \mathcal{G}_1] + \cdots + \mathbb{E}[C_T \mid \mathcal{G}_T].$$

Souvent plus facile à calculer que  $X_e$ .

Supposons pour simplifier que  $T = t$ , une constante.

## CMC étendu et MC filtré

Soit

$$X = C_1 + \cdots + C_T$$

où  $T$  est aléatoire. L'estimateur CMC est  $X_e = \mathbb{E}[C_1 + \cdots + C_T \mid \mathcal{G}]$ .

L'estimateur CMC étendu (ECMC) a la forme

$$X_{ee} = \mathbb{E}[C_1 \mid \mathcal{G}_1] + \cdots + \mathbb{E}[C_T \mid \mathcal{G}_T].$$

Souvent plus facile à calculer que  $X_e$ .

Supposons pour simplifier que  $T = t$ , une constante. On a alors

$$\mathbb{E}[X_{ee}] = \mathbb{E} \left[ \sum_{j=1}^t \mathbb{E}[C_j \mid \mathcal{G}_j] \right] = \sum_{j=1}^t \mathbb{E}[C_j] = \mathbb{E} \left[ \sum_{j=1}^t C_j \right] = \mathbb{E}[X]$$

et la variance est:

$$\text{Var}[X_{\text{ee}}] = \sum_{j=1}^t \text{Var}[\mathbb{E}[C_j \mid \mathcal{G}_j]] + 2 \sum_{j=1}^t \sum_{i=1}^{j-1} \text{Cov}[\mathbb{E}[C_j \mid \mathcal{G}_j], \mathbb{E}[C_i \mid \mathcal{G}_i]]$$

tandis que

$$\text{Var}[X] = \sum_{j=1}^t \text{Var}[C_j] + 2 \sum_{j=1}^t \sum_{i=1}^{j-1} \text{Cov}[C_j, C_i].$$

On a  $\text{Var}[\mathbb{E}[C_j \mid \mathcal{G}_j]] \leq \text{Var}[C_j]$  pour chaque  $j$ .

Mais pour les covariances, on ne sait pas!

$$\text{Var}[X_{\text{ee}}] = \sum_{j=1}^t \text{Var}[\mathbb{E}[C_j \mid \mathcal{G}_j]] + 2 \sum_{j=1}^t \sum_{i=1}^{j-1} \text{Cov}[\mathbb{E}[C_j \mid \mathcal{G}_j], \mathbb{E}[C_i \mid \mathcal{G}_i]]$$

tandis que

$$\text{Var}[X] = \sum_{j=1}^t \text{Var}[C_j] + 2 \sum_{j=1}^t \sum_{i=1}^{j-1} \text{Cov}[C_j, C_i].$$

On a  $\text{Var}[\mathbb{E}[C_j \mid \mathcal{G}_j]] \leq \text{Var}[C_j]$  pour chaque  $j$ .

Mais pour les covariances, on ne sait pas!

Lorsque  $T$  est aléatoire, les choses se compliquent davantage.

Lorsque  $\mathcal{G}_j \subset \mathcal{G}_{j+1}$  pour tout  $j$ ,  $\{\mathcal{G}_j, j \geq 1\}$  est un **filtrage** et ECMC s'appelle **Monte Carlo filtré**. Par exemple,  $\mathcal{G}_j$  peut représenter l'information disponible lors du  $j$ -ième événement de la simulation.

## CMC pour rendre un estimateur continu

Lorsqu'on estime une **dérivée par différences finies** avec des VAC, on a vu que la **continuité** de l'estimateur est importante.

Souvent, on peut le rendre continu en prenant une espérance conditionnelle.

## CMC pour rendre un estimateur continu

Lorsqu'on estime une **dérivée par différences finies** avec des VAC, on a vu que la **continuité** de l'estimateur est importante.

Souvent, on peut le rendre continu en prenant une espérance conditionnelle.

**Example:** File *GI/GI/1* (rappel).

Dist. de durée de service  $G_\theta$  dépend de  $\theta \in \Theta$ . Les temps d'attente obéissent à l'équation de Lindley:  $W_1(\theta) = 0$  et

$$W_{j+1}(\theta) = \max(0, W_j(\theta) + S_j(\theta) - A_j),$$

## CMC pour rendre un estimateur continu

Lorsqu'on estime une **dérivée par différences finies** avec des VAC, on a vu que la **continuité** de l'estimateur est importante.

Souvent, on peut le rendre continu en prenant une espérance conditionnelle.

**Exemple:** File  $GI/GI/1$  (rappel).

Dist. de durée de service  $G_\theta$  dépend de  $\theta \in \Theta$ . Les temps d'attente obéissent à l'équation de Lindley:  $W_1(\theta) = 0$  et

$$W_{j+1}(\theta) = \max(0, W_j(\theta) + S_j(\theta) - A_j),$$

On veut estimer  $\mu(\theta + \delta) - \mu(\theta)$  où  $\mu(\theta) = \mathbb{E}[f(\theta, \mathbf{U})]$ ,

$$f(\theta, \mathbf{U}) = \sum_{j=1}^t \mathbb{I}[X_j(\theta) > K]$$

où  $X_j(\theta) = W_j(\theta) + S_j(\theta)$ . Cette fonction n'est **pas continue** en  $\theta$ .



On a déjà examiné l'estimateur discontinu

$$Y^{(A)} = \sum_{j=1}^t (\mathbb{I}[X_j(\theta + \delta) > K] - \mathbb{I}[X_j(\theta) > K]).$$

On a déjà examiné l'estimateur discontinu

$$Y^{(A)} = \sum_{j=1}^t (\mathbb{I}[X_j(\theta + \delta) > K] - \mathbb{I}[X_j(\theta) > K]).$$

Pour le rendre continu, on va remplacer les indicateurs  $\mathbb{I}[X_j(\theta) > K]$  par des espérances conditionnelles.

On a déjà examiné l'estimateur discontinu

$$Y^{(A)} = \sum_{j=1}^t (\mathbb{I}[X_j(\theta + \delta) > K] - \mathbb{I}[X_j(\theta) > K]).$$

Pour le rendre continu, on va remplacer les indicateurs  $\mathbb{I}[X_j(\theta) > K]$  par des espérances conditionnelles. Un premier candidat:

$$\begin{aligned} \mathbb{E}[\mathbb{I}[X_j(\theta) > K] \mid W_j(\theta)] &= \mathbb{P}[X_j(\theta) > K \mid W_j(\theta)] \\ &= \mathbb{P}[S_j(\theta) > K - W_j(\theta) \mid W_j(\theta)] \\ &= 1 - G_\theta(K - W_j(\theta)). \end{aligned}$$

On a déjà examiné l'estimateur discontinu

$$Y^{(A)} = \sum_{j=1}^t (\mathbb{I}[X_j(\theta + \delta) > K] - \mathbb{I}[X_j(\theta) > K]).$$

Pour le rendre continu, on va remplacer les indicateurs  $\mathbb{I}[X_j(\theta) > K]$  par des espérances conditionnelles. Un premier candidat:

$$\begin{aligned} \mathbb{E}[\mathbb{I}[X_j(\theta) > K] \mid W_j(\theta)] &= \mathbb{P}[X_j(\theta) > K \mid W_j(\theta)] \\ &= \mathbb{P}[S_j(\theta) > K - W_j(\theta) \mid W_j(\theta)] \\ &= 1 - G_\theta(K - W_j(\theta)). \end{aligned}$$

L'estimateur de différence devient

$$Y^{(B)} = \sum_{j=1}^t [G_\theta(K - W_j(\theta)) - G_{\theta+\delta}(K - W_j(\theta + \delta))].$$

En posant des conditions (raisonnables) sur  $G_\theta$ , on peut montrer que les conditions du corollaire 6.6 pour les VAC sont satisfaites; donc  $\text{Var}[Y^{(B)}/\delta]$  demeure bornée quand  $\delta \rightarrow 0$  (L'Ecuyer et Perron 1994).

En posant des conditions (raisonnables) sur  $G_\theta$ , on peut montrer que les conditions du corollaire 6.6 pour les VAC sont satisfaites; donc  $\text{Var}[Y^{(B)}/\delta]$  demeure bornée quand  $\delta \rightarrow 0$  (L'Ecuyer et Perron 1994).

On a

$$\begin{aligned} \frac{\partial}{\partial \theta} \mathbb{E}[\mathbb{I}[X_j(\theta) > K] \mid W_j(\theta)] &= -\frac{\partial}{\partial \theta} [G_\theta(K - W_j(\theta))] \\ &= -(\partial G_\theta / \partial \theta)(K - W_j(\theta)) \\ &\quad + g_\theta(K - W_j(\theta))(\partial W_j(\theta) / \partial \theta). \end{aligned}$$

En posant des conditions (raisonnables) sur  $G_\theta$ , on peut montrer que les conditions du corollaire 6.6 pour les VAC sont satisfaites; donc  $\text{Var}[Y^{(B)}/\delta]$  demeure bornée quand  $\delta \rightarrow 0$  (L'Ecuyer et Perron 1994).

On a

$$\begin{aligned} \frac{\partial}{\partial \theta} \mathbb{E}[\mathbb{I}[X_j(\theta) > K] \mid W_j(\theta)] &= -\frac{\partial}{\partial \theta} [G_\theta(K - W_j(\theta))] \\ &= -(\partial G_\theta / \partial \theta)(K - W_j(\theta)) \\ &\quad + g_\theta(K - W_j(\theta))(\partial W_j(\theta) / \partial \theta). \end{aligned}$$

Ok si on suppose que  $(\partial G_\theta(x) / \partial \theta)$  est borné uniformément en  $\theta$  et  $x$ , et  $G_\theta$  a une densité  $g_\theta$  bornée.

On peut aussi essayer de conditionner sur **moins d'information**, par exemple sur  $X_{j-1}(\theta)$  à la place de  $W_j(\theta)$ , pour obtenir un estimateur  $Y^{(C)}$ .



On peut aussi essayer de conditionner sur **moins d'information**, par exemple sur  $X_{j-1}(\theta)$  à la place de  $W_j(\theta)$ , pour obtenir un estimateur  $Y^{(C)}$ .

Expérience avec file  $M/M/1$ ,  $\lambda = 1$ ,  $G_\theta$  exponentielle de moyenne  $\theta$ ,  $K = 2$ ,  $t = 20$ ,  $\theta = 1/2$ , et  $n = 10^4$ . **Variance** des estimateurs de la différence:

	$\delta = 10^{-3}$	$\delta = 10^{-4}$	$\delta = 10^{-5}$	$\delta = 10^{-6}$
$Y^{(A)}$	$1.47 \times 10^{-2}$	$2.09 \times 10^{-3}$	$1.99 \times 10^{-4}$	$1.50 \times 10^{-5}$
$Y^{(B)}$	$2.96 \times 10^{-4}$	$3.05 \times 10^{-6}$	$2.94 \times 10^{-8}$	$2.92 \times 10^{-10}$
$Y^{(C)}$	$1.75 \times 10^{-4}$	$1.66 \times 10^{-6}$	$1.65 \times 10^{-8}$	$1.63 \times 10^{-10}$

## Estimation indirecte

**Exemple:** Pour une file  $GI/GI/1$  soient

$S_j$  = durée de service du client  $j$ ;

$W_j$  = temps d'attente (dans la file) pour le client  $j$ ;

$Q(t)$  = nombre de clients dans le système au temps  $t$ .

## Estimation indirecte

**Exemple:** Pour une file  $GI/GI/1$  soient

$S_j$  = durée de service du client  $j$ ;

$W_j$  = temps d'attente (dans la file) pour le client  $j$ ;

$Q(t)$  = nombre de clients dans le système au temps  $t$ .

Considérons aussi les moyennes théoriques à l'état stationnaire:

$w$  = attente moyenne dans la file par client;

$q$  = longueur moyenne de la file d'attente;

$\lambda = 1/(\text{moyenne des inter-arrivées}) = \text{taux d'arrivée}.$

## Estimation indirecte

**Exemple:** Pour une file  $GI/GI/1$  soient

$S_j$  = durée de service du client  $j$ ;

$W_j$  = temps d'attente (dans la file) pour le client  $j$ ;

$Q(t)$  = nombre de clients dans le système au temps  $t$ .

Considérons aussi les moyennes théoriques à l'état stationnaire:

$w$  = attente moyenne dans la file par client;

$q$  = longueur moyenne de la file d'attente;

$\lambda = 1/(\text{moyenne des inter-arrivées}) = \text{taux d'arrivée}$ .

La loi de Little nous dit que

$$q = \lambda w.$$

## Estimation indirecte

**Exemple:** Pour une file  $GI/GI/1$  soient

$S_j$  = durée de service du client  $j$ ;

$W_j$  = temps d'attente (dans la file) pour le client  $j$ ;

$Q(t)$  = nombre de clients dans le système au temps  $t$ .

Considérons aussi les moyennes théoriques à l'état stationnaire:

$w$  = attente moyenne dans la file par client;

$q$  = longueur moyenne de la file d'attente;

$\lambda = 1/(\text{moyenne des inter-arrivées}) = \text{taux d'arrivée}$ .

La **loi de Little** nous dit que

$$q = \lambda w.$$

En effet, on peut écrire de deux façons la somme des durées d'attente durant  $[0, t]$ , pour  $t$  très grand est:

$$tq \approx \int_0^t Q(t)dt \approx \sum_{i=1}^{N(t)} W_i \approx \lambda wt.$$

Pour estimer  $q$ , on peut prendre

$$\bar{Q}_t = \frac{1}{t} \int_0^t Q(t) dt$$

pour un grand  $t$  ou encore

$$\lambda \bar{W}_n$$

pour un grand  $n$ .

Pour estimer  $q$ , on peut prendre

$$\bar{Q}_t = \frac{1}{t} \int_0^t Q(t) dt$$

pour un grand  $t$  ou encore

$$\lambda \bar{W}_n$$

pour un grand  $n$ .

En supposant que les  $W_j$  sont négativement corrélés avec les  $A_j$  (ce qui est quasiment toujours le cas), on peut montrer que la variance asymptotique du second estimateur est plus petite que celle du premier (si on connaît  $\lambda$ ).

Pour estimer  $q$ , on peut prendre

$$\bar{Q}_t = \frac{1}{t} \int_0^t Q(t) dt$$

pour un grand  $t$  ou encore

$$\lambda \bar{W}_n$$

pour un grand  $n$ .

En supposant que les  $W_j$  sont négativement corrélés avec les  $A_j$  (ce qui est quasiment toujours le cas), on peut montrer que la variance asymptotique du second estimateur est plus petite que celle du premier (si on connaît  $\lambda$ ).

**Mieux:** utiliser une moyenne pondérée de  $\bar{Q}_t$  et  $\lambda \bar{W}_{N(t)}$ :

$$\beta \bar{Q}_t + (1 - \beta) \lambda \bar{W}_{N(t)} = \bar{Q}_t - \beta (\lambda \bar{W}_{N(t)} - \bar{Q}_t).$$

Équivaut à choisir  $\bar{Q}_t$  comme estimateur et la différence  $\lambda \bar{W}_{N(t)} - \bar{Q}_t$  comme variable de contrôle.



Pour estimer  $q$ , on peut prendre

$$\bar{Q}_t = \frac{1}{t} \int_0^t Q(t) dt$$

pour un grand  $t$  ou encore

$$\lambda \bar{W}_n$$

pour un grand  $n$ .

En supposant que les  $W_j$  sont négativement corrélés avec les  $A_j$  (ce qui est quasiment toujours le cas), on peut montrer que la variance asymptotique du second estimateur est plus petite que celle du premier (si on connaît  $\lambda$ ).

**Mieux:** utiliser une moyenne pondérée de  $\bar{Q}_t$  et  $\lambda \bar{W}_{N(t)}$ :

$$\beta \bar{Q}_t + (1 - \beta) \lambda \bar{W}_{N(t)} = \bar{Q}_t - \beta (\lambda \bar{W}_{N(t)} - \bar{Q}_t).$$

Équivaut à choisir  $\bar{Q}_t$  comme estimateur et la différence  $\lambda \bar{W}_{N(t)} - \bar{Q}_t$  comme variable de contrôle.

Cette approche est particulièrement efficace lorsque le taux d'utilisation du serveur est élevé ( “heavy traffic” ).

Cette approche est particulièrement efficace lorsque le taux d'utilisation du serveur est élevé ( “heavy traffic” ).

Lorsque  $\lambda$  est connu, on peut en plus utiliser la moyenne empirique des inter-arrivées et/ou des durées de services comme variables de contrôle.

Cette approche est particulièrement efficace lorsque le taux d'utilisation du serveur est élevé ( “heavy traffic” ).

Lorsque  $\lambda$  est connu, on peut en plus utiliser la moyenne empirique des inter-arrivées et/ou des durées de services comme variables de contrôle.

Pour estimer  $w$ , il vaut mieux prendre  $\bar{W}_n$  que  $\bar{Q}_t/\lambda$ .

Cette approche est particulièrement efficace lorsque le taux d'utilisation du serveur est élevé ( “heavy traffic” ).

Lorsque  $\lambda$  est connu, on peut en plus utiliser la moyenne empirique des inter-arrivées et/ou des durées de services comme variables de contrôle.

Pour estimer  $w$ , il vaut mieux prendre  $\bar{W}_n$  que  $\bar{Q}_t/\lambda$ .

Pour estimer la durée espérée de **séjour**, il vaut mieux prendre  $\bar{W}_n + \mathbb{E}[S_j]$ .

