Chapitre 3
Génération de variables aléatoires

2025-2026

Niveau : L2-INFO

hajmessaoud.nada@gmail.com

Dr. Nada Haj Messaoud

Simulation à événements discrets

- Les paramètres d'entrée tels que les temps inter-arrivées (le temps entre les événements) et les temps de service (le temps nécessaire pour qu'un service soit complété) sont souvent modélisés à l'aide de variables aléatoires.

- Ces variables aléatoires sont associées à certaines distributions de probabilité, selon le système modélisé.

- Cela peut être réalisé à l'aide d'une séquence de nombres aléatoires qui sont indépendants les uns des autres et uniformément distribués entre 0 et 1.

Simulation à événements discrets

- **Variable de Bernoulli (Bernoulli Variate)**

  - C'est une expérience aléatoire qui n'a que deux résultats possibles : succès (1) ou échec (0).
    - Exemple : la transmission réussie ou non d'un paquet dans un réseau Wi-Fi
    - Exemple : pile ou face avec une pièce
  - Elle représente tous les phénomènes binaires (vrai/faux, oui/non, défaut/bon, etc.), et son paramètre p est simplement la probabilité de succès.

- Bernoulli trial (with parameter p)
  $p(1) = p, p(0) = 1 - p$

- Random variate generation
  Generate u
  If $0 < u \leq p$, $x = 1$;
  Otherwise $x = 0$

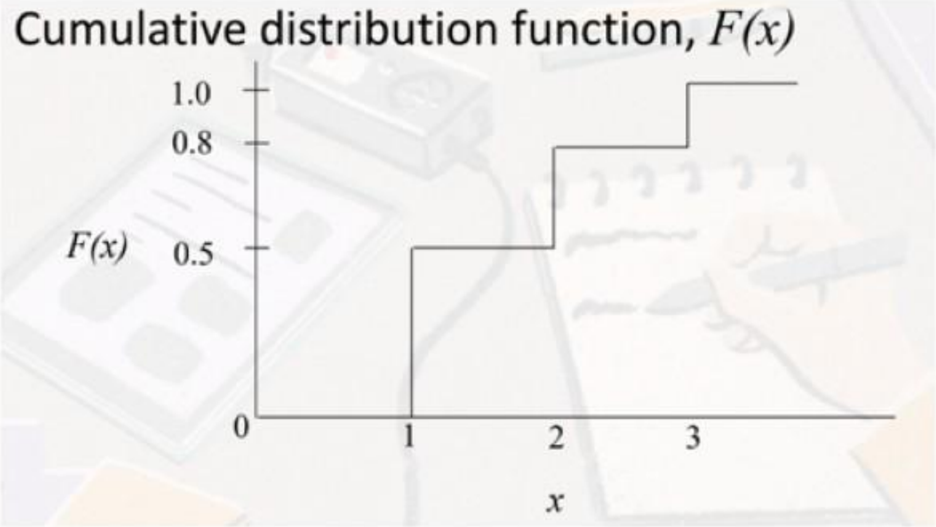
Méthodes : Génération de variables aléatoires

- Méthode d'inversion
- Méthode de la convolution
- Distribution empirique
- Autres techniques

Méthode d'inversion : Distribution Discrète

- C'est une technique permettant de générer des variables aléatoires qui suivent une distribution de probabilité donnée.
- L'idée centrale consiste à transformer des nombres aléatoires uniformes (entre 0 et 1) en valeurs correspondant à la distribution voulue, en se basant sur les probabilités cumulées.
- La fonction de répartition cumulative (CDF) indique la probabilité qu'une variable aléatoire X prenne une valeur inférieure ou égale à un certain x.

- Example: p(1) = 0.5, p(2) = 0.3, p(3) = 0.2



Méthode d'inversion : Distribution Discrète

Algorithme

- Generate random number u
- Random variate $x = i$ if $F(i-1) < u \leq F(i)$

Example : Considérons une distribution discrète avec les probabilités suivantes $\boldsymbol{p(1)=0.5}$, $\boldsymbol{p(2)=0.3}$, $\boldsymbol{p(3)=0.2}$. L'objectif est de générer des variables aléatoires (des échantillons) qui suivent exactement cette distribution.

