

Chapitre 1: Introduction

Lisez d'abord l'appendice, qui rappelle certaines notions de base en probabilité et statistique, disponible sur la page web du cours.

Systemes, Modeles et Simulation

Modèle: description simplifiée d'un système, dans le but de d'évaluer sa performance ou l'effet de certaines décisions.

Systemes, Modeles et Simulation

Modèle: description simplifiée d'un système, dans le but de d'évaluer sa performance ou l'effet de certaines décisions.

Simulation: faire évoluer le modèle d'un système en fournissant les entrées appropriées et observer les résultats.

Systemes, Modeles et Simulation

Modèle: description simplifiée d'un système, dans le but de d'évaluer sa performance ou l'effet de certaines décisions.

Simulation: faire évoluer le modèle d'un système en fournissant les entrées appropriées et observer les résultats.

La simulation permet d'expérimenter avec un système sans payer le véritable prix de nos erreurs.

Systemes, Modeles et Simulation

Modèle: description simplifiée d'un système, dans le but de d'évaluer sa performance ou l'effet de certaines décisions.

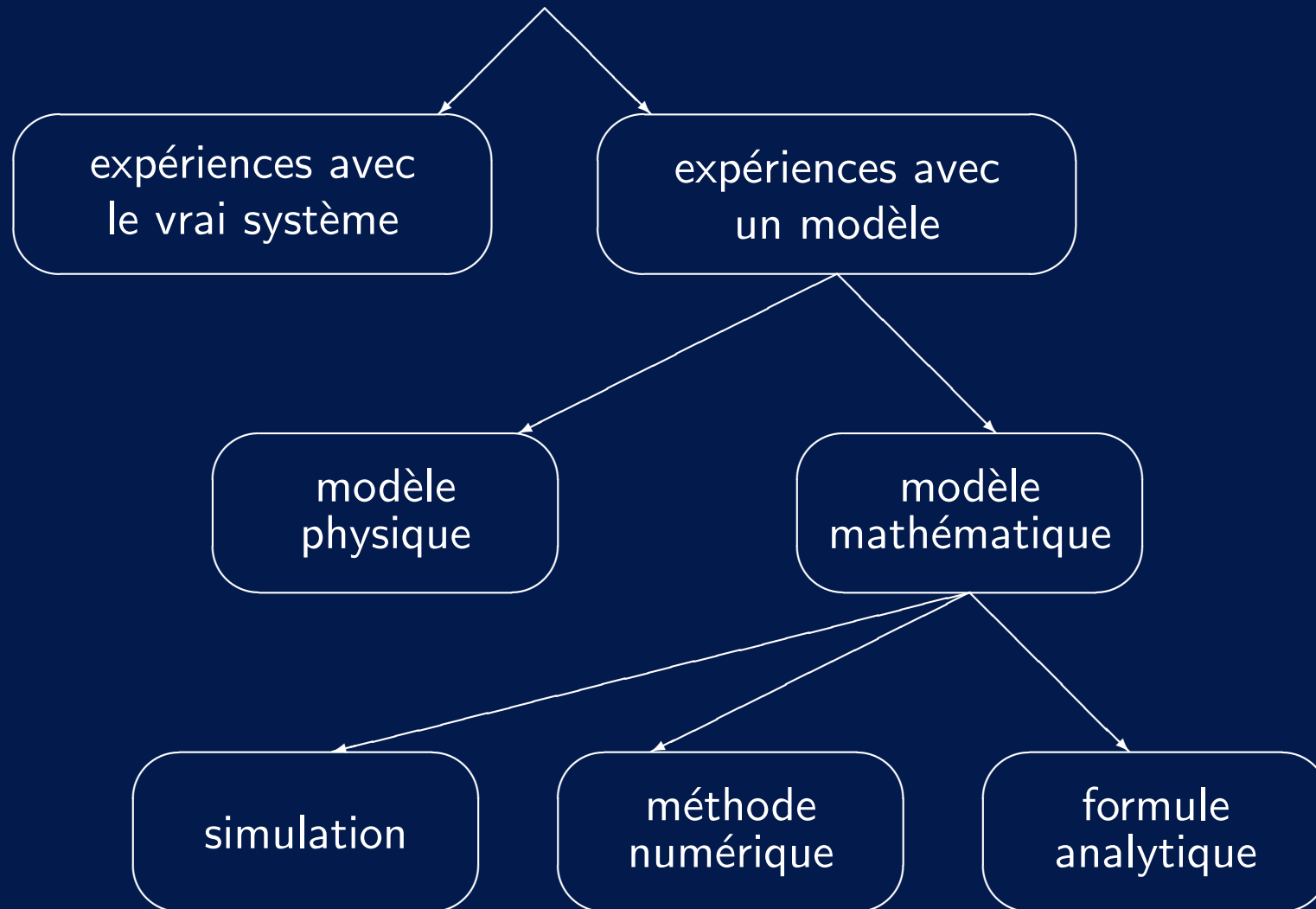
Simulation: faire évoluer le modèle d'un système en fournissant les entrées appropriées et observer les résultats.

La simulation permet d'expérimenter avec un système sans payer le véritable prix de nos erreurs.

Exemples:

- Prévisions météo; Simulateurs de vol; jeux vidéo;
- Outils pour la prise de décision: conception d'une usine; planification de la production; gestion des soins de santé; transport; télécommunications; décisions financières; etc.

Types de modèles



Avantages des modèles mathématiques et de la simulation:

- La simulation est non destructrice. Les erreurs ne sont pas coûteuses.
- Le système considéré n'a même pas besoin d'exister.
- On peut répéter à volonté des expériences identiques ou similaires dans les mêmes conditions.
- On peut souvent simuler un système beaucoup plus rapidement qu'il évolue dans la réalité (par ex., 2 années d'opération d'une usine en quelques minutes).
- On peut simuler des modèles très complexes, plus réalistes que ceux que l'on peut résoudre par des formules analytiques ou par les méthodes d'optimisation classiques.
- L'animation graphique peut permettre de voir évoluer le modèle.

Inconvénients:

- **Coût:** La modélisation et la programmation peuvent demander beaucoup d'effort, de temps et d'argent.
- Les **temps d'exécution** (CPU) peuvent devenir excessifs.
- L'**analyse statistique** des résultats n'est pas toujours simple.
- Ne fournit habituellement que des **estimations**; l'optimisation est beaucoup plus difficile par simulation que via les outils habituels de programmation mathématique.

Notre objectif dans ce cours: apprendre à **simuler efficacement** et à **analyser correctement** les résultats (du point de vue statistique).

Inconvénients:

- **Coût:** La modélisation et la programmation peuvent demander beaucoup d'effort, de temps et d'argent.
- Les **temps d'exécution** (CPU) peuvent devenir excessifs.
- L'**analyse statistique** des résultats n'est pas toujours simple.
- Ne fournit habituellement que des **estimations**; l'optimisation est beaucoup plus difficile par simulation que via les outils habituels de programmation mathématique.

Notre objectif dans ce cours: apprendre à **simuler efficacement** et à **analyser correctement** les résultats (du point de vue statistique).

Les méthodes Monte Carlo sont aussi utilisées pour des applications qui ne correspondent pas nécessairement à la simulation du comportement d'un système

complexe. Exemples: en combinatoire, pour compter des configurations; en statistique, pour estimer le niveau ou la p -valeur d'un test d'hypothèse; etc.

Types de modèles mathématiques

Modèle analytique: on obtient une formule.

Difficile à mettre au point mais facile à utiliser.

Exemples: Black-Scholes, files d'attente $M/M/s$, etc.

Types de modèles mathématiques

Modèle analytique: on obtient une formule.

Difficile à mettre au point mais facile à utiliser.

Exemples: Black-Scholes, files d'attente $M/M/s$, etc.

Méthodes numériques: itératives et/ou approximatives. E.g., programmation linéaire et non linéaire, programmation dynamique, systèmes linéaires pour grosses chaînes de Markov, méthodes éléments finis pour équations différentielles.

Types de modèles mathématiques

Modèle analytique: on obtient une formule.

Difficile à mettre au point mais facile à utiliser.

Exemples: Black-Scholes, files d'attente $M/M/s$, etc.

Méthodes numériques: itératives et/ou approximatives. E.g., programmation linéaire et non linéaire, programmation dynamique, systèmes linéaires pour grosses chaînes de Markov, méthodes éléments finis pour équations différentielles.

Simulation stochastique: implique de l'incertitude.

Outils de base: probabilités et statistique.

Types de modèles mathématiques

Modèle analytique: on obtient une formule.

Difficile à mettre au point mais facile à utiliser.

Exemples: Black-Scholes, files d'attente $M/M/s$, etc.

Méthodes numériques: itératives et/ou approximatives. E.g., programmation linéaire et non linéaire, programmation dynamique, systèmes linéaires pour grosses chaînes de Markov, méthodes éléments finis pour équations différentielles.

Simulation stochastique: implique de l'incertitude.

Outils de base: probabilités et statistique.

Un modèle de simulation sera habituellement plus détaillé et réaliste qu'un modèle analytique. Par contre, une formule analytique peut donner une meilleure intuition des principales propriétés du système et sera moins coûteuse à utiliser.

Modèle déterministe: Aucun aléa.

Ex.: Équations différentielles; programmation linéaire; . . .

Modèle Stochastique: Comportement aléatoire.

Ex.: Files d'attente; modèles de fiabilité; marchés financiers;. . .

Modèle déterministe: Aucun aléa.

Ex.: Équations différentielles; programmation linéaire; . . .

Modèle Stochastique: Comportement aléatoire.

Ex.: Files d'attente; modèles de fiabilité; marchés financiers;. . .

Modèle statique: Le "temps" ne joue pas un rôle prépondérant.

Ex.: Fiabilité d'un réseau; . . .

Modèle dynamique: On s'intéresse à l'évolution dans le temps.

Ex.: Trafic routier ou téléphonique; Système d'inventaire; . . .

Simulation des modèles dynamiques:

État du système: Les valeurs d'un ensemble de variables décrivant le système à un instant donné.

Simulation des modèles dynamiques:

État du système: Les valeurs d'un ensemble de variables décrivant le système à un instant donné.

Modèle en temps continu (simulation continue): L'état change continûment dans le temps, habituellement selon des équations différentielles.

Ex.: Un objet qui se déplace; prédictions météo; simulateurs de vol; . . .

Simulation des modèles dynamiques:

État du système: Les valeurs d'un ensemble de variables décrivant le système à un instant donné.

Modèle en temps continu (simulation continue): L'état change continûment dans le temps, habituellement selon des équations différentielles.

Ex.: Un objet qui se déplace; prédictions météo; simulateurs de vol; . . .

Modèle à événements discrets: L'état change par sauts, lors de l'occurrence d'événements, qui se produisent à des instants habituellement aléatoires

$$0 = t_0 \leq t_1 \leq t_2 \leq t_3 \cdots$$

Simulation des modèles dynamiques:

État du système: Les valeurs d'un ensemble de variables décrivant le système à un instant donné.

Modèle en temps continu (simulation continue): L'état change continûment dans le temps, habituellement selon des équations différentielles.

Ex.: Un objet qui se déplace; prédictions météo; simulateurs de vol; . . .

Modèle à événements discrets: L'état change par sauts, lors de l'occurrence d'événements, qui se produisent à des instants habituellement aléatoires

$$0 = t_0 \leq t_1 \leq t_2 \leq t_3 \dots$$

Modèle en temps discret: $t_i = i$ pour tout i .

Simulation des modèles dynamiques:

État du système: Les valeurs d'un ensemble de variables décrivant le système à un instant donné.

Modèle en temps continu (simulation continue): L'état change continûment dans le temps, habituellement selon des équations différentielles.

Ex.: Un objet qui se déplace; prédictions météo; simulateurs de vol; . . .

Modèle à événements discrets: L'état change par sauts, lors de l'occurrence d'événements, qui se produisent à des instants habituellement aléatoires

$$0 = t_0 \leq t_1 \leq t_2 \leq t_3 \dots$$

Modèle en temps discret: $t_i = i$ pour tout i .

Modèle hybride: Mélange d'évolution continue et d'événements discrets.

Simulation des modèles dynamiques:

État du système: Les valeurs d'un ensemble de variables décrivant le système à un instant donné.

Modèle en temps continu (simulation continue): L'état change continûment dans le temps, habituellement selon des équations différentielles.

Ex.: Un objet qui se déplace; prédictions météo; simulateurs de vol; . . .

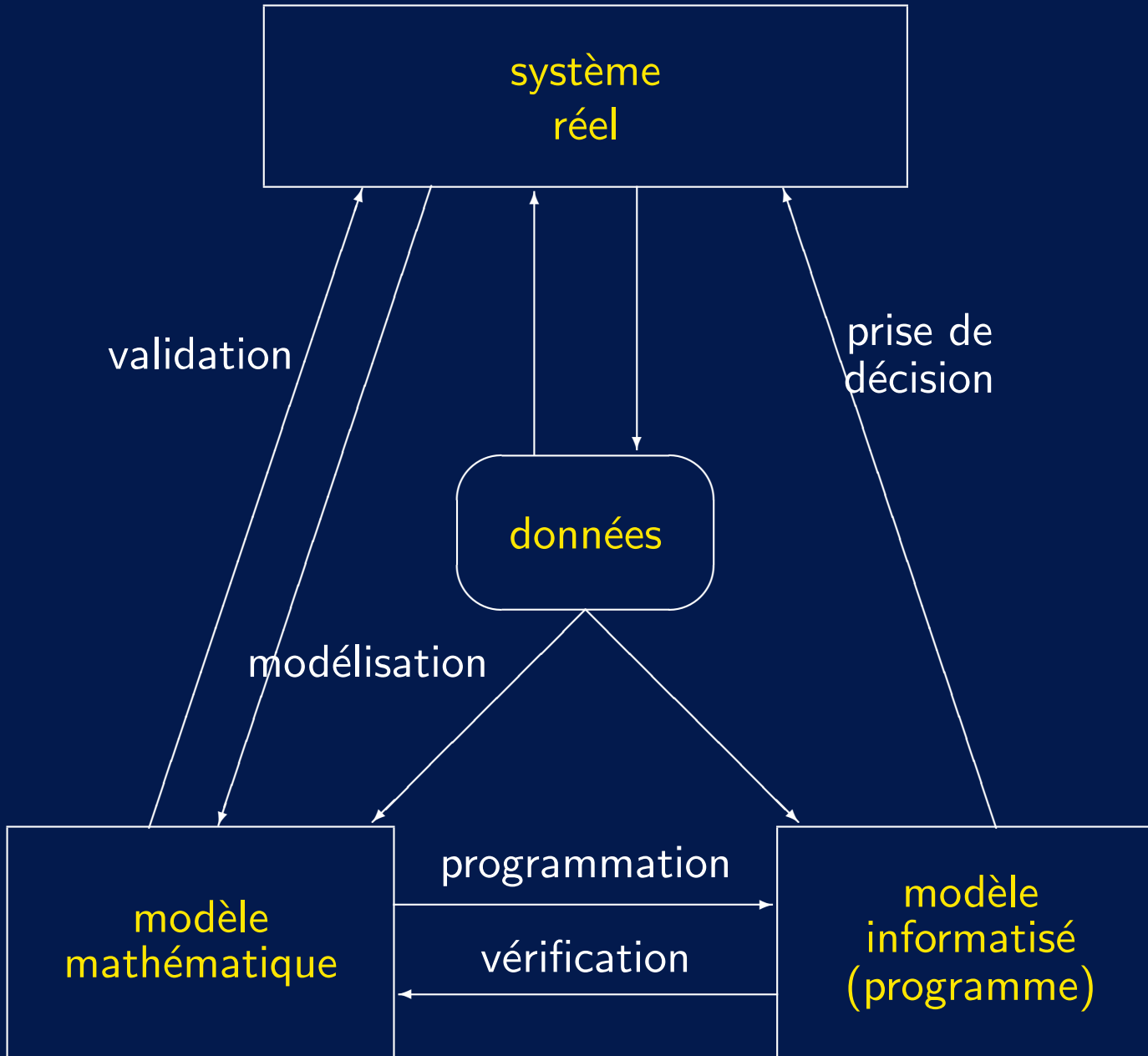
Modèle à événements discrets: L'état change par sauts, lors de l'occurrence d'événements, qui se produisent à des instants habituellement aléatoires

$$0 = t_0 \leq t_1 \leq t_2 \leq t_3 \dots$$

Modèle en temps discret: $t_i = i$ pour tout i .

Modèle hybride: Mélange d'évolution continue et d'événements discrets.

Intérêt principal dans ce cours: **la simulation stochastique à événements discrets.**



Validité d'un modèle: doit reproduire les aspects importants du comportement du système **pour l'application visée**.

La notion de validité est relative et dépend du niveau de précision voulu. Il n'y a pas de modèle "100% valide". Un modèle est toujours une **approximation** de la réalité, construit en faisant des compromis.

Validité d'un modèle: doit reproduire les aspects importants du comportement du système **pour l'application visée**.

La notion de validité est relative et dépend du niveau de précision voulu. Il n'y a pas de modèle "100% valide". Un modèle est toujours une **approximation** de la réalité, construit en faisant des compromis.

Les modèles utiles sont aussi fréquemment modifiés avec le temps. On doit les concevoir et les programmer en tenant compte de cela.

Validité d'un modèle: doit reproduire les aspects importants du comportement du système **pour l'application visée**.

La notion de validité est relative et dépend du niveau de précision voulu. Il n'y a pas de modèle "100% valide". Un modèle est toujours une **approximation** de la réalité, construit en faisant des compromis.

Les modèles utiles sont aussi fréquemment modifiés avec le temps. On doit les concevoir et les programmer en tenant compte de cela.

Exemple: modélisation des activités d'un aéroport.

Validité d'un modèle: doit reproduire les aspects importants du comportement du système **pour l'application visée**.

La notion de validité est relative et dépend du niveau de précision voulu. Il n'y a pas de modèle "100% valide". Un modèle est toujours une **approximation** de la réalité, construit en faisant des compromis.

Les modèles utiles sont aussi fréquemment modifiés avec le temps. On doit les concevoir et les programmer en tenant compte de cela.

Exemple: modélisation des activités d'un aéroport.

On ne peut pas tout modéliser. Qu'est-ce qu'on veut étudier?

Validité d'un modèle: doit reproduire les aspects importants du comportement du système **pour l'application visée**.

La notion de validité est relative et dépend du niveau de précision voulu. Il n'y a pas de modèle "100% valide". Un modèle est toujours une **approximation** de la réalité, construit en faisant des compromis.

Les modèles utiles sont aussi fréquemment modifiés avec le temps. On doit les concevoir et les programmer en tenant compte de cela.

Exemple: modélisation des activités d'un aéroport.

On ne peut pas tout modéliser. Qu'est-ce qu'on veut étudier?

L'attente aux comptoirs d'enregistrement?

L'achalandage aux guichets des douanes?

Les retards des arrivées et départs?

Exemples de modèles simples

Collisions dans un système de hachage.

Très grand ensemble d'identificateurs Φ , pas tous utilisés. Exemple: 48 bits.
Fonction de hachage $h : \Phi \rightarrow \{1, \dots, k\}$ (ensemble d'adresses), où $k \ll |\Phi|$.

Exemples de modèles simples

Collisions dans un système de hachage.

Très grand ensemble d'identificateurs Φ , pas tous utilisés. Exemple: 48 bits.

Fonction de hachage $h : \Phi \rightarrow \{1, \dots, k\}$ (ensemble d'adresses), où $k \ll |\Phi|$.

On tire m éléments distincts de Φ .

C = nombre de collisions dans les adresses (une variable aléatoire).

Exemples de modèles simples

Collisions dans un système de hachage.

Très grand ensemble d'identificateurs Φ , pas tous utilisés. Exemple: 48 bits.

Fonction de hachage $h : \Phi \rightarrow \{1, \dots, k\}$ (ensemble d'adresses), où $k \ll |\Phi|$.

On tire m éléments distincts de Φ .

C = nombre de collisions dans les adresses (une variable aléatoire).

Modèle analytique simplifié:

Si k est grand et $\lambda = m^2/(2k)$ est petit, alors $C \approx$ Poisson de moyenne λ :

$\mathbb{P}[C = x] = e^{-\lambda} \lambda^x / x!$ pour $x = 0, 1, 2, \dots$

Exemples de modèles simples

Collisions dans un système de hachage.

Très grand ensemble d'identificateurs Φ , pas tous utilisés. Exemple: 48 bits.

Fonction de hachage $h : \Phi \rightarrow \{1, \dots, k\}$ (ensemble d'adresses), où $k \ll |\Phi|$.

On tire m éléments distincts de Φ .

C = nombre de collisions dans les adresses (une variable aléatoire).

Modèle analytique simplifié:

Si k est grand et $\lambda = m^2/(2k)$ est petit, alors $C \approx$ Poisson de moyenne λ :

$\mathbb{P}[C = x] = e^{-\lambda} \lambda^x / x!$ pour $x = 0, 1, 2, \dots$

Simulation:

Si k est petit, ou si on veut évaluer la qualité de l'approx. par la loi de Poisson, on peut simuler le vrai modèle.

Exemples de modèles simples

Collisions dans un système de hachage.

Très grand ensemble d'identificateurs Φ , pas tous utilisés. Exemple: 48 bits.

Fonction de hachage $h : \Phi \rightarrow \{1, \dots, k\}$ (ensemble d'adresses), où $k \ll |\Phi|$.

On tire m éléments distincts de Φ .

C = nombre de collisions dans les adresses (une variable aléatoire).

Modèle analytique simplifié:

Si k est grand et $\lambda = m^2/(2k)$ est petit, alors $C \approx$ Poisson de moyenne λ :

$\mathbb{P}[C = x] = e^{-\lambda} \lambda^x / x!$ pour $x = 0, 1, 2, \dots$

Simulation:

Si k est petit, ou si on veut évaluer la qualité de l'approx. par la loi de Poisson, on peut simuler le vrai modèle. On génère m v.a. i.i.d. uniformes sur $\{1, \dots, k\}$ et on compte le nombre C de collisions. On répète cela n fois et on peut examiner la répartition empirique des n valeurs de C , estimer $\mathbb{E}[C]$ par leur moyenne, etc.

Un réseau d'activités stochastique.

Graphe acyclique $(\mathcal{N}, \mathcal{A})$, de noeuds \mathcal{N} et arcs \mathcal{A} (les activités).

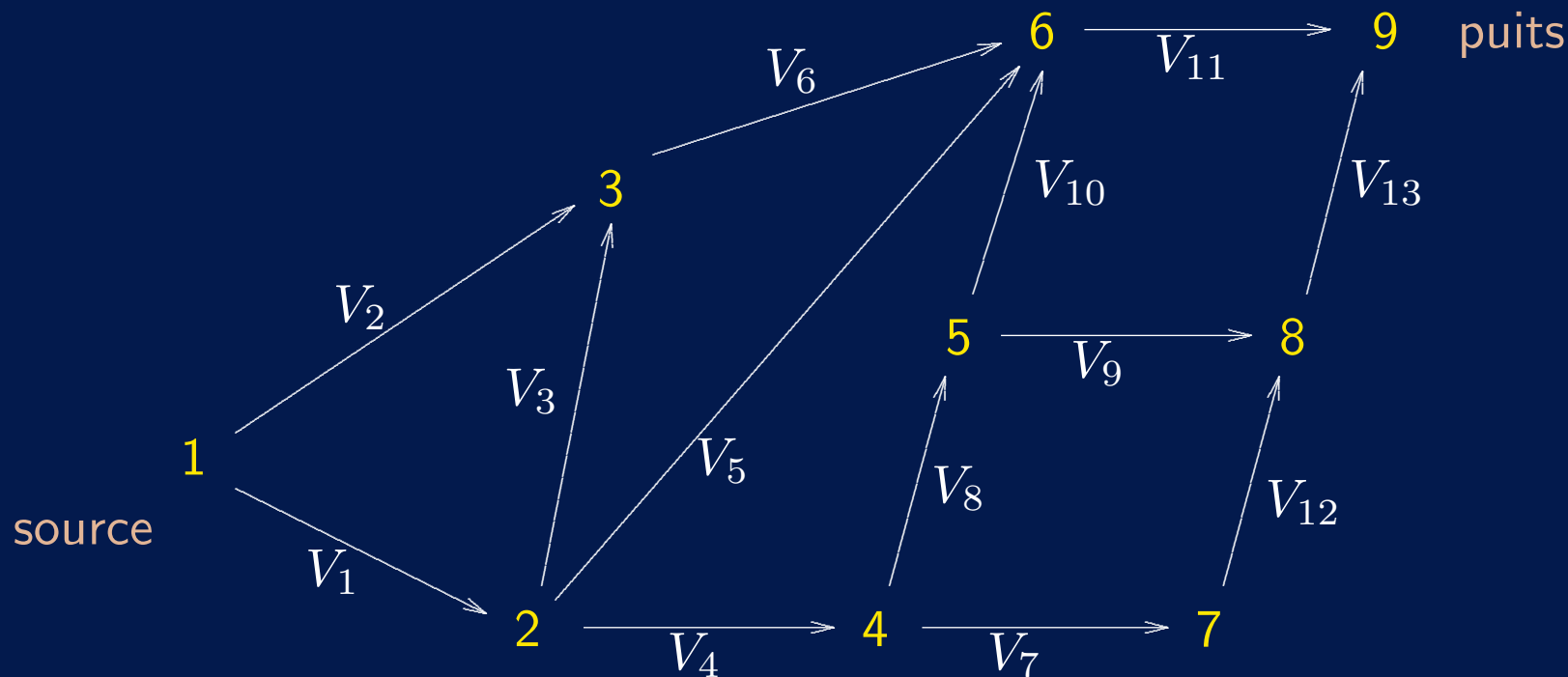
Le graphe donne les relations de précédence entre les activités.

