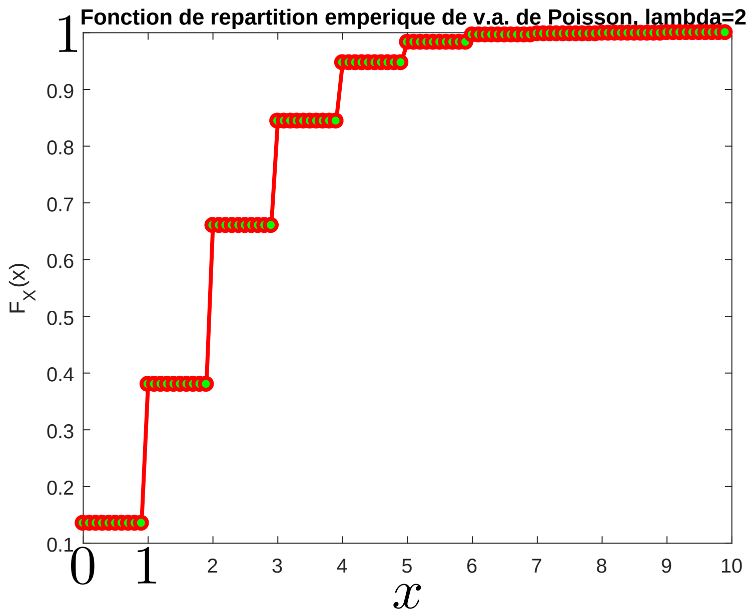
Markov Chain Monte Carlo

TP 1: Simulation de variables aléatoires discrètes



Théorie

- **Loi forte des grands nombres**

  Soient $X_1, X_2, ..., X_n$ des variables aléatoires de même loi. On note $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Alors

$$\lim_{n \to \infty} \overline{X}_n = \mathbb{E}(X) \quad \textit{presque} \quad \textit{surement}$$

- **Théorème Central Limite**

  Soient $X_1, X_2, ..., X_n$ des variables aléatoires indépendantes et identiquement distribuées d'espérance μ et de variance σ^2.
  Alors la loi de

$$Y_n = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}$$

  tend vers la loi normale centrée réduite. En d'autre termes, pour tous a et b réels,

$$\lim_{n \to \infty} \mathbb{P}[a \leq \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \leq b] = \mathbb{P}[a \leq Z \leq b]$$

  ou Z est une variable gaussienne centrée réduite, $Z \sim \mathcal{N}(0, 1)$.

Intervalle de confiance

- On choisi $a = -1.96, b = 1.96$. Donc

$$\lim_{n \to \infty} \mathbb{P}[|\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}| \leq 1.96] = 0.95$$

- Un échantillon de n valeurs de v.a. X_i est utilisé pour estimer l'espérance inconnue $\mathbb{E}(X) = \mu$.
  Pour grand n la probabilité que μ se trouve dans l'intervalle de confiance:

$$\overline{X}_n - \frac{1.96 \cdot \sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + \frac{1.96 \cdot \sigma}{\sqrt{n}}$$

  est égale à 0.95.

- Pour grand n ce résultat reste vrai quand σ est remplacé par l'quart-type de variance empirique $\sqrt{\mathbb{V}ar_{emp}(X)} = (\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2)^{\frac{1}{2}}$:

$$\overline{X}_n - \frac{1.96\sqrt{\mathbb{V}ar_{emp}(X)}}{\sqrt{n}} \leq \mu \leq \overline{X}_n + \frac{1.96\sqrt{\mathbb{V}ar_{emp}(X)}}{\sqrt{n}}$$

Simulation d'une v.a. de Bernoulli X

- V.a. de Bernoulli X modélise des succès et des échecs:

$$X = \begin{cases} 1\ (succes) & avec\ \ proba\ \ p_1 = 1/3 \\ 0\ (echec) & avec\ \ proba\ \ p_2 = 2/3 \end{cases}$$