$F(1)=0.5$,
$F(2)=0.8$,
$F(3)=1.0$.

•If $0 \leq U<0.5$, the random variate is 1.
•If $0.5 \leq U<0.8$, the random variate is 2.
•If $0.8 \leq U<1.0$, the random variate is 3.

Méthode d'inversion : Distribution Discrète

Exemple : Chaque pièce qui passe sur la chaîne peut être : bonne, Défectueuse mineure, ou Défectueuse grave.
je tire 10 nombres U et je te montre exactement où ils tombent.

Valeur x	Probabilité individuelle	CDF F(x)
1 = Bonne	0.92	F(1) = 0,92
2 = Défectueuse mineure	0.06	F(2) =0,98
3 = Défectueuse grave	0.02	F(3) = 1,0

U tiré	Où il tombe sur la CDF ?	Valeur générée
0,73	0 ≤ 0,73 < 0,92	1
0,12	0 ≤ 0,12 < 0,92	1
0,96	0,92 ≤ 0,96 < 0,98	2
0,994	0,98 ≤ 0,994 < 1,00	3
0,45	0 ≤ 0,45 < 0,92	1
0,83	0 ≤ 0,83 < 0,92	1
0,951	0,92 ≤ 0,951 < 0,98	2
0,9992	0,98 ≤ 0,9992 < 1,00	3
0,01	0 ≤ 0,01 < 0,92	1
0,89	0 ≤ 0,89 < 0,92	1

Méthode d'inversion : Variable uniforme discrète

- C'est quand tu tires un numéro entier au hasard entre deux bornes et que tous les numéros ont rigoureusement la même chance de sortir, comme un vrai tirage au sort équitable.
- Uniforme discrète (de paramètres a et b).

- Probabilité : chaque entier compris entre a et b (inclus) a rigoureusement la même probabilité de sortir :

$$P(x) = \frac{1}{b-a+1}$$

La variable aléatoire X est uniformément répartie sur les entiers compris dans l'intervalle[a,b].
La fonction de répartition (CDF) F(n) pour tout entier n∈[a,b] est donnée par :

$$F(n) = P(X \leq n) = \frac{n-a+1}{b-a+1}$$

Méthode d'inversion : Variable uniforme discrète

Nombre total de valeurs possibles = b − a + 1

→ Probabilité de chaque valeur = 1 / (b − a + 1)

$$P(x) = \frac{1}{b-a+1}$$

→ Probabilité d'avoir ≤ n = (n − a + 1) / (b − a + 1)

$$F(n) = P(X \leq n) = \frac{n-a+1}{b-a+1}$$

Méthode d'inversion : Variable uniforme discrète

- Random variate generation
  - Generate u
  - $x = a + floor(u * (b - a + 1))$

Exemple : une usine possède 6 machines identiques numérotées de 5 à 10 (c'est-à-dire 5, 6, 7, 8, 9, 10). Quand une pièce arrive, on choisit au hasard et de façon parfaitement équiprobable la machine qui va la traiter.

- Quelle est la probabilité que la pièce soit traitée par la machine n°7 ?
- Calculer $P(X = k)$ pour chaque $k = 5, 6, \ldots, 10$.
- Calculer $F(n)$ pour $n = 5, 6, 7, 8, 9, 10$,
- On tire $U = 0,42$. Quelle machine est choisie ?

Machine k	P(X = k)	F(k) exact (cumulée)	Borne inférieure	Borne supérieure (stricte)
5	1/6 ≈ 0,1667	1/6 ≈ 0,1667	0,0000	0,1667
6	0,1667	2/6 = 0,3333	0,1667	0,3333
7	0,1667	3/6 = 0,5000	0,3333	0,5000
8	0,1667	4/6 ≈ 0,6667	0,5000	0,6667
9	0,1667	5/6 ≈ 0,8333	0,6667	0,8333
10	0,1667	6/6 = 1,0000	0,8333	1,0000

Variable (aléatoire) géométrique

- Modélise le nombre d'essais nécessaires pour obtenir **le premier succès** dans une suite d'épreuves de Bernoulli indépendantes.
  - Nombre de lancers d'une pièce jusqu'à obtenir le premier « Face ».
  - Nombre de clients contactés jusqu'à la première vente.