Chaque activité $k \in \mathcal{A}$ a une durée aléatoire V_k de fonction de répartition F_k .

Si V_k est la longueur de l'arc k , la durée minimale du projet est la longueur T du plus long chemin dans le réseau.

On peut vouloir estimer $\mathbb{P}[T > x]$ pour un certain x , par exemple.

Simulation: On répète n fois: générer les V_k et calculer T ; puis on calcule la proportion des réalisations de T qui dépassent x .



Mouvement Brownien géométrique.

Modèle de Black et Scholes: sous la mesure de risque neutre, la valeur au marché d'une action au temps t est

$$S(t) = S(0)e^{(r-\sigma^2/2)t+\sigma B(t)}$$

où r est le taux d'intérêt sans risque, σ est la volatilité, et $B(\cdot)$ est un mouvement Brownien standard.

Mouvement Brownien géométrique.

Modèle de Black et Scholes: sous la mesure de risque neutre, la valeur au marché d'une action au temps t est

$$S(t) = S(0)e^{(r-\sigma^2/2)t+\sigma B(t)}$$

où r est le taux d'intérêt sans risque, σ est la volatilité, et $B(\cdot)$ est un mouvement Brownien standard.

Pour $t_2 > t_1 \geq 0$, $B(t_2) - B(t_1)$ est une v.a. normale de moyenne 0 et variance $t_2 - t_1$, indépendante de $B(\cdot)$ en dehors de $[t_1, t_2]$.

Mouvement Brownien géométrique.

Modèle de Black et Scholes: sous la mesure de risque neutre, la valeur au marché d'une action au temps t est

$$S(t) = S(0)e^{(r-\sigma^2/2)t+\sigma B(t)}$$

où r est le taux d'intérêt sans risque, σ est la volatilité, et $B(\cdot)$ est un mouvement Brownien standard.

Pour $t_2 > t_1 \geq 0$, $B(t_2) - B(t_1)$ est une v.a. normale de moyenne 0 et variance $t_2 - t_1$, indépendante de $B(\cdot)$ en dehors de $[t_1, t_2]$.

Supposons qu'un contrat financier nous assure un revenu $g(S(t_1), \dots, S(t_d))$ au temps $T \geq t_d$, où t_1, \dots, t_d sont des instants d'observation fixes.

La juste valeur au temps 0 d'un tel contrat, si $S(0) = s$, est

$$v(s, T) = \mathbb{E} \left[e^{-rT} g(S(t_1), \dots, S(t_d)) \mid S(0) = s \right],$$

où \mathbb{E} est l'espérance sous la mesure de risque neutre et e^{-rT} est le facteur d'actualisation.

Si g est simple, on peut disposer d'une formule analytique pour $v(s, T)$.
Par exemple, pour une option d'achat européenne (sans dividende),
 $g(S(T)) = \max[0, S(T) - K]$, où T est la date d'expiration et K est le prix d'achat au contrat.

Si g est simple, on peut disposer d'une formule analytique pour $v(s, T)$.

Par exemple, pour une option d'achat européenne (sans dividende),
 $g(S(T)) = \max[0, S(T) - K]$, où T est la date d'expiration et K est le prix d'achat au contrat.

Si $S(T) > K$, le détenteur achète l'action au prix K et revend immédiatement au prix du marché $S(T)$. Sinon il déchire le contrat.

Si g est simple, on peut disposer d'une formule analytique pour $v(s, T)$.

Par exemple, pour une option d'achat européenne (sans dividende),
 $g(S(T)) = \max[0, S(T) - K]$, où T est la date d'expiration et K est le prix d'achat au contrat.

Si $S(T) > K$, le détenteur achète l'action au prix K et revend immédiatement au prix du marché $S(T)$. Sinon il déchire le contrat.

Pour ce cas, la célèbre formule de Black-Scholes donne:

$$v(s, T) = s\Phi(-z_0 + \sigma\sqrt{T}) - Ke^{-rT}\Phi(-z_0),$$

où Φ est la f.r. normale standard et

$$z_0 = \ln(K/s) - (r - \sigma^2/2)T / (\sigma\sqrt{T}).$$

Si g est plus compliqué: pas de formule.

Exemple: option asiatique à observations discrètes (Bermudienne):

$$g(S(t_1), \dots, S(t_d)) = \max \left(0, \frac{1}{d} \sum_{j=1}^d S(t_j) - K \right).$$

Si g est plus compliqué: pas de formule.

Exemple: option asiatique à observations discrètes (Bermudienne):

$$g(S(t_1), \dots, S(t_d)) = \max \left(0, \frac{1}{d} \sum_{j=1}^d S(t_j) - K \right).$$

On peut alors estimer $v(s, T)$ par simulation:

Pour $i = 1, \dots, n$,

Poser $B(t_0) = t_0 = 0$;

Pour $j = 1, \dots, d$,

Générer $Z_j \sim N(0, 1)$;

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

et $S(t_j) = S(0) \exp \left[(r - \sigma^2/2)t_j + \sigma B(t_j) \right]$;

Calculer le revenu $X_i = g(S(t_1), \dots, S(t_d))$;

Calculer \bar{X}_n , la moyenne des X_i , et multiplier par e^{-rT} .

Si g est plus compliqué: pas de formule.

Exemple: option asiatique à observations discrètes (Bermudienne):

$$g(S(t_1), \dots, S(t_d)) = \max \left(0, \frac{1}{d} \sum_{j=1}^d S(t_j) - K \right).$$

On peut alors estimer $v(s, T)$ par simulation:

Pour $i = 1, \dots, n$,

Poser $B(t_0) = t_0 = 0$;

Pour $j = 1, \dots, d$,

Générer $Z_j \sim N(0, 1)$;

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

et $S(t_j) = S(0) \exp \left[(r - \sigma^2/2)t_j + \sigma B(t_j) \right]$;

Calculer le revenu $X_i = g(S(t_1), \dots, S(t_d))$;

Calculer \bar{X}_n , la moyenne des X_i , et multiplier par e^{-rT} .

Raffinements: volatilité stochastique, sauts pour $S(\cdot)$, $S(t)$ est un vecteur, etc.

Estimation d'un quantile.

Ce que l'on veut estimer n'est pas toujours une espérance mathématique.

Estimation d'un quantile.

Ce que l'on veut estimer n'est pas toujours une espérance mathématique.

Par exemple, soit Y le gain net obtenu d'un investissement donné durant une période fixée.

Estimation d'un quantile.

Ce que l'on veut estimer n'est pas toujours une espérance mathématique.

Par exemple, soit Y le gain net obtenu d'un investissement donné durant une période fixée. Pour p fixé (e.g., $p = 0.01$), on veut estimer la constante

$$\xi_p = \inf\{y : \mathbb{P}[Y \leq y] \geq p\} < 0.$$

Cette constante est le p -quantile de la distribution de Y .

Estimation d'un quantile.

Ce que l'on veut estimer n'est pas toujours une espérance mathématique.

Par exemple, soit Y le gain net obtenu d'un investissement donné durant une période fixée. Pour p fixé (e.g., $p = 0.01$), on veut estimer la constante

$$\xi_p = \inf\{y : \mathbb{P}[Y \leq y] \geq p\} < 0.$$

Cette constante est le p -quantile de la distribution de Y .

Dans ce contexte, $-\xi_p > 0$ s'appelle la valeur à risque au niveau p .

Interprétation: la probabilité de perdre ξ_p ou plus ne dépasse pas p .

Estimation d'un quantile.

Ce que l'on veut estimer n'est pas toujours une espérance mathématique.

Par exemple, soit Y le gain net obtenu d'un investissement donné durant une période fixée. Pour p fixé (e.g., $p = 0.01$), on veut estimer la constante

$$\xi_p = \inf\{y : \mathbb{P}[Y \leq y] \geq p\} < 0.$$

Cette constante est le p -quantile de la distribution de Y .

Dans ce contexte, $-\xi_p > 0$ s'appelle la valeur à risque au niveau p .

Interprétation: la probabilité de perdre ξ_p ou plus ne dépasse pas p .

Si la loi de Y est complexe, on ne peut pas calculer ξ_p exactement, mais on peut simuler n copies de Y , disons Y_1, \dots, Y_n , et estimer ξ_p par

$$\hat{\xi}_{p,n} = \inf \left\{ y \in \mathbb{R} : \frac{1}{n} \sum_{i=1}^n \mathbb{I}[Y_i \leq y] \geq p \right\} = Y_{(\lceil np \rceil)}.$$

Cet estimateur est biaisé pour n fini, mais fortement consistant.

Estimation d'un quantile.

Ce que l'on veut estimer n'est pas toujours une espérance mathématique.

Par exemple, soit Y le gain net obtenu d'un investissement donné durant une période fixée. Pour p fixé (e.g., $p = 0.01$), on veut estimer la constante

$$\xi_p = \inf\{y : \mathbb{P}[Y \leq y] \geq p\} < 0.$$

Cette constante est le p -quantile de la distribution de Y .

Dans ce contexte, $-\xi_p > 0$ s'appelle la valeur à risque au niveau p .

Interprétation: la probabilité de perdre ξ_p ou plus ne dépasse pas p .

Si la loi de Y est complexe, on ne peut pas calculer ξ_p exactement, mais on peut simuler n copies de Y , disons Y_1, \dots, Y_n , et estimer ξ_p par

$$\hat{\xi}_{p,n} = \inf \left\{ y \in \mathbb{R} : \frac{1}{n} \sum_{i=1}^n \mathbb{I}[Y_i \leq y] \geq p \right\} = Y_{(\lceil np \rceil)}.$$

Cet estimateur est biaisé pour n fini, mais fortement consistant.

Nous reviendrons plus tard sur l'estimation de quantiles.

Génération de V.A. Uniformes

Qu'est-ce qu'on veut?

Produire des suites de nombres qui ont l'air d'être choisis complètement au hasard:

- suites de bits: 011010100110110101001101100101000111...
- suite d'entiers de 0 à 100: 31, 83, 02, 72, 54, 26, ...
- suites d'entiers de 1 à n , 1 à $n - 1$, 1 à $n - 2$, ... (permutation aléatoire)
- suite de nombres réels entre 0 et 1.

Génération de V.A. Uniformes

Qu'est-ce qu'on veut?

Produire des suites de nombres qui ont l'air d'être choisis complètement au hasard:

- suites de bits: 011010100110110101001101100101000111...
- suite d'entiers de 0 à 100: 31, 83, 02, 72, 54, 26, ...
- suites d'entiers de 1 à n , 1 à $n - 1$, 1 à $n - 2$, ... (permutation aléatoire)
- suite de nombres réels entre 0 et 1.

Mécanismes physiques: lancer les dés, bouliers, roulettes,

bruit thermique dans les résistances de circuits électroniques,
capteurs de radiations, autres mécanismes basés sur la physique quantique,
microsecondes de l'horloge de l'ordinateur, ou d'un temps d'accès au disque, etc.

Génération de V.A. Uniformes

Qu'est-ce qu'on veut?

Produire des suites de nombres qui ont l'air d'être choisis complètement au hasard:

- suites de bits: 011010100110110101001101100101000111...
- suite d'entiers de 0 à 100: 31, 83, 02, 72, 54, 26, ...
- suites d'entiers de 1 à n , 1 à $n - 1$, 1 à $n - 2$, ... (permutation aléatoire)
- suite de nombres réels entre 0 et 1.

Mécanismes physiques: lancer les dés, bouliers, roulettes,

bruit thermique dans les résistances de circuits électroniques, capteurs de radiations, autres mécanismes basés sur la physique quantique, microsecondes de l'horloge de l'ordinateur, ou d'un temps d'accès au disque, etc.

Contient de la vraie entropie (incertitude), mais encombrant, pas facilement reproductible, pas toujours fiable, peu ou pas d'analyse mathématique possible.

Certains de ces mécanismes sont brevetés.

Plusieurs sont disponibles commercialement.

Pour améliorer les propriétés statistiques: combiner des blocs de bits (XOR).

Générateurs algorithmiques (ou pseudo-aléatoires) (GPA). Une fois les paramètres et l'état initial du GPA choisis, la suite produite est complètement déterministe.

Générateurs algorithmiques (ou pseudo-aléatoires) (GPA). Une fois les paramètres et l'état initial du GPA choisis, la suite produite est complètement déterministe.

Avantages: pas de matériel à installer, un logiciel suffit; souvent plus rapide; on peut facilement répéter la même séquence.

Générateurs algorithmiques (ou pseudo-aléatoires) (GPA). Une fois les paramètres et l'état initial du GPA choisis, la suite produite est complètement déterministe.

Avantages: pas de matériel à installer, un logiciel suffit; souvent plus rapide; on peut facilement répéter la même séquence.

Qualités requises? Dépend des applications.

1. Jeux d'ordinateurs personnels: L'apparence suffit.

Générateurs algorithmiques (ou pseudo-aléatoires) (GPA). Une fois les paramètres et l'état initial du GPA choisis, la suite produite est complètement déterministe.

Avantages: pas de matériel à installer, un logiciel suffit; souvent plus rapide; on peut facilement répéter la même séquence.

Qualités requises? Dépend des applications.

1. Jeux d'ordinateurs personnels: L'apparence suffit.

2. Simulation stochastique (Monte Carlo):

On veut que les propriétés statistiques du modèle mathématique soient bien reproduites par le simulateur.

Générateurs algorithmiques (ou pseudo-aléatoires) (GPA). Une fois les paramètres et l'état initial du GPA choisis, la suite produite est complètement déterministe.

Avantages: pas de matériel à installer, un logiciel suffit; souvent plus rapide; on peut facilement répéter la même séquence.

Qualités requises? Dépend des applications.

1. Jeux d'ordinateurs personnels: L'apparence suffit.

2. Simulation stochastique (Monte Carlo):

On veut que les propriétés statistiques du modèle mathématique soient bien reproduites par le simulateur.

3. Loteries, machines de casinos, casinos sur Internet, ...

Il ne faut pas que quiconque puisse obtenir un avantage pour inférer les prochains numéros ou encore des combinaisons plus probables. Conditions plus exigeantes que pour la simulation.

Générateurs algorithmiques (ou pseudo-aléatoires) (GPA). Une fois les paramètres et l'état initial du GPA choisis, la suite produite est complètement déterministe.

Avantages: pas de matériel à installer, un logiciel suffit; souvent plus rapide; on peut facilement répéter la même séquence.

Qualités requises? Dépend des applications.

1. Jeux d'ordinateurs personnels: L'apparence suffit.

2. Simulation stochastique (Monte Carlo):

On veut que les propriétés statistiques du modèle mathématique soient bien reproduites par le simulateur.

3. Loteries, machines de casinos, casinos sur Internet, ...

Il ne faut pas que quiconque puisse obtenir un avantage pour inférer les prochains numéros ou encore des combinaisons plus probables. Conditions plus exigeantes que pour la simulation.

4. Cryptologie: Encore plus exigeant. L'observation d'une partie de l'output ne doit nous aider d'aucune manière à deviner quoi que ce soit dans le reste.

Besoins pour la Simulation Stochastique

On utilise habituellement un GPA qui imite une suite U_0, U_1, U_2, \dots de variables aléatoires indépendantes de loi uniforme sur l'intervalle $(0, 1)$.

Besoins pour la Simulation Stochastique

On utilise habituellement un GPA qui imite une suite U_0, U_1, U_2, \dots de variables aléatoires indépendantes de loi uniforme sur l'intervalle $(0, 1)$.

Pour générer des v.a. selon d'autres lois, on applique des transformations à ces U_j .

Besoins pour la Simulation Stochastique

On utilise habituellement un GPA qui imite une suite U_0, U_1, U_2, \dots de variables aléatoires indépendantes de loi uniforme sur l'intervalle $(0, 1)$.

Pour générer des v.a. selon d'autres lois, on applique des transformations à ces U_j .

Comparaison de systèmes semblables avec valeurs aléatoires communes.

On simule un réseau de communication, ou un centre d'appels téléphoniques, ou un réseau de distribution de biens, ou une usine, ou le trafic automobile dans une ville, ou la gestion dynamique d'un portefeuille d'investissements (finance), etc.

On veut comparer deux configurations (ou politiques de gestion) semblables du système.

Une partie de la différence de performance sera due à la différence de configuration, et une autre partie sera due au bruit stochastique. On veut minimiser cette seconde partie.

Idée de base: simuler les deux configurations avec les mêmes valeurs uniformes U_j , utilisées exactement aux mêmes endroits.

On verra plus tard des résultats théoriques sur l'amélioration d'efficacité (réduction de variance) que cela apporte.

Mais l'implantation, avec synchronisation des v.a., peut être compliquée lorsque les deux configurations n'utilisent pas le même nombre de U_j (e.g., parfois on doit générer une v.a. dans un cas et pas dans l'autre).

Idée de base: simuler les deux configurations avec les mêmes valeurs uniformes U_j , utilisées exactement aux mêmes endroits.

On verra plus tard des résultats théoriques sur l'amélioration d'efficacité (réduction de variance) que cela apporte.

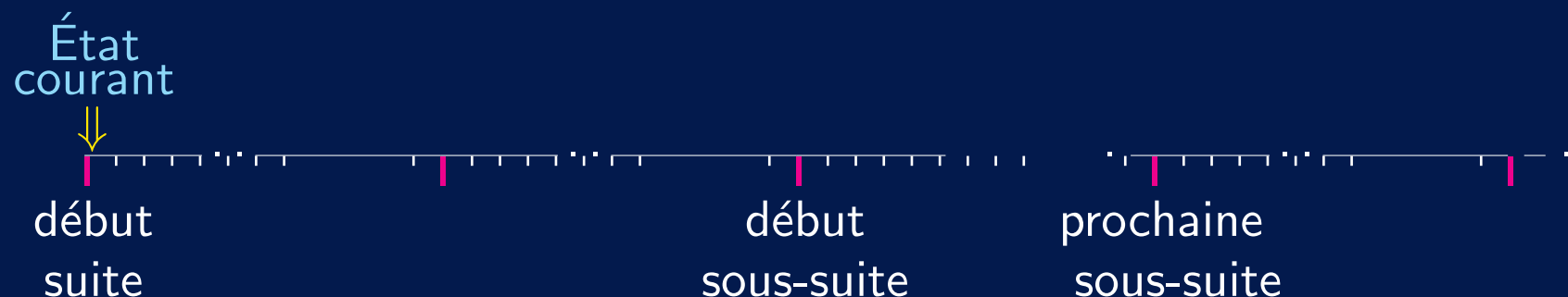
Mais l'implantation, avec synchronisation des v.a., peut être compliquée lorsque les deux configurations n'utilisent pas le même nombre de U_j (e.g., parfois on doit générer une v.a. dans un cas et pas dans l'autre).

Une solution: GPA avec suites et sous-suites multiples.

Chaque suite peut être vue comme un GPA virtuel.

Elle est partitionnée en sous-suites.

On peut créer autant de suites (distinctes et "indépendantes") que l'on veut.



Idée de base: simuler les deux configurations avec les mêmes valeurs uniformes U_j , utilisées exactement aux mêmes endroits.

On verra plus tard des résultats théoriques sur l'amélioration d'efficacité (réduction de variance) que cela apporte.

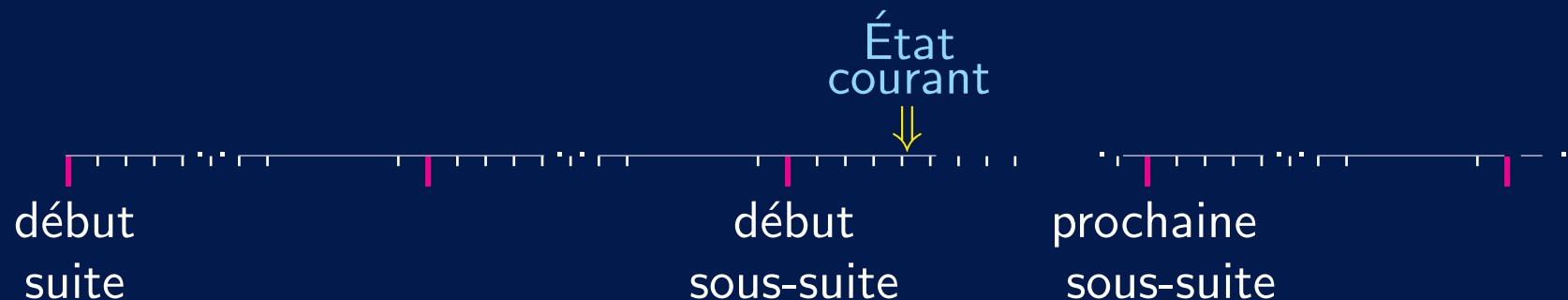
Mais l'implantation, avec synchronisation des v.a., peut être compliquée lorsque les deux configurations n'utilisent pas le même nombre de U_j (e.g., parfois on doit générer une v.a. dans un cas et pas dans l'autre).

Une solution: GPA avec suites et sous-suites multiples.

Chaque suite peut être vue comme un GPA virtuel.

Elle est partitionnée en sous-suites.

On peut créer autant de suites (distinctes et "indépendantes") que l'on veut.



Idée de base: simuler les deux configurations avec les mêmes valeurs uniformes U_j , utilisées exactement aux mêmes endroits.

On verra plus tard des résultats théoriques sur l'amélioration d'efficacité (réduction de variance) que cela apporte.

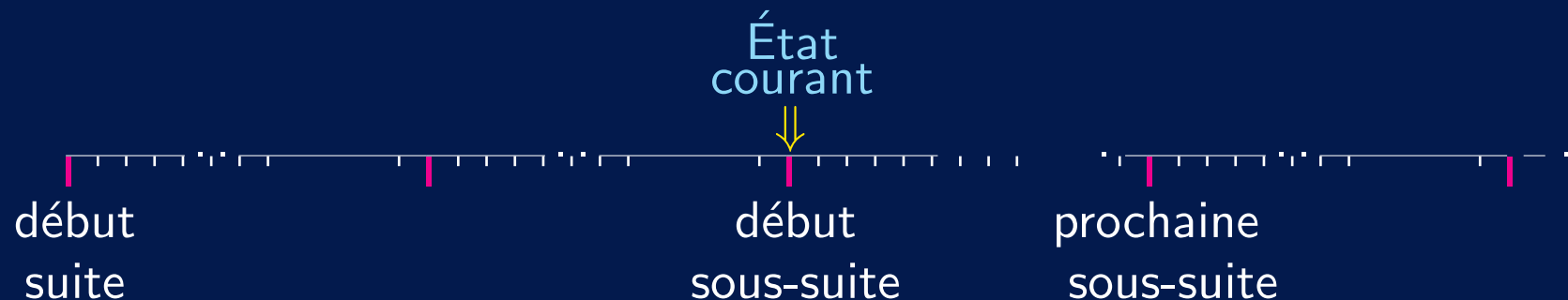
Mais l'implantation, avec synchronisation des v.a., peut être compliquée lorsque les deux configurations n'utilisent pas le même nombre de U_j (e.g., parfois on doit générer une v.a. dans un cas et pas dans l'autre).

Une solution: GPA avec suites et sous-suites multiples.

Chaque suite peut être vue comme un GPA virtuel.

Elle est partitionnée en sous-suites.

On peut créer autant de suites (distinctes et "indépendantes") que l'on veut.



Idée de base: simuler les deux configurations avec les mêmes valeurs uniformes U_j , utilisées exactement aux mêmes endroits.