- Simuler une v.a. X signifie simuler un échantillon des valeurs (réalisations indépendantes) de X.
- Comment simuler une chaîne de valeurs (N_{mc} valeurs) 0 et 1 telle que les fréquences d'apparition de 1 et de 0 sont respectivement 1/3 et 2/3?
- Soit une probabilité $0 \leq p \leq 1$
  IDEE PRINCIPALE $\mathbb{P}[rand() \leq p] = p$
  C'est la définition de la v.a. uniforme U par sa fonction de répartition: $F_U(p) = p$

Simulation d'une v.a. de Bernoulli X

- Intuitivement:
  - for $n = 1 : N_{mc}$
  - U= rand()
  - if U <1/3

    $X(n) = 1$

    else $X(n) = 0$
  - end if
  - end for
- Display(X)

- L'autre possibilité
  - U= rand(N_{mc})
  - $X = (U < 1/3)$

Vérification de la simulation. Convergence.

- On vérifie que les probabilités des des succès et des échecs dans la chaîne sont en effet 1/3 et 2/3.

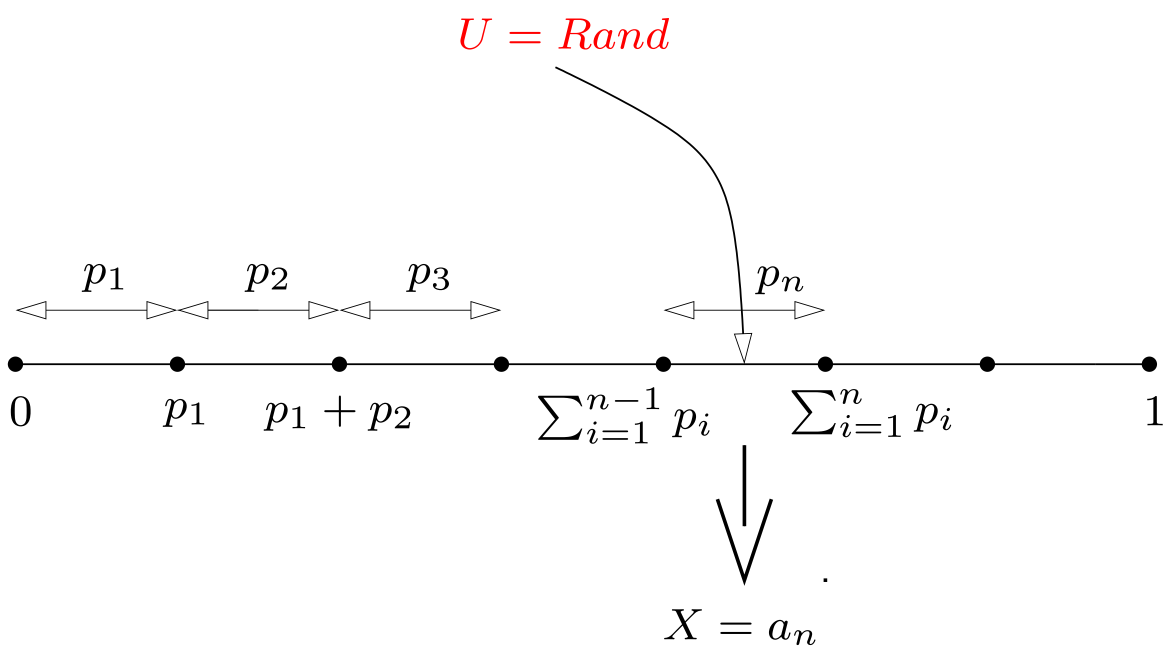
- • function[proba1, proba2]=Freq(N_{mc})
  ○ counter1=0; counter2=0
  • for $n = 1 : N_{mc}$
  • U= rand()
  •   if U <1/3
  ○ $X(n) = 1$
  ○   counter1=counter1+1
  ○   else $X(n) = 0$
  ○ counter2=counter2+1
  • end if
  • end for
  • proba1=counter1/N_{mc}
  • proba2=counter2/N_{mc}
  • endfunction