# Stratification

**Idée générale:** partitionner la “population” en  $k$  strates, de manière à ce que la variance à l’intérieur des strates soit inférieure à la variance générale, puis estimer la quantité désirée séparément à l’intérieur de chaque strate.

# Stratification

**Idée générale:** partitionner la “population” en  $k$  strates, de manière à ce que la variance à l’intérieur des strates soit inférieure à la variance générale, puis estimer la quantité désirée séparément à l’intérieur de chaque strate.

$p_s$  = fraction de la population dans la strate  $s$ .

$N_s$  = nombre d’observations dans la strate  $s$ .

$n = \sum_{s=1}^k N_s$ .

$X_{s,i}$  =  $i$ -ième observation de la strate  $s$ .

# Stratification

**Idée générale:** partitionner la “population” en  $k$  strates, de manière à ce que la variance à l’intérieur des strates soit inférieure à la variance générale, puis estimer la quantité désirée séparément à l’intérieur de chaque strate.

$p_s$  = fraction de la population dans la strate  $s$ .

$N_s$  = nombre d’observations dans la strate  $s$ .

$n = \sum_{s=1}^k N_s$ .

$X_{s,i}$  =  $i$ -ième observation de la strate  $s$ .

L’estimateur de

$$\mu = \mathbb{E}[X] = \sum_{s=1}^k p_s \mathbb{E}[X_{s,i}] = \sum_{s=1}^k p_s \mu_s$$

est alors

$$\bar{X}_{s,n} = \sum_{s=1}^k p_s \left( \frac{1}{N_s} \sum_{i=1}^{N_s} X_{s,i} \right) = \sum_{s=1}^k p_s \hat{\mu}_s.$$



**Cas où on peut décider des strates:**  $N_s = n_s$ .

Soit  $\sigma_s^2 = \text{Var}[X \mid S = s]$ , la variance dans la strate  $s$ . On a

$$\text{Var}[\bar{X}_{s,n}] = \sum_{s=1}^k p_s^2 \text{Var}[\hat{\mu}_s] = \sum_{s=1}^k \frac{p_s^2 \sigma_s^2}{n_s},$$

dont un estimateur sans biais est:

$$S_{s,n}^2 = \sum_{s=1}^k \frac{p_s^2 \hat{\sigma}_s^2}{n_s},$$

où  $\hat{\sigma}_s^2$  est la variance empirique  $X_{s,1}, \dots, X_{s,n_s}$ .

**Cas où on peut décider des strates:**  $N_s = n_s$ .

Soit  $\sigma_s^2 = \text{Var}[X \mid S = s]$ , la variance dans la strate  $s$ . On a

$$\text{Var}[\bar{X}_{s,n}] = \sum_{s=1}^k p_s^2 \text{Var}[\hat{\mu}_s] = \sum_{s=1}^k \frac{p_s^2 \sigma_s^2}{n_s},$$

dont un estimateur sans biais est:

$$S_{s,n}^2 = \sum_{s=1}^k \frac{p_s^2 \hat{\sigma}_s^2}{n_s},$$

où  $\hat{\sigma}_s^2$  est la variance empirique  $X_{s,1}, \dots, X_{s,n_s}$ .

**Allocation proportionnelle:**  $n_s \approx np_s$ .

**Cas où on peut décider des strates:**  $N_s = n_s$ .

Soit  $\sigma_s^2 = \text{Var}[X \mid S = s]$ , la variance dans la strate  $s$ . On a

$$\text{Var}[\bar{X}_{s,n}] = \sum_{s=1}^k p_s^2 \text{Var}[\hat{\mu}_s] = \sum_{s=1}^k \frac{p_s^2 \sigma_s^2}{n_s},$$

dont un estimateur sans biais est:

$$S_{s,n}^2 = \sum_{s=1}^k \frac{p_s^2 \hat{\sigma}_s^2}{n_s},$$

où  $\hat{\sigma}_s^2$  est la variance empirique  $X_{s,1}, \dots, X_{s,n_s}$ .

**Allocation proportionnelle:**  $n_s \approx np_s$ .

$$\text{Var}[\bar{X}_{sp,n}] = \frac{1}{n} \sum_{s=1}^k p_s \sigma_s^2.$$

**Allocation optimale:** On veut minimiser  $\text{Var}[\bar{X}_{s,n}]$  p.r. à  $n_1, \dots, n_k$ , sous les contraintes que  $n_s \geq 0$  et  $n_1 + \dots + n_k = n$ .

Intuition: à l'optimum, la dérivée

$$\frac{\partial \text{Var}[\bar{X}_{s,n}]}{\partial n_s} = -\frac{p_s^2 \sigma_s^2}{n_s^2}$$

doit être la même pour tout  $s$ .

On doit donc prendre  $n_s$  proportionnel à  $p_s \sigma_s$ .

**Allocation optimale:** On veut minimiser  $\text{Var}[\bar{X}_{s,n}]$  p.r. à  $n_1, \dots, n_k$ , sous les contraintes que  $n_s \geq 0$  et  $n_1 + \dots + n_k = n$ .

Intuition: à l'optimum, la dérivée

$$\frac{\partial \text{Var}[\bar{X}_{s,n}]}{\partial n_s} = -\frac{p_s^2 \sigma_s^2}{n_s^2}$$

doit être la même pour tout  $s$ .

On doit donc prendre  $n_s$  proportionnel à  $p_s \sigma_s$ .

Pour que la somme donne  $n$ , on doit donc prendre (en négligeant les arrondis):

$$n_s^* = \frac{np_s \sigma_s}{\bar{\sigma}} \quad \text{où} \quad \bar{\sigma} = \sum_{s=1}^k p_s \sigma_s.$$

**Allocation optimale:** On veut minimiser  $\text{Var}[\bar{X}_{s,n}]$  p.r. à  $n_1, \dots, n_k$ , sous les contraintes que  $n_s \geq 0$  et  $n_1 + \dots + n_k = n$ .

Intuition: à l'optimum, la dérivée

$$\frac{\partial \text{Var}[\bar{X}_{s,n}]}{\partial n_s} = -\frac{p_s^2 \sigma_s^2}{n_s^2}$$

doit être la même pour tout  $s$ .

On doit donc prendre  $n_s$  proportionnel à  $p_s \sigma_s$ .

Pour que la somme donne  $n$ , on doit donc prendre (en négligeant les arrondis):

$$n_s^* = \frac{np_s \sigma_s}{\bar{\sigma}} \quad \text{où} \quad \bar{\sigma} = \sum_{s=1}^k p_s \sigma_s.$$

On peut le montrer rigoureusement en utilisant un multiplicateur de Lagrange.

Cela donne

$$\text{Var}[\bar{X}_{\text{so},n}] = \sum_{s=1}^k \frac{p_s^2 \sigma_s^2}{n_s} = \sum_{s=1}^k \frac{p_s^2 \sigma_s^2}{np_s \sigma_s / \bar{\sigma}} = \frac{\bar{\sigma}^2}{n}.$$

En décomposant la variance, on peut comparer:

$$\text{Var}[\bar{X}_n] = \frac{1}{n} (\mathbb{E}[\text{Var}[X|S]] + \text{Var}[\mathbb{E}[X|S]])$$

En décomposant la variance, on peut comparer:

$$\begin{aligned}\text{Var}[\bar{X}_n] &= \frac{1}{n} (\mathbb{E}[\text{Var}[X|S]] + \text{Var}[\mathbb{E}[X|S]]) \\ &= \frac{1}{n} \left( \sum_{s=1}^k p_s \sigma_s^2 + \sum_{s=1}^k p_s (\mu_s - \mu)^2 \right)\end{aligned}$$



En décomposant la variance, on peut comparer:

$$\begin{aligned}\text{Var}[\bar{X}_n] &= \frac{1}{n} (\mathbb{E}[\text{Var}[X|S]] + \text{Var}[\mathbb{E}[X|S]]) \\ &= \frac{1}{n} \left( \sum_{s=1}^k p_s \sigma_s^2 + \sum_{s=1}^k p_s (\mu_s - \mu)^2 \right) \\ &= \text{Var}[\bar{X}_{\text{sp},n}] + \frac{1}{n} \sum_{s=1}^k p_s (\mu_s - \mu)^2\end{aligned}$$

En décomposant la variance, on peut comparer:

$$\begin{aligned}
 \text{Var}[\bar{X}_n] &= \frac{1}{n} (\mathbb{E}[\text{Var}[X|S]] + \text{Var}[\mathbb{E}[X|S]]) \\
 &= \frac{1}{n} \left( \sum_{s=1}^k p_s \sigma_s^2 + \sum_{s=1}^k p_s (\mu_s - \mu)^2 \right) \\
 &= \text{Var}[\bar{X}_{\text{sp},n}] + \frac{1}{n} \sum_{s=1}^k p_s (\mu_s - \mu)^2 \\
 &= \text{Var}[\bar{X}_{\text{so},n}] + \frac{1}{n} \sum_{s=1}^k p_s (\sigma_s - \bar{\sigma})^2 + \frac{1}{n} \sum_{s=1}^k p_s (\mu_s - \mu)^2.
 \end{aligned}$$

**Exemple.** On peut stratifier directement par rapport aux uniformes utilisées dans une simulation. Regardons le cas (simpliste) d'un problème à une dimension:

$$\mu = \mathbb{E}[f(U)] = \int_0^1 f(u) du.$$

**Exemple.** On peut stratifier directement par rapport aux uniformes utilisées dans une simulation. Regardons le cas (simpliste) d'un problème à une dimension:

$$\mu = \mathbb{E}[f(U)] = \int_0^1 f(u) du.$$

On peut partitionner  $(0, 1)$  en  $k$  intervalles de longueur  $1/k$ , générer  $U_s$  uniformément sur l'intervalle  $s$  pour chaque  $s$ , et estimer  $\mu$  par

$$X_s = \frac{1}{k} \sum_{s=1}^k f(U_s).$$

On peut répéter cela  $m$  fois pour un budget de  $n = mk$  simulations, et calculer la moyenne générale  $\bar{X}_{s,n}$ .

**Exemple.** On peut stratifier directement par rapport aux uniformes utilisées dans une simulation. Regardons le cas (simpliste) d'un problème à une dimension:

$$\mu = \mathbb{E}[f(U)] = \int_0^1 f(u) du.$$

On peut partitionner  $(0, 1)$  en  $k$  intervalles de longueur  $1/k$ , générer  $U_s$  uniformément sur l'intervalle  $s$  pour chaque  $s$ , et estimer  $\mu$  par

$$X_s = \frac{1}{k} \sum_{s=1}^k f(U_s).$$

On peut répéter cela  $m$  fois pour un budget de  $n = mk$  simulations, et calculer la moyenne générale  $\bar{X}_{s,n}$ . L'estimateur de variance devient

$$S_{s,n}^2 = \sum_{s=1}^k \frac{p_s^2 \hat{\sigma}_s^2}{n_s} = \sum_{s=1}^k \frac{\hat{\sigma}_s^2}{k^2 (n/k)} = \frac{1}{nk} \sum_{s=1}^k \hat{\sigma}_s^2.$$

Il s'agit ici de l'allocation proportionnelle et on a

$$\text{Var}[\bar{X}_n] = \text{Var}[\bar{X}_{s,n}] + \frac{1}{nk} \sum_{s=1}^k (\mu_s - \mu)^2$$

où

$$\mu_s = \int_{(s-1)/k}^{s/k} k f(u) du$$

est la moyenne de  $f$  sur le sous-intervalle  $s$ .

Plus la moyenne de  $f$  diffère entre les sous-intervalles, plus la variance est réduite.

Il s'agit ici de l'allocation proportionnelle et on a

$$\text{Var}[\bar{X}_n] = \text{Var}[\bar{X}_{s,n}] + \frac{1}{nk} \sum_{s=1}^k (\mu_s - \mu)^2$$

où

$$\mu_s = \int_{(s-1)/k}^{s/k} k f(u) du$$

est la moyenne de  $f$  sur le sous-intervalle  $s$ .

Plus la moyenne de  $f$  diffère entre les sous-intervalles, plus la variance est réduite.

Se généralise en principe à  $t$  dimensions: on partitionne l'hypercube unitaire en  $k^t$  sous-cubes. Mais cela devient vite impraticable lorsque  $t$  augmente.

Par contre, on peut l'appliquer aux quelques variables les plus importantes.

Exemple: centre d'appels.

**Post-stratification:** Ne pas choisir (fixer) les  $N_s$ ; laisser la simulation les générer comme d'habitude, puis utiliser  $\bar{X}_{s,n}$ .

Avantages: Dans plusieurs situations, on ne peut pas (ou il est trop difficile) de choisir la strate.

On peut appliquer la post-stratification seulement après coup.



## Variables antithétiques (AV)

Semblable à l'idée des VAC, sauf que l'on veut maintenant estimer **une seule espérance** par un moyenne de plusieurs estimateurs **négativement corrélés**.

Supposons que l'on a  $k$  estimateurs sans biais de  $\mu$ ,  $(X^{(1)}, \dots, X^{(k)})$ , avec  $\text{Var}[X^{(1)}] = \dots = \text{Var}[X^{(k)}]$ .

## Variables antithétiques (AV)

Semblable à l'idée des VAC, sauf que l'on veut maintenant estimer **une seule espérance** par un moyenne de plusieurs estimateurs **négativement corrélés**.

Supposons que l'on a  $k$  estimateurs sans biais de  $\mu$ ,  $(X^{(1)}, \dots, X^{(k)})$ , avec  $\text{Var}[X^{(1)}] = \dots = \text{Var}[X^{(k)}]$ .

La moyenne

$$X_a = \frac{1}{k} \sum_{\ell=1}^k X^{(\ell)}$$

est un estimateur **sans biais** de  $\mu$ , avec variance

$$\text{Var}[X_a] = \frac{1}{k^2} \sum_{j=1}^k \sum_{\ell=1}^k \text{Cov}[X^{(j)} X^{(\ell)}] = \frac{\text{Var}[X_{(1)}]}{k} + \frac{2}{k^2} \sum_{j < \ell} \text{Cov}[X^{(j)} X^{(\ell)}].$$

## Variables antithétiques (AV)

Semblable à l'idée des VAC, sauf que l'on veut maintenant estimer **une seule espérance** par un moyenne de plusieurs estimateurs **négativement corrélés**.

Supposons que l'on a  $k$  estimateurs sans biais de  $\mu$ ,  $(X^{(1)}, \dots, X^{(k)})$ , avec  $\text{Var}[X^{(1)}] = \dots = \text{Var}[X^{(k)}]$ .

La moyenne

$$X_a = \frac{1}{k} \sum_{\ell=1}^k X^{(\ell)}$$

est un estimateur **sans biais** de  $\mu$ , avec variance

$$\text{Var}[X_a] = \frac{1}{k^2} \sum_{j=1}^k \sum_{\ell=1}^k \text{Cov}[X^{(j)} X^{(\ell)}] = \frac{\text{Var}[X^{(1)}]}{k} + \frac{2}{k^2} \sum_{j < \ell} \text{Cov}[X^{(j)} X^{(\ell)}].$$

Si les  $X^{(\ell)}$  sont indépendants, les covariances sont nulles.

La variance est réduite ssi  $\text{Var}[X_a] < \text{Var}[X^{(1)}]/k$ ,

ssi la somme des covariances est négative.

Pour un total de  $n = mk$  répétitions, on définit

$$\bar{X}_{a,n} = \frac{1}{m} \sum_{i=1}^m X_{a,i} = \frac{1}{n} \sum_{i=1}^m \sum_{\ell=1}^k X_i^{(\ell)},$$

où  $X_{a,1}, \dots, X_{a,m}$  sont des copies i.i.d. de  $X_a$ .

Pour un total de  $n = mk$  répétitions, on définit

$$\bar{X}_{a,n} = \frac{1}{m} \sum_{i=1}^m X_{a,i} = \frac{1}{n} \sum_{i=1}^m \sum_{\ell=1}^k X_i^{(\ell)},$$

où  $X_{a,1}, \dots, X_{a,m}$  sont des copies i.i.d. de  $X_a$ .

Estimateur sans biais de  $\text{Var}[X_a]$ : variance empirique des  $m$  copies,

$$S_{a,m}^2 = \frac{1}{m-1} \sum_{i=1}^m (X_{a,i} - \bar{X}_{a,n})^2.$$

Pour un total de  $n = mk$  répétitions, on définit

$$\bar{X}_{a,n} = \frac{1}{m} \sum_{i=1}^m X_{a,i} = \frac{1}{n} \sum_{i=1}^m \sum_{\ell=1}^k X_i^{(\ell)},$$

où  $X_{a,1}, \dots, X_{a,m}$  sont des copies i.i.d. de  $X_a$ .

Estimateur sans biais de  $\text{Var}[X_a]$ : variance empirique des  $m$  copies,

$$S_{a,m}^2 = \frac{1}{m-1} \sum_{i=1}^m (X_{a,i} - \bar{X}_{a,n})^2.$$

**Méthode AV générale:**

Soit  $\mu = \mathbb{E}[X]$ , où  $X = f(\mathbf{U})$  et  $\mathbf{U} = (U_1, U_2, \dots)$  une suite i.i.d.  $U(0, 1)$ .

Pour  $\ell = 1, \dots, k$ , soit  $X^{(\ell)} = f(\mathbf{U}_\ell)$  où  $\mathbf{U}_\ell$  est une suite i.i.d.  $U(0, 1)$ .