Geometric (with parameter p)

$$p(n) = p(1-p)^{n-1}, \text{n} = 1,2,3, \ldots$$

Where:

- n = nombre d'essais jusqu'au premier succès (n = 1, 2, 3, …).
- p = probabilité de succès à chaque essai
- (1−p) =probabilité d'échec à chaque essai

- Random variate generation
  - Generate u
  - Geometric variate $X = \left\lceil \frac{\ln(u)}{\ln(1-p)} \right\rceil$

Variable (aléatoire) géométrique

- Exemple : au bout de combien d'emails envoyés aura-t-on le premier clic / la première vente ?
- Probabilité qu'un destinataire clique sur le lien ou achète : p = 4 % (0,04).

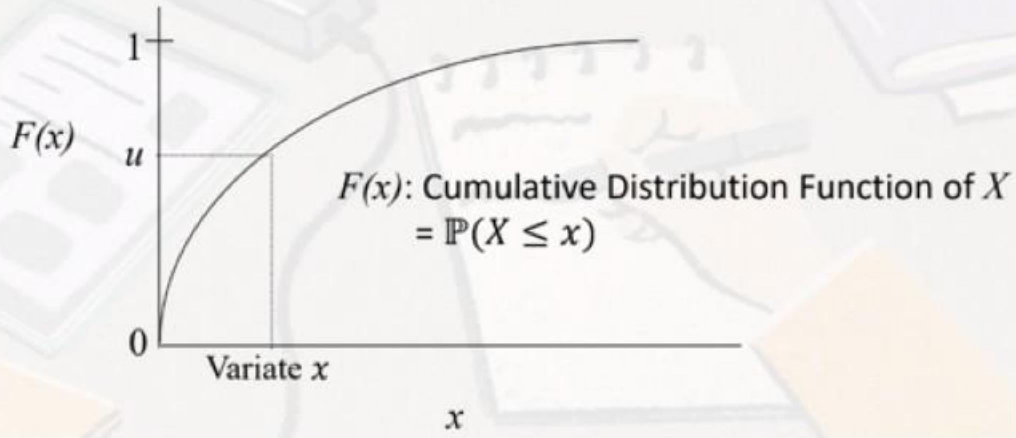
- Random variate generation
  - Generate u=0,27
  - Geometric variate $X = \left\lceil \frac{\ln(u)}{\ln(1-p)} \right\rceil$

- X= 32

Méthode de transformation inverse : Distributions continues

- C'est une technique très puissante pour générer des variables aléatoires qui suivent n'importe **quelle loi continue** (exponentielle, triangulaire, etc.) à partir d'un simple tirage uniforme.

- Algorithm
  - Generate uniform random number u
  - Solve $F(x) = u$ for random variate x



- **Générer** u ~ Uniforme(0,1)

- **Résoudre l'équation F(x) = u pour trouver x**→ c'est-à-dire : « quel est le x tel que la probabilité cumulée atteigne exactement u ? »

- La solution x obtenue suit automatiquement la loi voulue.

Génération d'une variable uniforme continue

- C'est l'une des lois de probabilité les plus simples. Dans une loi uniforme continue, tous les résultats compris dans un intervalle donné ont exactement la même probabilité (sont équiprobables).

1- Fonction de densité de probabilité (PDF) :

$$f(x) = \frac{1}{b-a}, \quad \text{for} \quad a \leq x \leq b$$

➡ Cela indique que la probabilité que X prenne n'importe quelle valeur dans l'intervalle [a, b] est la même pour chaque valeur au sein de cet intervalle.

2- Fonction de répartition cumulée (CDF) :

$$F(x) = \frac{x-a}{b-a}, \quad \text{for} \quad a \leq x \leq b$$

➡ La CDF représente la probabilité que X soit inférieur ou égal à une certaine valeur x.