On verra plus tard des résultats théoriques sur l'amélioration d'efficacité (réduction de variance) que cela apporte.

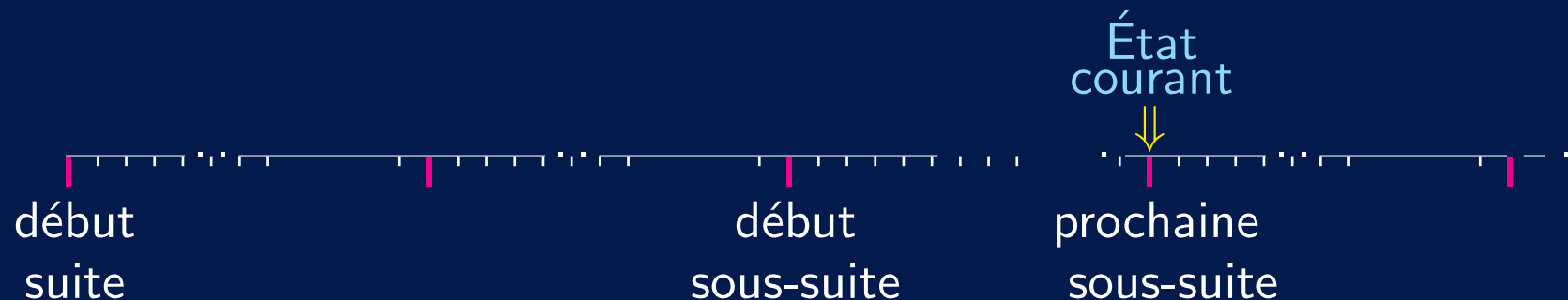
Mais l'implantation, avec synchronisation des v.a., peut être compliquée lorsque les deux configurations n'utilisent pas le même nombre de U_j (e.g., parfois on doit générer une v.a. dans un cas et pas dans l'autre).

Une solution: GPA avec suites et sous-suites multiples.

Chaque suite peut être vue comme un GPA virtuel.

Elle est partitionnée en sous-suites.

On peut créer autant de suites (distinctes et "indépendantes") que l'on veut.



Exemple d'interface Java (dans SSJ)

```
public interface RandomStream {  
    public void resetStartStream ();  
        Réinitialise la suite à son état initial.  
    public void resetStartSubstream ();  
        Réinitialise la suite au début de sa sous-suite courante. substream.  
    public void resetNextSubstream ();  
        Réinitialise la suite au début de sa prochaine sous-suite. substream.  
    public double nextDouble ();  
        Retourne une v.a.  $U(0, 1)$  de cette suite et avance d'un pas.  
    public int nextInt (int i, int j);  
        Retourne une v.a. uniforme sur  $\{i, i + 1, \dots, j\}$ .  
}
```

```
public class MRG32k3a implements RandomStream {
```

Une implantation particulière, basée sur un générateur de période $\approx 2^{181}$, partitionnée en suites disjointes de longueur 2^{127} , et sous-suites de longueur 2^{76} .

```
public MRG32k3a();
```

Construit une nouvelle suite.

```
}
```

```
public class LFSR113 implements RandomStream {
```

Une implantation basée sur une combinaison de "LFSR", de période $\approx 2^{113}$.

```
public LFSR113();
```

Construit une nouvelle suite.

```
}
```

Ce système, basé sur **MRG32k3a**, a été récemment adopté dans plusieurs logiciels de simulation et de statistique, tels que SAS, Arena, Witness, Simul8, Automod, ns2, MATLAB, ...

Génération de valeurs non-uniformes

Lois continues ou discrètes. Méthode par défaut: `inversion`.

```
public class RandomVariateGen {  
    public RandomVariateGen (RandomStream s, Distribution dist)
```

Crée un générateur pour la loi `dist`, avec la suite `s`.

```
    public double nextDouble()
```

Génère une nouvelle valeur.

Génération de valeurs non-uniformes

Lois continues ou discrètes. Méthode par défaut: `inversion`.

```
public class RandomVariateGen {  
    public RandomVariateGen (RandomStream s, Distribution dist)
```

Crée un générateur pour la loi `dist`, avec la suite `s`.

```
    public double nextDouble()
```

Génère une nouvelle valeur.

Il y a aussi des générateurs spécialisés pour plusieurs distributions.

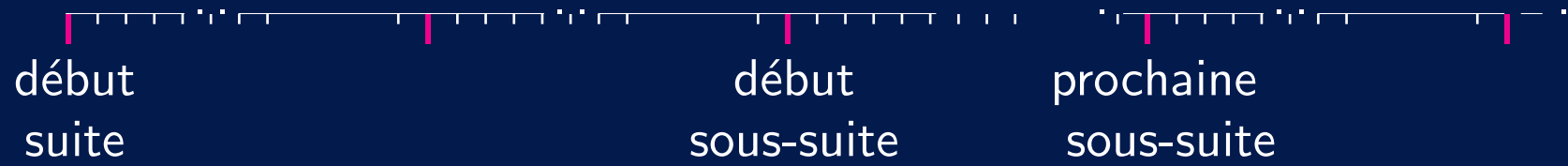
```
public class Normal extends RandomVariateGen {  
    public Normal (RandomStream s, double mu, double sigma);  
  
    public static double nextDouble (RandomStream s, double mu,  
                                     double sigma);
```

Génère une nouvelle v.a. normale, en utilisant la suite `s`.

Supposons que pour une suite de clients, on doit générer:

- (1) un instant d'arrivée pour chaque client,
- (2) une durée de service au serveur A pour chaque client
- (3) une durée de service au serveur B pour des groupes de 10 à 20 clients.

Alors on peut utiliser une suite pour chacun de ces trois usages.



Exemple de comparaison de deux configuration:

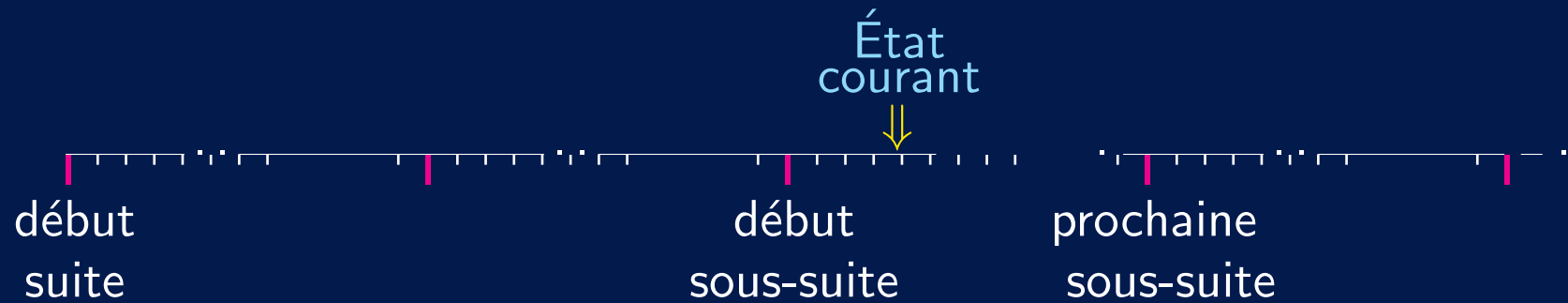
```

    ⋮
RandomStream genArr    = new RandomStream(); // Inter-arriv.
RandomStream genServA = new RandomStream(); // Serveur A.
RandomStream genServB = new RandomStream(); // Serveur B.

    ⋮
for (int rep = 0; rep < n; rep++) {
    genArr.resetNextSubstream();
    genServA.resetNextSubstream();
    genServB.resetNextSubstream();
    --- simuler configuration 1 ---

    genArr.resetStartSubstream();
    genServA.resetStartSubstream();
    genServB.resetStartSubstream();
    --- simuler configuration 2 ---
}

```



Définition et conception d'un GPA

\mathcal{S} , espace d'états fini;

$f : \mathcal{S} \rightarrow \mathcal{S}$, fonction de transition;

s_0 , germe (état initial);

$$s_n = f(s_{n-1})$$

Définition et conception d'un GPA

\mathcal{S} , espace d'états fini;

$f : \mathcal{S} \rightarrow \mathcal{S}$, fonction de transition;

\mathcal{U} , espace des valeurs de sortie;

$g : \mathcal{S} \rightarrow \mathcal{U}$, fonction de sortie.

s_0 , germe (état initial);

$$s_n = f(s_{n-1})$$

$$u_n = g(s_n)$$

Définition et conception d'un GPA

\mathcal{S} , espace d'états fini;

$f : \mathcal{S} \rightarrow \mathcal{S}$, fonction de transition;

\mathcal{U} , espace des valeurs de sortie;

$g : \mathcal{S} \rightarrow \mathcal{U}$, fonction de sortie.

s_0 , germe (état initial);

$$s_n = f(s_{n-1})$$

$$u_n = g(s_n)$$

s_0

Définition et conception d'un GPA

\mathcal{S} , espace d'états fini;

$f : \mathcal{S} \rightarrow \mathcal{S}$, fonction de transition;

\mathcal{U} , espace des valeurs de sortie;

$g : \mathcal{S} \rightarrow \mathcal{U}$, fonction de sortie.

s_0 , germe (état initial);

$$s_n = f(s_{n-1})$$

$$u_n = g(s_n)$$

$$\begin{array}{c} s_0 \\ \downarrow g \\ u_0 \end{array}$$

Définition et conception d'un GPA

\mathcal{S} , espace d'états fini;

$f : \mathcal{S} \rightarrow \mathcal{S}$, fonction de transition;

\mathcal{U} , espace des valeurs de sortie;

$g : \mathcal{S} \rightarrow \mathcal{U}$, fonction de sortie.

s_0 , germe (état initial);

$$s_n = f(s_{n-1})$$

$$u_n = g(s_n)$$

$$\begin{array}{ccc} s_0 & \xrightarrow{f} & s_1 \\ g \downarrow & & \\ u_0 & & \end{array}$$

Définition et conception d'un GPA

\mathcal{S} , espace d'états fini;

$f : \mathcal{S} \rightarrow \mathcal{S}$, fonction de transition;

\mathcal{U} , espace des valeurs de sortie;

$g : \mathcal{S} \rightarrow \mathcal{U}$, fonction de sortie.

s_0 , germe (état initial);

$$s_n = f(s_{n-1})$$

$$u_n = g(s_n)$$

$$\begin{array}{ccc} s_0 & \xrightarrow{f} & s_1 \\ g \downarrow & & g \downarrow \\ u_0 & & u_1 \end{array}$$

Définition et conception d'un GPA

\mathcal{S} , espace d'états fini;

$f : \mathcal{S} \rightarrow \mathcal{S}$, fonction de transition;

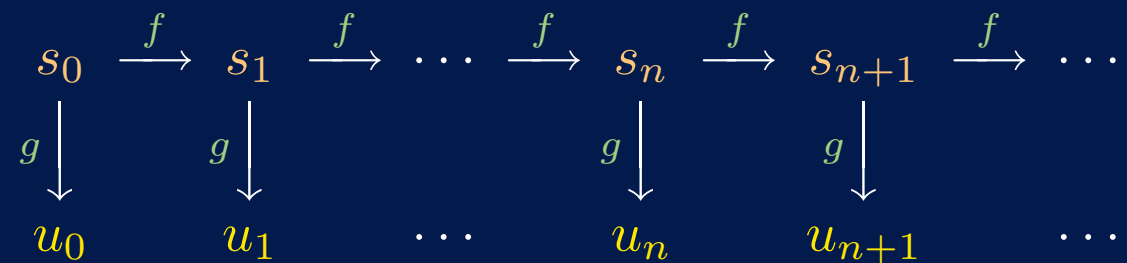
\mathcal{U} , espace des valeurs de sortie;

$g : \mathcal{S} \rightarrow \mathcal{U}$, fonction de sortie.

s_0 , germe (état initial);

$$s_n = f(s_{n-1})$$

$$u_n = g(s_n)$$



Définition et conception d'un GPA

\mathcal{S} , espace d'états fini;

$f : \mathcal{S} \rightarrow \mathcal{S}$, fonction de transition;

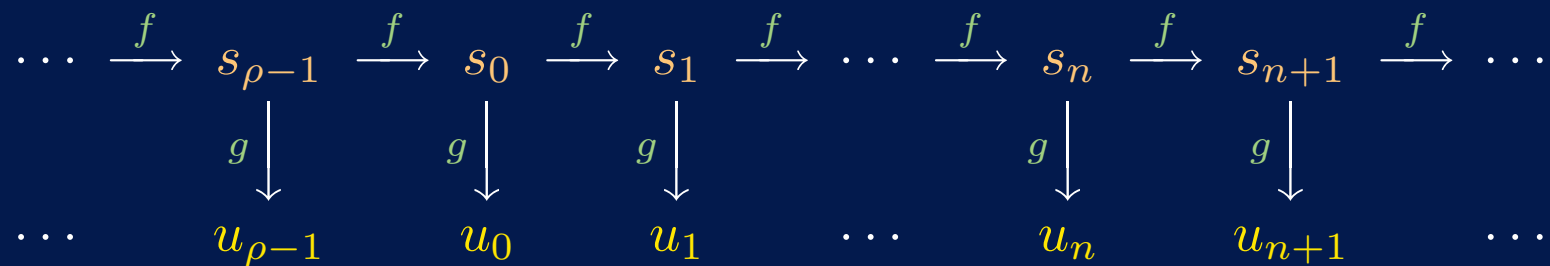
\mathcal{U} , espace des valeurs de sortie;

$g : \mathcal{S} \rightarrow \mathcal{U}$, fonction de sortie.

s_0 , germe (état initial);

$$s_n = f(s_{n-1})$$

$$u_n = g(s_n)$$



Définition et conception d'un GPA

\mathcal{S} , espace d'états fini;

$f : \mathcal{S} \rightarrow \mathcal{S}$, fonction de transition;

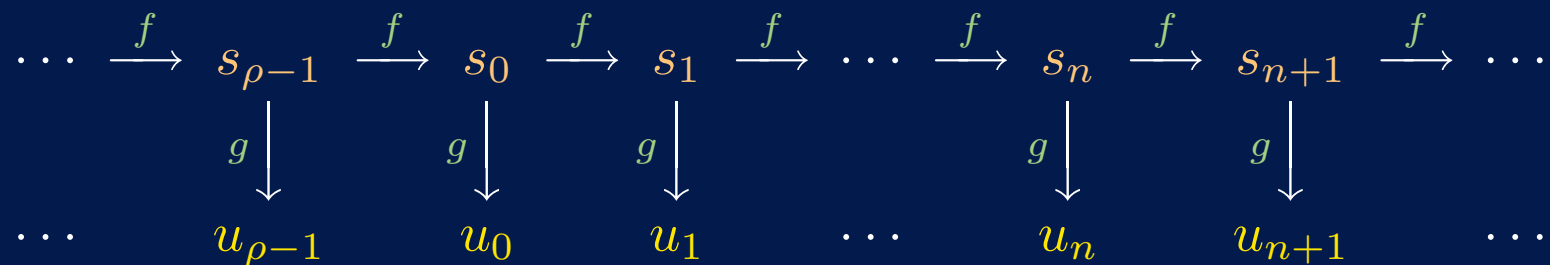
\mathcal{U} , espace des valeurs de sortie;

$g : \mathcal{S} \rightarrow \mathcal{U}$, fonction de sortie.

s_0 , germe (état initial);

$$s_n = f(s_{n-1})$$

$$u_n = g(s_n)$$



Période: $\rho \leq \text{card}(\mathcal{S})$. $s_{n+\rho} = s_n \quad \forall n \geq \tau$.

Définition et conception d'un GPA

\mathcal{S} , espace d'états fini;

$f : \mathcal{S} \rightarrow \mathcal{S}$, fonction de transition;

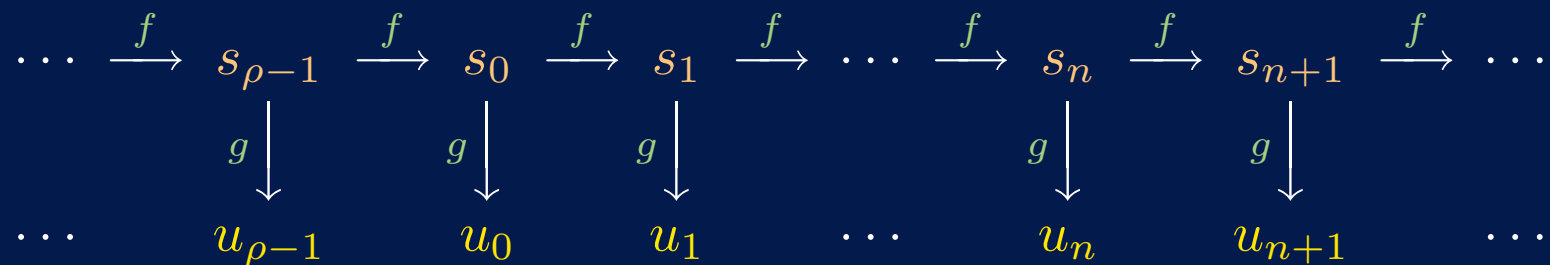
\mathcal{U} , espace des valeurs de sortie;

$g : \mathcal{S} \rightarrow \mathcal{U}$, fonction de sortie.

s_0 , germe (état initial);

$$s_n = f(s_{n-1})$$

$$u_n = g(s_n)$$

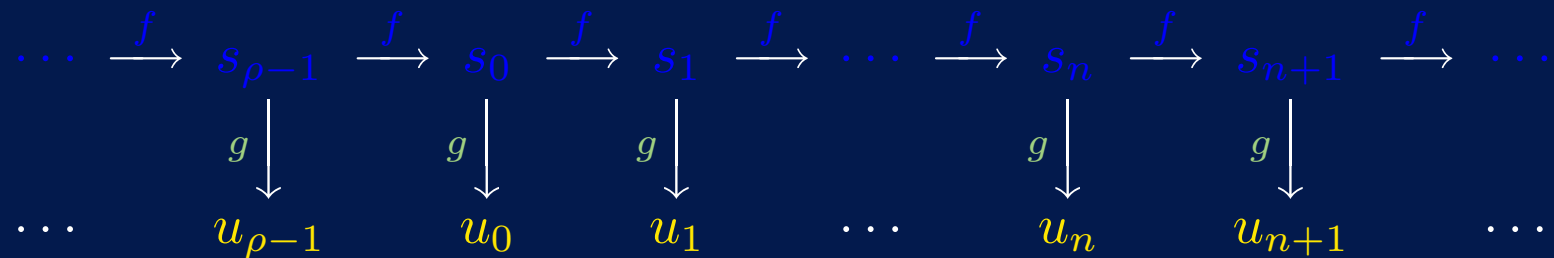


Période: $\rho \leq \text{card}(\mathcal{S})$. $s_{n+\rho} = s_n \quad \forall n \geq \tau$.

On supposera que $\tau = 0$ et $\mathcal{U} = [0, 1)$.

$$\begin{array}{ccccccccccc}
 \dots & \xrightarrow{f} & s_{\rho-1} & \xrightarrow{f} & s_0 & \xrightarrow{f} & s_1 & \xrightarrow{f} & \dots & \xrightarrow{f} & s_n & \xrightarrow{f} & s_{n+1} & \xrightarrow{f} & \dots \\
 & & \downarrow g & & \downarrow g & & \downarrow g & & & & \downarrow g & & \downarrow g & & \\
 \dots & & u_{\rho-1} & & u_0 & & u_1 & & \dots & & u_n & & u_{n+1} & & \dots
 \end{array}$$

Objectif: en observant seulement (u_0, u_1, \dots) , il doit être difficile de distinguer cette suite de la réalisation d'une suite de v.a. i.i.d. uniformes sur \mathcal{U} .

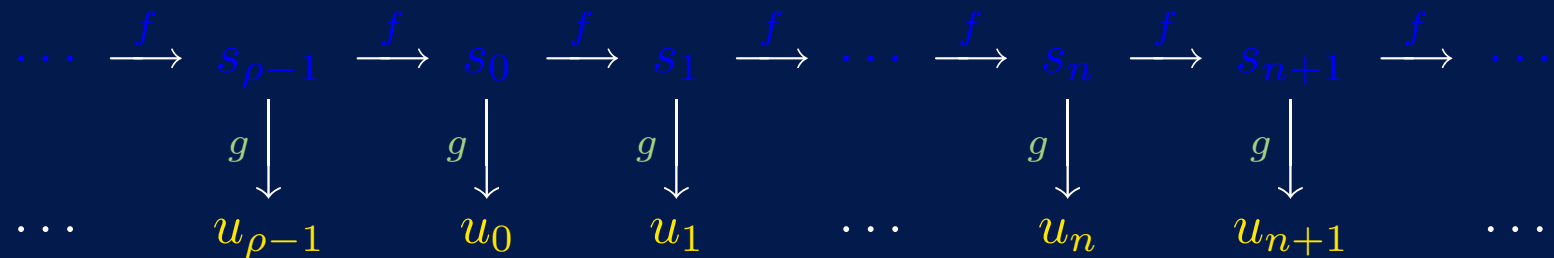


Objectif: en observant seulement (u_0, u_1, \dots) , il doit être difficile de distinguer cette suite de la réalisation d'une suite de v.a. i.i.d. uniformes sur \mathcal{U} .

Utopie: ne pas pouvoir distinguer mieux qu'en tirant à pile ou face.

Autrement dit, que la suite passe **tous** les tests statistiques imaginables.

Cela est impossible! On y reviendra plus loin.



Objectif: en observant seulement (u_0, u_1, \dots) , il doit être difficile de distinguer cette suite de la réalisation d'une suite de v.a. i.i.d. uniformes sur \mathcal{U} .

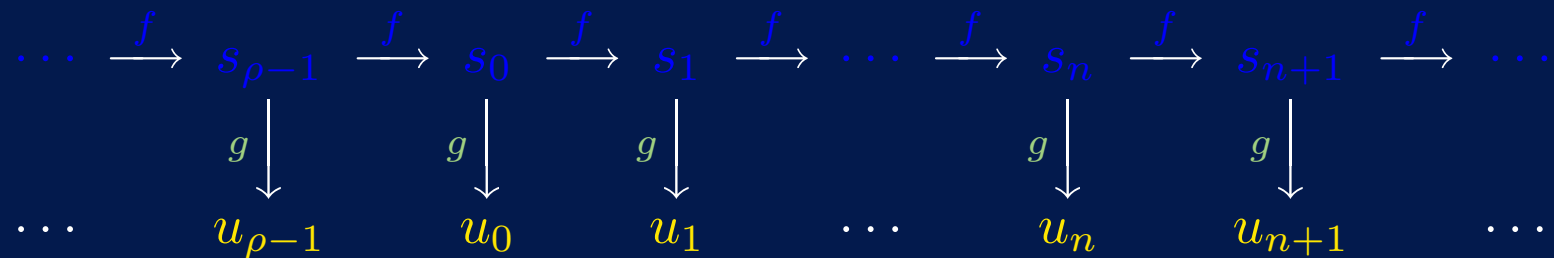
Utopie: ne pas pouvoir distinguer mieux qu'en tirant à pile ou face.

Autrement dit, que la suite passe **tous** les tests statistiques imaginables.

Cela est impossible! On y reviendra plus loin.

On veut aussi: vitesse, facilité d'implantation, suites reproductibles.

Compromis entre vitesse / bonnes propriétés statistiques / (im)prévisibilité.



Objectif: en observant seulement (u_0, u_1, \dots) , il doit être difficile de distinguer cette suite de la réalisation d'une suite de v.a. i.i.d. uniformes sur \mathcal{U} .

Utopie: ne pas pouvoir distinguer mieux qu'en tirant à pile ou face.

Autrement dit, que la suite passe **tous** les tests statistiques imaginables.

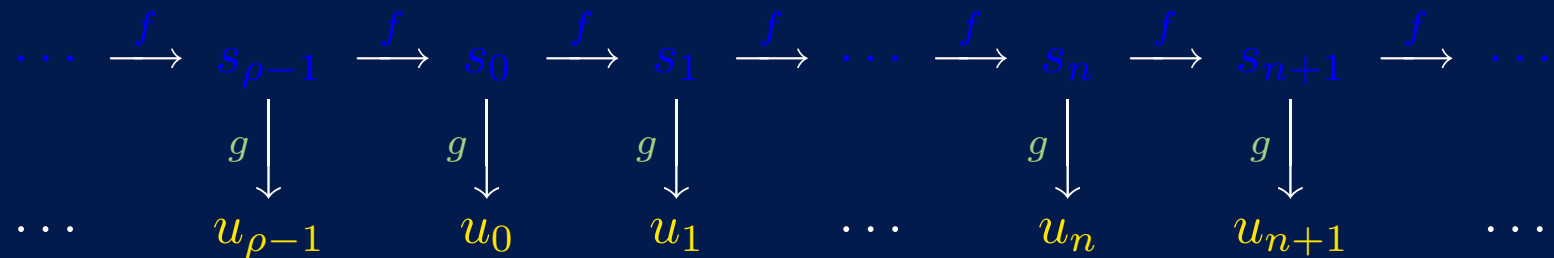
Cela est impossible! On y reviendra plus loin.