- • for $N_{mc} = 1 : 200$
  • [proba1(N_{mc}),proba2(N_{mc})]= Freq(N_{mc})
  • end for
  • plot(proba1,'r')
  • plot(proba2,'g')

Simulation d'une v.a. discrete X

- Variable aléatoire X est définie par un tableau:

$\mathbb{P}$	$p_1 = 0.25$	$p_2 = 0.3$	$p_3 = 0.1$	$p_4 = 0.18$	$p_5 = 0.05$	$p_6 = 0.08$	$p_7 = 0.02$	$p_8 = 0.02$
X	$a_1 = 1$	$a_2 = 2$	$a_3 = 3$	$a_4 = 4$	$a_5 = 5$	$a_6 = 6$	$a_7 = 7$	$a_8 = 8$



- Idée de simulation: on cherche dans quel intervalle tombe une v.a. uniforme U.

Inversion de la fonction de répartition.

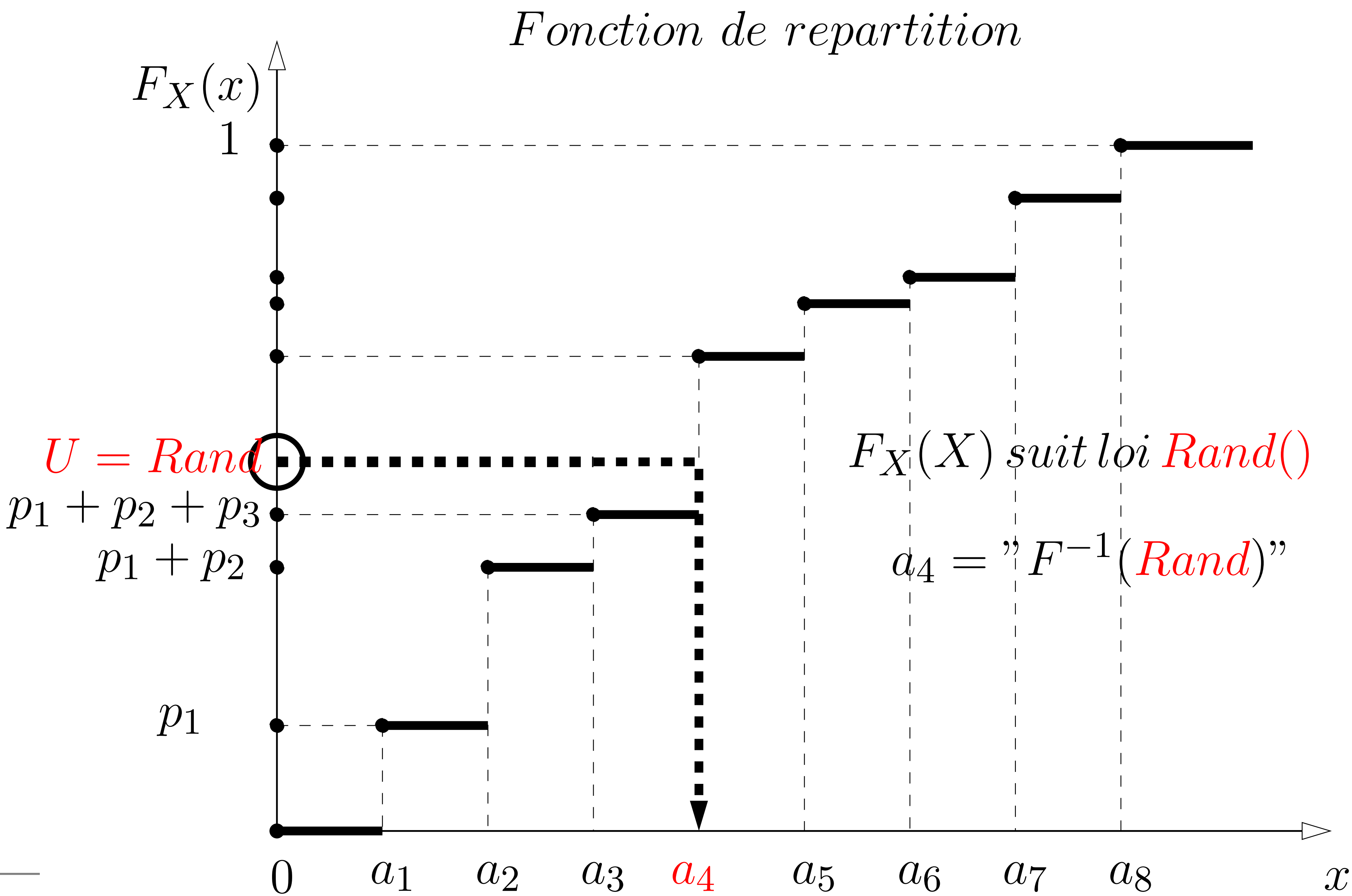
- $$\sum_{i=1}^{n-1} p_i < Rand(\,) \leq \sum_{i=1}^{n} p_i \quad \Rightarrow X = a_n$$