3- Inverse de la CDF :

$$x = a + (b-a) \cdot u$$

➡ u est un nombre aléatoire tiré d'une distribution uniforme entre 0 et 1. Cette formule nous permet de générer des nombres aléatoires qui suivent une distribution uniforme continue.

Génération d'une variable uniforme continue

- Exemple : Horaire d'arrivée d'un bus

Le bus arrive de façon totalement aléatoire entre 9h00 et 9h30 du matin. Le temps d'arrivée du bus (en minutes après 9h00) suit une loi uniforme continue sur l'intervalle [0, 30] minutes.

Étape 1 : Fonction de densité (PDF) : $f(x) = \frac{1}{30-0} = \frac{1}{30}, \quad 0 \leq x \leq 30$ → Toutes les minutes entre 9h00 et 9h30 ont exactement la même probabilité d'être l'heure d'arrivée du bus.

Étape 2 : Fonction de répartition (CDF) : $F(x) = \frac{x-0}{30-0} = \frac{x}{30}, \quad 0 \leq x \leq 30$ → Si x = 10 minutes, il y a 33 % de chances que le bus arrive dans les 10 premières minutes (avant 9h10).

Étape 3 : Inverse de la CDF : On tire u = 0,6

→ x = 30 × 0,6 = **18 minutes**

→ Le bus arrive à **9h18**

Variable aléatoire exponentielle

- C'est une variable aléatoire qui suit la loi exponentielle. Elle sert à modéliser le temps entre deux événements aléatoires dans un processus de Poisson
- Elle est définie par le paramètre λ (lambda) = taux d'événements par unité de temps
- Sa fonction de densité (PDF) est :

$$f(x) = \lambda e^{-\lambda x}$$

- Pour générer des valeurs qui suivent cette loi, la méthode la plus simple et la plus efficace consiste à inverser la fonction de répartition (CDF)

$$F(x; \lambda) = 1 - e^{-\lambda x}$$

- Random variate generation
  - Generate U
  - $$X = -\frac{1}{\lambda} \ln(1 - U)$$

Exemple : Un guichet de banque reçoit en moyenne **12 clients par heure** → λ = 12 clients/heure → Temps moyen entre deux clients = 1/λ = **5 minutes**
On veut simuler : **au bout de combien de temps arrive le prochain client ?**
U = 0,05 ≈ 0,26 minute

Variable de Poisson

- Modélise le nombre d'événements qui se produisent dans un intervalle de temps (ou d'espace) fixe, quand ces événements arrivent de façon totalement aléatoire à un rythme moyen constant λ.

$$p_k = P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},\ k = 0, 1, 2, \ldots$$

- Génération d'une variable de Poisson λ en utilisant la loi exponentielle

1.Initialisation

s ← 0 (temps cumulé)

c ← 0 (compteur d'événements)

2. Boucle tant que s ≤ 1 (ou ≤ T si on veut un intervalle de longueur T)

a. Générer y ~ Exponentielle(λ) → temps jusqu'au prochain événement

b. s ← s + y → on avance dans le temps

c. si s ≤ 1 alors c ← c + 1 → l'événement tombe dans l'intervalle

3. Résultat

X = c − 1 → nombre d'événements dans l'intervalle [0,1]

Variable de Poisson

Exemple : Un café reçoit en moyenne **λ = 5 clients par heure**. Le nombre de clients qui arrivent pendant **1 heure** suit une **loi de Poisson de paramètre λ = 5**.

Objectif : simuler ce nombre grâce à l'algorithme basé sur l'exponentielle.

- $y_1 = 0.2$: $s = 0.2$, $c = 1$.
- $y_2 = 0.15$: $s = 0.35$, $c = 2$.
- $y_3 = 0.3$: $s = 0.65$, $c = 3$.
- $y_4 = 0.25$: $s = 0.9$, $c = 4$.
- $y_5 = 0.2$: $s = 1.1$, $c = 5$.
- Stop as $s > 1$.