On veut aussi: vitesse, facilité d'implantation, suites reproductibles.

Compromis entre vitesse / bonnes propriétés statistiques / (im)prévisibilité.

Si l'état initial s_0 est choisi au hasard, le GPA est comme une roulette géante:

Pour générer t nombres aléatoires, on tourne la roulette pour choisir s_0 , puis on retient $\mathbf{u} = (u_0, \dots, u_{t-1})$.



Objectif: en observant seulement (u_0, u_1, \dots) , il doit être difficile de distinguer cette suite de la réalisation d'une suite de v.a. i.i.d. uniformes sur \mathcal{U} .

Utopie: ne pas pouvoir distinguer mieux qu'en tirant à pile ou face.

Autrement dit, que la suite passe **tous** les tests statistiques imaginables.

Cela est impossible! On y reviendra plus loin.

On veut aussi: vitesse, facilité d'implantation, suites reproductibles.

Compromis entre vitesse / bonnes propriétés statistiques / (im)prévisibilité.

Si l'état initial s_0 est choisi au hasard, le GPA est comme une roulette géante:

Pour générer t nombres aléatoires, on tourne la roulette pour choisir s_0 , puis on retient $\mathbf{u} = (u_0, \dots, u_{t-1})$.

Machines de casinos et loteries: on réinitialise s_0 très souvent.

La loi uniforme sur $[0, 1]^t$.

Choisir s_0 au hasard correspond à choisir un point au hasard dans l'espace échantillonnal $\Psi_t = \{\mathbf{u} = (u_0, \dots, u_{t-1}) = (g(s_0), \dots, g(s_{t-1})), s_0 \in \mathcal{S}\}$, qui peut être interprété comme une approximation de $[0, 1]^t$.

La loi uniforme sur $[0, 1]^t$.

Choisir s_0 au hasard correspond à choisir un point au hasard dans l'espace échantillonnal $\Psi_t = \{\mathbf{u} = (u_0, \dots, u_{t-1}) = (g(s_0), \dots, g(s_{t-1})), s_0 \in \mathcal{S}\}$, qui peut être interprété comme une approximation de $[0, 1]^t$.

Critère: Ψ_t doit recouvrir $[0, 1]^t$ très uniformément pour t jusqu'à (disons) t_0 .

La loi uniforme sur $[0, 1]^t$.

Choisir s_0 au hasard correspond à choisir un point au hasard dans l'espace échantillonnal $\Psi_t = \{\mathbf{u} = (u_0, \dots, u_{t-1}) = (g(s_0), \dots, g(s_{t-1})), s_0 \in \mathcal{S}\}$, qui peut être interprété comme une approximation de $[0, 1]^t$.

Critère: Ψ_t doit recouvrir $[0, 1]^t$ très uniformément pour t jusqu'à (disons) t_0 .

Il nous faut une mesure d'uniformité de Ψ_t (ou une mesure de **discrépance** entre la loi empirique de Ψ_t et la loi uniforme). Plusieurs définitions possibles.

Important: doit être facilement **calculable** sans générer les points.

Pour cela, il faut bien comprendre la structure mathématique de Ψ_t .

Pour cette raison, la plupart des GPA utilisés en simulation sont basés sur des récurrences linéaires (simples).

La loi uniforme sur $[0, 1]^t$.

Choisir s_0 au hasard correspond à choisir un point au hasard dans l'espace échantillonnal $\Psi_t = \{\mathbf{u} = (u_0, \dots, u_{t-1}) = (g(s_0), \dots, g(s_{t-1})), s_0 \in \mathcal{S}\}$, qui peut être interprété comme une approximation de $[0, 1]^t$.

Critère: Ψ_t doit recouvrir $[0, 1]^t$ très uniformément pour t jusqu'à (disons) t_0 .

Il nous faut une mesure d'uniformité de Ψ_t (ou une mesure de **discrépance** entre la loi empirique de Ψ_t et la loi uniforme). Plusieurs définitions possibles.

Important: doit être facilement **calculable** sans générer les points.

Pour cela, il faut bien comprendre la structure mathématique de Ψ_t .

Pour cette raison, la plupart des GPA utilisés en simulation sont basés sur des récurrences linéaires (simples).

Pourquoi ne pas insister que Ψ_t lui-même ressemble à un ensemble de points choisis au hasard (e.g., ne soit pas trop uniforme)?

En fait, on veut cela seulement pour la fraction infime de Ψ_t que l'on utilise.

La loi uniforme sur $[0, 1]^t$.

Choisir s_0 au hasard correspond à choisir un point au hasard dans l'espace échantillonnal $\Psi_t = \{\mathbf{u} = (u_0, \dots, u_{t-1}) = (g(s_0), \dots, g(s_{t-1})), s_0 \in \mathcal{S}\}$, qui peut être interprété comme une approximation de $[0, 1]^t$.

Critère: Ψ_t doit recouvrir $[0, 1]^t$ très uniformément pour t jusqu'à (disons) t_0 .

Il nous faut une mesure d'uniformité de Ψ_t (ou une mesure de **discrépance** entre la loi empirique de Ψ_t et la loi uniforme). Plusieurs définitions possibles.

Important: doit être facilement **calculable** sans générer les points.

Pour cela, il faut bien comprendre la structure mathématique de Ψ_t .

Pour cette raison, la plupart des GPA utilisés en simulation sont basés sur des récurrences linéaires (simples).

Pourquoi ne pas insister que Ψ_t lui-même ressemble à un ensemble de points choisis au hasard (e.g., ne soit pas trop uniforme)?

En fait, on veut cela seulement pour la fraction infime de Ψ_t que l'on utilise.

Généralisation: mesurer l'uniformité de $\Psi_I = \{(u_{i_1}, \dots, u_{i_t}) \mid s_0 \in \mathcal{S}\}$

pour une classe choisie d'ensembles d'indices (non successifs) de forme

$$I = \{i_1, i_2, \dots, i_t\}.$$

Pour une famille donnée \mathcal{J} de sous-ensembles de $\{0, 1, \dots, t\}$, on s'assure que Ψ_I est suffisamment uniforme pour tout $I \in \mathcal{J}$.

Un exemple (trop) simple: générateur à congruence linéaire.

L'état $x_n \in \{0, \dots, m - 1\}$ suit la récurrence

$$x_n = ax_{n-1} \bmod m, \quad \text{et l'output est } u_n = x_n/m.$$

Un exemple (trop) simple: générateur à congruence linéaire.

L'état $x_n \in \{0, \dots, m - 1\}$ suit la récurrence

$$x_n = ax_{n-1} \bmod m, \quad \text{et l'output est } u_n = x_n/m.$$

Période $\rho \leq m - 1$ (peut être atteinte si m est premier).

Un exemple (trop) simple: générateur à congruence linéaire.

L'état $x_n \in \{0, \dots, m - 1\}$ suit la récurrence

$$x_n = ax_{n-1} \bmod m, \quad \text{et l'output est } u_n = x_n/m.$$

Période $\rho \leq m - 1$ (peut être atteinte si m est premier).

Exemple jouet: $m = 101$ et $a = 12$. Si $x_0 = 10$, alors

$$x_0 = 10$$

$$u_0 = 10/101 \approx 0.09901,$$

Un exemple (trop) simple: générateur à congruence linéaire.

L'état $x_n \in \{0, \dots, m - 1\}$ suit la récurrence

$$x_n = ax_{n-1} \bmod m, \quad \text{et l'output est } u_n = x_n/m.$$

Période $\rho \leq m - 1$ (peut être atteinte si m est premier).

Exemple jouet: $m = 101$ et $a = 12$. Si $x_0 = 10$, alors

$$\begin{array}{ll} x_0 & = 10 & u_0 & = 10/101 \approx 0.09901, \\ x_1 & = 12 \times 10 \bmod 101 = 19, & u_1 & = 19/101 \approx 0.18812, \end{array}$$

Un exemple (trop) simple: générateur à congruence linéaire.

L'état $x_n \in \{0, \dots, m - 1\}$ suit la récurrence

$$x_n = ax_{n-1} \bmod m, \quad \text{et l'output est } u_n = x_n/m.$$

Période $\rho \leq m - 1$ (peut être atteinte si m est premier).

Exemple jouet: $m = 101$ et $a = 12$. Si $x_0 = 10$, alors

x_0	$=$	10	u_0	$=$	$10/101 \approx 0.09901,$
x_1	$=$	$12 \times 10 \bmod 101 = 19,$	u_1	$=$	$19/101 \approx 0.18812,$
x_2	$=$	$12 \times 19 \bmod 101 = 26,$	u_2	$=$	$26/101 \approx 0.25742,$

Un exemple (trop) simple: générateur à congruence linéaire.

L'état $x_n \in \{0, \dots, m - 1\}$ suit la récurrence

$$x_n = ax_{n-1} \bmod m, \quad \text{et l'output est } u_n = x_n/m.$$

Période $\rho \leq m - 1$ (peut être atteinte si m est premier).

Exemple jouet: $m = 101$ et $a = 12$. Si $x_0 = 10$, alors

x_0	$=$	10	u_0	$=$	$10/101 \approx 0.09901,$
x_1	$=$	$12 \times 10 \bmod 101 = 19,$	u_1	$=$	$19/101 \approx 0.18812,$
x_2	$=$	$12 \times 19 \bmod 101 = 26,$	u_2	$=$	$26/101 \approx 0.25742,$
x_3	$=$	$12 \times 26 \bmod 101 = 9,$	u_3	$=$	$9/101 \approx 0.08911,$

etc. La suite est périodique de période $\rho = 100$.

Un exemple (trop) simple: générateur à congruence linéaire.

L'état $x_n \in \{0, \dots, m - 1\}$ suit la récurrence

$$x_n = ax_{n-1} \bmod m, \quad \text{et l'output est } u_n = x_n/m.$$

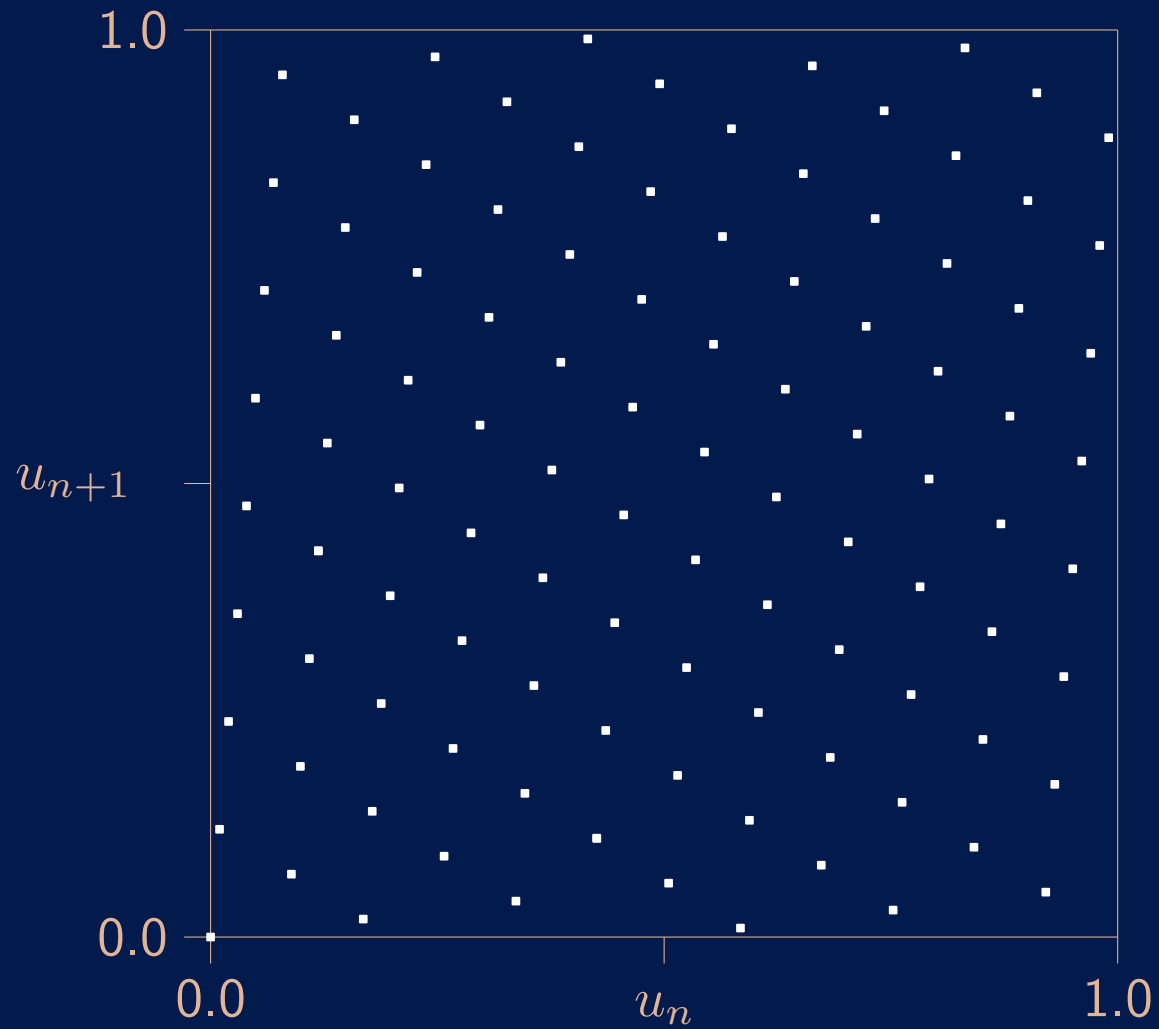
Période $\rho \leq m - 1$ (peut être atteinte si m est premier).

Exemple jouet: $m = 101$ et $a = 12$. Si $x_0 = 10$, alors

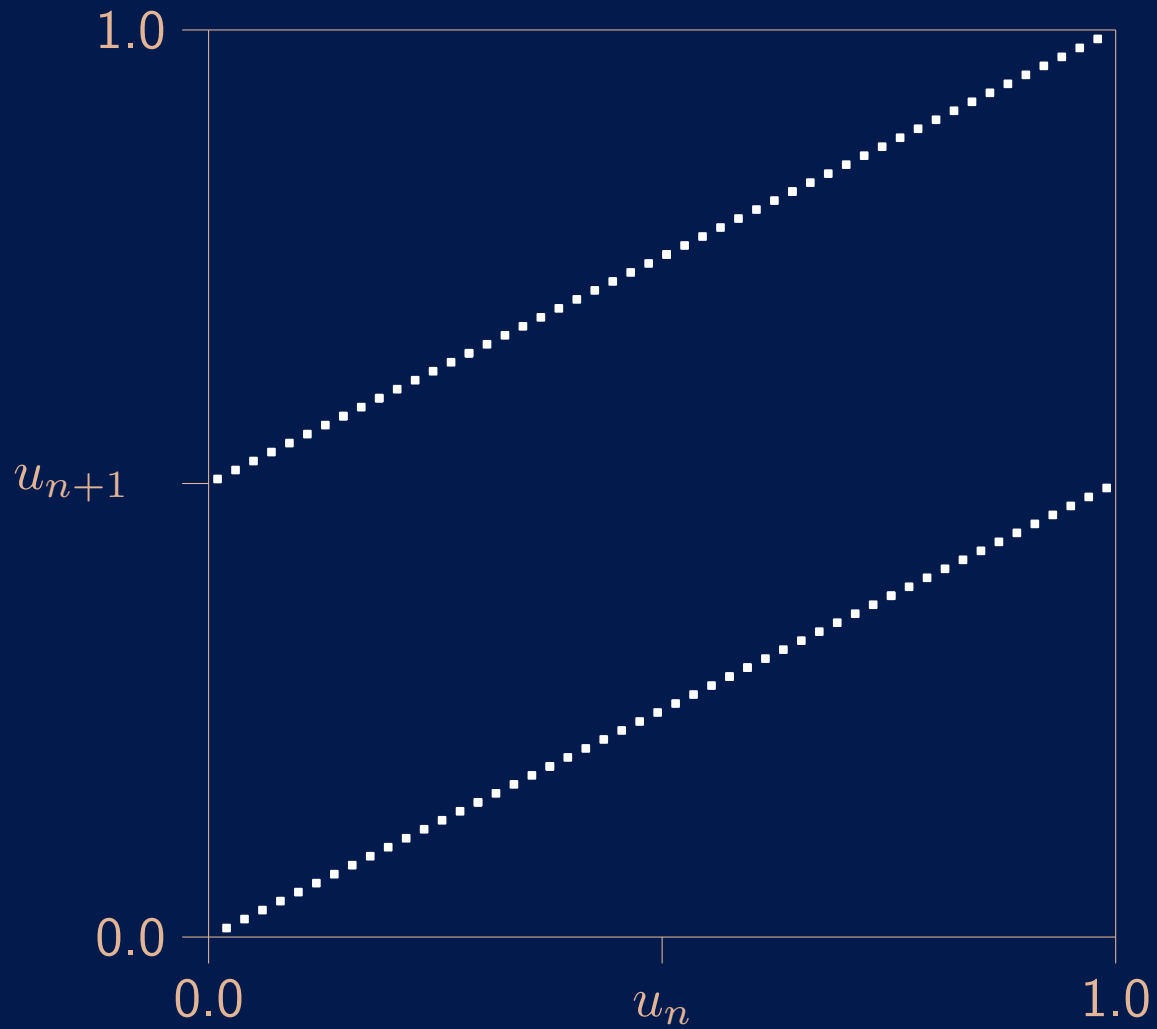
x_0	$=$	10	u_0	$=$	$10/101 \approx 0.09901,$
x_1	$=$	$12 \times 10 \bmod 101 = 19,$	u_1	$=$	$19/101 \approx 0.18812,$
x_2	$=$	$12 \times 19 \bmod 101 = 26,$	u_2	$=$	$26/101 \approx 0.25742,$
x_3	$=$	$12 \times 26 \bmod 101 = 9,$	u_3	$=$	$9/101 \approx 0.08911,$

etc. La suite est périodique de période $\rho = 100$.

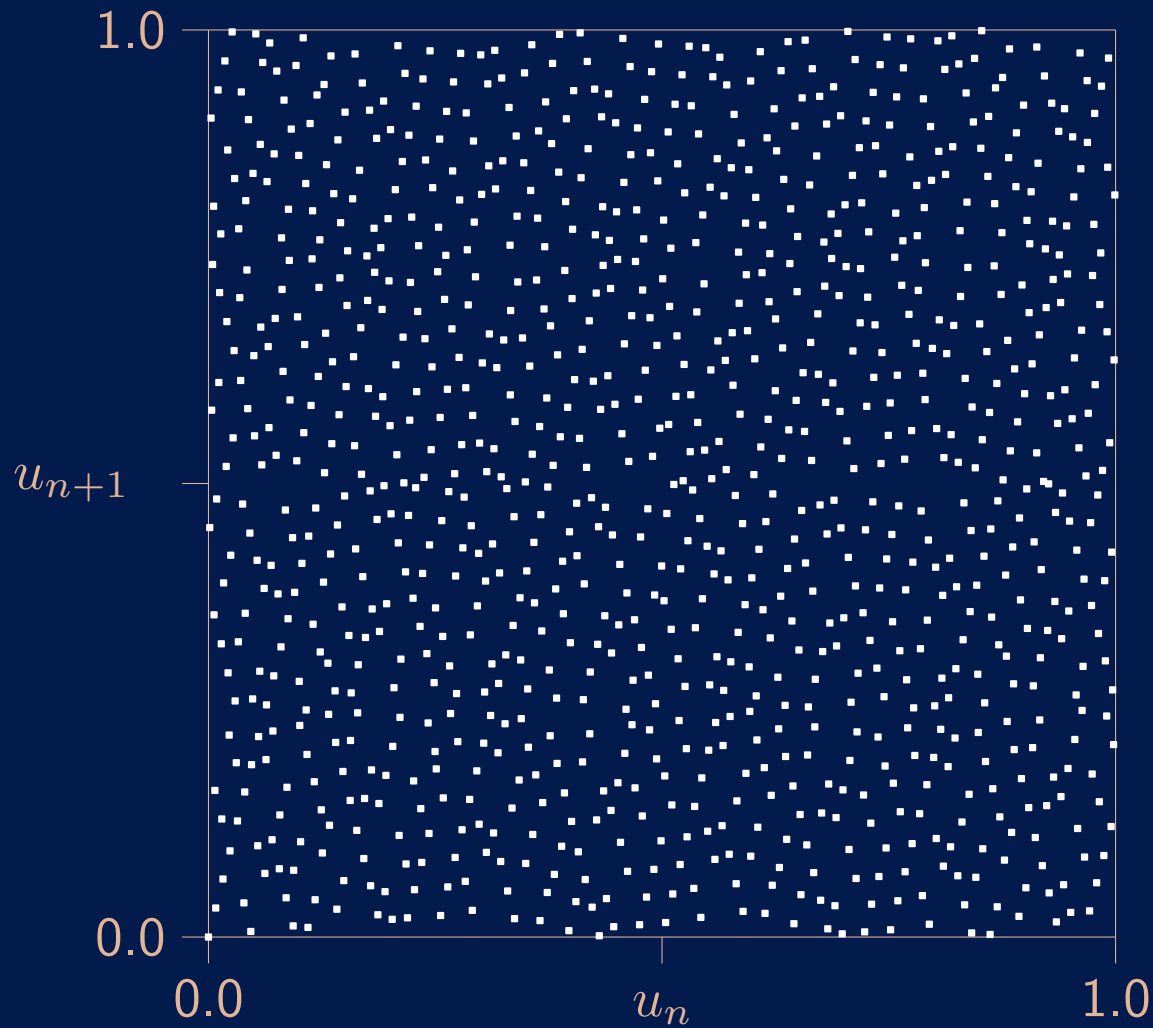
Populaire pour $m = 2^{31} - 1$ ou $m = 2^{32}$ ou $m = 2^{48}$. Mais à rejeter.



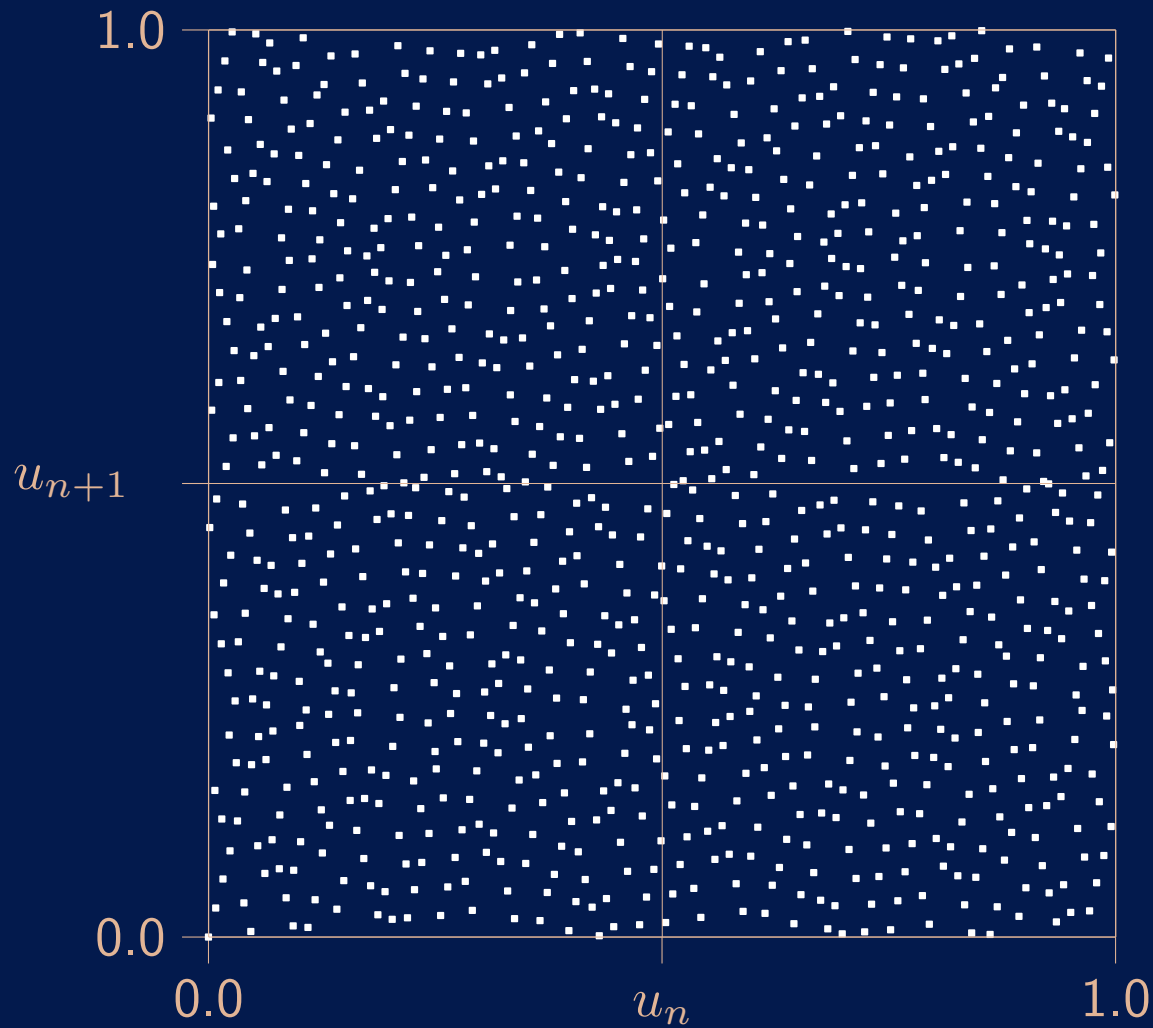
Exemple jouet: LCG avec $m = 101$ et $a = 12$.