Le but de AV (général) est d'induire une dépendance entre les  $\mathbf{U}_\ell$  de manière à rendre  $\sum_{j < \ell} \text{Cov}[X^{(j)}, X^{(\ell)}] < 0$ .

## Méthode AV “classique”: Paires antithétiques.

On a  $k = 2$ ,

$$\mathbf{U}_1 = \mathbf{U} = (U_1, U_2, \dots),$$

$$\mathbf{U}_2 = \mathbf{1} - \mathbf{U} = (1 - U_1, 1 - U_2, \dots), \text{ la suite antithétique.}$$

## Méthode AV “classique”: Paires antithétiques.

On a  $k = 2$ ,

$$\mathbf{U}_1 = \mathbf{U} = (U_1, U_2, \dots),$$

$$\mathbf{U}_2 = \mathbf{1} - \mathbf{U} = (1 - U_1, 1 - U_2, \dots), \text{ la suite antithétique.}$$

Posons  $X_a = [X^{(1)} + X^{(2)}]/2$ ,

$$\begin{aligned} \text{Var}[X_a] &= (\text{Var}[X^{(1)}] + \text{Var}[X^{(2)}] + 2 \text{Cov}[X^{(1)}, X^{(2)}])/4 \\ &= (\text{Var}[X^{(1)}] + \text{Cov}[X^{(1)}, X^{(2)}])/2, \end{aligned}$$

et  $\text{Var}[X_a] < \text{Var}[X]/2$  ssi  $\text{Cov}[X^{(1)}, X^{(2)}] < 0$ .

**Intuition:** Les événements désastreux pour  $X^{(1)}$  seront compensés par des événements antithétiques chanceux pour  $X^{(2)}$ , et vice-versa.



Un échantillonnage est **AV-concordant** si  $U_{2,j} = 1 - U_{1,j}$  pour certaines coordonnées  $j$  telles que  $f$  est monotone par rapport à  $U_j$ , tandis que  $U_{2,j}$  est indépendant de  $U_{1,j}$  pour les autres coordonnées.

**Théorème.** Pour un échantillonnage AV-concordant, on a  $\text{Cov}[X^{(1)}, X^{(2)}] \leq 0$ .

Un échantillonnage est **AV-concordant** si  $U_{2,j} = 1 - U_{1,j}$  pour certaines coordonnées  $j$  telles que  $f$  est monotone par rapport à  $U_j$ , tandis que  $U_{2,j}$  est indépendant de  $U_{1,j}$  pour les autres coordonnées.

**Théorème.** Pour un échantillonnage AV-concordant, on a  $\text{Cov}[X^{(1)}, X^{(2)}] \leq 0$ .

**Preuve:** en prenant  $f_1 = f_2 = f$ , VAC-discordant est équivalent à AV-concordant, et le résultat découle du théorème sur les VAC.

Un échantillonnage est **AV-concordant** si  $U_{2,j} = 1 - U_{1,j}$  pour certaines coordonnées  $j$  telles que  $f$  est monotone par rapport à  $U_j$ , tandis que  $U_{2,j}$  est indépendant de  $U_{1,j}$  pour les autres coordonnées.

**Théorème.** Pour un échantillonnage AV-concordant, on a  $\text{Cov}[X^{(1)}, X^{(2)}] \leq 0$ .

**Preuve:** en prenant  $f_1 = f_2 = f$ , VAC-discordant est équivalent à AV-concordant, et le résultat découle du théorème sur les VAC.

Meilleur cas possible: pour une fonction linéaire la variance est réduite à zéro.

Un échantillonnage est **AV-concordant** si  $U_{2,j} = 1 - U_{1,j}$  pour certaines coordonnées  $j$  telles que  $f$  est monotone par rapport à  $U_j$ , tandis que  $U_{2,j}$  est indépendant de  $U_{1,j}$  pour les autres coordonnées.

**Théorème.** Pour un échantillonnage AV-concordant, on a  $\text{Cov}[X^{(1)}, X^{(2)}] \leq 0$ .

**Preuve:** en prenant  $f_1 = f_2 = f$ , VAC-discordant est équivalent à AV-concordant, et le résultat découle du théorème sur les VAC.

Meilleur cas possible: pour une fonction linéaire la variance est réduite à zéro.  
Pire cas: si  $X^{(1)}$  et  $X^{(2)}$  sont parfaitement corrélées, la variance est doublée (e.g.,  $f(\mathbf{U}) = |U_1 - 1/2|$ ).

Pour **minimiser la covariance**, il faudrait générer directement  $X^{(1)} = F^{-1}(U)$  et  $X^{(2)} = F^{-1}(1 - U)$ , où  $U \sim U(0, 1)$  et  $F$  est la fonct. de répartition de  $X$ .

Pas pratique.

Comme pour les VAC, on appliquera l'inversion pour maximiser la covariance à un plus bas niveau.

**Exemple.** Réseau d'activités stochastique.

On veut estimer  $\mu = \mathbb{P}[T > x]$ , où  $T$  est la longueur du plus long chemin et  $x$  une constante.

On génère  $V_j = F_j^{-1}(U_j)$  pour  $X^{(1)}$  et  $V_j = F_j^{-1}(1 - U_j)$  pour  $X^{(2)}$ .

Puisque  $T$  est non-décroissante en  $V_j = F_j^{-1}(U_j)$ , qui est non-décroissante en  $U_j$ , on a un échantillonnage AV-concordant et la variance est réduite.

**Exemple.** Réseau d'activités stochastique.

On veut estimer  $\mu = \mathbb{P}[T > x]$ , où  $T$  est la longueur du plus long chemin et  $x$  une constante.

On génère  $V_j = F_j^{-1}(U_j)$  pour  $X^{(1)}$  et  $V_j = F_j^{-1}(1 - U_j)$  pour  $X^{(2)}$ .

Puisque  $T$  est non-décroissante en  $V_j = F_j^{-1}(U_j)$ , qui est non-décroissante en  $U_j$ , on a un échantillonnage AV-concordant et la variance est réduite.

Expérience avec l'exemple numérique déjà vu:  
la variance est réduite environ de moitié.

**Exemple.** Option asiatique.

Soit un mouvement brownien géométrique  $S(\cdot)$  généré à partir d'un mouvement brownien standard  $B(\cdot)$  dont les valeurs aux instants d'observation  $t_j$  sont générées par

$$B(t_j) = B(t_{j-1}) + \sqrt{t_j - t_{j-1}} Z_j,$$

où  $t_0 = 0$  et  $Z_1, \dots, Z_d$  sont i.i.d.  $N(0, 1)$ .

**Exemple.** Option asiatique.

Soit un mouvement brownien géométrique  $S(\cdot)$  généré à partir d'un mouvement brownien standard  $B(\cdot)$  dont les valeurs aux instants d'observation  $t_j$  sont générées par

$$B(t_j) = B(t_{j-1}) + \sqrt{t_j - t_{j-1}} Z_j,$$

où  $t_0 = 0$  et  $Z_1, \dots, Z_d$  sont i.i.d.  $N(0, 1)$ .

Chaque  $S(t_j)$  est non-décroissant par rapport à chaque  $Z_\ell$ . Ainsi la fonction est **monotone** si le revenu de l'option est monotone en fonction des  $S(t_j)$  et si les  $Z_j$  sont générés par inversion (ou son équivalent).



### Exemple. Option asiatique.

Soit un mouvement brownien géométrique  $S(\cdot)$  généré à partir d'un mouvement brownien standard  $B(\cdot)$  dont les valeurs aux instants d'observation  $t_j$  sont générées par

$$B(t_j) = B(t_{j-1}) + \sqrt{t_j - t_{j-1}} Z_j,$$

où  $t_0 = 0$  et  $Z_1, \dots, Z_d$  sont i.i.d.  $N(0, 1)$ .

Chaque  $S(t_j)$  est non-décroissant par rapport à chaque  $Z_\ell$ . Ainsi la fonction est **monotone** si le revenu de l'option est monotone en fonction des  $S(t_j)$  et si les  $Z_j$  sont générés par inversion (ou son équivalent).

Pour la répétition **antithétique**, il suffit de remplacer  $(Z_1, \dots, Z_t)$  par  $(-Z_1, \dots, -Z_t)$ , car si  $U_j = \Phi(Z_j)$ , alors  $U_j \sim U(0, 1)$  et  $1 - U_j = 1 - \Phi(Z_j) = \Phi(-Z_j)$ .

Le mouvement brownien correspondant est une copie-miroir de l'autre!

Fonctionne aussi si on utilise le **pont brownien**.

**Combinaison VAC-AV.** Supposons que l'on veut comparer deux systèmes en utilisant des VAC, et aussi des paires de répétitions antithétiques.

Si ces deux méthodes fonctionnent bien séparément, fera-t-on nécessairement mieux en les combinant ?

**Combinaison VAC-AV.** Supposons que l'on veut comparer deux systèmes en utilisant des VAC, et aussi des paires de répétitions antithétiques.

Si ces deux méthodes fonctionnent bien séparément, fera-t-on nécessairement mieux en les combinant ? **Non.**

**Combinaison VAC-AV.** Supposons que l'on veut comparer deux systèmes en utilisant des VAC, et aussi des paires de répétitions antithétiques.

Si ces deux méthodes fonctionnent bien séparément, fera-t-on nécessairement mieux en les combinant ? **Non.**

Soit  $(X_k^{(1)}, X_k^{(2)}) = f_k(\mathbf{U}), f_k(\mathbf{1} - \mathbf{U}))$  la paire antithétique pour le système  $k$ .  
On a

$$\begin{aligned}
 & \text{Var} \left[ \frac{X_2^{(1)} + X_2^{(2)}}{2} - \frac{X_1^{(1)} + X_1^{(2)}}{2} \right] \\
 = & \frac{\text{Var}[X_1^{(1)}] + \text{Var}[X_1^{(2)}] + \text{Var}[X_2^{(1)}] + \text{Var}[X_2^{(2)}]}{4} \\
 & + \frac{\text{Cov}[X_1^{(1)}, X_1^{(2)}] + \text{Cov}[X_2^{(1)}, X_2^{(2)}]}{2} - \frac{\text{Cov}[X_1^{(1)}, X_2^{(1)}] + \text{Cov}[X_1^{(2)}, X_2^{(2)}]}{2} \\
 & - \frac{\text{Cov}[X_1^{(2)}, X_2^{(1)}] + \text{Cov}[X_1^{(1)}, X_2^{(2)}]}{2}. \\
 = & \text{Var}[X_1^{(1)}] + \underbrace{\text{Cov}[X_1^{(1)}, X_1^{(2)}]}_{\leq 0} - \underbrace{\text{Cov}[X_1^{(1)}, X_2^{(1)}]}_{\geq 0} - \underbrace{\text{Cov}[X_1^{(2)}, X_2^{(1)}]}_{?}.
 \end{aligned}$$

## Échantillonnage par hypercube Latin (LHS)

C'est une forme particulière de **stratification** avec allocation proportionnelle. On stratifie indépendamment par rapport à chaque uniforme. Le nombre **s** de dimensions peut être très grand: stratification **unidimensionnelle** seulement.

## Échantillonnage par hypercube Latin (LHS)

C'est une forme particulière de **stratification** avec allocation proportionnelle.

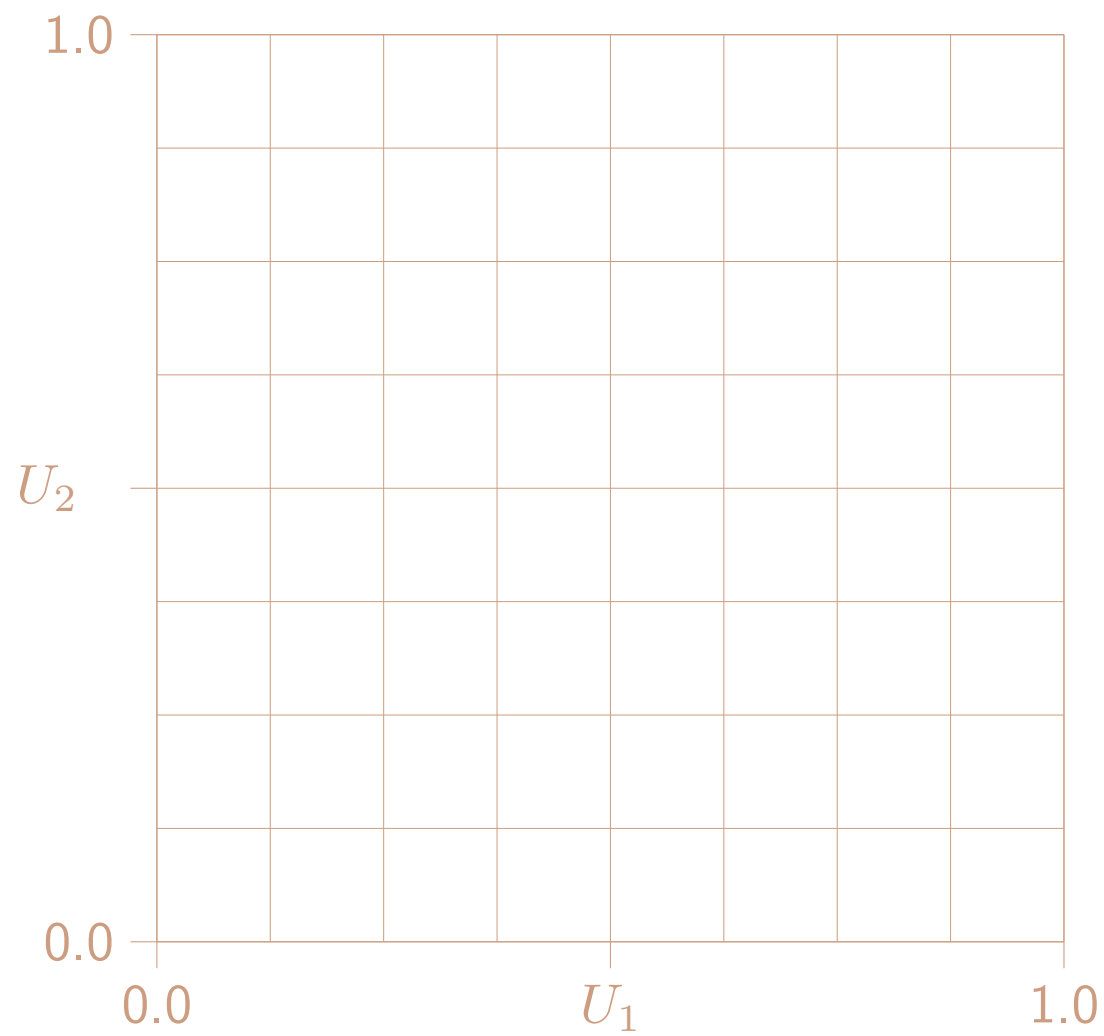
On stratifie indépendamment par rapport à chaque uniforme. Le nombre  $s$  de dimensions peut être très grand: stratification **unidimensionnelle** seulement.

On partitionne l'hypercube en  $k^s$  sous-cubes, mais au lieu de générer un point par sous-cube, on génère  $k$  points au total de manière à ce que pour chaque coordonnée  $j$ , chacun des sous-intervalles de  $(0, 1)$  contient exactement un point ayant sa coordonnée  $j$  dans ce sous-intervalle.