Résultats :

- **x=c−1=5−1=4.**
- 4 clients se sont présentés au café pendant cette période d'une heure.

Méthode de la convolution

- Consiste à additionner plusieurs variables aléatoires.
- Permet de générer la loi de la somme de plusieurs variables aléatoires indépendantes.

1. Générer n variables aléatoires indépendantes $y_1, y_2, \ldots, y_n$
2. Faire $x = y_1 + y_2 + \ldots + y_n \rightarrow x$ suit automatiquement la convolution des lois des y_i

Exemple classique : somme de deux dés à 6 faces (dé équilibré)

P(x=2) = 1/36; (Combinaisons possibles: (1, 1))
P(x=3) = 2/36; (Combinaisons possibles : (1, 2) and (2, 1))
P(x=4) = 3/36; (Combinaisons possibles : (1, 3), (2, 2), and (3, 1))
P(x=5) = 4/36; (Combinaisons possibles : (1, 4), (2, 3), (3, 2), and (4, 1))
P(x=6) = 5/36; (Combinaisons possibles : (1, 5), (2, 4), (3, 3), (4, 2), and (5, 1))
P(x=7) = 6/36; (Combinaisons possibles : (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1))
P(x=8) = 5/36; (Combinaisons possibles : (2, 6), (3, 5), (4, 4), (5, 3), and (6, 2))
P(x=9) = 4/36; (Combinaisons possibles : (3, 6), (4, 5), (5, 4), and (6, 3))
P(x=10) = 3/36; (Combinaisons possibles : (4, 6), (5, 5), and (6, 4))
P(x=11) = 2/36; (Combinaisons possibles : (5, 6) and (6, 5))
P(x=12) = 1/36; (Combinaisons possibles : (6, 6))

Méthode de la convolution

- Consiste à additionner plusieurs variables aléatoires.
- Permet de générer la loi de la somme de plusieurs variables aléatoires indépendantes.

1. Générer n variables aléatoires indépendantes $y_1, y_2, \dots, y_n$
2. Faire $x = y_1 + y_2 + \dots + y_n$ → x suit automatiquement la convolution des lois des y_i

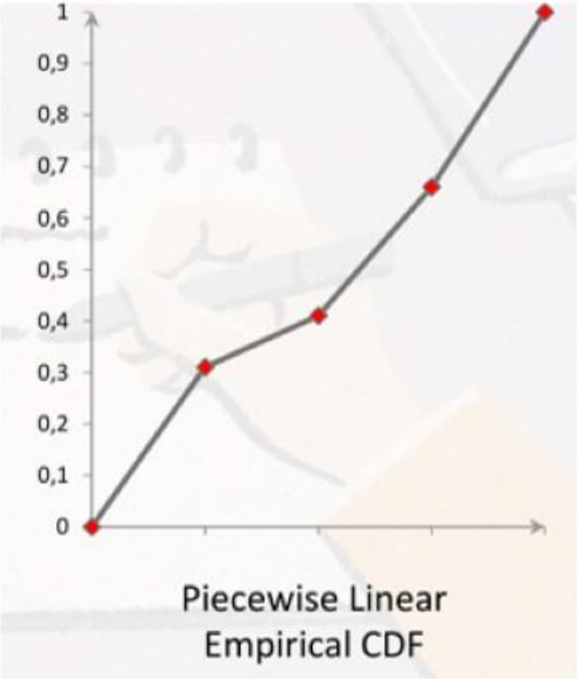
Exemple : combien de temps le client reste-t-il au total au guichet ?

Distribution empirique

Example: Distribution empirique linéaire par morceaux

- On l'utilise quand aucune loi théorique classique (exponentielle, normale, etc.) ne colle bien aux données réelles.
- Convient pour des données continues.
- Parfait quand on dispose d'un gros échantillon de mesures réelles.