- $$F_X(a_{n-1}) < Rand(\,) \leq F_X(a_n) \quad \Rightarrow X = a_n$$

- $$F_X(X) \quad suit \quad la \quad loi \quad Rand(\,)$$

- $$F_X^{-1}(Rand(\,)) = X$$

Simulation d'une v.a. discrete X



Simulation d’une v.a. discrete X

- Algorithme 1A: Simulation d’une réalisation de variable aléatoire discrete
  ◦ Définir le vecteur de probabilité $p(n)$
  ◦ Définir le vecteur a des réalisations $a(n)$ de la variable aléatoire X.
  • function[X]=V_A_Discrete(p, a)
  ◦ Set $n = 1, \quad F = p(1)$
  ◦ Générer U % U=rand()
  ◦ While $U > F$
  ◦ Set $F = F + p(n+1), \quad n = n+1$
  ◦ end while
  ◦ Set $X = a(n)$
  • endfunction
- On applique l’Algorithme 1A N_{mc} fois pour générer N_{mc} réalisations de X

Algo 1B: Simulation d’un échantillon

- On répète l’algorithme 1A N_{mc} fois pour simuler une chaîne de N_{mc} valeurs.
- Algorithme 1B de simulation d’une chaîne de valeurs de v.a. discrete.

  - function [X]= Chaine_valeurs_V_A_Discrete (p, a)
    - for $n = 1 : N_{mc}$
    - $X(n)$=V_A_Discrete(p, a)
    - end
  - endfunction
- Simuler et visualiser une chaîne de réalisations de X.

Verification de simulation.

- On vérifie que la loi simulée est bien la loi demandée.
  Il existe trois tests de validation.
- Test1   On compare l'espérance d'un échantillon simulé avec l'espérance théorique.
- Test2   On compare la variance d'un échantillon simulé avec la variance théorique.
- Test3   On construit la fonction de repartition et la fonction de densité d'un échantillon simulé (les fonctions de repartition et de densité empiriques) et on les compare avec les fonctions de repartition et de densité analytiques da la loi demandée.

Verification de simulation.