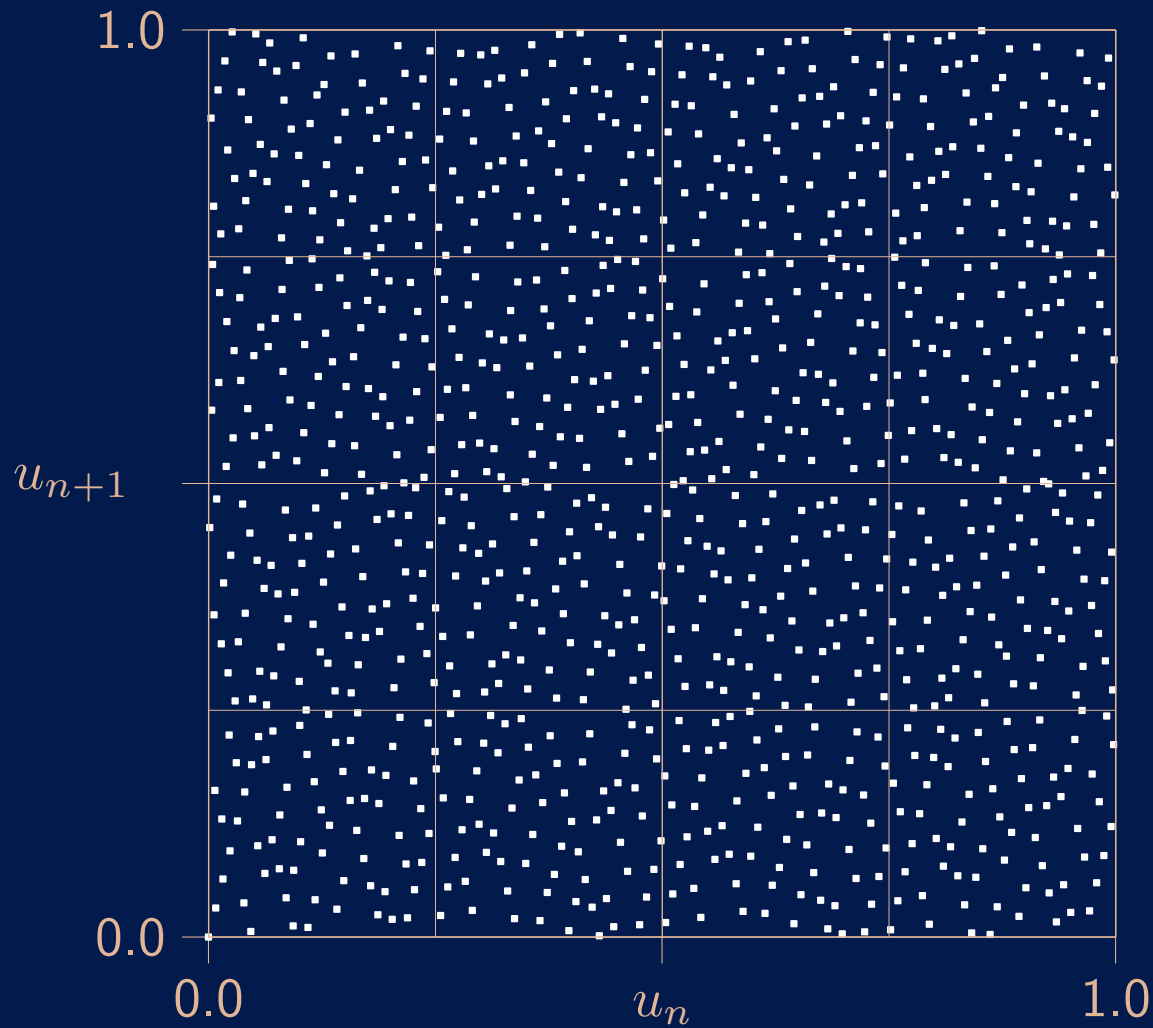
Exemple jouet: LCG avec $m = 101$ et $a = 51$.



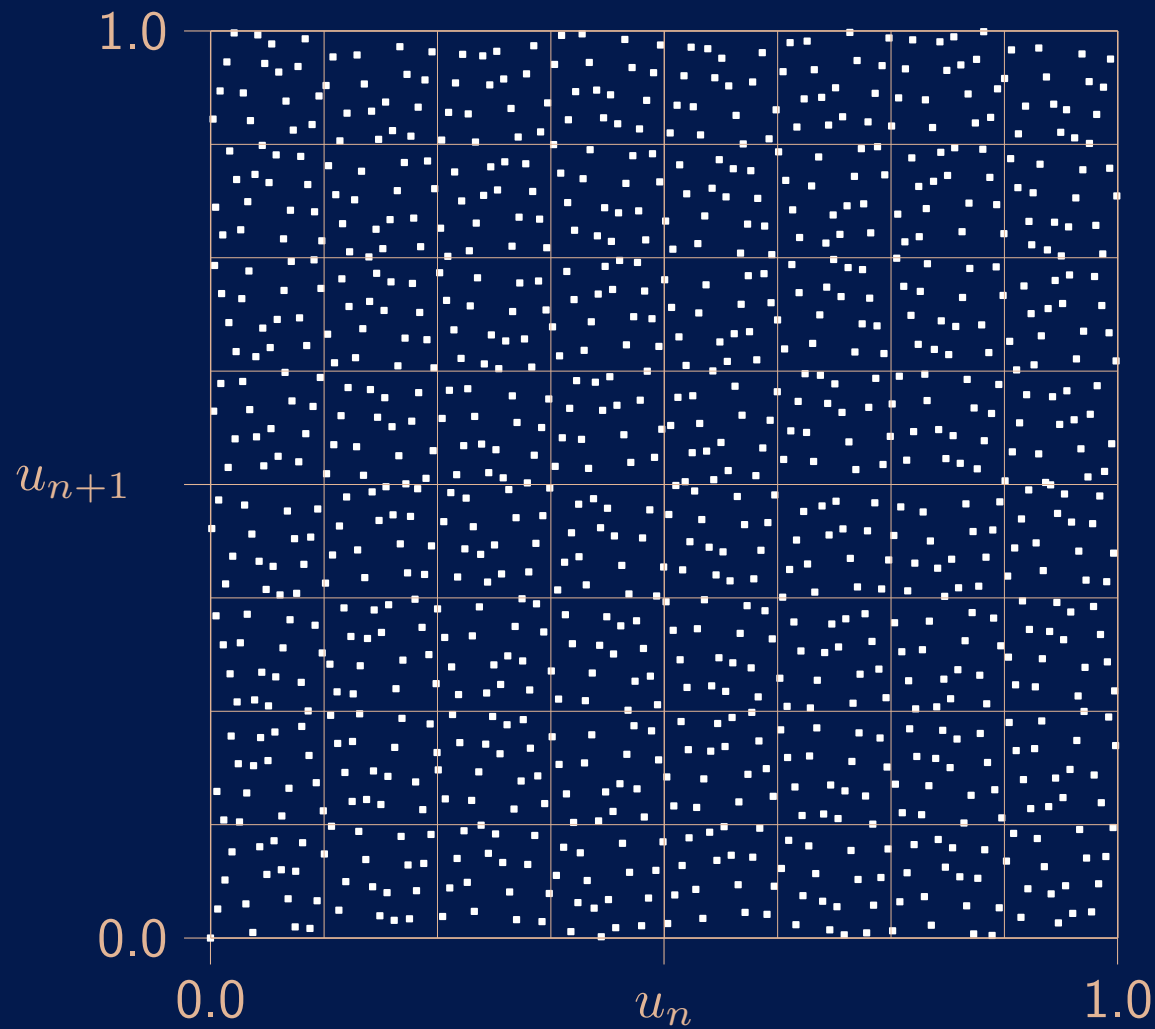
Exemple jouet: LFSR (Tausworthe combiné) avec $n = 1024 = 2^{10}$.



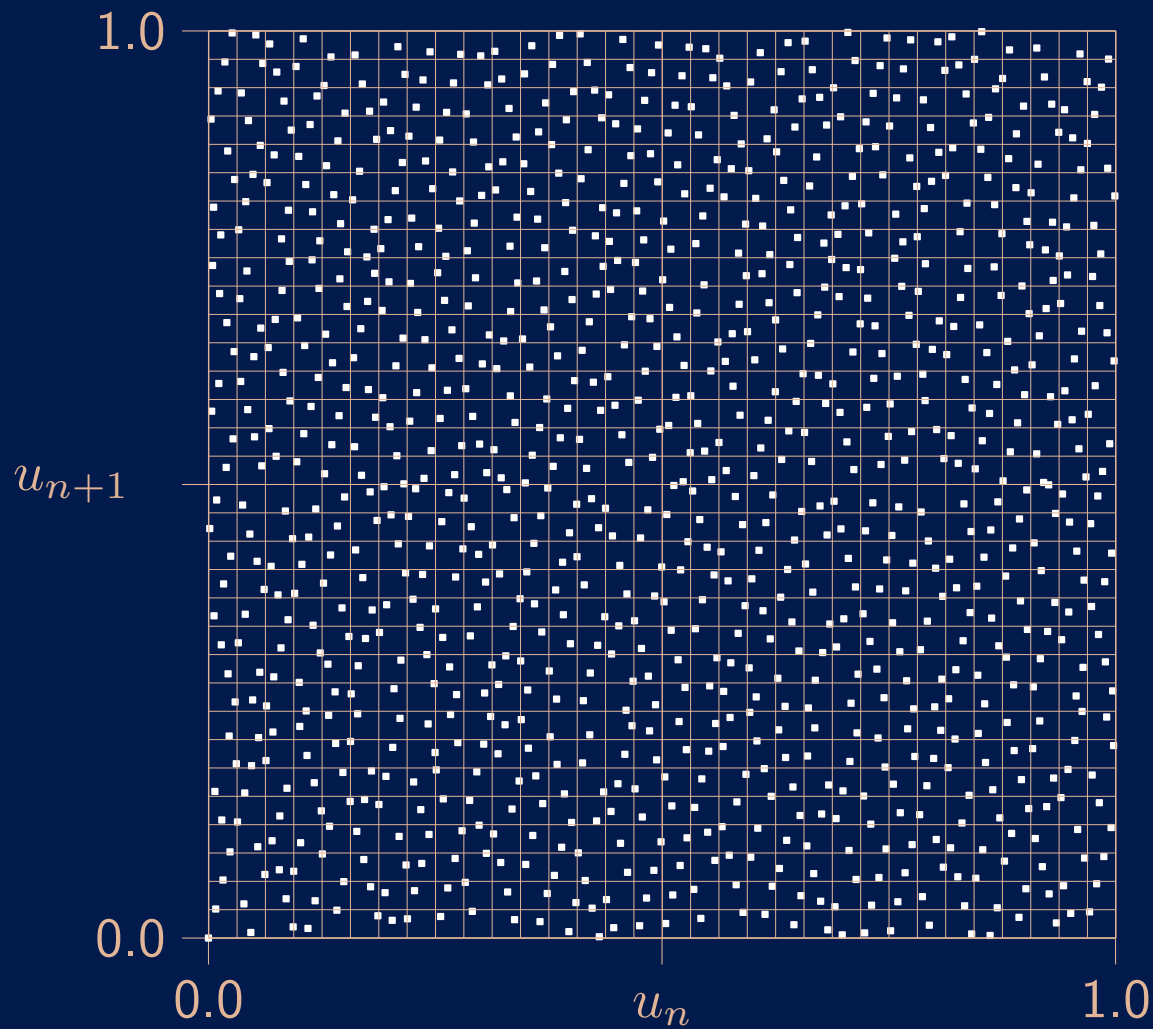
Exemple jouet: LFSR (Tausworthe combiné) avec $n = 1024 = 2^{10}$.



Exemple jouet: LFSR (Tausworthe combiné) avec $n = 1024 = 2^{10}$.



Exemple jouet: LFSR (Tausworthe combiné) avec $n = 1024 = 2^{10}$.



Exemple jouet: LFSR (Tausworthe combiné) avec $n = 1024 = 2^{10}$.

Vrai aussi en 3, 4, 5, ... 10 dimensions.

V.A. Non Uniformes: Inversion

Soit F la fonction de répartition de X , i.e., $F(x) = \mathbb{P}[X \leq x]$ pour tout $x \in \mathbb{R}$.

V.A. Non Uniformes: Inversion

Soit F la fonction de répartition de X , i.e., $F(x) = \mathbb{P}[X \leq x]$ pour tout $x \in \mathbb{R}$.
Pour générer X , générer $U \sim U(0, 1)$ et poser:

$$X = F^{-1}(U) = \min\{x \mid F(x) \geq U\}.$$

V.A. Non Uniformes: Inversion

Soit F la fonction de répartition de X , i.e., $F(x) = \mathbb{P}[X \leq x]$ pour tout $x \in \mathbb{R}$.
Pour générer X , générer $U \sim U(0, 1)$ et poser:

$$X = F^{-1}(U) = \min\{x \mid F(x) \geq U\}.$$

On a en effet $\mathbb{P}[X \leq x] = \mathbb{P}[F^{-1}(U) \leq x]$

V.A. Non Uniformes: Inversion

Soit F la fonction de répartition de X , i.e., $F(x) = \mathbb{P}[X \leq x]$ pour tout $x \in \mathbb{R}$.
Pour générer X , générer $U \sim U(0, 1)$ et poser:

$$X = F^{-1}(U) = \min\{x \mid F(x) \geq U\}.$$

On a en effet $\mathbb{P}[X \leq x] = \mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[U \leq F(x)]$

V.A. Non Uniformes: Inversion

Soit F la fonction de répartition de X , i.e., $F(x) = \mathbb{P}[X \leq x]$ pour tout $x \in \mathbb{R}$.
Pour générer X , générer $U \sim U(0, 1)$ et poser:

$$X = F^{-1}(U) = \min\{x \mid F(x) \geq U\}.$$

On a en effet $\mathbb{P}[X \leq x] = \mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x)$.

V.A. Non Uniformes: Inversion

Soit F la fonction de répartition de X , i.e., $F(x) = \mathbb{P}[X \leq x]$ pour tout $x \in \mathbb{R}$.
Pour générer X , générer $U \sim U(0, 1)$ et poser:

$$X = F^{-1}(U) = \min\{x \mid F(x) \geq U\}.$$

On a en effet $\mathbb{P}[X \leq x] = \mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x)$.

Contrainte: il faut une formule ou une bonne approximation pour F^{-1} .

Exemple: $X \sim \text{Weibull}(\alpha, \lambda)$.

On a $F(x) = 1 - \exp[-(\lambda x)^\alpha]$ pour $x > 0$.

Exemple: $X \sim \text{Weibull}(\alpha, \lambda)$.

On a $F(x) = 1 - \exp[-(\lambda x)^\alpha]$ pour $x > 0$.

Inversion:

$$U = 1 - \exp[-(\lambda X)^\alpha]$$

Exemple: $X \sim \text{Weibull}(\alpha, \lambda)$.

On a $F(x) = 1 - \exp[-(\lambda x)^\alpha]$ pour $x > 0$.

Inversion:

$$\begin{aligned} U &= 1 - \exp[-(\lambda X)^\alpha] \\ \exp[-(\lambda X)^\alpha] &= 1 - U \end{aligned}$$

Exemple: $X \sim \text{Weibull}(\alpha, \lambda)$.

On a $F(x) = 1 - \exp[-(\lambda x)^\alpha]$ pour $x > 0$.

Inversion:

$$\begin{aligned}U &= 1 - \exp[-(\lambda X)^\alpha] \\ \exp[-(\lambda X)^\alpha] &= 1 - U \\ (\lambda X)^\alpha &= -\ln(1 - U)\end{aligned}$$

Exemple: $X \sim \text{Weibull}(\alpha, \lambda)$.

On a $F(x) = 1 - \exp[-(\lambda x)^\alpha]$ pour $x > 0$.

Inversion:

$$\begin{aligned}U &= 1 - \exp[-(\lambda X)^\alpha] \\ \exp[-(\lambda X)^\alpha] &= 1 - U \\ (\lambda X)^\alpha &= -\ln(1 - U) \\ \lambda X &= [-\ln(1 - U)]^{1/\alpha}\end{aligned}$$

Exemple: $X \sim \text{Weibull}(\alpha, \lambda)$.

On a $F(x) = 1 - \exp[-(\lambda x)^\alpha]$ pour $x > 0$.

Inversion:

$$\begin{aligned}U &= 1 - \exp[-(\lambda X)^\alpha] \\ \exp[-(\lambda X)^\alpha] &= 1 - U \\ (\lambda X)^\alpha &= -\ln(1 - U) \\ \lambda X &= [-\ln(1 - U)]^{1/\alpha} \\ X &= [-\ln(1 - U)]^{1/\alpha} / \lambda = F^{-1}(U).\end{aligned}$$

Exemple: $X \sim \text{Weibull}(\alpha, \lambda)$.

On a $F(x) = 1 - \exp[-(\lambda x)^\alpha]$ pour $x > 0$.

Inversion:

$$\begin{aligned}U &= 1 - \exp[-(\lambda X)^\alpha] \\ \exp[-(\lambda X)^\alpha] &= 1 - U \\ (\lambda X)^\alpha &= -\ln(1 - U) \\ \lambda X &= [-\ln(1 - U)]^{1/\alpha} \\ X &= [-\ln(1 - U)]^{1/\alpha} / \lambda = F^{-1}(U).\end{aligned}$$

Cas particulier: si $\alpha = 1$, $X \sim \text{Exponentielle}(\lambda)$.

Pour générer: $X = F^{-1}(U) = -\ln(1 - U) / \lambda$.

Exemple: Soit $\mathbb{P}[X = i] = p_i$ où $p_0 = 1/2$, $p_2 = 3/8$, $p_4 = 1/8$, et $p_i = 0$ ailleurs. ³⁷

Exemple: Soit $\mathbb{P}[X = i] = p_i$ où $p_0 = 1/2$, $p_2 = 3/8$, $p_4 = 1/8$, et $p_i = 0$ ailleurs. 37

La fonction F correspondante est constante par morceaux, avec un saut de hauteur p_i à $x = i$.

Exemple: Soit $\mathbb{P}[X = i] = p_i$ où $p_0 = 1/2$, $p_2 = 3/8$, $p_4 = 1/8$, et $p_i = 0$ ailleurs.

La fonction F correspondante est constante par morceaux, avec un saut de hauteur p_i à $x = i$.

Inversion: retourner 0 si $U < 1/2$, 2 si $1/2 \leq U < 7/8$, et 4 si $U \geq 7/8$.

Exemple: Soit $\mathbb{P}[X = i] = p_i$ où $p_0 = 1/2$, $p_2 = 3/8$, $p_4 = 1/8$, et $p_i = 0$ ailleurs.

La fonction F correspondante est constante par morceaux, avec un saut de hauteur p_i à $x = i$.

Inversion: retourner 0 si $U < 1/2$, 2 si $1/2 \leq U < 7/8$, et 4 si $U \geq 7/8$.

Loi discrète: $\mathbb{P}[X = x_i] = p_i$ pour $i = 0, 1, \dots$

On a $F(x_i) = p_0 + \dots + p_i$ et (disons) $F(x_{-1}) = 0$.

L'inversion retourne $X = x_i$ ssi

$$F(x_i) \geq U > F(x_{i-1}).$$

Exemple: $X \sim \text{Géométrique}(p)$, où $0 < p < 1$.

$\mathbb{P}[X = x] = p(1 - p)^x$ pour $x = 0, 1, 2, \dots$ et

$F(x) = 1 - (1 - p)^{\lfloor x+1 \rfloor}$ pour $x \geq 0$.

Exemple: $X \sim \text{Géométrique}(p)$, où $0 < p < 1$.

$\mathbb{P}[X = x] = p(1 - p)^x$ pour $x = 0, 1, 2, \dots$ et

$F(x) = 1 - (1 - p)^{\lfloor x+1 \rfloor}$ pour $x \geq 0$.

Pour $x \geq 0$ entier, $F(x) = 1 - (1 - p)^{x+1}$.

L'inversion doit retourner $X = x$ ssi $F(x) \geq U > F(x - 1)$,

Exemple: $X \sim \text{Géométrique}(p)$, où $0 < p < 1$.

$\mathbb{P}[X = x] = p(1 - p)^x$ pour $x = 0, 1, 2, \dots$ et

$F(x) = 1 - (1 - p)^{\lfloor x+1 \rfloor}$ pour $x \geq 0$.

Pour $x \geq 0$ entier, $F(x) = 1 - (1 - p)^{x+1}$.

L'inversion doit retourner $X = x$ ssi $F(x) \geq U > F(x - 1)$, que l'on peut réécrire

$$1 - (1 - p)^{x+1} \geq U > 1 - (1 - p)^x,$$

Exemple: $X \sim \text{Géométrique}(p)$, où $0 < p < 1$.

$\mathbb{P}[X = x] = p(1 - p)^x$ pour $x = 0, 1, 2, \dots$ et

$F(x) = 1 - (1 - p)^{\lfloor x+1 \rfloor}$ pour $x \geq 0$.

Pour $x \geq 0$ entier, $F(x) = 1 - (1 - p)^{x+1}$.

L'inversion doit retourner $X = x$ ssi $F(x) \geq U > F(x - 1)$, que l'on peut réécrire

$$\begin{aligned} 1 - (1 - p)^{x+1} &\geq U > 1 - (1 - p)^x, && \text{ou} \\ -(x + 1) \ln(1 - p) &\geq -\ln(1 - U) > -x \ln(1 - p), \end{aligned}$$

Exemple: $X \sim \text{Géométrique}(p)$, où $0 < p < 1$.

$\mathbb{P}[X = x] = p(1 - p)^x$ pour $x = 0, 1, 2, \dots$ et

$F(x) = 1 - (1 - p)^{\lfloor x+1 \rfloor}$ pour $x \geq 0$.

Pour $x \geq 0$ entier, $F(x) = 1 - (1 - p)^{x+1}$.

L'inversion doit retourner $X = x$ ssi $F(x) \geq U > F(x - 1)$, que l'on peut réécrire

$$\begin{aligned} 1 - (1 - p)^{x+1} &\geq U > 1 - (1 - p)^x, && \text{ou} \\ -(x + 1) \ln(1 - p) &\geq -\ln(1 - U) > -x \ln(1 - p), && \text{ou} \\ (x + 1) &\geq \ln(1 - U) / \ln(1 - p) > x, \end{aligned}$$

ce qui donne $x = \lceil \ln(1 - U) / \ln(1 - p) \rceil - 1$.

Avec probabilité 1, c'est la même chose que retourner $X = \lfloor \ln(1 - U) / \ln(1 - p) \rfloor$.

- Dans plusieurs cas (normale, Student, chi-deux, etc.), pas de formule pour F^{-1} ,³⁹ mais approximation numérique.

- Dans plusieurs cas (normale, Student, chi-deux, etc.), pas de formule pour F^{-1} , mais approximation numérique.
- Plus difficile lorsque la forme de F dépend des paramètres (beta, gamma, par exemple).