- Axe horizontal (x) → les **valeurs possibles** de la variable.
- Axe vertical (y) → la **probabilité cumulée** (de 0 à 1)
- Les **points rouges** = les vraies données triées
- Les **segments noirs** relient ces points → on obtient une **courbe en escalier lissée** (linéaire par morceaux)
- Cette courbe représente la **fonction de répartition empirique (CDF estimée)** à partir des données réelles

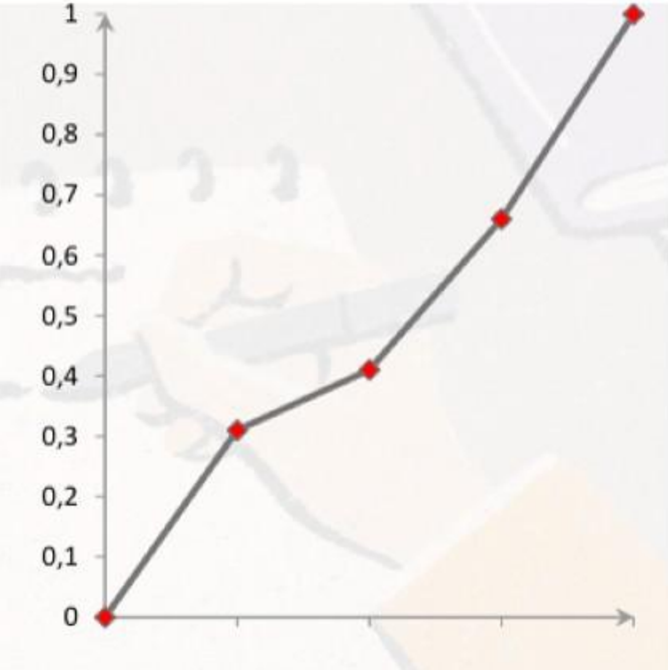


Distribution empirique

Example: Distribution empirique linéaire par morceaux

Étapes pour générer une valeur x

1. **Identifier l'intervalle : Si x dans l'intervalle i.** $a_{i-1} < x \leq a_i$



Piecewise Linear
Empirical CDF

- X : la valeur générée à l'intérieur de l'intervalle i

Distribution empirique

Example: Distribution empirique linéaire par morceaux

Calcul de la pente α_i de l'intervalle i :

$$\alpha_i = \frac{c_i - c_{i-1}}{a_i - a_{i-1}}$$

- *ai*: pente de la droite dans l'intervalle i,
- *ci*: probabilité cumulée à la fin de l'intervalle i,
- *ci*−1: probabilité cumulée au début de l'intervalle i,
- ai: borne supérieure de l'intervalle i,
- *ai*−1: borne inférieure de l'intervalle i.

2. **Calculer x** :

Pour l'intervalle identifié i, utiliser cette formule :

$$x = a_{i-1} + \frac{u - c_{i-1}}{\alpha_i}$$

Exemple : Distribution empirique

Exemple : Temps de réparation de machines

Imaginons que nous ayons collecté les temps de réparation de 100 machines en panne. Ces données réelles sont résumées dans le tableau ci-dessous :

Interval (Hours)	Frequency	Relative Frequency	Cumulative Frequency	Slope (α_i)
0.0 < x ≤ 0.5	31	0.31	0.31	?
0.5 < x ≤ 1.0	10	0.10	0.41	?
1.0 < x ≤ 1.5	25	0.25	0.66	?
1.5 < x ≤ 2.0	34	0.34	1.00	?

1. **Intervalle (en heures)** → la tranche de temps (ex. : 0,5 à 1,0 h)
2. **Fréquence** → combien de machines ont été réparées dans cette tranche
3. **Fréquence relative** → quelle part (en proportion ou %) des 100 machines tombe dans cette tranche.

$$\text{Relative Frequency} = \frac{\text{Frequency}}{\text{Total Machines}}$$

Exemple : Distribution empirique

Interval (Hours)	Frequency	Relative Frequency	Cumulative Frequency	Slope (α_i)
0.0 < x ≤ 0.5	31	0.31	0.31	?
0.5 < x ≤ 1.0	10	0.10	0.41	?
1.0 < x ≤ 1.5	25	0.25	0.66	?
1.5 < x ≤ 2.0	34	0.34	1.00	?