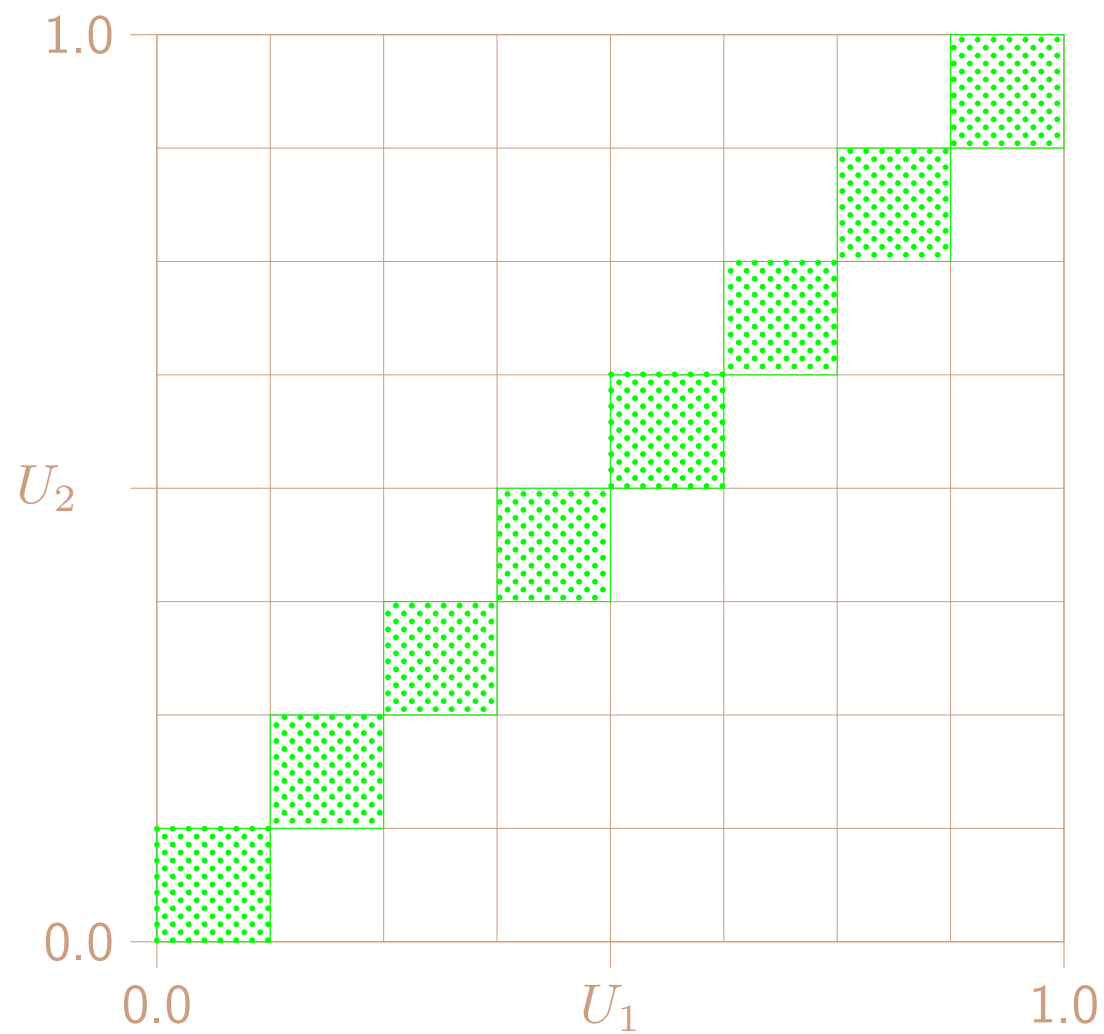
## Échantillonnage par hypercube Latin (LHS)

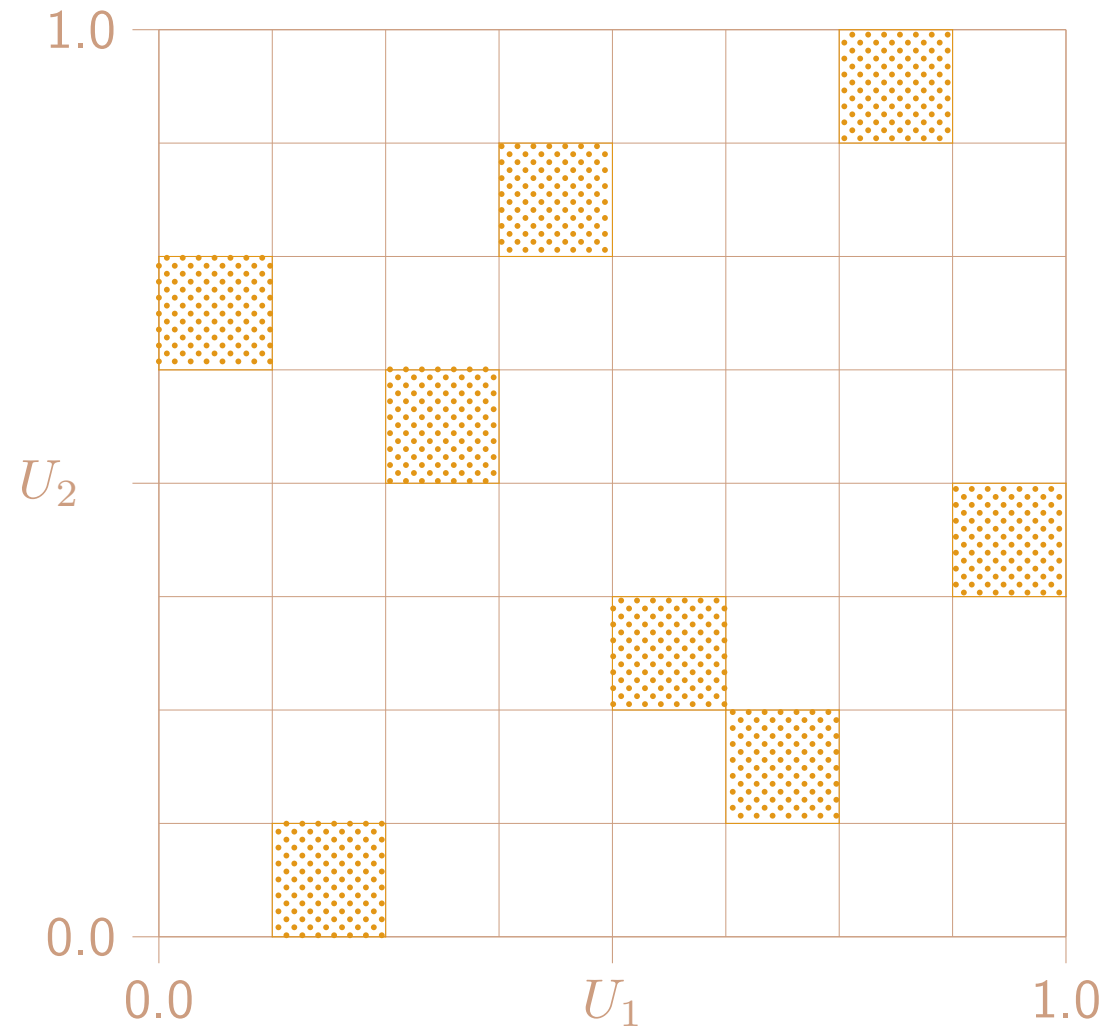
C'est une forme particulière de **stratification** avec allocation proportionnelle. On stratifie indépendamment par rapport à chaque uniforme. Le nombre  $s$  de dimensions peut être très grand: stratification **unidimensionnelle** seulement. On partitionne l'hypercube en  $k^s$  sous-cubes, mais au lieu de générer un point par sous-cube, on génère  $k$  points au total de manière à ce que pour chaque coordonnée  $j$ , chacun des sous-intervalles de  $(0, 1)$  contient exactement un point ayant sa coordonnée  $j$  dans ce sous-intervalle.

(Si on préfère, on peut appliquer la méthode seulement pour certaines coordonnées de  $\mathbf{U}$ , disons pour  $j \in \Psi$ .)









On fait une permutation aléatoire des colonnes.

Les  $k$  points

$$\mathbf{U}_1 = (U_{1,1}, \dots, U_{1,j}, \dots, U_{1,s}), \quad \dots, \quad \mathbf{U}_k = (U_{k,1}, \dots, U_{k,j}, \dots, U_{k,s}),$$

sont générés comme suit:

Les  $k$  points

$$\mathbf{U}_1 = (U_{1,1}, \dots, U_{1,j}, \dots, U_{1,s}), \quad \dots, \quad \mathbf{U}_k = (U_{k,1}, \dots, U_{k,j}, \dots, U_{k,s}),$$

sont générés comme suit:

1. Choisir un ensemble de coordonnées  $\Psi \subseteq \{1, \dots, s\}$  pour lesquelles on va appliquer LHS.

Les  $k$  points

$$\mathbf{U}_1 = (U_{1,1}, \dots, U_{1,j}, \dots, U_{1,s}), \quad \dots, \quad \mathbf{U}_k = (U_{k,1}, \dots, U_{k,j}, \dots, U_{k,s}),$$

sont générés comme suit:

1. Choisir un ensemble de coordonnées  $\Psi \subseteq \{1, \dots, s\}$  pour lesquelles on va appliquer LHS.
2. Pour chaque  $j \in \Psi$ , générer une **permutation aléatoire**  $(\pi_{1,j}, \dots, \pi_{k,j})$  des entiers  $\{1, \dots, k\}$ .

## Les $k$ points

$$\mathbf{U}_1 = (U_{1,1}, \dots, U_{1,j}, \dots, U_{1,s}), \quad \dots, \quad \mathbf{U}_k = (U_{k,1}, \dots, U_{k,j}, \dots, U_{k,s}),$$

sont générés comme suit:

1. Choisir un ensemble de coordonnées  $\Psi \subseteq \{1, \dots, s\}$  pour lesquelles on va appliquer LHS.
2. Pour chaque  $j \in \Psi$ , générer une **permutation aléatoire**  $(\pi_{1,j}, \dots, \pi_{k,j})$  des entiers  $\{1, \dots, k\}$ .
3. Pour chaque  $(i, j)$ , générer  $U_{i,j}$  uniformément dans  $((\pi_{i,j} - 1)/k, \pi_{i,j}/k)$  si  $j \in \Psi$ , et  $U_{i,j} \sim U(0, 1)$  indépendant du reste si  $j \notin \Psi$ .

## Les $k$ points

$$\mathbf{U}_1 = (U_{1,1}, \dots, U_{1,j}, \dots, U_{1,s}), \quad \dots, \quad \mathbf{U}_k = (U_{k,1}, \dots, U_{k,j}, \dots, U_{k,s}),$$

sont générés comme suit:

1. Choisir un ensemble de coordonnées  $\Psi \subseteq \{1, \dots, s\}$  pour lesquelles on va appliquer LHS.
2. Pour chaque  $j \in \Psi$ , générer une **permutation aléatoire**  $(\pi_{1,j}, \dots, \pi_{k,j})$  des entiers  $\{1, \dots, k\}$ .
3. Pour chaque  $(i, j)$ , générer  $U_{i,j}$  uniformément dans  $((\pi_{i,j} - 1)/k, \pi_{i,j}/k)$  si  $j \in \Psi$ , et  $U_{i,j} \sim U(0, 1)$  indépendant du reste si  $j \notin \Psi$ .
4. Calculer  $X_{\text{lh}} = [f(\mathbf{U}_1) + \dots + f(\mathbf{U}_k)]/k$ .

On a  $\mathbf{U}_i \sim U[0, 1]^s$  pour tout  $i$ .



On a  $\mathbf{U}_i \sim U[0, 1]^s$  pour tout  $i$ .

Pour chaque  $j \in \Psi$ ,  $U_{1,j}, \dots, U_{k,j}$  est un échantillon stratifié de la loi  $U(0, 1)$ .

On a  $\mathbf{U}_i \sim U[0, 1]^s$  pour tout  $i$ .

Pour chaque  $j \in \Psi$ ,  $U_{1,j}, \dots, U_{k,j}$  est un échantillon stratifié de la loi  $U(0, 1)$ .

On peut répliquer la procédure  $m$  fois, indépendamment, estimer la moyenne par la moyenne globale  $\bar{X}_{\text{lh},n}$  (où  $n = mk$ ), et la variance de  $X_{\text{lh}}$  par la variance empirique des  $m$  moyennes.

On a  $\mathbf{U}_i \sim U[0, 1]^s$  pour tout  $i$ .

Pour chaque  $j \in \Psi$ ,  $U_{1,j}, \dots, U_{k,j}$  est un échantillon stratifié de la loi  $U(0, 1)$ .

On peut répliquer la procédure  $m$  fois, indépendamment, estimer la moyenne par la moyenne globale  $\bar{X}_{\text{lh},n}$  (où  $n = mk$ ), et la variance de  $X_{\text{lh}}$  par la variance empirique des  $m$  moyennes.

En général, LHS ne réduit pas toujours la variance (voir exemple dans les notes), mais la réduit dans la plupart des cas.

**Théorème.** Si  $f(U_1, \dots, U_s)$  est monotone en  $U_j$  pour tout  $j \in \Psi$ , alors  $\text{Var}[X_{\text{lh}}] \leq \text{Var}[\bar{X}_k]$ .

Nous avons vu plusieurs exemples où  $f$  est une fonction monotone de chaque  $U_j$ . Le théorème s'applique dans tous ces cas.

## Quasi-Monte Carlo Randomisé

Voir articles:

P. L'Ecuyer, "Quasi-Monte Carlo Methods with Applications in Finance," *Finance and Stochastics*, 13, 3, 2009, 307-349.

[www.springerlink.com/content/384785458w558462/fulltext.pdf](http://www.springerlink.com/content/384785458w558462/fulltext.pdf)

P. L'Ecuyer and D. Munger, "On Figures of Merit for Randomly-Shifted Lattice Rules", article invité pour le livre: *Monte Carlo and Quasi Monte Carlo Methods 2010*, H. Wozniakowski and L. Plaskota, Eds., Springer-Verlag, Berlin, 2011.

Voir <http://www.iro.umontreal.ca/~lecuyer/myftp/papers/>

## Quasi-Monte Carlo Randomisé

Voir articles:

P. L'Ecuyer, "Quasi-Monte Carlo Methods with Applications in Finance," *Finance and Stochastics*, 13, 3, 2009, 307-349.

[www.springerlink.com/content/384785458w558462/fulltext.pdf](http://www.springerlink.com/content/384785458w558462/fulltext.pdf)

P. L'Ecuyer and D. Munger, "On Figures of Merit for Randomly-Shifted Lattice Rules", article invité pour le livre: *Monte Carlo and Quasi Monte Carlo Methods 2010*, H. Wozniakowski and L. Plaskota, Eds., Springer-Verlag, Berlin, 2011.

Voir <http://www.iro.umontreal.ca/~lecuyer/myftp/papers/>

On veut estimer

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

On choisit un ensemble de points très uniforme,  $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subset [0, 1)^s$ .

# Quasi-Monte Carlo Randomisé

Voir articles:

P. L'Ecuyer, "Quasi-Monte Carlo Methods with Applications in Finance," *Finance and Stochastics*, 13, 3, 2009, 307-349.

[www.springerlink.com/content/384785458w558462/fulltext.pdf](http://www.springerlink.com/content/384785458w558462/fulltext.pdf)

P. L'Ecuyer and D. Munger, "On Figures of Merit for Randomly-Shifted Lattice Rules", article invité pour le livre: *Monte Carlo and Quasi Monte Carlo Methods 2010*, H. Wozniakowski and L. Plaskota, Eds., Springer-Verlag, Berlin, 2011.

Voir <http://www.iro.umontreal.ca/~lecuyer/myftp/papers/>

On veut estimer

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

On choisit un ensemble de points très uniforme,  $P_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subset [0, 1)^s$ .

On randomise  $P_n$  en  $\tilde{P}_n = \{\mathbf{U}_1, \dots, \mathbf{U}_n\} \subset [0, 1)^s$  tel que

- (1) on préserve l'uniformité de  $P_n$  comme ensemble;
- (2) chaque point  $\mathbf{U}_i$  de  $\tilde{P}_n$  suit la loi uniforme sur  $[0, 1)^s$ .

On estime  $\mu$  par la moyenne

$$X_{\text{rqmc}} = \frac{1}{n} \sum_{i=1}^n f(\mathbf{U}_i).$$

On estime  $\mu$  par la moyenne

$$X_{\text{rqmc}} = \frac{1}{n} \sum_{i=1}^n f(\mathbf{U}_i).$$

Pour pouvoir estimer la variance, on répète cela  $m$  fois, indépendamment, pour obtenir  $X_{\text{rqmc},1}, \dots, X_{\text{rqmc},m}$ .

On calcule la moyenne  $\bar{X}_{\text{rqmc},mn}$  et la variance  $S_{\text{rqmc},m}^2$  de ces  $m$  copies.

**Proposition.**

$$\mathbb{E}[\bar{X}_{\text{rqmc},mn}] = \mu$$

et

$$\mathbb{E}[S_{\text{rqmc},m}^2] = \text{Var}[X_{\text{rqmc}}] = m \text{Var}[\bar{X}_{\text{rqmc},mn}].$$

Permet de calculer un intervalle de confiance pour  $\mu$ .

Choix de  $m$ ? Habituellement de 5 à 25.



On estime  $\mu$  par la moyenne

$$X_{\text{rqmc}} = \frac{1}{n} \sum_{i=1}^n f(\mathbf{U}_i).$$

Pour pouvoir estimer la variance, on répète cela  $m$  fois, indépendamment, pour obtenir  $X_{\text{rqmc},1}, \dots, X_{\text{rqmc},m}$ .

On calcule la moyenne  $\bar{X}_{\text{rqmc},mn}$  et la variance  $S_{\text{rqmc},m}^2$  de ces  $m$  copies.

**Proposition.**

$$\mathbb{E}[\bar{X}_{\text{rqmc},mn}] = \mu$$

et

$$\mathbb{E}[S_{\text{rqmc},m}^2] = \text{Var}[X_{\text{rqmc}}] = m \text{Var}[\bar{X}_{\text{rqmc},mn}].$$

Permet de calculer un intervalle de confiance pour  $\mu$ .

Choix de  $m$ ? Habituellement de 5 à 25.

RQMC est un cas particulier de GAV.

La variance est réduite p.r. à MC ssi  $\sum_{i \neq j} \text{Cov}[f(\mathbf{U}_i), f(\mathbf{U}_j)] < 0$ .

## Constructions digitales

En une dimension ( $s = 1$ ), pour  $n$  fixé, on peut prendre simplement  $P_n = \{0, 1/n, \dots, (n-1)/n\}$ .

# Constructions digitales

En une dimension ( $s = 1$ ), pour  $n$  fixé, on peut prendre simplement  $P_n = \{0, 1/n, \dots, (n-1)/n\}$ .

$3/8, 7/8, 1/16, 9/16, \dots,$



## Constructions digitales

En une dimension ( $s = 1$ ), pour  $n$  fixé, on peut prendre simplement  $P_n = \{0, 1/n, \dots, (n-1)/n\}$ .

Si  $n$  n'est pas fixé, il faut une suite infinie.

Par exemple: 0,  $3/8$ ,  $7/8$ ,  $1/16$ ,  $9/16$ ,  $\dots$ ,

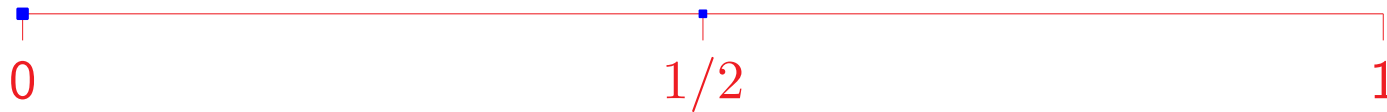


# Constructions digitales

En une dimension ( $s = 1$ ), pour  $n$  fixé, on peut prendre simplement  $P_n = \{0, 1/n, \dots, (n-1)/n\}$ .