- Dans plusieurs cas (normale, Student, chi-deux, etc.), pas de formule pour F^{-1} , mais approximation numérique.
- Plus difficile lorsque la forme de F dépend des paramètres (beta, gamma, par exemple).
- Inversion préférable car monotone (on verra pourquoi plus tard).

- Dans plusieurs cas (normale, Student, chi-deux, etc.), pas de formule pour F^{-1} , mais approximation numérique.
- Plus difficile lorsque la forme de F dépend des paramètres (beta, gamma, par exemple).
- Inversion préférable car monotone (on verra pourquoi plus tard).
- Mais autres méthodes parfois beaucoup plus rapides.

Intégration par Monte Carlo

Les problèmes de simulation dont nous avons discuté à date se ramènent tous à estimer une intégrale de la forme

$$\mu = \mathbb{E}[X] = \int_0^1 \cdots \int_0^1 f(u_1, \dots, u_s) du_1 \cdots du_s = \int_{[0,1]^s} f(\mathbf{u}) d\mathbf{u} = \mathbb{E}[f(\mathbf{U})],$$

où $f : [0, 1)^s \rightarrow \mathbb{R}$ et \mathbf{u} est dans $[0, 1)^s$.

La dimension s peut parfois être aléatoire ou infinie.

La fonction f est souvent très compliquée.

μ est la valeur moyenne de f sur $[0, 1)^s$.

Intégration par Monte Carlo

Les problèmes de simulation dont nous avons discuté à date se ramènent tous à estimer une intégrale de la forme

$$\mu = \mathbb{E}[X] = \int_0^1 \cdots \int_0^1 f(u_1, \dots, u_s) du_1 \cdots du_s = \int_{[0,1]^s} f(\mathbf{u}) d\mathbf{u} = \mathbb{E}[f(\mathbf{U})],$$

où $f : [0, 1)^s \rightarrow \mathbb{R}$ et \mathbf{u} est dans $[0, 1)^s$.

La dimension s peut parfois être aléatoire ou infinie.

La fonction f est souvent très compliquée.

μ est la valeur moyenne de f sur $[0, 1)^s$.

Exemple: réseau d'activités stochastique à 13 activités.

On a $s = 13$. À un point $(u_1, \dots, u_{13}) \in [0, 1)^{13}$ correspond un vecteur $(V_1, \dots, V_{13}) = (F_1^{-1}(u_1), \dots, F_{13}^{-1}(u_{13}))$ de durées des activités, et une durée du projet T .

Intégration par Monte Carlo

Les problèmes de simulation dont nous avons discuté à date se ramènent tous à estimer une intégrale de la forme

$$\mu = \mathbb{E}[X] = \int_0^1 \cdots \int_0^1 f(u_1, \dots, u_s) du_1 \cdots du_s = \int_{[0,1]^s} f(\mathbf{u}) d\mathbf{u} = \mathbb{E}[f(\mathbf{U})],$$

où $f : [0, 1]^s \rightarrow \mathbb{R}$ et \mathbf{u} est dans $[0, 1]^s$.

La dimension s peut parfois être aléatoire ou infinie.

La fonction f est souvent très compliquée.

μ est la valeur moyenne de f sur $[0, 1]^s$.

Exemple: réseau d'activités stochastique à 13 activités.

On a $s = 13$. À un point $(u_1, \dots, u_{13}) \in [0, 1]^{13}$ correspond un vecteur $(V_1, \dots, V_{13}) = (F_1^{-1}(u_1), \dots, F_{13}^{-1}(u_{13}))$ de durées des activités, et une durée du projet T . On a $f(u_1, \dots, u_{13}) = 1$ si $T > x$ et $f(u_1, \dots, u_{13}) = 0$ sinon.

Exemple: option asiatique.

Supposons que $Z_j = \Phi^{-1}(U_j)$ où les U_j sont i.i.d. $U(0, 1)$. On a ici $s = d$ et

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

Si s ne dépasse pas 3 ou 4: méthodes d'intégration numériques.

Si s ne dépasse pas 3 ou 4: méthodes d'intégration numériques.

Si s est grand, estimateur de Monte Carlo:

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n f(\mathbf{U}_i),$$

où n est la taille de l'échantillon et $\mathbf{U}_1, \dots, \mathbf{U}_n$ sont i.i.d. uniformes sur $[0, 1)^s$.

Si s ne dépasse pas 3 ou 4: méthodes d'intégration numériques.

Si s est grand, estimateur de Monte Carlo:

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n f(\mathbf{U}_i),$$

où n est la taille de l'échantillon et $\mathbf{U}_1, \dots, \mathbf{U}_n$ sont i.i.d. uniformes sur $[0, 1)^s$.

On a

$$\mathbb{E}[\hat{\mu}_n] = \mathbb{E}[f(\mathbf{U}_i)] = \int_{[0,1)^s} f(\mathbf{u}) d\mathbf{u} = \mu$$

(sans bias) et

$$\text{Var}[\hat{\mu}_n] = \frac{1}{n} [\mathbb{E}[f^2(\mathbf{U}_i)] - \mu^2] = \frac{\sigma^2}{n}$$

où

$$\sigma^2 = \int_{[0,1]^s} f^2(\mathbf{u}) d\mathbf{u} - \mu^2$$

est la variance de $f(\mathbf{U}_i)$.

où

$$\sigma^2 = \int_{[0,1]^s} f^2(\mathbf{u}) d\mathbf{u} - \mu^2$$

est la variance de $f(\mathbf{U}_i)$.

La variance dépend de la moyenne de f^2 .

où

$$\sigma^2 = \int_{[0,1]^s} f^2(\mathbf{u}) d\mathbf{u} - \mu^2$$

est la variance de $f(\mathbf{U}_i)$.

La variance dépend de la moyenne de f^2 .

Elle peut être infinie. Exemple: $s = 1$ et $f(u) = u^{-1/2}$, on a

$$\int_0^1 f^2(u) du = \int_0^1 (1/u) du = \infty.$$

Convergence

Théorème. Supposons que $\sigma^2 < \infty$. Lorsque $n \rightarrow \infty$:

(i) Loi forte des grands nombres: $\lim_{n \rightarrow \infty} \hat{\mu}_n = \mu$ avec probabilité 1.

Convergence

Théorème. Supposons que $\sigma^2 < \infty$. Lorsque $n \rightarrow \infty$:

(i) Loi forte des grands nombres: $\lim_{n \rightarrow \infty} \hat{\mu}_n = \mu$ avec probabilité 1.

(ii) Théorème de la limite centrale:

$$\frac{\sqrt{n}(\hat{\mu}_n - \mu)}{\sigma} \Rightarrow N(0, 1), \text{ i.e.,}$$
$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{\sqrt{n}(\hat{\mu}_n - \mu)}{\sigma} \leq x \right] = \Phi(x) = \mathbb{P}[Z \leq x]$$

pour tout $x \in \mathbb{R}$, où $Z \sim N(0, 1)$ et $\Phi(\cdot)$ sa fonction de répartition.

Convergence

Théorème. Supposons que $\sigma^2 < \infty$. Lorsque $n \rightarrow \infty$:

- (i) Loi forte des grands nombres: $\lim_{n \rightarrow \infty} \hat{\mu}_n = \mu$ avec probabilité 1.
- (ii) Théorème de la limite centrale:

$$\frac{\sqrt{n}(\hat{\mu}_n - \mu)}{\sigma} \Rightarrow N(0, 1), \text{ i.e.,}$$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{\sqrt{n}(\hat{\mu}_n - \mu)}{\sigma} \leq x \right] = \Phi(x) = \mathbb{P}[Z \leq x]$$

pour tout $x \in \mathbb{R}$, où $Z \sim N(0, 1)$ et $\Phi(\cdot)$ sa fonction de répartition.

Les deux propriétés tiennent aussi si on remplace σ^2 par son estimateur sans biais

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n(\bar{X}_n)^2 \right),$$

où $X_i = f(\mathbf{U}_i)$.

Ainsi, pour n grand et un niveau de confiance $1 - \alpha$,

$$\mathbb{P}[|\hat{\mu}_n - \mu| \leq xS_n/\sqrt{n}] \approx \Phi(x) = 1 - \alpha.$$

Intervalle de confiance au niveau α :

$$(\hat{\mu}_n \pm z_{\alpha/2}S_n/\sqrt{n}), \text{ où } z_{\alpha/2} = \Phi^{-1}(\alpha/2).$$

Exemple: $z_{\alpha/2} \approx 1.96$ pour $\alpha = 0.05$.

Ainsi, pour n grand et un niveau de confiance $1 - \alpha$,

$$\mathbb{P}[|\hat{\mu}_n - \mu| \leq xS_n/\sqrt{n}] \approx \Phi(x) = 1 - \alpha.$$

Intervalle de confiance au niveau α :

$$(\hat{\mu}_n \pm z_{\alpha/2}S_n/\sqrt{n}), \text{ où } z_{\alpha/2} = \Phi^{-1}(\alpha/2).$$

Exemple: $z_{\alpha/2} \approx 1.96$ pour $\alpha = 0.05$.

La largeur de l'intervalle de confiance est asymptotiquement proportionnelle à σ/\sqrt{n} , donc converge en $O(n^{-1/2})$.

Ainsi, pour n grand et un niveau de confiance $1 - \alpha$,

$$\mathbb{P}[|\hat{\mu}_n - \mu| \leq xS_n/\sqrt{n}] \approx \Phi(x) = 1 - \alpha.$$

Intervalle de confiance au niveau α :

$$(\hat{\mu}_n \pm z_{\alpha/2}S_n/\sqrt{n}), \text{ où } z_{\alpha/2} = \Phi^{-1}(\alpha/2).$$

Exemple: $z_{\alpha/2} \approx 1.96$ pour $\alpha = 0.05$.

La largeur de l'intervalle de confiance est asymptotiquement proportionnelle à σ/\sqrt{n} , donc converge en $O(n^{-1/2})$.

Si n est petit et les X_i suivent la loi normale, alors $\sqrt{n}(\hat{\mu}_n - \mu)/S_n \sim \text{Student}(n - 1)$.

Ainsi, pour n grand et un niveau de confiance $1 - \alpha$,

$$\mathbb{P}[|\hat{\mu}_n - \mu| \leq xS_n/\sqrt{n}] \approx \Phi(x) = 1 - \alpha.$$

Intervalle de confiance au niveau α :

$$(\hat{\mu}_n \pm z_{\alpha/2}S_n/\sqrt{n}), \text{ où } z_{\alpha/2} = \Phi^{-1}(\alpha/2).$$

Exemple: $z_{\alpha/2} \approx 1.96$ pour $\alpha = 0.05$.

La largeur de l'intervalle de confiance est asymptotiquement proportionnelle à σ/\sqrt{n} , donc converge en $O(n^{-1/2})$.

Si n est petit et les X_i suivent la loi normale, alors $\sqrt{n}(\hat{\mu}_n - \mu)/S_n \sim \text{Student}(n - 1)$.

Mais si la loi des X_i est très asymétrique, ces intervalles ne plus valides même comme approximations.

Example: On veut estimer $\mu = \mathbb{P}[0 \leq Z \leq 2]$ où $Z \sim N(0, 1)$.
Si ϕ est la densité normale standard, on a

$$\mu = \int_0^2 \phi(y) dy = \int_0^1 2\phi(2u) du = \int_0^1 f(u) du$$

via le changement de variable $y = 2u$ et $f(u) = 2\phi(2u)$.

Example: On veut estimer $\mu = \mathbb{P}[0 \leq Z \leq 2]$ où $Z \sim N(0, 1)$.
Si ϕ est la densité normale standard, on a

$$\mu = \int_0^2 \phi(y) dy = \int_0^1 2\phi(2u) du = \int_0^1 f(u) du$$

via le changement de variable $y = 2u$ et $f(u) = 2\phi(2u)$.

Pour estimer μ par MC:

Pour $i = 1, \dots, n$,

Générer $U_i \sim U(0, 1)$;

Poser $X_i = 2\phi(2U_i)$;

Calculer $\hat{\mu}_n = \bar{X}_n$, puis un intervalle de confiance pour μ .

Exemple: Réseau d'activités stochastique. On a

$$\begin{aligned}X_i &= \mathbb{I}[T_i > x], \\ \hat{\mu}_n &= \frac{1}{n} \sum_{i=1}^n X_i = \frac{Y}{n}, \\ S_n^2 &= \frac{Y(1 - Y/n)}{n - 1}.\end{aligned}$$

Intervalle de confiance à 95%: $(Y/n \pm 1.96S_n/\sqrt{n})$.
Valide si $n\mu$ n'est pas trop proche de 0 ou 1.

Exemple: Réseau d'activités stochastique. On a

$$\begin{aligned}X_i &= \mathbb{I}[T_i > x], \\ \hat{\mu}_n &= \frac{1}{n} \sum_{i=1}^n X_i = \frac{Y}{n}, \\ S_n^2 &= \frac{Y(1 - Y/n)}{n - 1}.\end{aligned}$$

Intervalle de confiance à 95%: $(Y/n \pm 1.96S_n/\sqrt{n})$.

Valide si $n\mu$ n'est pas trop proche de 0 ou 1.

En fait, $Y \sim \text{Binomiale}(n, \mu)$.

Supposons par exemple que l'on observe $Y = 882$. On a alors:

$$\bar{X}_n = 882/1000 = 0.882;$$

$$S_n^2 = \bar{X}_n(1 - \bar{X}_n)n/(n - 1) \approx 0.1042.$$

Supposons par exemple que l'on observe $Y = 882$. On a alors:

$$\bar{X}_n = 882/1000 = 0.882;$$

$$S_n^2 = \bar{X}_n(1 - \bar{X}_n)n/(n - 1) \approx 0.1042.$$

On obtient l'intervalle de confiance à 95%:

$$(\bar{X}_n \pm 1.96S_n/\sqrt{n}) \approx (0.882 \pm 0.020) = (0.862, 0.902).$$

Supposons par exemple que l'on observe $Y = 882$. On a alors:

$$\bar{X}_n = 882/1000 = 0.882;$$

$$S_n^2 = \bar{X}_n(1 - \bar{X}_n)n/(n - 1) \approx 0.1042.$$

On obtient l'intervalle de confiance à 95%:

$$(\bar{X}_n \pm 1.96S_n/\sqrt{n}) \approx (0.882 \pm 0.020) = (0.862, 0.902).$$

Notre estimateur de μ a donc deux chiffres significatifs: $\mu \approx 0.88$.

Le "2" dans 0.882 n'est pas significatif.

Supposons par exemple que l'on observe $Y = 882$. On a alors:

$$\bar{X}_n = 882/1000 = 0.882;$$

$$S_n^2 = \bar{X}_n(1 - \bar{X}_n)n/(n - 1) \approx 0.1042.$$

On obtient l'intervalle de confiance à 95%:

$$(\bar{X}_n \pm 1.96S_n/\sqrt{n}) \approx (0.882 \pm 0.020) = (0.862, 0.902).$$

Notre estimateur de μ a donc deux chiffres significatifs: $\mu \approx 0.88$.

Le "2" dans 0.882 n'est pas significatif.

Il faut éviter de donner des chiffres non significatifs dans les résultats, car cela peut induire en erreur.

Avantages de Monte Carlo pour l'intégration:

- Ne requiert qu'une hypothèse très faible sur f .
- Le taux de convergence de l'erreur ne dépend pas de la dimension t , contrairement aux méthodes d'intégration numérique classiques.
- On peut estimer l'erreur de manière probabiliste (intervalles de confiance). Les méthodes d'intégration numériques donnent des bornes sur l'erreur, mais peu pratiques.

Efficacité des estimateurs

Soit X estimateur de μ .

$$\beta = \mathbb{E}[X] - \mu \quad \text{biais}$$

$$\sigma^2 = \text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \quad \text{variance}$$

$$\text{MSE}[X] = \mathbb{E}[(X - \mu)^2] = \beta^2 + \sigma^2 \quad \text{erreur quadratique moyenne}$$

$$\sqrt{\text{MSE}[X]} \quad \text{erreur absolue}$$

$$\text{RE}[X] = \sqrt{\text{MSE}[X]}/|\mu|, \quad \text{pour } \mu \neq 0 \quad \text{erreur relative}$$

Soit $C(X)$ l'espérance mathématique du temps de calcul de X .
On définit l'efficacité de l'estimateur X par

$$\text{Eff}(X) = \frac{1}{C(X) \cdot \text{MSE}(X)}.$$

On dit que X est plus efficace que Y si $\text{Eff}(X) > \text{Eff}(Y)$.

L'amélioration de l'efficacité consiste à trouver des estimateurs plus efficaces, en ce sens.

Peut être obtenu en diminuant la variance, ou le biais, ou le temps de calcul.

D'où vient cette définition?

Soit X_1, \dots, X_n i.i.d. et $C(\bar{X}_n) = \kappa n$.

Soit X_1, \dots, X_n i.i.d. et $C(\bar{X}_n) = \kappa n$.
On a $\text{Var}[\bar{X}_n] = \text{Var}[X_i]/n = \sigma^2/n$.

Soit X_1, \dots, X_n i.i.d. et $C(\bar{X}_n) = \kappa n$.

On a $\text{Var}[\bar{X}_n] = \text{Var}[X_i]/n = \sigma^2/n$.

Si $\mathbb{E}[X_i] = \mu$, alors la mesure de l'efficacité est

$$\text{Eff}[\bar{X}_n] = \frac{1}{C(\bar{X}_n) \cdot \text{MSE}(\bar{X}_n)} = \frac{1}{\kappa n \sigma^2 / n} = \frac{1}{\kappa \sigma^2}.$$

Elle ne dépend pas de n .

Soit X_1, \dots, X_n i.i.d. et $C(\bar{X}_n) = \kappa n$.

On a $\text{Var}[\bar{X}_n] = \text{Var}[X_i]/n = \sigma^2/n$.

Si $\mathbb{E}[X_i] = \mu$, alors la mesure de l'efficacité est

$$\text{Eff}[\bar{X}_n] = \frac{1}{C(\bar{X}_n) \cdot \text{MSE}(\bar{X}_n)} = \frac{1}{\kappa n \sigma^2 / n} = \frac{1}{\kappa \sigma^2}.$$

Elle ne dépend pas de n .

Par contre, si $\mathbb{E}[X_i] - \mu = \beta \neq 0$,

$$\text{Eff}[\bar{X}_n] = \frac{1}{\kappa n (\beta^2 + \sigma^2/n)} = \frac{1}{\kappa (n\beta^2 + \sigma^2)}$$

diminue avec n , et on ne peut généralement pas estimer β .

On pourra s'intéresser à étudier le comportement asymptotique de l'efficacité.

Soit X_1, \dots, X_n i.i.d. et $C(\bar{X}_n) = \kappa n$.

On a $\text{Var}[\bar{X}_n] = \text{Var}[X_i]/n = \sigma^2/n$.

Si $\mathbb{E}[X_i] = \mu$, alors la mesure de l'efficacité est

$$\text{Eff}[\bar{X}_n] = \frac{1}{C(\bar{X}_n) \cdot \text{MSE}(\bar{X}_n)} = \frac{1}{\kappa n \sigma^2/n} = \frac{1}{\kappa \sigma^2}.$$

Elle ne dépend pas de n .

Par contre, si $\mathbb{E}[X_i] - \mu = \beta \neq 0$,

$$\text{Eff}[\bar{X}_n] = \frac{1}{\kappa n (\beta^2 + \sigma^2/n)} = \frac{1}{\kappa (n\beta^2 + \sigma^2)}$$

diminue avec n , et on ne peut généralement pas estimer β .

On pourra s'intéresser à étudier le comportement asymptotique de l'efficacité.

En principe on peut estimer σ^2 par S_n^2 , mais $\text{RE}[S_n^2]$ est parfois énorme.

Ainsi, le coût de calcul, la variance et le biais ne disent pas tout.

Elles ne tiennent pas compte, par exemple, des moments d'ordre plus élevé de l'estimateur, ou encore de la qualité de l'estimateur de variance.

Notre définition d'efficacité est un compromis.

Exemple.

On veut estimer $p = \mathbb{P}\{A\}$ où A est un événement rare (p est proche de 0).
La variable binaire $X = \mathbb{I}[A]$ est un estimateur sans biais de p ,
de variance (et MSE) $\text{Var}[X] = \text{MSE}[X] = p(1 - p)$.

Exemple.

On veut estimer $p = \mathbb{P}\{A\}$ où A est un événement rare (p est proche de 0).

La variable binaire $X = \mathbb{I}[A]$ est un estimateur sans biais de p ,
de variance (et MSE) $\text{Var}[X] = \text{MSE}[X] = p(1 - p)$.

Si p est petit, $\text{Var}[X] \approx p$ est petite.

Exemple.

On veut estimer $p = \mathbb{P}\{A\}$ où A est un événement rare (p est proche de 0).

La variable binaire $X = \mathbb{I}[A]$ est un estimateur sans biais de p ,
de variance (et MSE) $\text{Var}[X] = \text{MSE}[X] = p(1 - p)$.

Si p est petit, $\text{Var}[X] \approx p$ est petite.

Mais on veut $\text{Var}[X] \ll p$.

En effet, l'estimateur trivial $Y = 0$ donne $\text{Var}[Y] = 0$ et $\text{MSE}[Y] = p^2$.

Exemple.

On veut estimer $p = \mathbb{P}\{A\}$ où A est un événement rare (p est proche de 0).

La variable binaire $X = \mathbb{I}[A]$ est un estimateur sans biais de p ,
de variance (et MSE) $\text{Var}[X] = \text{MSE}[X] = p(1 - p)$.

Si p est petit, $\text{Var}[X] \approx p$ est petite.

Mais on veut $\text{Var}[X] \ll p$.

En effet, l'estimateur trivial $Y = 0$ donne $\text{Var}[Y] = 0$ et $\text{MSE}[Y] = p^2$.

Si on prend \bar{X}_n , alors $\text{Var}[\bar{X}_n] = p(1 - p)/n$, de sorte que $\text{MSE}[\bar{X}_n] < \text{MSE}[Y]$
ssi $n > (1 - p)/p$.

Exemple.

On veut estimer $p = \mathbb{P}\{A\}$ où A est un événement rare (p est proche de 0).

La variable binaire $X = \mathbb{I}[A]$ est un estimateur sans biais de p , de variance (et MSE) $\text{Var}[X] = \text{MSE}[X] = p(1 - p)$.

Si p est petit, $\text{Var}[X] \approx p$ est petite.

Mais on veut $\text{Var}[X] \ll p$.

En effet, l'estimateur trivial $Y = 0$ donne $\text{Var}[Y] = 0$ et $\text{MSE}[Y] = p^2$.

Si on prend \bar{X}_n , alors $\text{Var}[\bar{X}_n] = p(1 - p)/n$, de sorte que $\text{MSE}[\bar{X}_n] < \text{MSE}[Y]$ ssi $n > (1 - p)/p$.

Si $|p|$ est petit, il est plus approprié de considérer le **MSE relatif** $\text{MSE}[X]/p^2$, ou l'**erreur relative** $\text{RE}[X]$, car la largeur relative d'un intervalle de confiance sur p est à peu près proportionnelle à $\text{RE}[X]/\sqrt{n}$.

On a ici $\text{RE}[X] = \sqrt{(1 - p)/p} \rightarrow \infty$ lorsque $p \rightarrow 0$.

Exemple.

On veut estimer $p = \mathbb{P}\{A\}$ où A est un événement rare (p est proche de 0).

La variable binaire $X = \mathbb{I}[A]$ est un estimateur sans biais de p , de variance (et MSE) $\text{Var}[X] = \text{MSE}[X] = p(1 - p)$.

Si p est petit, $\text{Var}[X] \approx p$ est petite.

Mais on veut $\text{Var}[X] \ll p$.

En effet, l'estimateur trivial $Y = 0$ donne $\text{Var}[Y] = 0$ et $\text{MSE}[Y] = p^2$.

Si on prend \bar{X}_n , alors $\text{Var}[\bar{X}_n] = p(1 - p)/n$, de sorte que $\text{MSE}[\bar{X}_n] < \text{MSE}[Y]$ ssi $n > (1 - p)/p$.

Si $|p|$ est petit, il est plus approprié de considérer le **MSE relatif** $\text{MSE}[X]/p^2$, ou l'**erreur relative** $\text{RE}[X]$, car la largeur relative d'un intervalle de confiance sur p est à peu près proportionnelle à $\text{RE}[X]/\sqrt{n}$.

On a ici $\text{RE}[X] = \sqrt{(1 - p)/p} \rightarrow \infty$ lorsque $p \rightarrow 0$.