4. Cumulative Frequency: proportion de réparations terminées jusqu'à la borne supérieure de l'intervalle.

Formule :

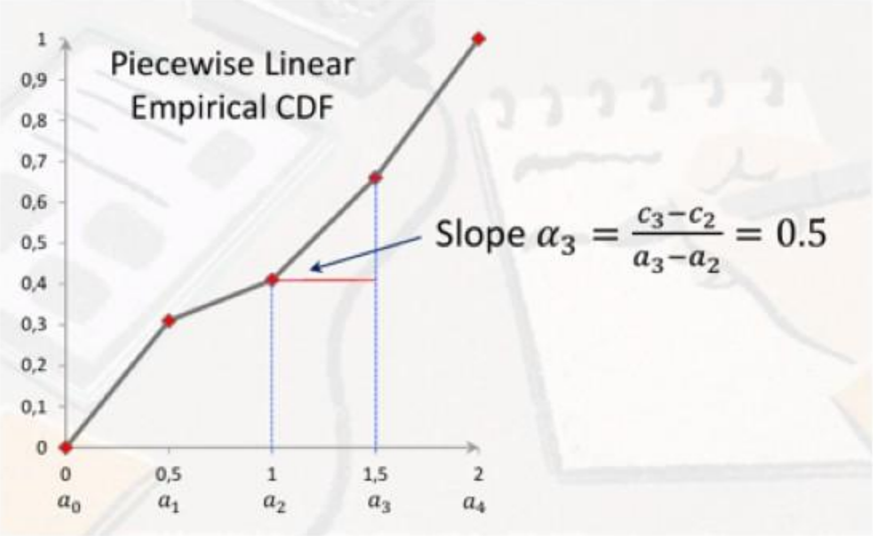
Fréquence cumulée à l'intervalle i = somme de toutes les fréquences relatives des intervalles précédents + la fréquence relative de l'intervalle i.

5. The slope indique à quelle vitesse la proportion cumulée augmente dans chaque intervalle.

Exemple : Distribution empirique

Calculer : Slope (αi):

Interval (Hours)	Frequency	Relative Frequency	Cumulative Frequency (c_i)	Slope (α_i)
$0.0 < x \leq 0.5$	31	0.31	0.31	$\frac{0.31-0}{0.5-0.0} = 0.62$
$0.5 < x \leq 1.0$	10	0.10	0.41	$\frac{0.41-0.31}{1.0-0.5} = 0.2$
$1.0 < x \leq 1.5$	25	0.25	0.66	$\frac{0.66-0.41}{1.5-1.0} = 0.5$
$1.5 < x \leq 2.0$	34	0.34	1.00	$\frac{1.00-0.66}{2.0-1.5} = 0.68$



- **31% de pannes rapides** (< 30 min) → problèmes simples
- **34% de pannes longues** (1.5-2h) → problèmes complexes

Example : Distribution empirique

Interval (Hours)	Frequency	Relative Frequency	Cumulative Frequency (c_i)	Slope (α_i)
$0.0 < x \leq 0.5$	31	0.31	0.31	$\frac{0.31-0}{0.5-0.0} = 0.62$
$0.5 < x \leq 1.0$	10	0.10	0.41	$\frac{0.41-0.31}{1.0-0.5} = 0.2$
$1.0 < x \leq 1.5$	25	0.25	0.66	$\frac{0.66-0.41}{1.5-1.0} = 0.5$
$1.5 < x \leq 2.0$	34	0.34	1.00	$\frac{1.00-0.66}{2.0-1.5} = 0.68$

Supposant u=0.83 : $$x = a_{i-1} + \frac{u - c_{i-1}}{\alpha_i}$$

puisque $0.66 < 0.83 \leq 1.0$, on peut conclure que i=4.

$$x = 1.5 + \frac{1}{0.68} \cdot (0.83 - 0.66)$$

$$x \approx 1.75$$

La variable aléatoire x représente un temps de réparation simulé d'une machine, généré à partir de la distribution empirique construite sur les données réelles.

Dans cette simulation, la machine serait réparée en environ 1,75 heures.