Si  $n$  n'est pas fixé, il faut une suite infinie.

Par exemple:  $0, 1/2,$   $3/8, 7/8,$   $1/16, 9/16, \dots,$

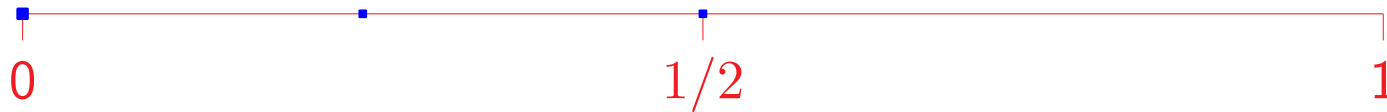


# Constructions digitales

En une dimension ( $s = 1$ ), pour  $n$  fixé, on peut prendre simplement  $P_n = \{0, 1/n, \dots, (n-1)/n\}$ .

Si  $n$  n'est pas fixé, il faut une suite infinie.

Par exemple:  $0, 1/2, 1/4,$   $3/8, 7/8, 1/16, 9/16, \dots,$

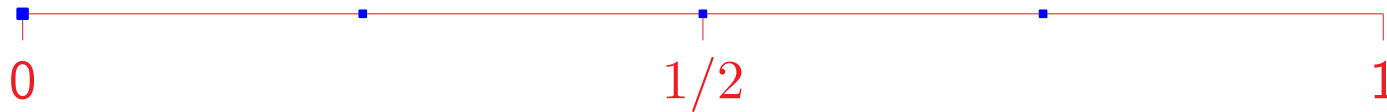


# Constructions digitales

En une dimension ( $s = 1$ ), pour  $n$  fixé, on peut prendre simplement  $P_n = \{0, 1/n, \dots, (n-1)/n\}$ .

Si  $n$  n'est pas fixé, il faut une suite infinie.

Par exemple:  $0, 1/2, 1/4, 3/4, \quad 3/8, 7/8, 1/16, 9/16, \dots$ ,

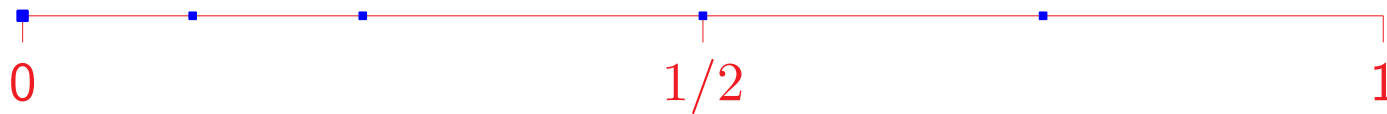


## Constructions digitales

En une dimension ( $s = 1$ ), pour  $n$  fixé, on peut prendre simplement  $P_n = \{0, 1/n, \dots, (n-1)/n\}$ .

Si  $n$  n'est pas fixé, il faut une suite infinie.

Par exemple:  $0, 1/2, 1/4, 3/4, 1/8, 3/8, 7/8, 1/16, 9/16, \dots$ ,



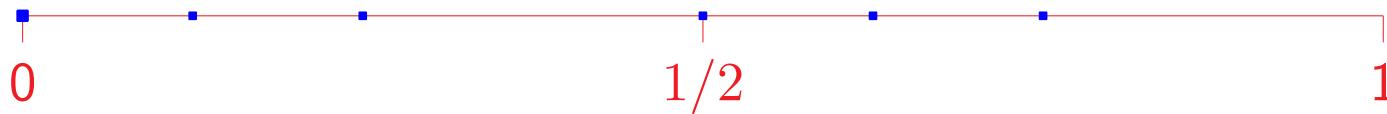


## Constructions digitales

En une dimension ( $s = 1$ ), pour  $n$  fixé, on peut prendre simplement  $P_n = \{0, 1/n, \dots, (n-1)/n\}$ .

Si  $n$  n'est pas fixé, il faut une suite infinie.

Par exemple:  $0, 1/2, 1/4, 3/4, 1/8, 5/8, 3/8, 7/8, 1/16, 9/16, \dots$ ,

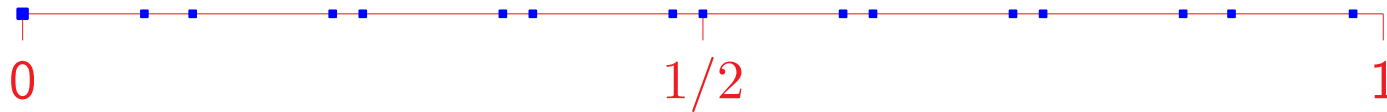


# Constructions digitales

En une dimension ( $s = 1$ ), pour  $n$  fixé, on peut prendre simplement  $P_n = \{0, 1/n, \dots, (n-1)/n\}$ .

Si  $n$  n'est pas fixé, il faut une suite infinie.

Par exemple:  $0, 1/2, 1/4, 3/4, 1/8, 5/8, 3/8, 7/8, 1/16, 9/16, \dots$ ,

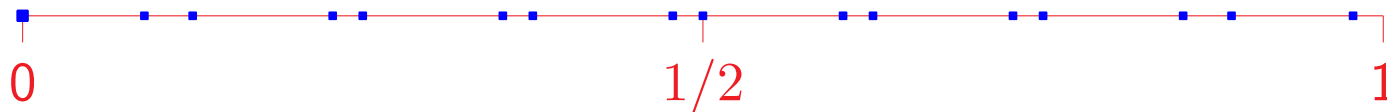


## Constructions digitales

En une dimension ( $s = 1$ ), pour  $n$  fixé, on peut prendre simplement  $P_n = \{0, 1/n, \dots, (n-1)/n\}$ .

Si  $n$  n'est pas fixé, il faut une suite infinie.

Par exemple:  $0, 1/2, 1/4, 3/4, 1/8, 5/8, 3/8, 7/8, 1/16, 9/16, \dots$ ,



C'est la suite de van der Corput en base 2.

Plus généralement, prenons une base  $b \geq 2$ . Si

$$i = a_0 + a_1b + \dots + a_{k-1}b^{k-1},$$

alors le radical inverse de  $i$  en base  $b$  est

$$\psi_b(i) = a_0b^{-1} + a_1b^{-2} + \dots + a_{k-1}b^{-k}.$$

Plus généralement, prenons une base  $b \geq 2$ . Si

$$i = a_0 + a_1b + \dots + a_{k-1}b^{k-1},$$

alors le radical inverse de  $i$  en base  $b$  est

$$\psi_b(i) = a_0b^{-1} + a_1b^{-2} + \dots + a_{k-1}b^{-k}.$$

Pour un  $b$  donné,  $\psi_b(0), \psi_b(1), \psi_b(2), \dots$  est la suite de van der Corput en base  $b$ .

Pour  $b = 3$ :

$i$	en base 3	$\psi_3(i)$
0	0000	0.0000 = 0

Pour  $b = 3$ :

$i$	en base 3	$\psi_3(i)$
0	0000	$0.0000 = 0$
1	0001	$0.1000 = 1/3$

Pour  $b = 3$ :

$i$	en base 3	$\psi_3(i)$
0	0000	$0.0000 = 0$
1	0001	$0.1000 = 1/3$
2	0002	$0.2000 = 2/3$



Pour  $b = 3$ :

$i$	en base 3	$\psi_3(i)$
0	0000	$0.0000 = 0$
1	0001	$0.1000 = 1/3$
2	0002	$0.2000 = 2/3$
3	0010	$0.0100 = 1/9$

Pour  $b = 3$ :

$i$	en base 3	$\psi_3(i)$
0	0000	$0.0000 = 0$
1	0001	$0.1000 = 1/3$
2	0002	$0.2000 = 2/3$
3	0010	$0.0100 = 1/9$
4	0011	$0.1100 = 4/9$

Pour  $b = 3$ :

$i$	en base 3	$\psi_3(i)$
0	0000	$0.0000 = 0$
1	0001	$0.1000 = 1/3$
2	0002	$0.2000 = 2/3$
3	0010	$0.0100 = 1/9$
4	0011	$0.1100 = 4/9$
5	0012	$0.2100 = 7/9$

Pour  $b = 3$ :

$i$	en base 3	$\psi_3(i)$
0	0000	$0.0000 = 0$
1	0001	$0.1000 = 1/3$
2	0002	$0.2000 = 2/3$
3	0010	$0.0100 = 1/9$
4	0011	$0.1100 = 4/9$
5	0012	$0.2100 = 7/9$
6	0020	$0.0200 = 2/9$
7	0021	$0.1200 = 5/9$
8	0022	$0.2200 = 8/9$
9	0100	$0.0010 = 1/27$
10	0101	$0.1010 = 10/27$
$\vdots$	$\vdots$	$\vdots$

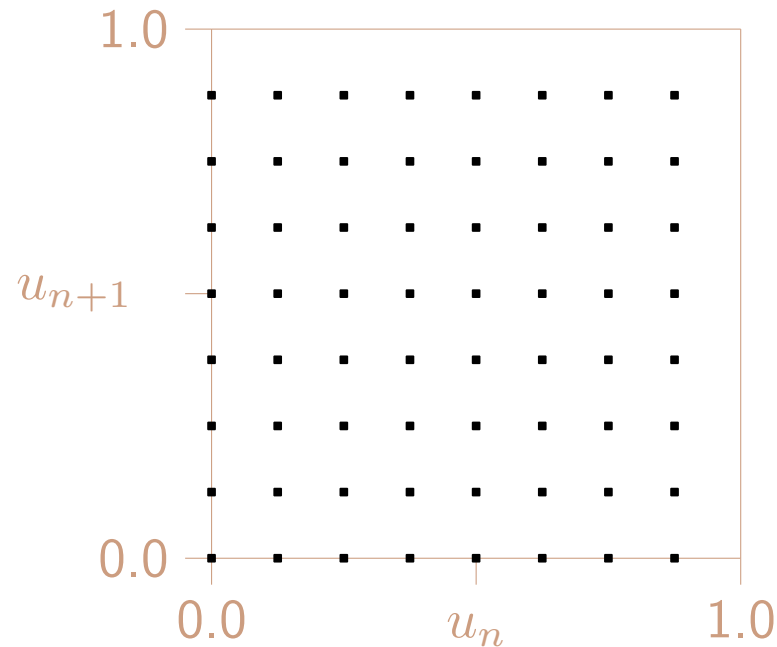
Pour  $b = 3$ :

$i$	en base 3	$\psi_3(i)$
0	0000	$0.0000 = 0$
1	0001	$0.1000 = 1/3$
2	0002	$0.2000 = 2/3$
3	0010	$0.0100 = 1/9$
4	0011	$0.1100 = 4/9$
5	0012	$0.2100 = 7/9$
6	0020	$0.0200 = 2/9$
7	0021	$0.1200 = 5/9$
8	0022	$0.2200 = 8/9$
9	0100	$0.0010 = 1/27$
10	0101	$0.1010 = 10/27$
$\vdots$	$\vdots$	$\vdots$

En général, si on s'arrête à  $n = b^k$  points, on a tous les multiples de  $1/n$ , mais ordonnés selon une **permutation** particulière.

En  $s > 1$  dimensions, une solution naive consiste à prendre le produit cartésien de  $s$  copies de  $\{0, 1/d, \dots, (d-1)/d\}$ . Cela donne  $n = d^s$  points qui forment une grille rectangulaire.

En  $s > 1$  dimensions, une solution naive consiste à prendre le produit cartésien de  $s$  copies de  $\{0, 1/d, \dots, (d-1)/d\}$ . Cela donne  $n = d^s$  points qui forment une grille rectangulaire.



**Mauvais** car (1) devient vite impracticable quand  $s$  augmente;  
 (2) les points se projettent les uns sur les autres (e.g., les projections unidimensionnelles ne contiennent que  $d$  points).

## Hammersley (1960) et Halton (1960):

Utiliser la suite de van der Corput avec une base  $b_j$  (première) différente pour chaque coordonnée  $j$ . On utilise les bases 2, 3, 5, 7, 11, ...



## Hammersley (1960) et Halton (1960):

Utiliser la suite de van der Corput avec une base  $b_j$  (première) différente pour chaque coordonnée  $j$ . On utilise les bases 2, 3, 5, 7, 11, ...

Mais  $b_j$  augmente vite avec  $j$ .

## Hammersley (1960) et Halton (1960):

Utiliser la suite de van der Corput avec une base  $b_j$  (première) différente pour chaque coordonnée  $j$ . On utilise les bases 2, 3, 5, 7, 11, ...

Mais  $b_j$  augmente vite avec  $j$ .

**Autre idée:** Prendre  $\{0, 1/n, \dots, (n-1)/n\}$  pour chaque coordonnée  $j$ , mais choisir une permutation différente pour chaque  $j$ .

## Hammersley (1960) et Halton (1960):

Utiliser la suite de van der Corput avec une base  $b_j$  (première) différente pour chaque coordonnée  $j$ . On utilise les bases 2, 3, 5, 7, 11, ...

Mais  $b_j$  augmente vite avec  $j$ .

**Autre idée:** Prendre  $\{0, 1/n, \dots, (n-1)/n\}$  pour chaque coordonnée  $j$ , mais choisir une permutation différente pour chaque  $j$ .

LHS fait cela, mais en choisissant les permutations au hasard, indépendamment les unes des autres.

## Hammersley (1960) et Halton (1960):

Utiliser la suite de van der Corput avec une base  $b_j$  (première) différente pour chaque coordonnée  $j$ . On utilise les bases 2, 3, 5, 7, 11, ...

Mais  $b_j$  augmente vite avec  $j$ .

**Autre idée:** Prendre  $\{0, 1/n, \dots, (n-1)/n\}$  pour chaque coordonnée  $j$ , mais choisir une permutation différente pour chaque  $j$ .

LHS fait cela, mais en choisissant les permutations au hasard, indépendamment les unes des autres.

**Question:** Qu'arrive-t-il si les permutations toutes les mêmes? Ou presque?

## Réseau digital

Un **réseau digital** linéaire sur  $\mathbb{Z}_b$ , avec  $n = b^k$  points: pour  $i = 0, \dots, b^k - 1$ ,

$$i = \sum_{\ell=0}^{k-1} a_{i,\ell} b^\ell,$$

$$\begin{pmatrix} u_{i,j,1} \\ u_{i,j,2} \\ \vdots \\ u_{i,j,w} \end{pmatrix} = \mathbf{C}^j \begin{pmatrix} a_{i,0} \\ a_{i,1} \\ \vdots \\ a_{i,k-1} \end{pmatrix},$$

$$u_{i,j} = \sum_{\ell=1}^w u_{i,j,\ell} b^{-\ell},$$

$$\mathbf{u}_i = (u_{i,1}, \dots, u_{i,s}),$$

où la **matrice génératrice**  $\mathbf{C}^j$  pour la coordonnée  $j$ ,  $1 \leq j \leq s$ , est  $w \times k$ , sur  $\mathbb{Z}_b$ .

## Réseau digital

Un **réseau digital** linéaire sur  $\mathbb{Z}_b$ , avec  $n = b^k$  points: pour  $i = 0, \dots, b^k - 1$ ,

$$i = \sum_{\ell=0}^{k-1} a_{i,\ell} b^\ell,$$

$$\begin{pmatrix} u_{i,j,1} \\ u_{i,j,2} \\ \vdots \\ u_{i,j,w} \end{pmatrix} = \mathbf{C}^j \begin{pmatrix} a_{i,0} \\ a_{i,1} \\ \vdots \\ a_{i,k-1} \end{pmatrix},$$

$$u_{i,j} = \sum_{\ell=1}^w u_{i,j,\ell} b^{-\ell},$$

$$\mathbf{u}_i = (u_{i,1}, \dots, u_{i,s}),$$

où la **matrice génératrice**  $\mathbf{C}^j$  pour la coordonnée  $j$ ,  $1 \leq j \leq s$ , est  $w \times k$ , sur  $\mathbb{Z}_b$ .

Pour la théorie on suppose  $w = \infty$ , mais  $w$  est fini en pratique.

La définition de Niederreiter (1992) est plus générale: en particulier, on peut aussi appliquer des permutations sur les digits de  $\mathbb{Z}_b$  avant et après la transformation linéaire.

Pour  $n$  fixé, la coord.  $j$  du point  $i$  est  $i/n$  si  $\mathbf{C}^j$  est l'identité réfléchie.

On a la suite de [van der Corput sequence](#) si  $\mathbf{C}^j$  est l'identité.

Pour  $n$  fixé, la coord.  $j$  du point  $i$  est  $i/n$  si  $\mathbf{C}^j$  est l'identité réfléchie.

On a la suite de [van der Corput sequence](#) si  $\mathbf{C}^j$  est l'identité.

[Suite digitale](#): suite infinie de points.

On peut utiliser les  $n = b^k$  premiers points pour un entier  $k$ .

Exemples: Sobol', Faure, Niederreiter, ...

Voir le package [hups](#) de SSJ.

Tout l'art est dans la construction des matrices  $\mathbf{C}^j$ .