Espérance et variance

- Calcul théorique pour l'espérance et la variance.
  - • Calculer l'espérance théorique : $\mathbb{E}(X) = \sum_{i=1}^{8} p_i a_i$
  - • Calculer la variance théorique $\mathbb{V}ar(X) = \sum_{i=1}^{8} p_i a_i^2 - \mathbb{E}(X)^2$
- Travail à faire. Calcul empirique. Soit $N_{mc} = 100$
  - • Calculer l'espérance empirique : $\mathbb{E}_{emp}(X) = \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} X_i$
  - • Calculer la variance empirique :

    $\mathbb{V}ar_{emp}(X) = \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} (X_i)^2 - (\frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} X_i)^2$
- Est-ce que le résultat est coherent avec la loi des Grands Nombres?

$$\lim_{N_{mc}\to\infty} \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} X_i = \mathbb{E}(X)?$$

$$\lim_{N_{mc}\to\infty} \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} (X_i)^2 - (\frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} X_i)^2 = \mathbb{V}ar(X)?$$

- Faites les mêmes calculs pour $N_{mc} = 1000$, $N_{mc} = 100000$. Conclure.

Simulation de v.a. de Poisson,1

- Soit X est v.a. de Poisson $\quad \mathbb{P}(X = n) = \frac{e^{-\lambda}\lambda^n}{n!}.$
- Montrer $\mathbb{P}(X = n + 1) = \frac{\lambda}{(n+1)} \cdot \mathbb{P}(X = n)$
- Algorithme 2A de simulation de v.a. de Poisson
  - function[X]=V_A_Poisson(λ)
    - Set $n = 0, \quad proba = e^{-\lambda}, \quad F = proba;$
    - Générer U
    - While $U > F$
    - Set $proba = \frac{\lambda}{(n+1)} \cdot proba, \quad F = F + proba, \quad n = n + 1$
    - end
    - Set $X = n$
    - end
  - endfunction

Simulation de v.a. de Poisson,2

- Algorithme 2B de simulation de N_{mc} valeurs de v.a. de Poisson
  - function [X]= Chaine_valeurs_V_A_Poisson (λ)
  - ∘ for $n = 1 : N_{mc}$
  - ∘ $X(n)$=V_A_Poisson(λ)
  - ∘ end
  - endfunction
- Visualiser un échantillon simulé: $N_{mc} = 100$ terms, $\lambda = 2$
- Calcul théorique.
  - L'espérance théorique : $\mathbb{E}(X) = \sum_{n=1}^{\infty} n \cdot \frac{e^{-\lambda}\lambda^n}{n!} = \lambda$
  - La variance théorique $\mathbb{V}ar(X) = \sum_{n=1}^{\infty} n^2 \cdot \frac{e^{-\lambda}\lambda^n}{n!} = \lambda$
- Travail à faire. Calcul empirique. Soit $N_{mc} = 10000$, $\lambda = 2$
  - Calculer l'espérance empirique : $\mathbb{E}_{empirique}(X) = \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} X_i$ et comparer avec λ
  - Calculer la variance empirique : $\mathbb{V}ar_{empirique}(X) = \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} (X_i)^2 - (\frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} X_i)^2$ et comparer avec λ.
- Conclure.

Verification de simulation.

Fonction de repartition.

- Il y a un échantillon de N_{mc} v.a. de Poisson. Nous utilisons la définition de la fonction de répartition $F_X(x) = \mathbb{P}[X \leq x], \quad x \in R$ est une variable classique,
- On choisit un intervalle $[a, b]$ sur lequel on veut tracer la fonction de repartition, on discrétise cet intervalle en N_x parties :

$$\Delta x = \frac{b-a}{N_x}, \quad x_i = a + \Delta x \cdot i, \quad 0 \leq i \leq N_x$$