Par exemple, si $p \approx 10^{-10}$, il faut $n \approx 10^{12}$ pour une erreur relative de 10%.

Autre problème relié: Si p est très petit (np est petit), on a de fortes chances d'avoir $X_1 = \dots = X_n = 0$, ce qui donne $\bar{X}_n = S_n^2 = 0$ et un intervalle de confiance de largeur 0.

Autre problème relié: Si p est très petit (np est petit), on a de fortes chances d'avoir $X_1 = \dots = X_n = 0$, ce qui donne $\bar{X}_n = S_n^2 = 0$ et un intervalle de confiance de largeur 0.

La probabilité d'un échec étant $1 - p$, on a $\mathbb{P}[\bar{X}_n = S_n^2 = 0] = (1 - p)^n \approx 1 - np$.

Autre problème relié: Si p est très petit (np est petit), on a de fortes chances d'avoir $X_1 = \dots = X_n = 0$, ce qui donne $\bar{X}_n = S_n^2 = 0$ et un intervalle de confiance de largeur 0.

La probabilité d'un échec étant $1 - p$, on a $\mathbb{P}[\bar{X}_n = S_n^2 = 0] = (1 - p)^n \approx 1 - np$.

Ici, $n\bar{X}_n$ suit la loi binomiale de paramètres (n, p) et l'approximation normale est bonne seulement si n et np sont grands.

Autre problème relié: Si p est très petit (np est petit), on a de fortes chances d'avoir $X_1 = \dots = X_n = 0$, ce qui donne $\bar{X}_n = S_n^2 = 0$ et un intervalle de confiance de largeur 0.

La probabilité d'un échec étant $1 - p$, on a $\mathbb{P}[\bar{X}_n = S_n^2 = 0] = (1 - p)^n \approx 1 - np$.

Ici, $n\bar{X}_n$ suit la loi binomiale de paramètres (n, p) et l'approximation normale est bonne seulement si n et np sont grands.

Si n est grand et np est petit: approx. par la loi de Poisson.

Autre problème relié: Si p est très petit (np est petit), on a de fortes chances d'avoir $X_1 = \dots = X_n = 0$, ce qui donne $\bar{X}_n = S_n^2 = 0$ et un intervalle de confiance de largeur 0.

La probabilité d'un échec étant $1 - p$, on a $\mathbb{P}[\bar{X}_n = S_n^2 = 0] = (1 - p)^n \approx 1 - np$.

Ici, $n\bar{X}_n$ suit la loi binomiale de paramètres (n, p) et l'approximation normale est bonne seulement si n et np sont grands.

Si n est grand et np est petit: approx. par la loi de Poisson.

On verra plus loin des techniques adaptées à ce contexte d'événements rares.

Estimateur succès/échec (“hit-or-miss”)

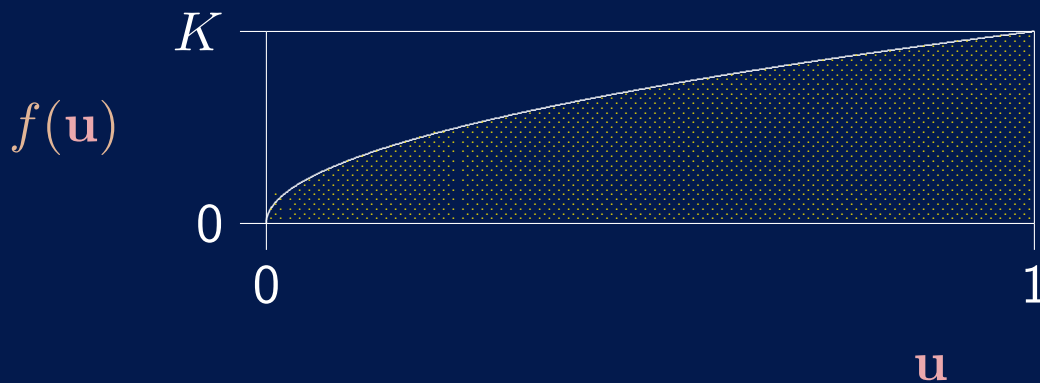
On suppose $0 \leq f(\mathbf{u}) \leq K$.

Soient $\mathbf{U}_i \sim U[0, 1)^s$ et $V_i \sim U(0, 1)$, $i = 1, \dots, n$, mutuellement indépendantes,

$$B_i = \mathbb{I}[V_i K \leq f(\mathbf{U}_i)] = \begin{cases} 1 & \text{si } V_i K \leq f(\mathbf{U}_i); \\ 0 & \text{sinon,} \end{cases}$$

et

$$\tilde{\mu}_n = \frac{K}{n} \sum_{i=1}^n B_i.$$



Les B_i sont i.i.d. Bernoulli de paramètre $p = \mu/K$. Ainsi $\mathbb{E}[\tilde{\mu}_n] = Kp = \mu$ et

$$\text{Var}[\tilde{\mu}_n] = \frac{1}{n} [\mathbb{E}[(KB_i)^2] - \mu^2] = \frac{1}{n} [K^2p - \mu^2] = \mu(K - \mu)/n.$$

Les B_i sont i.i.d. Bernoulli de paramètre $p = \mu/K$. Ainsi $\mathbb{E}[\tilde{\mu}_n] = Kp = \mu$ et

$$\text{Var}[\tilde{\mu}_n] = \frac{1}{n} [\mathbb{E}[(KB_i)^2] - \mu^2] = \frac{1}{n} [K^2p - \mu^2] = \mu(K - \mu)/n.$$

On a $\int_{[0,1]^s} f^2(\mathbf{u})d\mathbf{u} \leq \int_{[0,1]^s} Kf(\mathbf{u})d\mathbf{u} = K\mu$, et donc

$$\text{Var}[\hat{\mu}_n] \leq (K\mu - \mu^2)/n = \text{Var}[\tilde{\mu}_n].$$

Les B_i sont i.i.d. Bernoulli de paramètre $p = \mu/K$. Ainsi $\mathbb{E}[\tilde{\mu}_n] = Kp = \mu$ et

$$\text{Var}[\tilde{\mu}_n] = \frac{1}{n} [\mathbb{E}[(KB_i)^2] - \mu^2] = \frac{1}{n} [K^2p - \mu^2] = \mu(K - \mu)/n.$$

On a $\int_{[0,1]^s} f^2(\mathbf{u})d\mathbf{u} \leq \int_{[0,1]^s} Kf(\mathbf{u})d\mathbf{u} = K\mu$, et donc

$$\text{Var}[\hat{\mu}_n] \leq (K\mu - \mu^2)/n = \text{Var}[\tilde{\mu}_n].$$

Mais il se peut que $\tilde{\mu}_n$ soit moins coûteux à calculer.

Choix de la loi d'échantillonnage

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

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

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

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

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

Choix de la loi d'échantillonnage

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

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

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

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

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

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

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

Par exemple, g pourrait être une densité normale.

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

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

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

Exemple (simple): estimer une probabilité.

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

Exemple (simple): estimer une probabilité.

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

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

Exemple (simple): estimer une probabilité.

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

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

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

Répéter n fois et calculer la moyenne.

Exemple (simple): estimer une probabilité.

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

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

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

Répéter n fois et calculer la moyenne.

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

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

où

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

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

Exemple (simple): estimer une probabilité.

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

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

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

Répéter n fois et calculer la moyenne.

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

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

où

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

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

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

On a aussi

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

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

On a aussi

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

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

On a aussi

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

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

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

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

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

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

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

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

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

On peut écrire

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

où

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

car on a toujours $Y \geq y_0$.

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

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

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

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

On peut écrire

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

où

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

car on a toujours $Y \geq y_0$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Générer Y_1 selon sa densité conditionnelle à $Y_1 < a$,

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

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

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

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

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

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

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

Générer Y_1 selon sa densité conditionnelle à $Y_1 < a$,

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

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

La nouvelle densité de Y_1 est

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

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

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

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

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

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

On a

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

où

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

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

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

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

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

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

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

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

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

On a ensuite

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

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

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

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

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

On a ensuite

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

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

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

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

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

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

On a ensuite

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

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

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

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

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

Importance sampling: cas général

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

Importance sampling: cas général

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

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

Importance sampling: cas général

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

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

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

Importance sampling: cas général

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

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

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

La façon directe d'estimer μ est:

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

Importance sampling: cas général

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

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

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

La façon directe d'estimer μ est:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Même principe pour les lois discrètes.

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

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

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

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

Même principe pour les lois discrètes.

Pourquoi changer la densité de π à g ?

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

Choix optimal de g .

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

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

Choix optimal de g .

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

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

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

Choix optimal de g .

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

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

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

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

Choix optimal de g .

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

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

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

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

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

Choix optimal de g .

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

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

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

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

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

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

Cette densité réduit la variance à zero.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Mais calculer τ peut prendre un temps infini!

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

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

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

Mais calculer τ peut prendre un temps infini!

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

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

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

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

Mais calculer τ peut prendre un temps infini!

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

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

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

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

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

Mais calculer τ peut prendre un temps infini!

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

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

Solution: IS.

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

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

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

Mais calculer τ peut prendre un temps infini!

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

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

Solution: IS.

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

On verra plus tard que le changement de mesure suivant est approprié dans ce cas: ⁷⁰
remplacer la densité $h(x)$ par

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

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

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

où

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

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

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

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

où

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

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

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

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

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

où

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

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

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

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

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

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

où

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

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

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

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

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

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

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

où

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Dans certaines situations, on connaît π seulement à un facteur près.
Si on génère $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ i.i.d. selon g , un estimateur sans biais de μ serait:

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

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

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

quand $n \rightarrow \infty$.

Dans certaines situations, on connaît π seulement à un facteur près.
Si on génère $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ i.i.d. selon g , un estimateur sans biais de μ serait:

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

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

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

quand $n \rightarrow \infty$.

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

Méthodes quasi-Monte Carlo (QMC)

On veut estimer

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

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

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

Méthodes quasi-Monte Carlo (QMC)

On veut estimer

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

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

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

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

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

Méthodes quasi-Monte Carlo (QMC)

On veut estimer

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

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

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

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

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

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

Méthodes quasi-Monte Carlo (QMC)

On veut estimer

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

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

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

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

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

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

QMC classique: E_n est déterministe.

Questions:

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

Questions:

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

Comment construire P_n et mesurer son uniformité?

Questions:

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

Comment construire P_n et mesurer son uniformité?

Facile en 1 dimension. Mais en plusieurs dimensions?

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

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

Questions:

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

Comment construire P_n et mesurer son uniformité?

Facile en 1 dimension. Mais en plusieurs dimensions?

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

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

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

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

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

Questions:

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

Comment construire P_n et mesurer son uniformité?

Facile en 1 dimension. Mais en plusieurs dimensions?

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

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

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

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

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

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

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

Solutions évidentes:

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

ou encore

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

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

Solutions évidentes:

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

ou encore

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

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

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

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

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

Solutions évidentes:

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

ou encore

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

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

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

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

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

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

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

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

Solution simpliste en $s > 1$ dimensions: grille rectangulaire

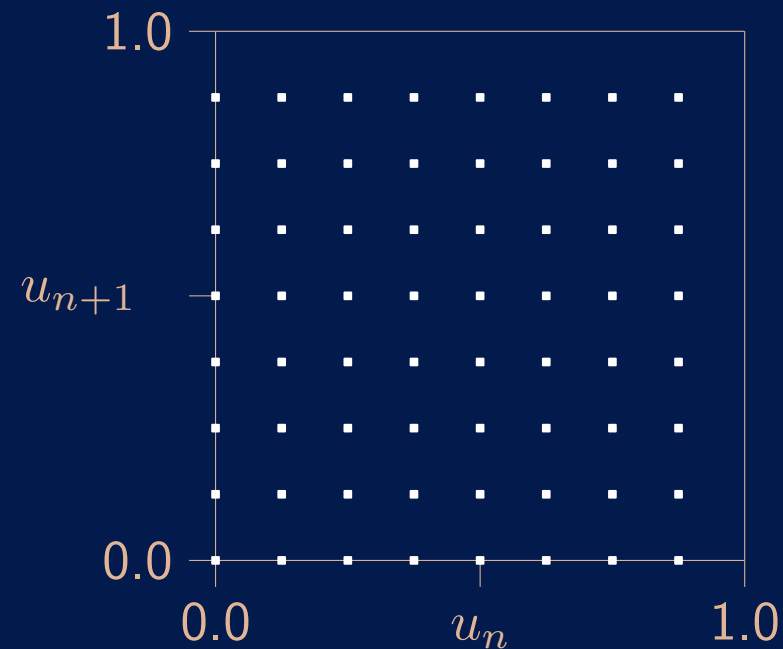
$$P_n = \{(i_1/d, \dots, i_t/d) \text{ tels que } 0 \leq i_j < d \ \forall j\},$$

où $n = d^s$.

Solution simpliste en $s > 1$ dimensions: grille rectangulaire

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

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



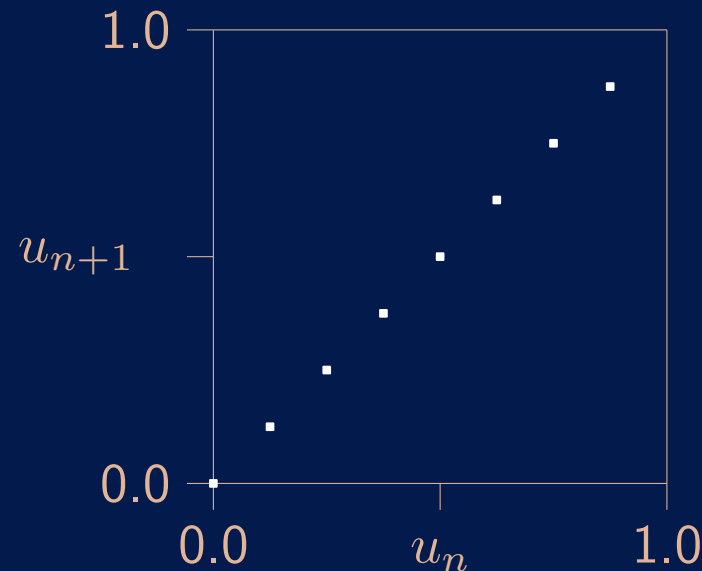
Idée: on voudrait construire P_n de manière à ce que chaque projection unidimensionnelle soit $\{0, 1/n, \dots, (n-1)/n\}$.

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

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

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

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

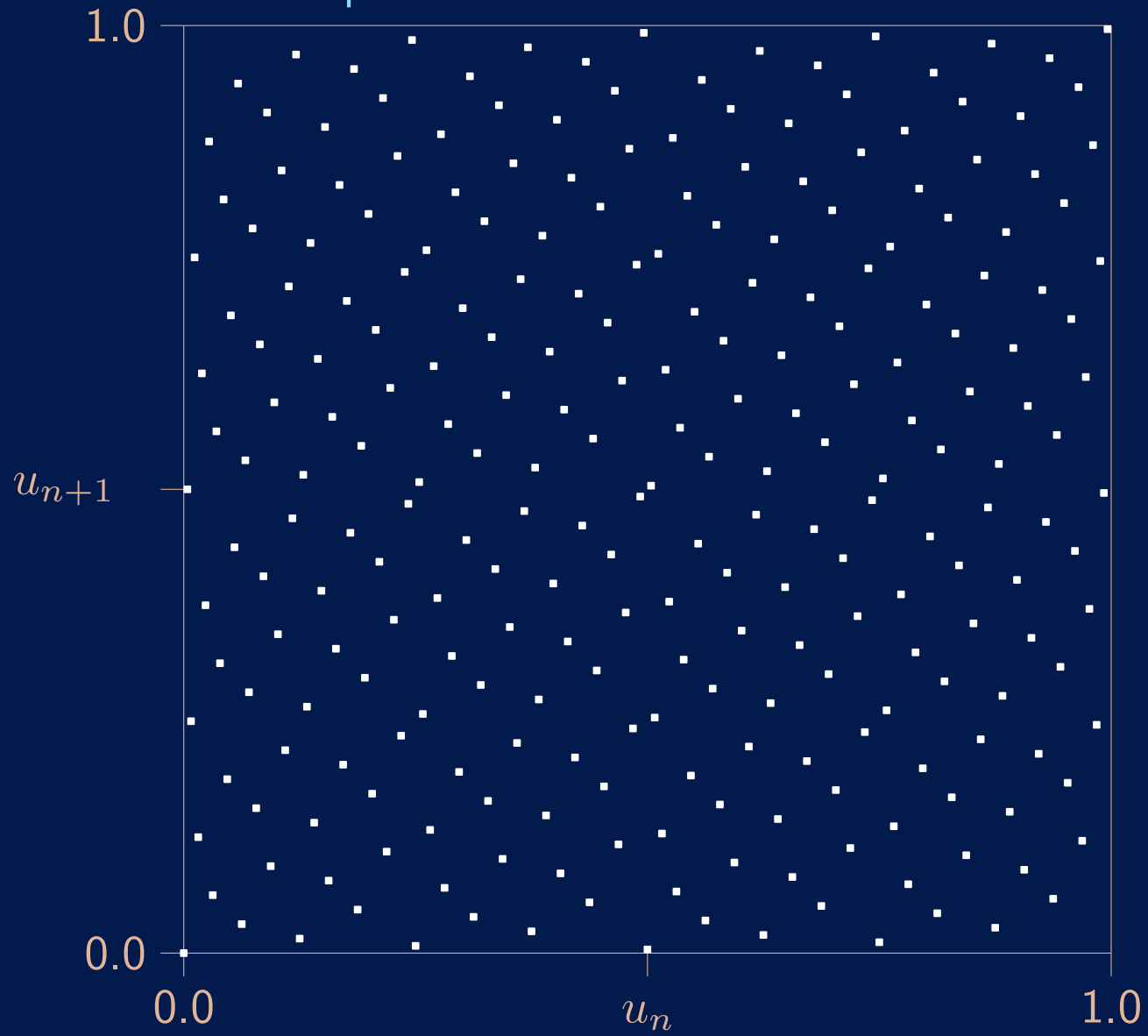


I.e., on cherche une permutation de \mathbb{Z}_n pour chacune des s coordonnées, et on veut les choisir pour que P_n soit très uniforme sur $[0, 1)^s$.

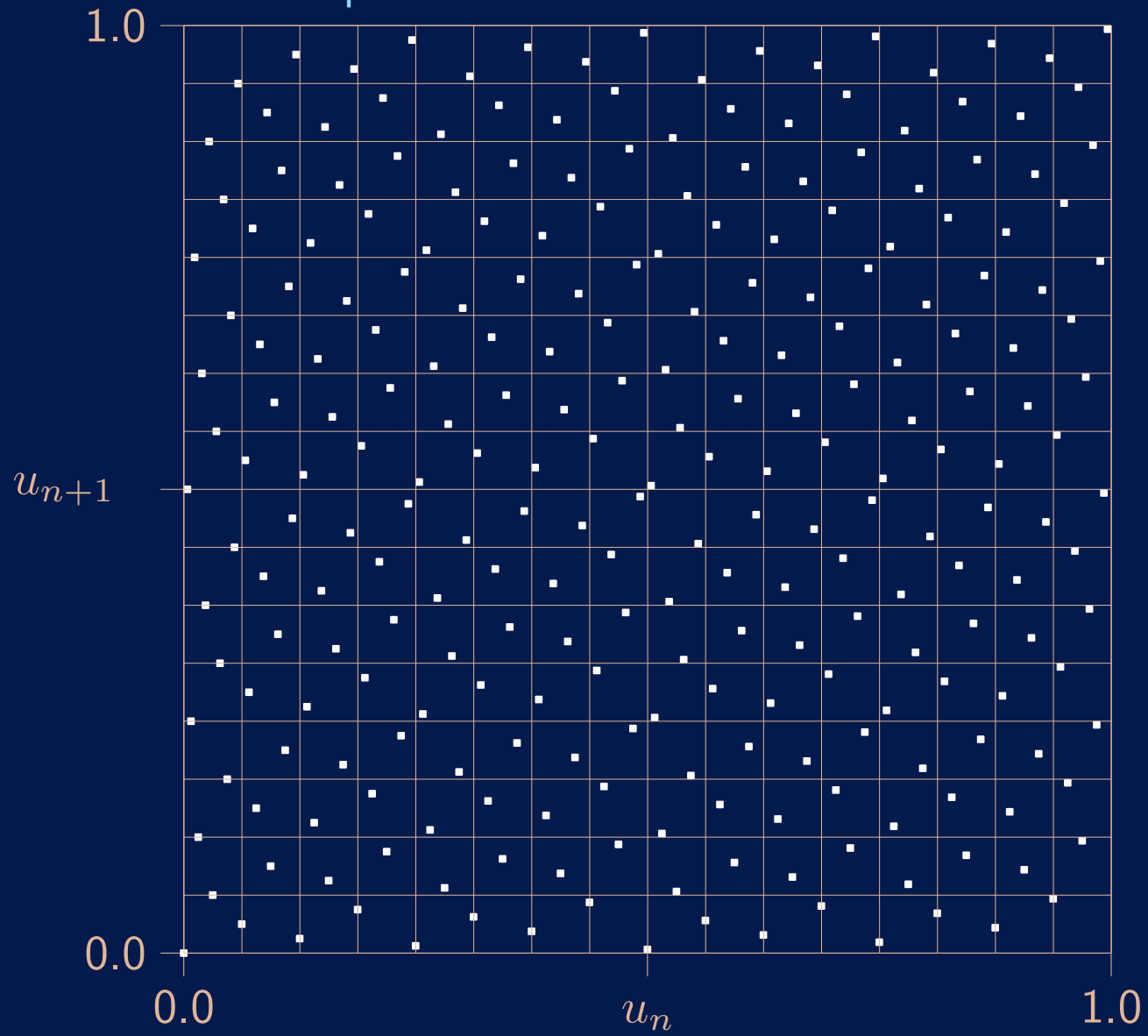
Exemple: Soient $n = 2^8 = 256$ et $s = 2$. Prenons les points (en binaire):

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
⋮	⋮	⋮
254	.11111110	.01111111
255	.11111111	.11111111

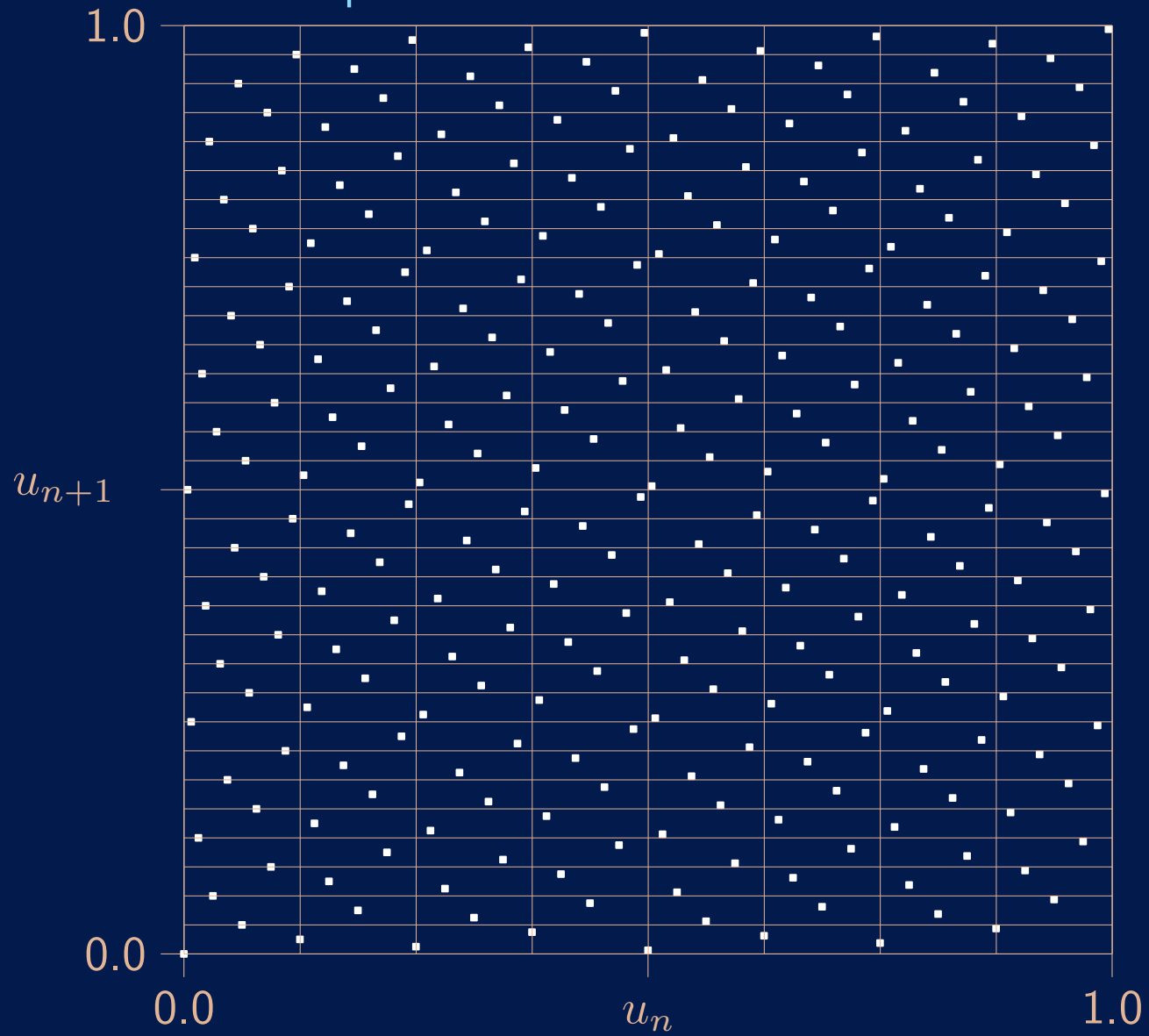
Example: $n = 2^8 = 256$ and $s = 2$.