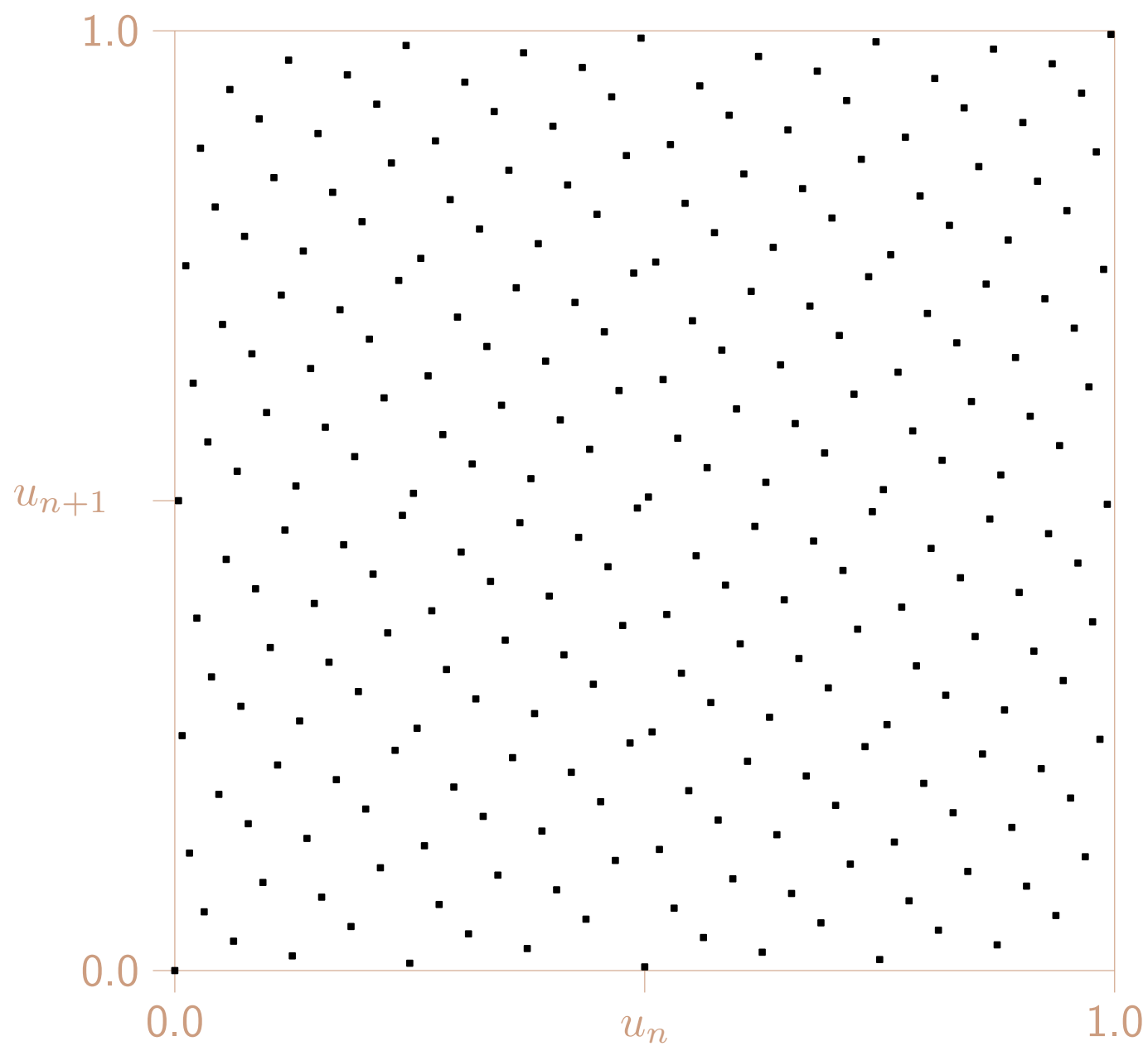


**Exemple:** Soient  $n = 2^8 = 256$  et  $s = 2$ . Prenons les points (en binaire):

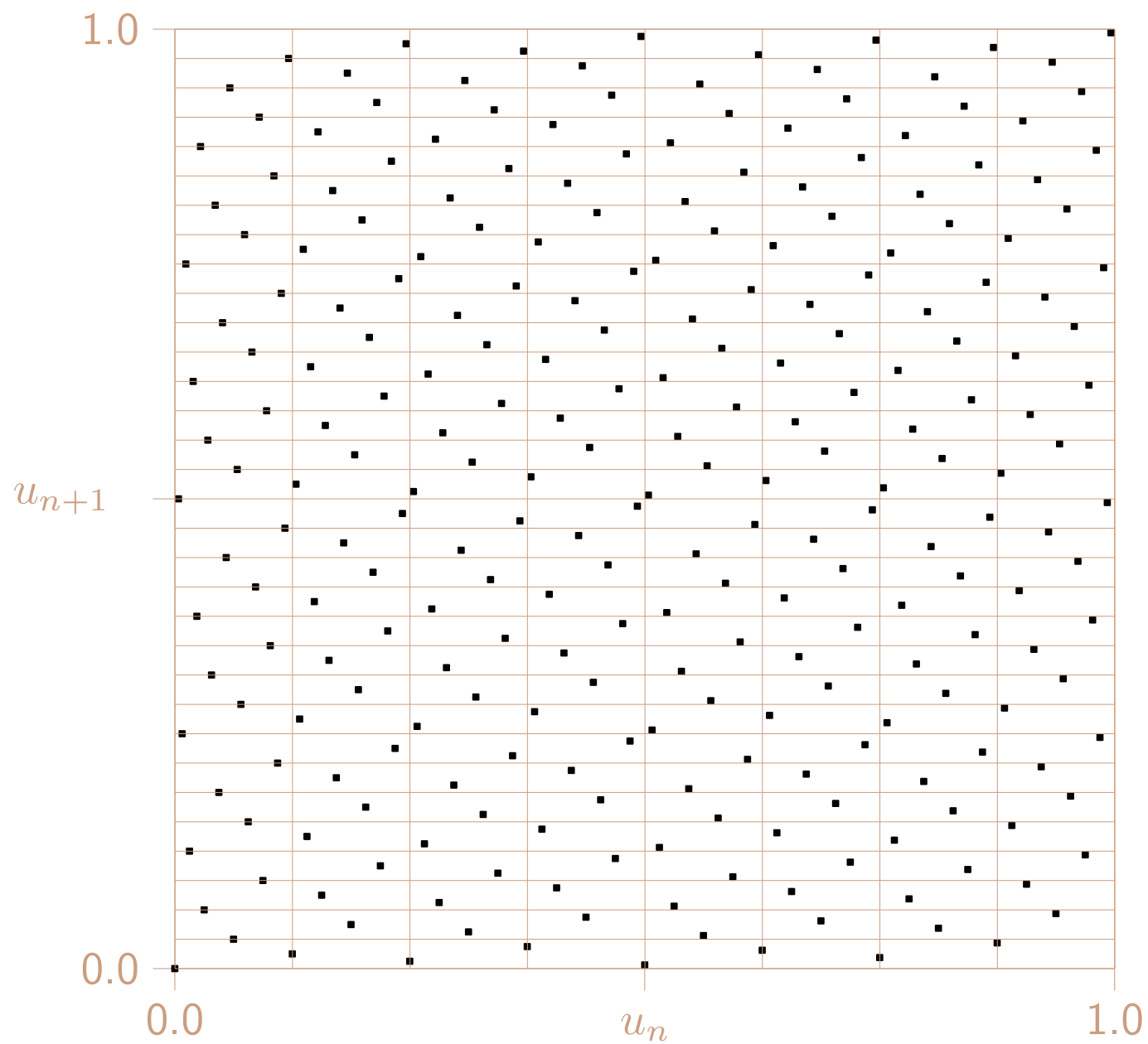
$i$	$u_{1,i}$	$u_{2,i}$
0	.00000000	.0
1	.00000001	.1
2	.00000010	.01
3	.00000011	.11
4	.00000100	.001
5	.00000101	.101
6	.00000110	.011
$\vdots$	$\vdots$	$\vdots$
254	.11111110	.01111111
255	.11111111	.11111111

Ici,  $\mathbf{C}^1$  est l'identité réfléchie et  $\mathbf{C}^2$  est l'identité.

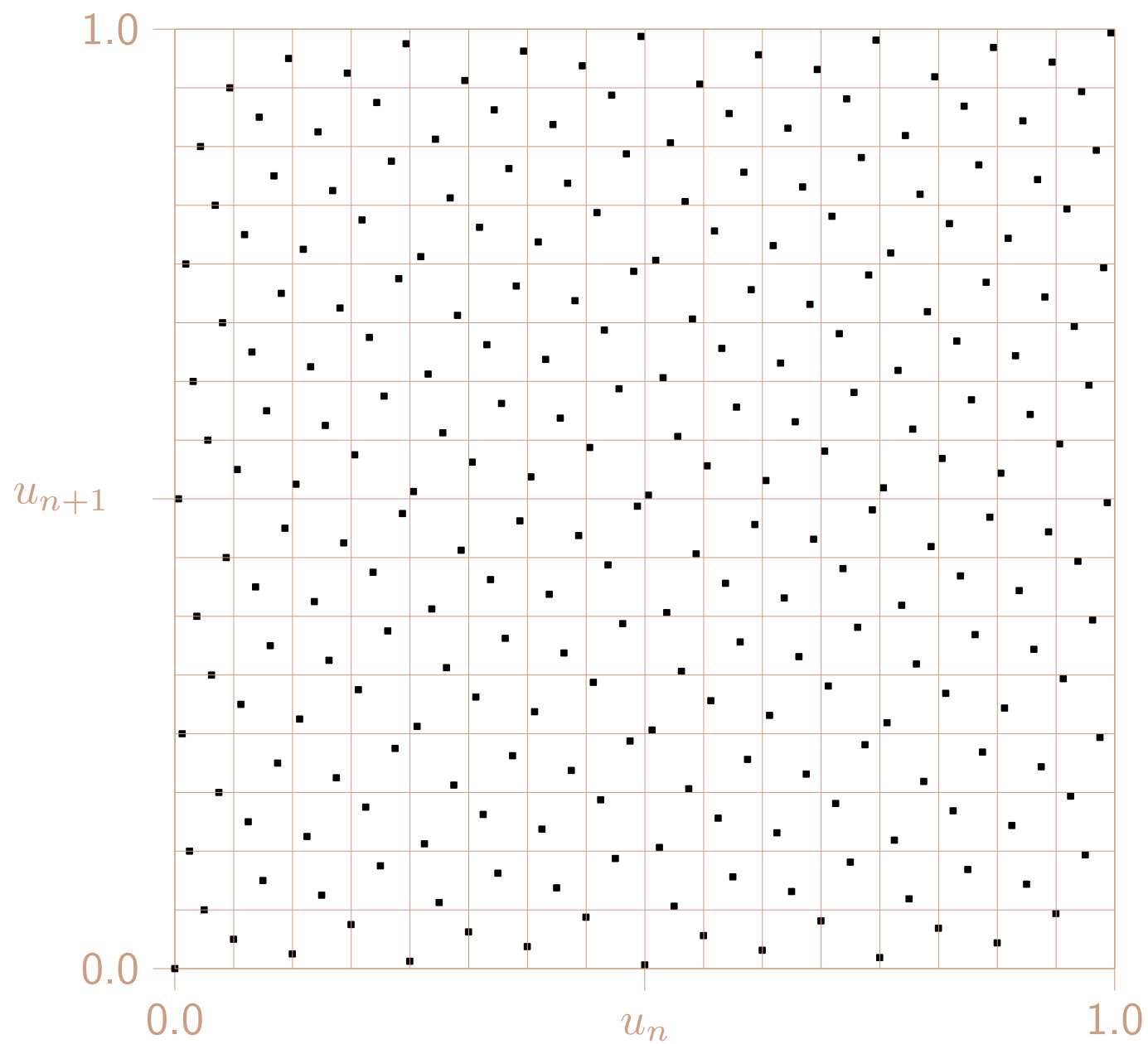
Ensemble  $P_n$  de cardinalité 256.



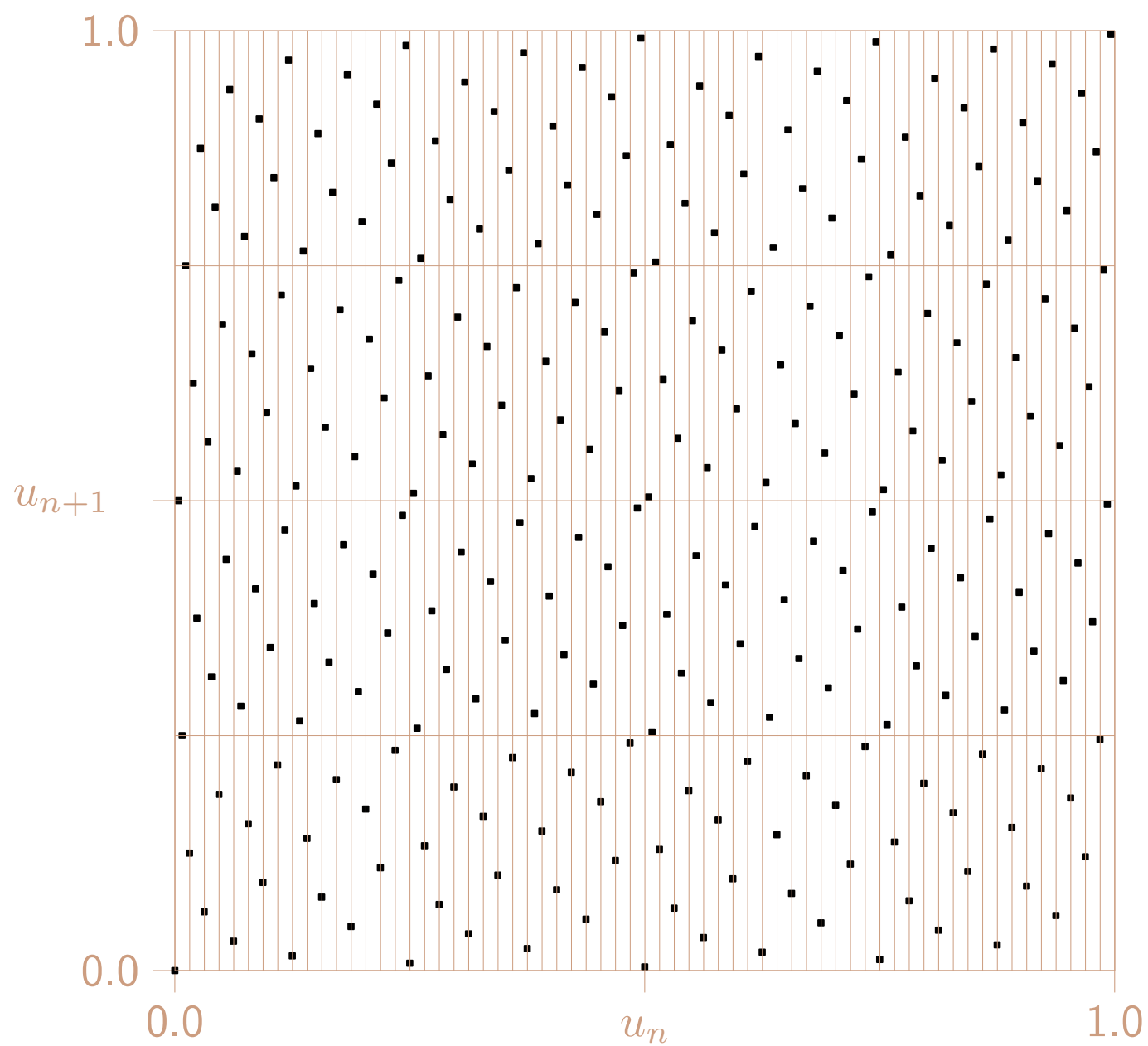
Ensemble  $P_n$  de cardinalité 256.



Ensemble  $P_n$  de cardinalité 256.



Ensemble  $P_n$  de cardinalité 256.



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.

Ces points forment un  $(0, k, 2)$ -réseau 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.

Ces points forment un  $(0, k, 2)$ -réseau en base 2.

On peut mesurer l'uniformité des réseaux digitaux exactement de la même façon que pour le  $\Psi_s$  d'un générateur de v.a. uniforme basé sur une récurrence modulo 2.

L'ensemble  $P_n$  est  $(q_1, \dots, q_s)$ -équidistribué en base  $b$  ssi la matrice formée des  $q_1$  premières lignes de  $\mathbf{C}^1$ ,  $q_2$  premières lignes de  $\mathbf{C}^2$ ,  $\dots$ ,  $q_s$  premières lignes de  $\mathbf{C}^s$ , est de plein rang (mod  $b$ ).

Pour vérifier l'équidistribution, on construit ces matrices et on calcule leur rang.

$P_n$  est un  $(q, k, s)$ -réseau ssi il est  $(q_1, \dots, q_s)$ -équidistribué dès que  $q_1 + \dots + q_s = q$ . N'est possible que si  $b \geq s - 1$ .

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.

Ces points forment un  $(0, k, 2)$ -réseau en base 2.

On peut mesurer l'uniformité des réseaux digitaux exactement de la même façon que pour le  $\Psi_s$  d'un générateur de v.a. uniforme basé sur une récurrence modulo 2.