On définit $F_X(x)$ en chaque point discrete x_i:

$$F_X(x_i) = \mathbb{P}[X \leq x_i].$$

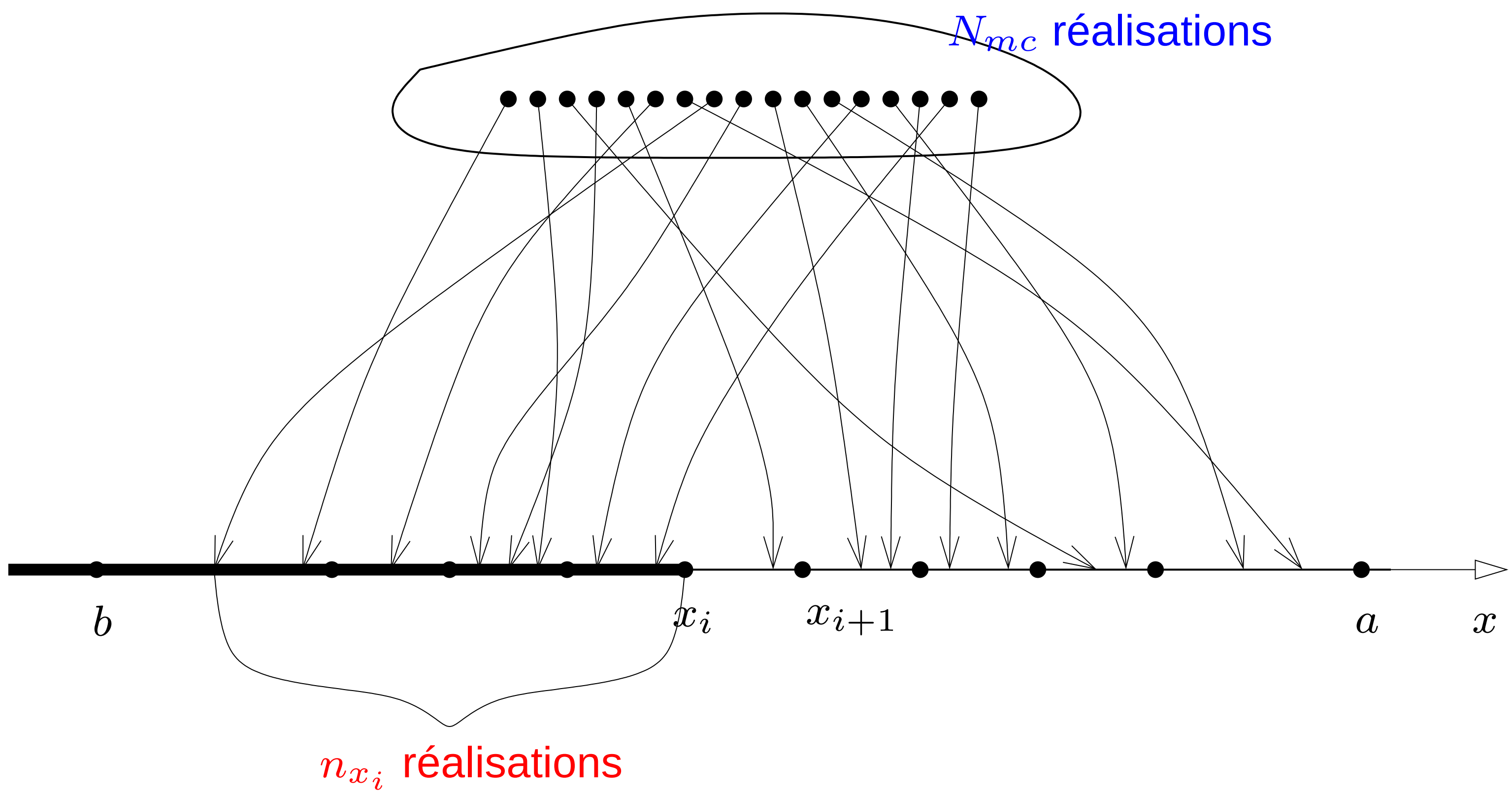
- On définit la fonction empirique de répartition $F_X(x)$ en chaque point discrete x_i:

$$F_X(x_i) = \mathbb{P}[X \leq x_i] = \mathbb{E}[1_{X \leq x_i}] = \frac{1}{N_{mc}} \sum_{n=1}^{N_{mc}} 1_{X_n \leq x_i}$$

Verification de simulation.