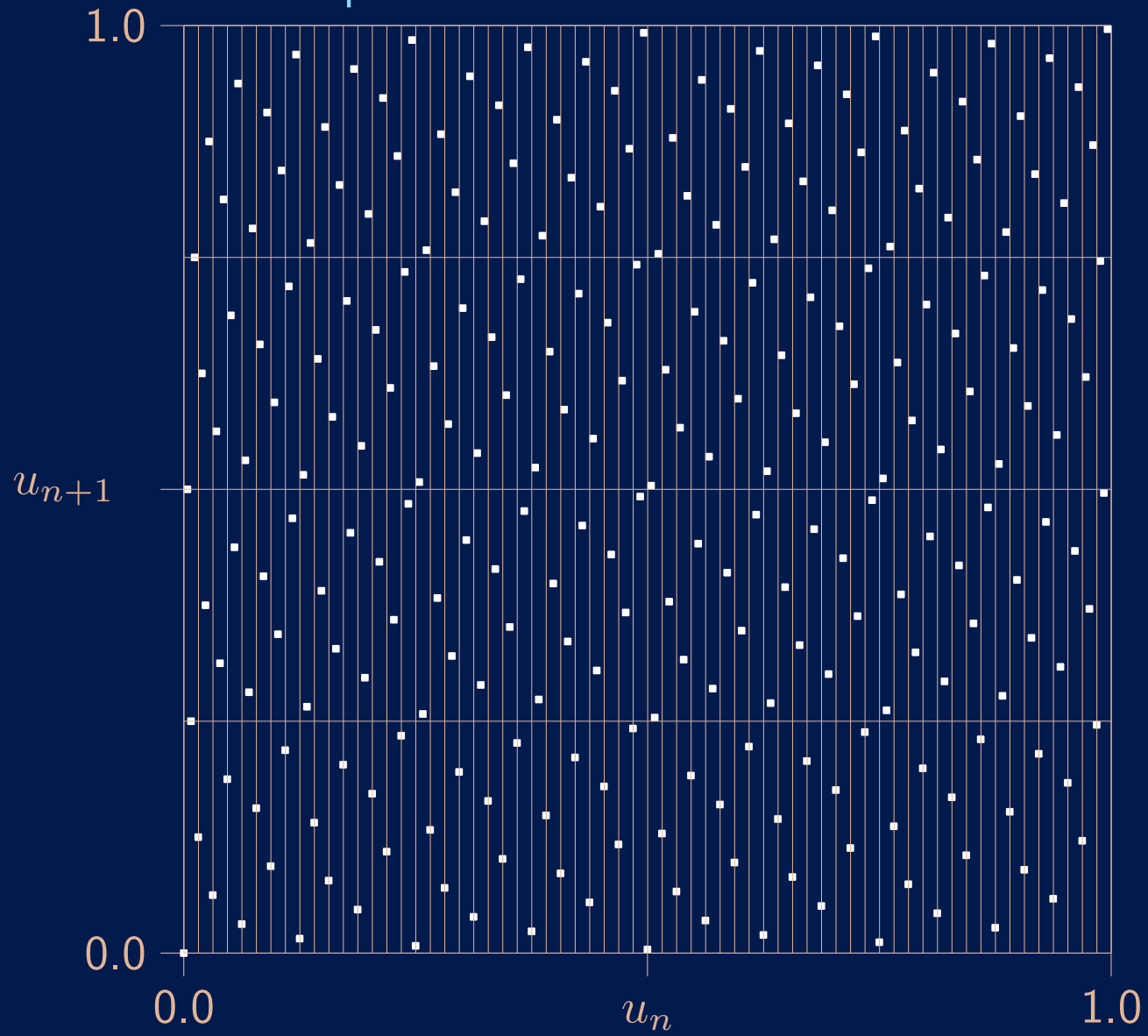
Example: $n = 2^8 = 256$ and $s = 2$.



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



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



Ces points forment un $(0, 8, 2)$ -réseau en base 2 (expliquer).

Ces points forment un $(0, 8, 2)$ -réseau en base 2 (expliquer).

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

Ces points forment un $(0, 8, 2)$ -réseau en base 2 (expliquer).

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 $(0, 8, 2)$ -réseau en base 2 (expliquer).

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.

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

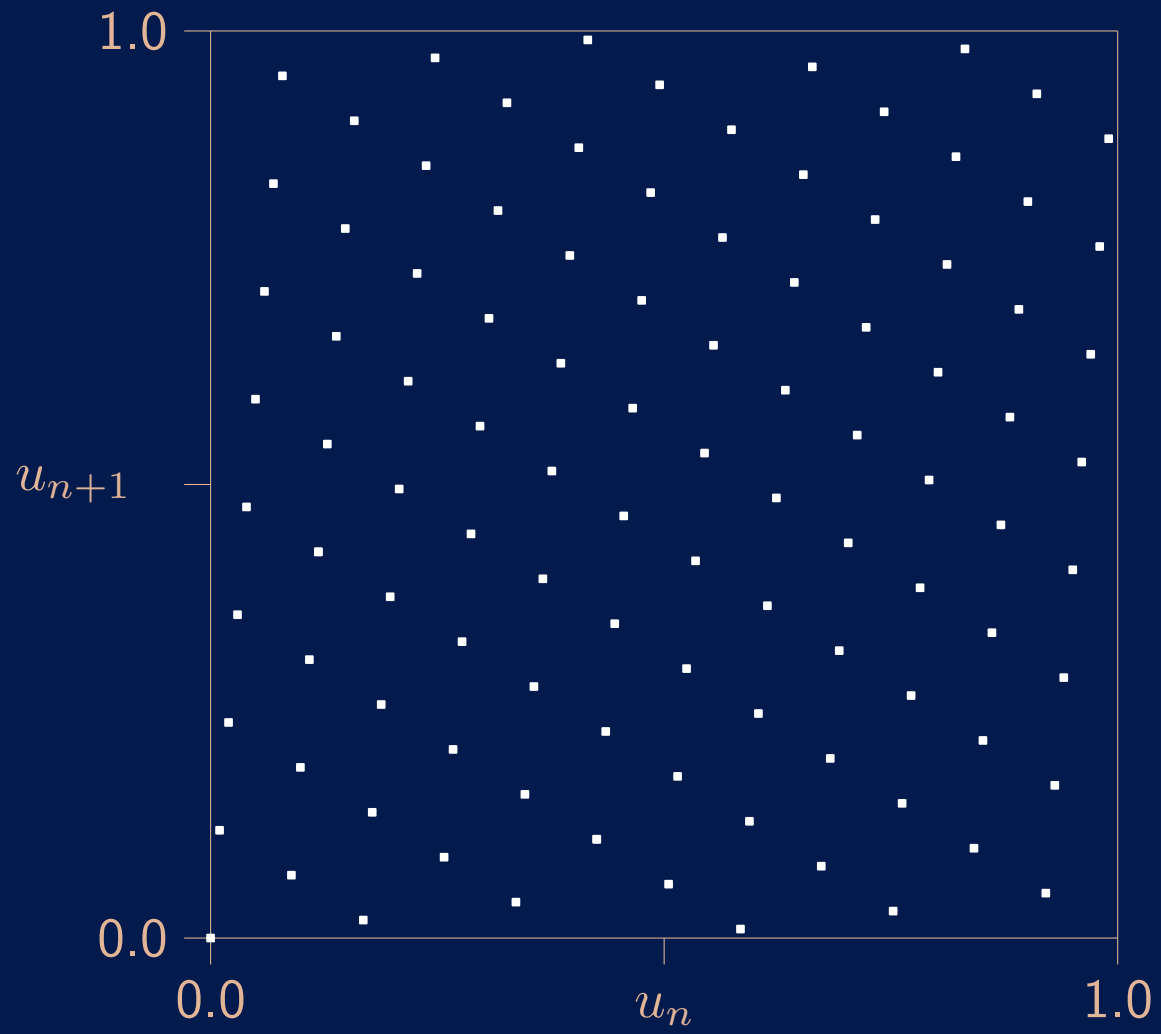
Ces points forment un $(0, 8, 2)$ -réseau en base 2 (expliquer).

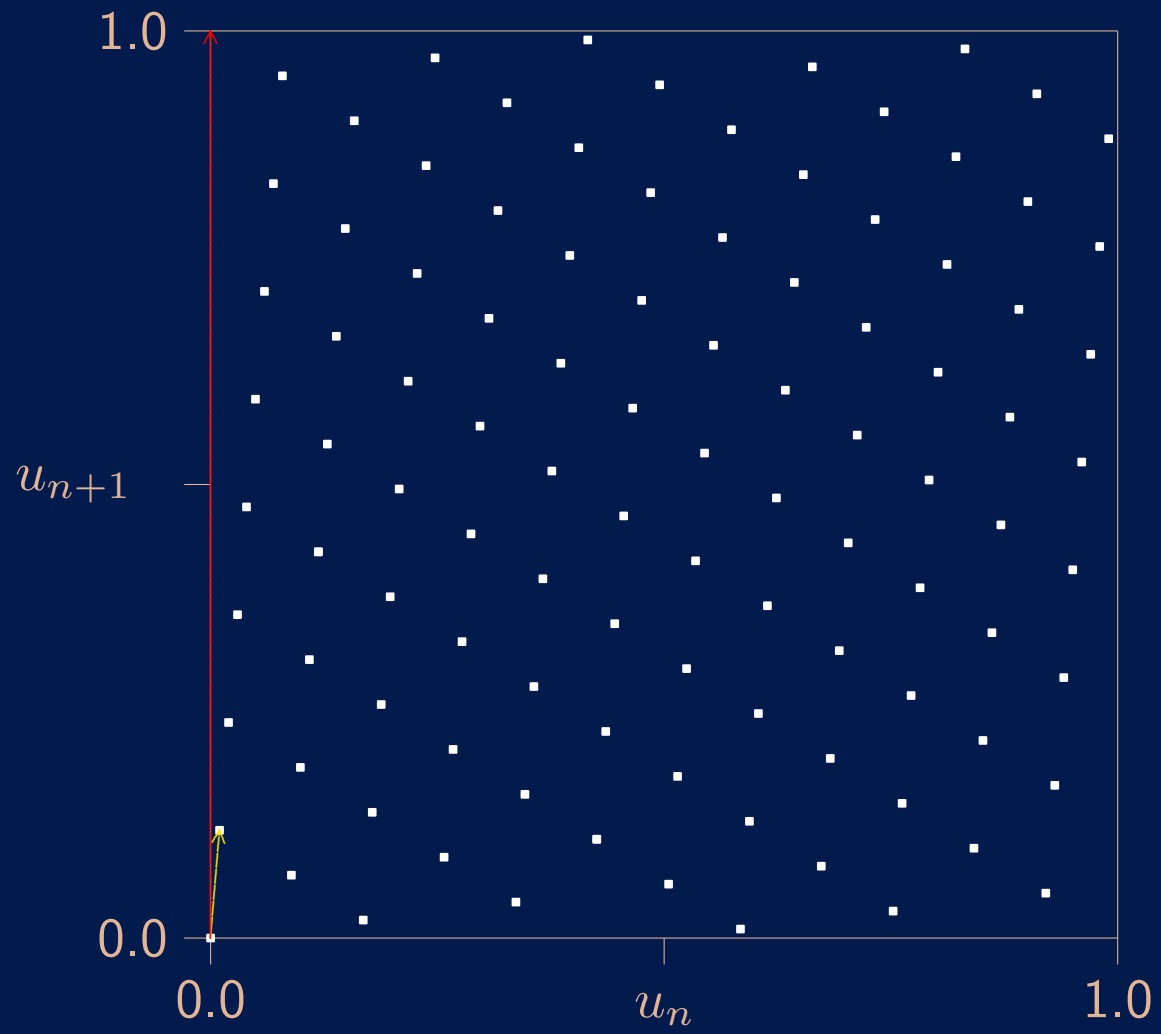
En général, on peut prendre $n = 2^k$ points.

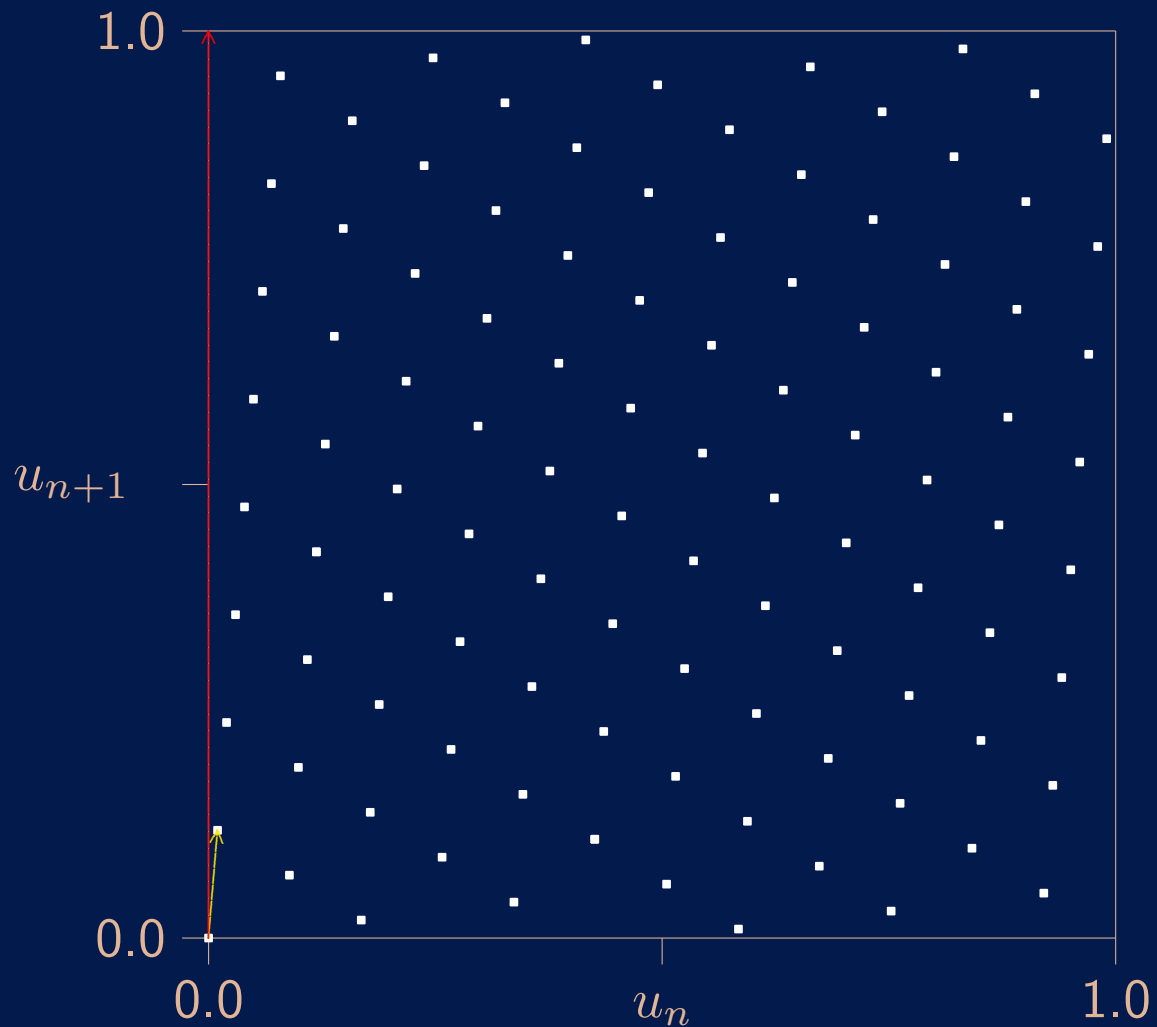
Si on partitionne $[0, 1)^2$ en rectangles de tailles 2^{-k_1} par 2^{-k_2} où $k_1 + k_2 \leq k$, chaque rectangle contiendra exactement le même nombre de points.

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

Autre exemple: "lattice" (réseau) avec $n = 101$.

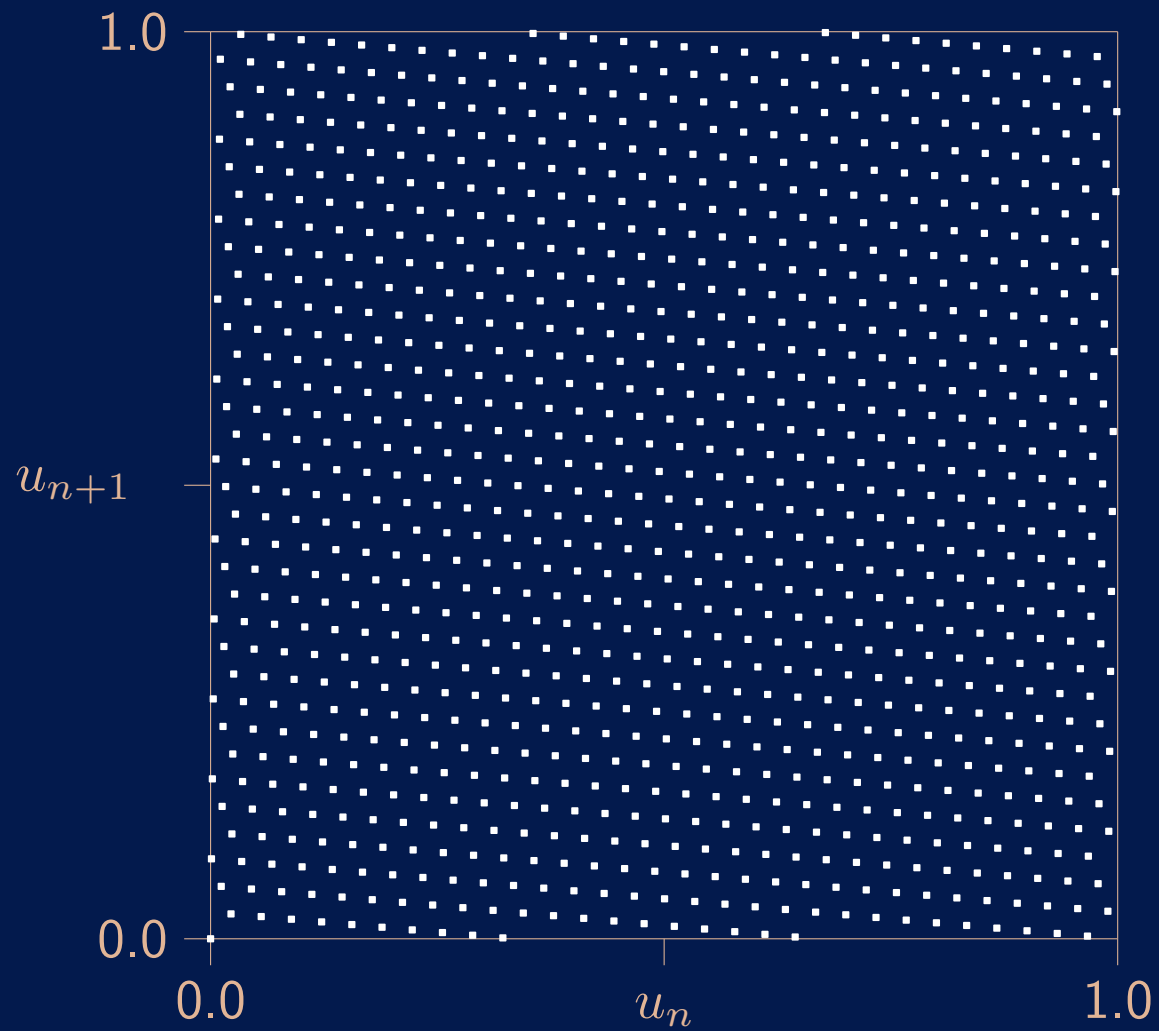






Exemple jouet: LCG avec $m = 101$ et $a = 12$.

Les u_i sont tous les multiples entiers de $(1/101, 12/101)$, modulo 1.



Exemple: LCG avec $s = 2$ et $n = 1021$:

Les réseaux digitaux (“digital nets”) et les règles de réseaux (“lattice rules”) se généralisent à un t arbitraire.

Les réseaux digitaux (“digital nets”) et les règles de réseaux (“lattice rules”) se généralisent à un t arbitraire.

Par exemple, on peut définir:

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

pour $\mathbf{v} = (1, v_1, \dots, v_{s-1})$ bien choisi.

Les réseaux digitaux (“digital nets”) et les règles de réseaux (“lattice rules”) se généralisent à un t arbitraire.

Par exemple, on peut définir:

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

pour $\mathbf{v} = (1, v_1, \dots, v_{s-1})$ bien choisi.

Pour un réseau digital en base 2, il suffit de choisir t permutations de $\{0, 1, \dots, 2^k - 1\}$.

Les réseaux digitaux (“digital nets”) et les règles de réseaux (“lattice rules”) se généralisent à un t arbitraire.

Par exemple, on peut définir:

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

pour $\mathbf{v} = (1, v_1, \dots, v_{s-1})$ bien choisi.

Pour un réseau digital en base 2, il suffit de choisir t permutations de $\{0, 1, \dots, 2^k - 1\}$.

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

Discrépance, convergence et borne de Koksma-Hlawka

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

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

Possible pour tout B ?

Discrépance, convergence et borne de Koksma-Hlawka

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

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

Possible pour tout B ? Non.

Discrédance, convergence et borne de Koksma-Hlawka

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

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

Possible pour tout B ? **Non.**

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

Discrépance, convergence et borne de Koksma-Hlawka

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

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

Possible pour tout B ? Non.

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

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

Discrédance, convergence et borne de Koksma-Hlawka

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

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

Possible pour tout B ? Non.

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

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

Mesure de discrédance pour P_n :

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

Discr ance, convergence et borne de Koksma-Hlawka

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

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

Possible pour tout B ? Non.

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

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

Mesure de discr ance pour P_n :

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

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

Inégalité de Koksma-Hlawka (erreur pire cas):

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

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

Inégalité de Koksma-Hlawka (erreur pire cas):

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

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

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

Inégalité de Koksma-Hlawka (erreur pire cas):

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

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

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

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

Inégalité de Koksma-Hlawka (erreur pire cas):

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

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

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

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

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

Estimation de l'erreur

L'idée est de randomiser P_n de manière à :

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

Estimation de l'erreur

L'idée est de randomiser P_n de manière à:

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

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

Estimation de l'erreur

L'idée est de randomiser P_n de manière à:

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

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

On répète la randomisation m fois, indépendamment, et on calcule la moyenne \bar{X}_m et la variance S_m^2 des m valeurs de $\bar{\mu}_n$.

On peut montrer que $\mathbb{E}[\bar{X}_m] = \mu$ et $\mathbb{E}[S_m^2] = \text{Var}[\bar{\mu}_n] = m\text{Var}[\bar{X}_m]$.

Estimation de l'erreur

L'idée est de randomiser P_n de manière à:

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

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

On répète la randomisation m fois, indépendamment, et on calcule la moyenne \bar{X}_m et la variance S_m^2 des m valeurs de $\bar{\mu}_n$.

On peut montrer que $\mathbb{E}[\bar{X}_m] = \mu$ et $\mathbb{E}[S_m^2] = \text{Var}[\bar{\mu}_n] = m\text{Var}[\bar{X}_m]$.

Permet de calculer un intervalle de confiance pour μ .

Choix de m ? Souvent autour de 10.