L'ensemble  $P_n$  est  $(q_1, \dots, q_s)$ -équidistribué en base  $b$  ssi la matrice formée des  $q_1$  premières lignes de  $\mathbf{C}^1$ ,  $q_2$  premières lignes de  $\mathbf{C}^2$ ,  $\dots$ ,  $q_s$  premières lignes de  $\mathbf{C}^s$ , est de plein rang (mod  $b$ ).

Pour vérifier l'équidistribution, on construit ces matrices et on calcule leur rang.

$P_n$  est un  $(q, k, s)$ -réseau ssi il est  $(q_1, \dots, q_s)$ -équidistribué dès que  $q_1 + \dots + q_s = q$ . N'est possible que si  $b \geq s - 1$ .

Une suite infinie  $\{\mathbf{u}_0, \mathbf{u}_1, \dots\}$  dans  $[0, 1)^s$  est une  $(q, s)$ -suite en base  $b$  si pour tout  $k > 0$  et  $\nu \geq 0$ ,  $Q(k, \nu) = \{\mathbf{u}_i : i = \nu b^k, \dots, (\nu + 1)b^k - 1\}$ , est un  $(q, k, s)$ -réseau en base  $b$ .



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.

Ces points forment un  $(0, k, 2)$ -réseau en base 2.

On peut mesurer l'uniformité des réseaux digitaux exactement de la même façon que pour le  $\Psi_s$  d'un générateur de v.a. uniforme basé sur une récurrence modulo 2.

L'ensemble  $P_n$  est  $(q_1, \dots, q_s)$ -équidistribué en base  $b$  ssi la matrice formée des  $q_1$  premières lignes de  $\mathbf{C}^1$ ,  $q_2$  premières lignes de  $\mathbf{C}^2$ ,  $\dots$ ,  $q_s$  premières lignes de  $\mathbf{C}^s$ , est de plein rang (mod  $b$ ).

Pour vérifier l'équidistribution, on construit ces matrices et on calcule leur rang.

$P_n$  est un  $(q, k, s)$ -réseau ssi il est  $(q_1, \dots, q_s)$ -équidistribué dès que  $q_1 + \dots + q_s = q$ . N'est possible que si  $b \geq s - 1$ .

Une suite infinie  $\{\mathbf{u}_0, \mathbf{u}_1, \dots\}$  dans  $[0, 1)^s$  est une  $(q, s)$ -suite en base  $b$  si pour tout  $k > 0$  et  $\nu \geq 0$ ,  $Q(k, \nu) = \{\mathbf{u}_i : i = \nu b^k, \dots, (\nu + 1)b^k - 1\}$ , est un  $(q, k, s)$ -réseau en base  $b$ . N'est possible que si  $b \geq s$ .

## Suites et réseaux de Faure en base $b$

Faure (1982) a proposé les matrices

$$\mathbf{C}_j = \mathbf{P}^j \bmod b = \mathbf{P}\mathbf{C}_{j-1} \bmod b$$

avec  $\mathbf{C}_0 = \mathbf{I}$  et  $\mathbf{P} = (p_{l,c})$  triangulaire sup. où

$$p_{l,c} = \binom{c}{l} = \frac{c!}{l!(c-l)!}$$

pour  $l \leq c$ .

## Suites et réseaux de Faure en base $b$

Faure (1982) a proposé les matrices

$$\mathbf{C}_j = \mathbf{P}^j \bmod b = \mathbf{P} \mathbf{C}_{j-1} \bmod b$$

avec  $\mathbf{C}_0 = \mathbf{I}$  et  $\mathbf{P} = (p_{l,c})$  triangulaire sup. où

$$p_{l,c} = \binom{c}{l} = \frac{c!}{l!(c-l)!}$$

pour  $l \leq c$ .

Faure a prouvé que si  $b$  est premier et  $b \geq s$ , cela donne une  $(0, s)$ -suite en base  $b$ .

Ainsi, pour tout  $k > 0$  et  $\nu \geq 0$ ,  $Q(k, \nu) = \{\mathbf{u}_i : i = \nu b^k, \dots, (\nu + 1)b^k - 1\}$  (qui contient  $n = b^k$  points) est un  $(0, k, s)$ -réseau en base  $b$ .

Dans cet ensemble, chaque coordonnée  $j$  visite toutes les valeurs  $\{0, 1/n, \dots, (n-1)/n\}$  une et une seule fois.

Si on fixe  $n = b^k$ , on peut gagner une dimension:  $\mathbf{C}_j$  devient  $\mathbf{C}_{j+1}$  pour tout  $j \geq 0$  et on prend l'identité réfléchie pour  $\mathbf{C}_0$  (la première coordonnée de chaque point  $i$  est  $i/n$ ). Cet ensemble de points en  $s + 1$  dimensions est encore un  $(0, k, s)$ -réseau en base  $b$ .

## Suites et réseaux de Sobol'

Sobol' (1967) a proposé une suite digitale en base  $b = 2$  où

$$\mathbf{C}_j = \begin{pmatrix} 1 & v_{j,2,1} & \cdots & v_{j,c,1} & \cdots \\ 0 & 1 & \cdots & v_{j,c,2} & \cdots \\ \vdots & 0 & \ddots & \vdots & \\ & \vdots & & 1 & \end{pmatrix}.$$

## Suites et réseaux de Sobol'

Sobol' (1967) a proposé une suite digitale en base  $b = 2$  où

$$\mathbf{C}_j = \begin{pmatrix} 1 & v_{j,2,1} & \dots & v_{j,c,1} & \dots \\ 0 & 1 & \dots & v_{j,c,2} & \dots \\ \vdots & 0 & \ddots & \vdots & \\ & \vdots & & 1 & \end{pmatrix}.$$

La colonne  $c$  de  $\mathbf{C}_j$  est représentée par un entier impair

$$m_{j,c} = \sum_{l=1}^c v_{j,c,l} 2^{c-l} = v_{j,c,1} 2^{c-1} + \dots + v_{j,c,c-1} 2 + 1 < 2^c.$$

Les entiers  $m_{j,c}$  sont choisis comme suit.

Pour chaque  $j$ , on choisit un polynôme primitif sur  $\mathbb{F}_2$ ,

$$f_j(z) = z^{d_j} + a_{j,1}z^{d_j-1} + \cdots + a_{j,d_j},$$

et on choisit les  $d_j$  entiers  $m_{j,0}, \dots, m_{j,d_j-1}$ .

Pour chaque  $j$ , on choisit un polynôme primitif sur  $\mathbb{F}_2$ ,

$$f_j(z) = z^{d_j} + a_{j,1}z^{d_j-1} + \cdots + a_{j,d_j},$$

et on choisit les  $d_j$  entiers  $m_{j,0}, \dots, m_{j,d_j-1}$ .

Ensuite,  $m_{j,d_j}, m_{j,d_j+1}, \dots$  sont déterminés par la récurrence

$$m_{j,c} = 2a_{j,1}m_{j,c-1} \oplus \cdots \oplus 2^{d_j-1}a_{j,d_j-1}m_{j,c-d_j+1} \oplus 2^{d_j}m_{j,c-d_j} \oplus m_{j,c-d_j}$$

**Proposition.** Si les polynômes  $f_j(z)$  sont tous distincts, on obtient une  $(q, s)$ -sequence avec  $q \leq d_0 + \cdots + d_{s-1} + 1 - s$ .



Pour chaque  $j$ , on choisit un polynôme primitif sur  $\mathbb{F}_2$ ,

$$f_j(z) = z^{d_j} + a_{j,1}z^{d_j-1} + \cdots + a_{j,d_j},$$

et on choisit les  $d_j$  entiers  $m_{j,0}, \dots, m_{j,d_j-1}$ .

Ensuite,  $m_{j,d_j}, m_{j,d_j+1}, \dots$  sont déterminés par la récurrence

$$m_{j,c} = 2a_{j,1}m_{j,c-1} \oplus \cdots \oplus 2^{d_j-1}a_{j,d_j-1}m_{j,c-d_j+1} \oplus 2^{d_j}m_{j,c-d_j} \oplus m_{j,c-d_j}$$

**Proposition.** Si les polynômes  $f_j(z)$  sont tous distincts, on obtient une  $(q, s)$ -sequence avec  $q \leq d_0 + \cdots + d_{s-1} + 1 - s$ .

Sobol' suggère donc de lister tous les polynômes primitifs sur  $\mathbb{F}_2$  par ordre croissant de degré, en commençant par  $f_0(z) \equiv 1$  (ce qui donne  $\mathbf{C}_0 = \mathbf{I}$ ), et de prendre  $f_j(z)$  comme le  $(j+1)$ -ième polynôme de la liste.

Pour les premiers  $m_{j,c}$ , il y a beaucoup de choix (on pourrait optimiser un critère...). Les valeurs proposées par Sobol donnent une  $(s, \ell)$ -equidistribution pour  $\ell = 1$  et  $\ell = 2$  (les 2 premiers bits).

Pour  $n = 2^k$  fixé, on peut gagner une dimension exactement comme pour Faure.

## Randomisation.

Décalage aléatoire digital en base  $b$ .

Avantage: préserve toutes les propriétés d'équidistribution.

## Randomisation.

Décalage aléatoire digital en base  $b$ .

**Avantage:** préserve toutes les propriétés d'équidistribution.

On a  $u_{i,j} = \sum_{\ell=1}^w u_{i,j,\ell} b^{-\ell}$ .

Soit  $\mathbf{U} = (U_1, \dots, U_s) \sim U[0, 1)^s$  où  $U_j = \sum_{\ell=1}^w U_{j,\ell} b^{-\ell}$ .

On remplace chaque  $u_{i,j}$  par  $\tilde{U}_{i,j} = \sum_{\ell=1}^w [(u_{i,j,\ell} + U_{j,\ell}) \bmod b] b^{-\ell}$ .

Pour  $b = 2$ , l'opération devient un ou-exclusif bit à bit.

**Proposition.**  $\tilde{P}_n$  est  $(q_1, \dots, q_s)$ -équidistribué en base  $b$  ssi  $P_n$  l'est.

Pour  $w = \infty$ , chaque point  $\tilde{\mathbf{U}}_i$  suit la loi uniforme sur  $(0, 1)^s$ .

Autres permutations qui préservent l'équidistribution et peuvent aider à réduire davantage la variance:

**Scrambling matriciel** (Matoušek, Hickernell et Hong, Tezuka, Owen):

On multiplie chaque matrice  $\mathbf{C}^j$  à gauche par une matrice aléatoire  $w \times w$  non-singulière et triangulaire inférieure,  $\mathbf{M}_j$ , mod  $b$ . Plusieurs variantes.

On applique ensuite un décalage aléatoire digital en base  $b$ .

Autres permutations qui préservent l'équidistribution et peuvent aider à réduire davantage la variance:

**Scrambling matriciel** (Matoušek, Hickernell et Hong, Tezuka, Owen):

On multiplie chaque matrice  $\mathbf{C}^j$  à gauche par une matrice aléatoire  $w \times w$  non-singulière et triangulaire inférieure,  $\mathbf{M}_j$ , mod  $b$ . Plusieurs variantes.

On applique ensuite un décalage aléatoire digital en base  $b$ .

**Scrambling uniforme emboîté** (Owen 1995).

Très coûteux. Mais variance dans  $O(n^{-3}(\log n)^s)$  si  $f$  est assez lisse !

## Règles de réseaux ( “lattice rules” ) dans $\mathbb{R}^s$

[Korobov 1959, Sloan et Joe 1994, etc.]

Réseau d'intégration:

$$L_s = \left\{ \mathbf{v} = \sum_{j=1}^s z_j \mathbf{v}_j \text{ tels que chaque } z_j \in \mathbb{Z} \right\},$$

où  $\mathbf{v}_1, \dots, \mathbf{v}_s$  sont linéairement indépendants dans  $\mathbb{R}^s$  et  $\mathbb{Z}^s \subseteq L_s$ .

## Règles de réseaux (“lattice rules”) dans $\mathbb{R}^s$

[Korobov 1959, Sloan et Joe 1994, etc.]

Réseau d'intégration:

$$L_s = \left\{ \mathbf{v} = \sum_{j=1}^s z_j \mathbf{v}_j \text{ tels que chaque } z_j \in \mathbb{Z} \right\},$$

où  $\mathbf{v}_1, \dots, \mathbf{v}_s$  sont linéairement indépendants dans  $\mathbb{R}^s$  et  $\mathbb{Z}^s \subseteq L_s$ .

Règle de réseau: on prend  $P_n = L_s \cap [0, 1)^s$ .

Règle de rang 1:  $\mathbf{v}_1 = (a_0, a_1, \dots, a_{s-1})/n$ ,  $\mathbf{v}_2 = \mathbf{e}_2$ ,  $\dots$ ,  $\mathbf{v}_s = \mathbf{e}_s$ .

## Règles de réseaux (“lattice rules”) dans $\mathbb{R}^s$

[Korobov 1959, Sloan et Joe 1994, etc.]

Réseau d'intégration:

$$L_s = \left\{ \mathbf{v} = \sum_{j=1}^s z_j \mathbf{v}_j \text{ tels que chaque } z_j \in \mathbb{Z} \right\},$$

où  $\mathbf{v}_1, \dots, \mathbf{v}_s$  sont linéairement indépendants dans  $\mathbb{R}^s$  et  $\mathbb{Z}^s \subseteq L_s$ .

Règle de réseau: on prend  $P_n = L_s \cap [0, 1)^s$ .

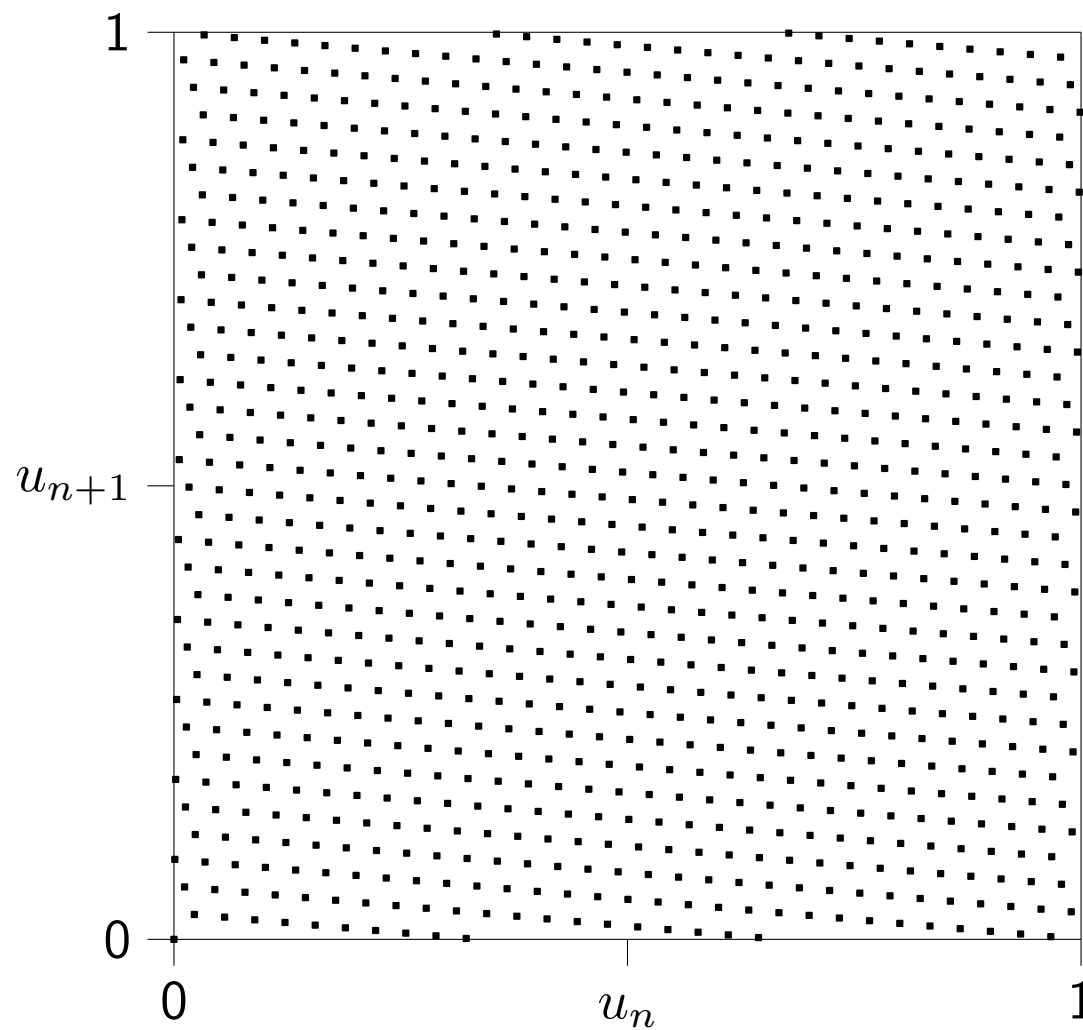
Règle de rang 1:  $\mathbf{v}_1 = (a_0, a_1, \dots, a_{s-1})/n$ ,  $\mathbf{v}_2 = \mathbf{e}_2$ ,  $\dots$ ,  $\mathbf{v}_s = \mathbf{e}_s$ .

Règle de Korobov (rang 1):  $\mathbf{v}_1 = (1, a, a^2 \bmod n, \dots, a^{s-1} \bmod n)/n$ .

Équivaut à prendre  $P_n$  égal au  $\Psi_s$  d'un LCG de module  $n$  et multiplicateur  $a$ :

$$P_n = \{(x_0, \dots, x_{s-1})/n \text{ tels que } x_0 \in \mathbb{Z}_n \text{ et } x_j = ax_{j-1} \bmod n \text{ pour tout } j > 1\}. \quad \text{Rapide!}$$





Règle de Korobov,  $s = 2$ ,  $n = 1021$ ,  $a = 90$ .

