Algorithmes randomisés

Nicolas Gast

Inria – Univ. Grenoble Alpes

2020. Maths701, Master MEEF.

Plan du cours

1. Rappel de Probabilités
   - Événements et espaces probabilisés
   - Variables aléatoires
   - Inégalités de concentration, intervales de confiance
2. Génération d'aléatoire et simulation
   - Comment générer de l'aléatoire?
   - Générer selon une loi donnée
   - Exemple de simulation: la formule de Lindley
3. Algorithmes randomisés
   - Algorithmes "Las Vegas": Quicksort / Quickselect
   - Algorithmes "Monte Carlo"
   - Algorithmes décentralisés / apprentissage par renforcement

Espace probabilisé

Un espace probabilité est $(\Omega, \mathcal{F}, \mathbb{P}[])$, où

- Ω est l'ensemble des possibles.
- $\mathcal{F} \subset 2^{\Omega}$ est l'ensemble des événements *mesurables*.
- $\mathbb{P}[] : \mathcal{F} \to [0, 1]$ est une mesure de probabilité.

$$\mathbb{P}[E] = \text{ probabilité de l'événement } E.$$

Exemple : lancé de dé à 6 faces

$\Omega = \{0; 1; 2; 3; 4; 5; 6\}$

$\mathcal{F} = \{\emptyset; \{1\}, \{2\}, \ldots, \{1,2\}, \ldots\}$

$\mathbb{P}(\{1\}) = \frac{1}{6} = \mathbb{P}(\{2\})$

$\mathbb{P}(\{1, 2, 5\}) = \frac{1}{2}$

$\boxed{\mathbb{P}(\{0\}) = 0}$

$\mathbb{P}(\{1; \ldots; 6\}) = 1$

$\mathbb{P}(\emptyset) = 0$;

Indépendance et probabilité conditionnelle

A et B sont indépendants

A et B' ne sont pas indépendants

Soit $A, B \in \mathcal{F}$ deux événements. A et B sont indépendents si $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$.

Ex: $A = \{1, 3, 5\}$ $B = \{1, 2\}$

$\Rightarrow \mathbb{P}(A) = \frac{1}{2}$ $\mathbb{P}(B) = \frac{1}{3}$

$\mathbb{P}(A \cap B) = \mathbb{P}(\{1\}) = \frac{1}{6}$

$A = \{1, 3, 5\}$ $\mathbb{P}(A) = \frac{1}{2}$

$B' = \{1\}$ $\mathbb{P}(B') = \frac{1}{6}$

$\mathbb{P}(A \cap B') = \frac{1}{6} \neq \frac{1}{2} \cdot \frac{1}{6}$

Si $A \in \mathcal{F}$ is un événement de probabilité non nulle, alors:

$$\mathbb{P}[B \mid A] = \frac{\mathbb{P}[B \cap A]}{\mathbb{P}[A]}$$

"proba de B sachant A"

$\mathbb{P}($dé est 1 $\mid$ dé est impair$)$

$= \frac{1/6}{1/2} = \frac{1}{3}$

$\mathbb{P}($dé est impair $\mid$ dé est 1$)$

$= \frac{1/6}{1/6} = 1$

Ex: lancé de deux dés.

Règle de Bayes

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \quad (1)$$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (2)$$

Bayes rule

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\,\mathbb{P}[B]}{\mathbb{P}[A]}.$$

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A) \quad (1)$$

$$= \mathbb{P}(A|B) \cdot \mathbb{P}(B) \quad (2)$$

Exercice

$\mathbb{P}(\text{test positif} \mid \text{malade}) = 0.95$

en général
$$\boxed{\mathbb{P}(A \mid B) \neq \mathbb{P}(B \mid A)}$$

Example: Une maladie a touché 10% de la population. Le laboratoire Testex a créé un nouveau test qu'il permet de savoir si on a été infecté ou non. Ce test est censé être $\boxed{\text{fiable à 95\%}}$ (c'est à dire qu'il donne un résultat juste 95% du temps.)

Question – Un individu est testé positif par le test. Quelle est la probabilité qu'il soit malade?

$$\mathbb{P}(\text{malade} \mid \text{test positif}) = \frac{\overbrace{\mathbb{P}(\text{test positif} \mid \text{malade})}^{0.95} \cdot \overbrace{\mathbb{P}(\text{malade})}^{0.1}}{\mathbb{P}(\text{test } \underbrace{\text{positif}}_{x})}$$

$$x = 0.1 \times 0.95 + 0.9 \times 0.05 = 0.14$$

$$\mathbb{P}(\text{malade} \mid \text{test positif}) \approx 0.68 \ll 0.95$$

Table des matières

Variables aléatoire $\omega \in \Omega \Rightarrow X(\omega)$

Soit $\mathcal{X}$ un sous-ensemble de $\mathbb{R}$ (typiquement, $\mathcal{X} = \mathbb{R}$ ou $\mathcal{X}$ est fini).

Une variable aléatoire à valeur dans $\mathcal{X}$ est une fonction mesurable, $X : \Omega \to \mathcal{X}$. On note

$$\underline{\mathbb{P}\left[X \in A\right]} = \mathbb{P}\left[\{\omega : X(\omega) \in A\}\right].$$

En règle général, on n'écrira pas $X(\omega)$ mais uniquement X.

Exemple: lancés de dés.