Fonctions de repartition.

- On calcule le nombre de réalisation n_{x_i} de v.a. X qui tombe dans l'intervalle $]-\infty, x_i]$, puis on calcule la probabilité: $\mathbb{P}[X \leq x_i] = \frac{n_{x_i}}{N_{mc}}$.
- On répète cette procedure pour chaque intervalle $]-\infty, x_i], \quad i = 1 : N_x$.



Verification de simulation.

- Algorithme de simulation de la fonction de repartition

• Simuler un échantillon de N_{mc} réalisations de v.a. X et utiliser le vecteur X comme l'entrée.

• function [ ]= Repartition_Poisson (X)

○ Définir a, Δx et N_x

○ for $i = 1 : N_x + 1$ % Par une boucle définir les points x_i d'un vecteur x

○ $x(i) = a + \Delta x \cdot (i - 1)$

○ counter =0 % Initialisation du Compteur

○ for $n = 1 : N_{mc}$

○ if $X_n \leq x_i$

○ counter = counter +1

○ end if

○ end for

Proba(i)= $\frac{Counter}{N_{mc}}$ % Calcul la probabilité empirique

○ end for

○ plot(x, Proba) % Tracer le graphe de la fonction de repartition

• endfunction

Verification de simulation.

Fonction de la densité.

- Nous utilisons la définition de la fonction de densité $f_X(x) = \frac{\mathbb{P}[x < X \leq x + \Delta x]}{\Delta x}$
- On choisit l'intervalle $[a, b]$, on discrétise cet intervalle en N_x parties :

$$\Delta x = \frac{b-a}{N_x}, \quad x_i = a + \Delta x \cdot i, \quad 0 \leq i \leq N_x$$

On définit $f_X(x)$ en chaque point discrete x_i:

$$f_X(x_i) = \frac{\mathbb{P}[x_i < X \leq x_i + \Delta x]}{\Delta x}.$$

On utilise l'échantillon de v.a. X, on fixe x_i, on calcule le nombre de réalisation $n_{x_i, x_i + \Delta x}$ qui tombe dans l'intervalle $[x_i, x_i + \Delta x]$, puis on calcule la probabilité:

$$\mathbb{P}[x_i < X \leq x_i + \Delta x] = \mathbb{E}[1_{x_i < X \leq x_i + \Delta x}] = \frac{1}{N_{mc}} \sum_{n=1}^{N_{mc}} 1_{x_i < X_n \leq x_i + \Delta x} = \frac{n_{x_i, x_i + \Delta x}}{N_{mc}}.$$

Cette procedure on répète pour chaque intervalle $[x_i, x_i + \Delta x], \quad i = 1 : N_x$.

Verification de simulation.

- Algorithme de simulation de fonction de densité

Pour un paramètre d'entré on utilise un échantillon de N_{mc} réalisations de v.a. X.

• function[Densite]= Densite_empirique(X)