Toutes les projections unidimensionnelles sont  $\{0, 1/n, \dots, (n-1)/n\}$ .

## Example: Pricing a Asian basket option

We have  $c$  assets,  $d$  observation times. Want to estimate  $\mathbb{E}[f(\mathbf{U})]$ , where

$$f(\mathbf{U}) = e^{-rT} \max \left[ 0, \frac{1}{cd} \sum_{i=1}^c \sum_{j=1}^d S_i(t_j) - K \right]$$

is the net discounted payoff and  $S_i(t_j)$  is the value of asset  $i$  at time  $t_j$ .

Suppose  $(S_1(t), \dots, S_c(t))$  obeys a geometric Brownian motion. Then,  $f(\mathbf{U}) = g(\mathbf{Y})$  where  $\mathbf{Y} = (Y_1, \dots, Y_s) \sim N(\mathbf{0}, \Sigma)$  and  $s = cd$ .

## Example: Pricing a Asian basket option

We have  $c$  assets,  $d$  observation times. Want to estimate  $\mathbb{E}[f(\mathbf{U})]$ , where

$$f(\mathbf{U}) = e^{-rT} \max \left[ 0, \frac{1}{cd} \sum_{i=1}^c \sum_{j=1}^d S_i(t_j) - K \right]$$

is the net discounted payoff and  $S_i(t_j)$  is the value of asset  $i$  at time  $t_j$ .

Suppose  $(S_1(t), \dots, S_c(t))$  obeys a geometric Brownian motion. Then,  $f(\mathbf{U}) = g(\mathbf{Y})$  where  $\mathbf{Y} = (Y_1, \dots, Y_s) \sim N(\mathbf{0}, \Sigma)$  and  $s = cd$ .

To generate  $\mathbf{Y}$ , we can decompose  $\Sigma = \mathbf{C}\mathbf{C}^t$ , generate  $\mathbf{Z} = (Z_1, \dots, Z_s) \sim N(\mathbf{0}, \mathbf{I})$ , and return  $\mathbf{Y} = \mathbf{C}\mathbf{Z}$ .

The  $Z_j$ 's are generated by inversion:  $Z_j = \Phi^{-1}(U_j)$ .

## Example: Pricing a Asian basket option

We have  $c$  assets,  $d$  observation times. Want to estimate  $\mathbb{E}[f(\mathbf{U})]$ , where

$$f(\mathbf{U}) = e^{-rT} \max \left[ 0, \frac{1}{cd} \sum_{i=1}^c \sum_{j=1}^d S_i(t_j) - K \right]$$

is the net discounted payoff and  $S_i(t_j)$  is the value of asset  $i$  at time  $t_j$ .

Suppose  $(S_1(t), \dots, S_c(t))$  obeys a geometric Brownian motion. Then,  $f(\mathbf{U}) = g(\mathbf{Y})$  where  $\mathbf{Y} = (Y_1, \dots, Y_s) \sim N(\mathbf{0}, \Sigma)$  and  $s = cd$ .

To generate  $\mathbf{Y}$ , we can decompose  $\Sigma = \mathbf{C}\mathbf{C}^t$ , generate  $\mathbf{Z} = (Z_1, \dots, Z_s) \sim N(\mathbf{0}, \mathbf{I})$ , and return  $\mathbf{Y} = \mathbf{C}\mathbf{Z}$ .

The  $Z_j$ 's are generated by inversion:  $Z_j = \Phi^{-1}(U_j)$ .

Possible decompositions of  $\Sigma$ : Cholesky, PCA, bridge sampling, etc.

Large impact on the effective dimension!

Price of option based on the geometric average is known exactly, so we can take its payoff  $\bar{S}^{(G)}$  as a CV.

Price of option based on the geometric average is known exactly, so we can take its payoff  $\bar{S}^{(G)}$  as a CV.

We use the following point sets:

- (a) Sobol' nets with a random digital shift only (Sob-S),
- (b) Sobol' nets with a left matrix scramble followed by a random digital shift (Sob-LMS-S),
- (c) Korobov lattice rules with a random shift modulo 1 (Kor-S), and
- (d) Korobov lattice rules with a random shift modulo 1 followed by a baker transformation (Kor-S-B).

All random variates were generated by [inversion](#).

The [variance reduction factor](#) (VRF) is defined as the Monte Carlo variance (per observation) divided by  $n$  times the variance of  $Q_n$  for RQMC. The RQMC variance was estimated by making  $m = 100$  independent replications of the randomization. These VRFs are noisy, with a standard error of about 20 percent or more.

## Example 1: $c$ assets, $d = 1$

Parameters from Glasserman (2004):

$\rho_{i,j} = 0$ ,  $T = 1$ ,  $\sigma = 0.5$ ,  $r = 0.05$ ,  $S(0) = 100$ , and  $K = 100$ .

## Example 1: $c$ assets, $d = 1$

Parameters from Glasserman (2004):

$\rho_{i,j} = 0$ ,  $T = 1$ ,  $\sigma = 0.5$ ,  $r = 0.05$ ,  $S(0) = 100$ , and  $K = 100$ .

Here,  $\Sigma$  is diagonal.



## Example 1: $c$ assets, $d = 1$

Parameters from Glasserman (2004):

$\rho_{i,j} = 0$ ,  $T = 1$ ,  $\sigma = 0.5$ ,  $r = 0.05$ ,  $S(0) = 100$ , and  $K = 100$ .

Here,  $\Sigma$  is diagonal.

Variance Reduction factors for  $c = 5$  (left) and  $c = 10$  (right):

### Sobol' Nets

	$n = 2^{14}$	$n = 2^{16}$	$n = 2^{18}$
Sob+S	953, 168	2363, 162	7156, 180
Sob+LMS+S	733, 112	2265, 174	7058, 253

### Korobov Lattice Rules

	$n = 16381$ $a = 5693$	$n = 65521$ $a = 944$	$n = 262139$ $a = 21876$
Kor+S	178, 74	312, 21	416, 117
Kor+S+B	376, 77	440, 89	3434, 425

Replace  $\rho_{i,j} = 0$  by  $\rho_{i,j} = 0.4$  for  $i \neq j$ . Let  $c = 10$ .

Variance reduction factors for Cholesky (left) and PCA (right).

### Sobol' Nets

	$n = 2^{14}$		$n = 2^{16}$		$n = 2^{18}$	
Sob+S	289	882	508	3567	1033	10299
Sob+LMS+S	381	4931	491	11452	593	39831

### Korobov Lattice Rules

	$n = 16381, a = 5693$		$n = 65521, a = 944$		$n = 262139, a = 21876$	
Kor+S	106	737	30	1614	193	4218
Kor+S+B	185	6820	217	6864	684	20984

## Example 2: $c = 10, d = 25$

Let  $c = 10, d = 25, \rho_{i,j} = 0.4$  for all  $i \neq j, T = 1, r = 0.04,$   
 $\sigma_i = 0.1 + 0.4(i - 1)/9$  for all  $i, S_i(0) = 100,$  and  $K = 100$ . (Imai and Tan 2002)

This gives a 250-dimensional integration problem.

## Example 2: $c = 10, d = 25$

Let  $c = 10, d = 25, \rho_{i,j} = 0.4$  for all  $i \neq j, T = 1, r = 0.04,$   
 $\sigma_i = 0.1 + 0.4(i - 1)/9$  for all  $i, S_i(0) = 100,$  and  $K = 100$ . (Imai and Tan 2002)

This gives a 250-dimensional integration problem.

Variance reduction factors for Cholesky (left) and PCA (right):

Sobol' Nets

	$n = 2^{14}$		$n = 2^{16}$		$n = 2^{18}$	
Sob+S	10	1299	17	3184	32	6046
Sob+LMS+S	6	4232	4	9219	35	16557

Korobov Lattice Rules

	$n = 16381, a = 5693$		$n = 65521, a = 944$		$n = 262139, a = 21876$	
Kor+S	18	878	18	1504	9	2643
Kor+S+B	50	4553	46	3657	43	7553

### Example 3: An Asian Option on a Single Asset

Let  $c = 1$ ,  $S(0) = 100$ ,  $r = \ln(1.09)$ ,  $\sigma_i = 0.2$ ,

$T = 120/365$ ,  $t_j = D_1/365 + (T - D_1/365)(j - 1)/(d - 1)$  for  $j = 1, \dots, d$ .

### Example 3: An Asian Option on a Single Asset

Let  $c = 1$ ,  $S(0) = 100$ ,  $r = \ln(1.09)$ ,  $\sigma_i = 0.2$ ,

$T = 120/365$ ,  $t_j = D_1/365 + (T - D_1/365)(j - 1)/(d - 1)$  for  $j = 1, \dots, d$ .

Kemna and Vorst (1990) suggested using the payoff based on the **geometric average** as a control variate (CV).

### Example 3: An Asian Option on a Single Asset

Let  $c = 1$ ,  $S(0) = 100$ ,  $r = \ln(1.09)$ ,  $\sigma_i = 0.2$ ,  
 $T = 120/365$ ,  $t_j = D_1/365 + (T - D_1/365)(j - 1)/(d - 1)$  for  $j = 1, \dots, d$ .

Kemna and Vorst (1990) suggested using the payoff based on the [geometric average](#) as a control variate (CV).

We estimated the [optimal CV coefficient](#) by pilot runs for MC and for each combination of sampling scheme and RQMC method.

### Example 3: An Asian Option on a Single Asset

Let  $c = 1$ ,  $S(0) = 100$ ,  $r = \ln(1.09)$ ,  $\sigma_i = 0.2$ ,  
 $T = 120/365$ ,  $t_j = D_1/365 + (T - D_1/365)(j - 1)/(d - 1)$  for  $j = 1, \dots, d$ .

Kemna and Vorst (1990) suggested using the payoff based on the [geometric average](#) as a control variate (CV).

We estimated the [optimal CV coefficient](#) by pilot runs for MC and for each combination of sampling scheme and RQMC method.

$d$	$D_1$	$K$	$\mu$	$\sigma^2$	VRF
10	111	90	13.008	105	$1.53 \times 10^6$
10	111	100	5.863	61	$1.07 \times 10^6$
10	12	90	11.367	46	5400
10	12	100	3.617	23	3950
120	1	90	11.207	41	5050
120	1	100	3.367	20	4100



Sequential sampling (left), bridge sampling (middle), and PCA (right).

$d$	$D_1$	$K$	$P_n$	without CV			with CV		
				SEQ	BBS	PCA	SEQ	BBS	PCA
10	111	90	Sob+DS	9572	12549	14279	63	183	4436
10	111	90	Kor+S	5943	6014	13751	18	29	291
10	111	90	Kor+S+B	88927	256355	563665	90	177	668
10	111	100	Sob+DS	5764	6638	10309	42	82	1913
10	111	100	Kor+S	2224	3682	8782	12	31	397
10	111	100	Kor+S+B	27214	29042	313724	29	61	635
10	12	90	Sob+DS	2205	9053	12175	27	67	434
10	12	90	Kor+S	442	1720	13790	13	50	71
10	12	90	Kor+S+B	1394	26883	446423	31	66	200
10	12	100	Sob+DS	368	2025	9506	21	42	274
10	12	100	Kor+S	63	909	5039	8	26	47
10	12	100	Kor+S+B	133	1317	123650	18	54	119
120	1	90	Sob+DS	325	7079	15101	3	48	483
120	1	90	Kor+S	192	2025	984	5	47	75
120	1	90	Kor+S+B	394	15575	474314	13	55	280
120	1	100	Sob+DS	39	1776	10244	3	48	217
120	1	100	Kor+S	24	672	5538	3	23	29
120	1	100	Kor+S+B	29	1101	162531	9	29	144

Example: for  $K = 90$  and  $d = 10$ , Sobol' with PCA combined with CV reduces the variance approximately by a factor of  $6.8 \times 10^9$ , without increasing the CPU time (less than 20%).

Example: for  $K = 90$  and  $d = 10$ , Sobol' with PCA combined with CV reduces the variance approximately by a factor of  $6.8 \times 10^9$ , without increasing the CPU time (less than 20%).

For  $d = 120$ , PCA sampling needs around three times the CPU time per run of SEQ sampling, in our SSJ implementation.

## Example 4: Asian Option Under a VG Process

$S(t)$  = value of a given asset at time  $t$ .

Variance gamma (VG) (e.g., Madan, Carr, and Chang 1998):

$$S(t) = S(0) \exp\{rt + B(G(t; 1, \nu), \theta, \sigma) + \omega t\},$$

where  $\omega = \ln(1 - \theta\nu - \sigma^2\nu/2)/\nu$ ,

$B$  is a Brownian process with drift and variance parameters  $\theta$  and  $\sigma$ ,

$G$  is a gamma process with mean and variance parameters 1 and  $\nu$ ,

$B$  and  $G$  are independent.

## Example 4: Asian Option Under a VG Process

$S(t)$  = value of a given asset at time  $t$ .

Variance gamma (VG) (e.g., Madan, Carr, and Chang 1998):

$$S(t) = S(0) \exp\{rt + B(G(t; 1, \nu), \theta, \sigma) + \omega t\},$$

where  $\omega = \ln(1 - \theta\nu - \sigma^2\nu/2)/\nu$ ,

$B$  is a Brownian process with drift and variance parameters  $\theta$  and  $\sigma$ ,

$G$  is a gamma process with mean and variance parameters 1 and  $\nu$ ,

$B$  and  $G$  are independent.

Asian call option has discounted payoff:

$$e^{-rt} \max \left( 0, \frac{1}{d} \sum_{j=1}^d S(t_j) - K \right).$$

This is an integration problem in  $s = 2d$  dimensions.

Sequential sampling (BGSS): Generate  
 $\tau_1 = G(t_1), B(\tau_1), \tau_2 = G(t_2), B(\tau_2), \dots$   
Can be done with either MC or QMC.

Sequential sampling (BGSS): Generate

$$\tau_1 = G(t_1), B(\tau_1), \tau_2 = G(t_2), B(\tau_2), \dots$$

Can be done with either MC or QMC.

Gamma bridge sampling (BGBS):

$$\text{Generate } \tau_d = G(t_d), B(\tau_d), \tau_{d/2} = G(t_{d/2}), B(\tau_{d/2}), \tau_{d/4} = G(t_{d/4}), B(\tau_{d/4}), \tau_{3d/4} = G(t_{3d/4}), B(\tau_{3d/4}), \dots$$

Sequential sampling (BGSS): Generate

$$\tau_1 = G(t_1), B(\tau_1), \tau_2 = G(t_2), B(\tau_2), \dots$$

Can be done with either MC or QMC.

Gamma bridge sampling (BGBS):

$$\text{Generate } \tau_d = G(t_d), B(\tau_d), \tau_{d/2} = G(t_{d/2}), B(\tau_{d/2}), \tau_{d/4} = G(t_{d/4}), B(\tau_{d/4}), \tau_{3d/4} = G(t_{3d/4}), B(\tau_{3d/4}), \dots$$

For  $t_a < t < t_b$  and  $\tau_a < \tau < \tau_b$ ,

the distribution of  $G(t)$  conditional on  $G(t_a), G(t_b)$  is known (beta) and the distribution of  $B(\tau)$  conditional on  $B(\tau_a), B(\tau_b)$  is known (normal).



Sequential sampling (**BGSS**): Generate

$$\tau_1 = G(t_1), B(\tau_1), \tau_2 = G(t_2), B(\tau_2), \dots$$

Can be done with either MC or QMC.

Gamma bridge sampling (**BGBS**):

$$\text{Generate } \tau_d = G(t_d), B(\tau_d), \tau_{d/2} = G(t_{d/2}), B(\tau_{d/2}), \tau_{d/4} = G(t_{d/4}), B(\tau_{d/4}), \tau_{3d/4} = G(t_{3d/4}), B(\tau_{3d/4}), \dots$$

For  $t_a < t < t_b$  and  $\tau_a < \tau < \tau_b$ ,

the distribution of  $G(t)$  conditional on  $G(t_a), G(t_b)$  is known (beta) and the distribution of  $B(\tau)$  conditional on  $B(\tau_a), B(\tau_b)$  is known (normal).

Difference of gammas bridge sampling (**DGBS**):

Write  $S$  as a difference of two gamma processes and use bridge sampling for each.

Let  $\theta = -0.1436$ ,  $\sigma = 0.12136$ ,  $\nu = 0.3$ ,  $r = 0.1$ ,  $T = 1$ ,  $K = 101$ , and  $S(0) = 100$ .

Let  $\theta = -0.1436$ ,  $\sigma = 0.12136$ ,  $\nu = 0.3$ ,  $r = 0.1$ ,  $T = 1$ ,  $K = 101$ , and  $S(0) = 100$ .

Variance reduction factors with BGSS (left), BGBS (middle), and DGBS (right):

Sobol' Nets

	$n = 2^{14}$		$n = 2^{16}$		$n = 2^{18}$	
Sob+S	37 359	585	41 421	1077	75 510	1154
Sob+LMS+S	29 530	557	49 565	995	77 735	1642

Korobov Lattice Rules

	$n = 16381, a = 5693$		$n = 65521, a = 944$		$n = 262139, a = 21876$	
Kor+S	17 54	119	24 138	263	22 285	557
Kor+S+B	52 53	57	44 44	433	92 93	1688