∘ for $i = 1 : N_x$ % Par une boucle définir les points x_i

∘ $x(i) = a + \Delta x \cdot (i - 1)$

∘ counter=0 %Initialisation du Compteur

∘ for $n = 1 : N_{mc}$

∘ if $X(n) \in [x(i), x(i) + \Delta x]$

∘ counter = counter +1

∘ end if

∘ end for

∘ Proba(i)= $\frac{counter}{N_{mc}}$ % probabilité empirique

∘ Densité(i)=Proba(i)/Δx % Δx=1 si la v. a. est discrete

∘ end for

∘ plot(x, Densité) % Tracer le graphe de la fonction de densité

• endfunction

Travail à faire pour Poisson

- Soient $\lambda = 2, N_{mc} = 10000$
  - Simuler la v.a. de Poisson
  - Tracer sa fonction de repartition: $[a, b] = [0, 10], \Delta = 0.1, N_x = 100$
  - Tracer sa fonction de densité: $[a, b] = [0, 10], \Delta = 0.1, N_x = 100$
- Soient $\lambda = 4, N_{mc} = 10000$
  - Simuler la v.a. de Poisson
  - Tracer sa fonction de repartition: $[a, b] = [0, 15], \Delta = 0.15, N_x = 100$
  - Tracer sa fonction de densité: $[a, b] = [0, 15], \Delta = 1, N_x = 15$
- Soient $\lambda = 50, N_{mc} = 10000$
  - Simuler la v.a. de Poisson
  - Tracer sa fonction de repartition: $[a, b] = [0, 200], \Delta = 2, N_x = 100$
  - Tracer sa fonction de densité: $[a, b] = [0, 200], \Delta = 1, N_x = 200$

Travail à faire pour v.a. Binomial

- Soient $N = 20, p = 0.5, N_{mc} = 10000$
- On définit la v.a. Binomiale X par la somme de N v.a.indépendantes de Bernoulli Y_i:

$$X = \sum_{i=1}^{N} Y_i, \qquad Y_i = \begin{cases} 1 \quad avec \quad proba \quad p \\ 0 \quad avec \quad proba \quad 1-p \end{cases}$$

• p est la probabilité d'un succès et les valeurs de X représente le nombre de succès parmi N essais

• Simuler la v.a. de Binomiale à partir de la définition à l'aide de v.a. de Bernoulli.

• Tracer sa fonction de repartition

• Tracer sa fonction de densité

Simulation 2 de v.a. Binomiale

- La v.a. de loi binomiale $Bin(N, p)$ s'écrit comme une somme de N v. a. de Bernoulli $B(p)$ indépendantes et représente k succès parmi N essais: $\mathbb{P}(X = k) = C_k^N p^k (1-p)^{(N-k)}$
- Monter $\mathbb{P}(X = k+1) = \frac{p(N-k)}{(1-p)(k+1)} \cdot \mathbb{P}(X = k)$

  • Algorithme 4 de simulation d'une seule variable aléatoire Binomiale.

  • function[X]=V_A_Binomiale(p, N)

  ◦ Set $k = 0, \quad proba = (1-p)^N, \quad F = proba;$

  ◦ Générer U

  ◦ While $U > F$

  ◦ Set $proba = \frac{p(N-k)}{(1-p)(k+1)} \cdot proba, \quad F = F + proba, \quad k = k+1$

  ◦ end

  ◦ Set $X = k$

  ◦ end

  • endfunction

Travail à faire pour Binomial

- Soient $N = 50, p = 0.2, N_{mc} = 10000$

• Simuler la v.a. Binomial par l'algorithme 4

• Calculer l'espérance et la variance empiriques, les comparer avec les valeurs théoriques

$\mathbb{E}(X) = Np, \quad \mathbb{V}ar(X) = Np(1-p)$

• Tracer sa fonction de repartition:

$[a, b] = [0, 50], \Delta = 0.5, N_x = 100$

• Tracer sa fonction de densité:

$[a, b] = [0, 20], \Delta = 1, N_x = 20$

Simulation de v.a. Géométrique

- Soit X v.a. géométrique avec le paramètre p telle que $\mathbb{P}(X = n) = (1-p)^{n-1} \cdot p$.
  V. a. géométrique représente le nombre d'essais avant d'arrivé d'un premiere succès. La probabilité d'un succès est p.
  • Simuler la v.a. Géométrique par l'algorithme 4 avec $p = 0.2$
  • Tracer sa fonction de densité
  • Simuler la v.a. Géométrique à l'aide de v.a. de Bernoulli.
  Simuler plusieurs fois la v.a. de Bernoulli. Compter le nombre des échecs ($X = 0$) avant d'arrivé d'un premiere succès ($X = 1$). Identifier une valeur de v.a. Géométrique à celle du compteur
  • Tracer la fonction de densité de v.a. Géométrique