X = le dé est pair $\Rightarrow X(\omega) = \begin{cases} 1 & \text{si } \omega \in \{0;2;4;6\} \\ 0 & \text{sinon} \end{cases}$

$\Omega = \{0;1;\dots;6\}$

Y = la valeur du dé $\quad Y(\omega) = \omega$.

Espérance, variance

Soit X une variable aléatoire à valeur dans un ensemble fini $\mathcal{X} \subset \mathbb{R}$.
L'espérance de X, notée $\mathbb{E}[X]$ est:

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x . \mathbb{P}(X = x)$$

La variance de X est:

$$\mathrm{var}(X) = \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2\right]$$

$$= \sum_{x \in \mathcal{X}} \left(x - \mathbb{E}[X]\right)^2 \mathbb{P}(X = x)$$

$$\geq 0.$$

Loi usuelles

	Définition	Espérance	Variance
• Loi de Bernoulli B	param $p \in [0;1]$ $\mathbb{P}(B=1)=p$; $\mathbb{P}(B=0)=1-p$	p	$\mathrm{var}(B) = (1-p)^2 p + p^2(1-p)$ $= p(1-p)$
Loi géométrique G	$p \in [0;1]$ G = nombre de tirages nécessaires avant d'arriver à 1. $\mathbb{P}(G=1) = p$ $\mathbb{P}(G=2) = (1-p)p$ $\mathbb{P}(G=3) = (1-p)^2 p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
• Loi binomiale N	n, p ⟹ on fait n tirages d'une loi de Bernoulli de param p on compte le nombre de 1 $\mathbb{P}(N=i) = \binom{n}{i} p^i (1-p)^{n-i}$ $N = \sum_{k=1}^{n} B_k$	np	$np(1-p)$
Loi de Poisson X	1 param $\lambda \geq 0$ $\mathbb{P}(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ

Variables aléatoires indépendantes

Deux variables sont indépendantes si $\forall A, B$

$$\mathbb{P}[X \in A \cap Y \in B] = \mathbb{P}[X \in A]\,\mathbb{P}[X \in B].$$

Exemple: lancés de dés.

on lance deux dés de façon indépendantes X_1, X_2

$\mathbb{P}(X_1 = i \text{ et } X_2 = j) = \frac{1}{36}$ pour tout $i, j \in \{1; \dots; 6\}$

Est-ce que pour toutes valeurs de s et d :

$\mathbb{P}(S = s \text{ et } D = d) = \mathbb{P}(S=s)\,\mathbb{P}(D=d)$

$S = X_1 + X_2$

$D = X_1 - X_2$

$\mathbb{P}(S = 12) = \frac{1}{36}$

$\mathbb{P}(D = 1) > 0$

$\mathbb{P}(S = 12 \text{ et } D = 1) = 0 \neq \mathbb{P}(S=12)\,\mathbb{P}(D=1)$

donc : S et D ne sont pas indépendantes.

Espérance, variance et indépendance

Soit X et Y deux variables aléatoires:

- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ → même si X et Y ~~ne~~ sont pas indépendantes.
- Si X et Y sont indépendantes, alors $\mathrm{var}[X+Y] = \mathrm{var}[X] + \mathrm{var}[Y]$.

Exemple: loi de Bernoulli.

$X \sim$ Bernoulli de param $\frac{1}{2}$

$Y = X \sim$ Bernoulli de param $\frac{1}{2}$

$Z = 1 - X \sim$ Bernoulli de param $\frac{1}{2}$

→ ne ~~sont~~ sont pas indépendantes.

$$X + Y = \begin{cases} 2 & \text{avec proba } 1/2 \\ 0 & \text{avec proba } 1/2 \end{cases}$$

$X + Z = 1$ $\quad$ $\mathrm{var}(X+Z) = 0$

$\mathbb{E}[X+Y] = 1$

$\mathrm{var}(X+Y) = (2-1)^2 \frac{1}{2} + (0-1)^2 \cdot \frac{1}{2} = 1$

$\mathrm{var}(X) = \frac{1}{2}\left(1-\frac{1}{2}\right) = \frac{1}{4}$

$\mathrm{var}(Y) = 1/4$

$\mathrm{var}(Z) = 1/4$

Table des matières

Inégalité de Markov et de Chebychev

Soit X une variable aléatoire (réelle) $\forall a > 0$:

Ineg de Markov : $\mathbb{P}[X \geq a] \leq \frac{1}{a}\mathbb{E}[|X|]$

preuve : $|X| = |X| \mathbb{1}_{\{|X| \geq a\}} + |X| \mathbb{1}_{\{|X| < a\}} \geq a \mathbb{1}_{\{|X| \geq a\}}$

$\mathbb{1}_{\{|X| \geq a\}} = \begin{cases} 1 & \text{si } |X| \geq a \\ 0 & \text{sinon} \end{cases}$ $\quad$ $\mathbb{1}_{\{|X| < a\}} = \begin{cases} 1 & \text{si } |X| < a \\ 0 & \text{sinon} \end{cases}$ $\quad$ $|X| \mathbb{1}_{\{|X| < a\}} \geq 0$

$|X| \mathbb{1}_{\{|X| \geq a\}} \geq a \mathbb{1}_{\{|X| \geq a\}}$

$\mathbb{E}[|X|] \geq a \underbrace{\mathbb{E}[\mathbb{1}_{\{|X| \geq a\}}]}_{= \mathbb{P}(|X| \geq a)}$

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq a] \leq \frac{1}{a^2}\text{var}[|X|]$$

$\mathbb{P}(X \geq a) \leq \mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a}$ ∎

Chebyshev :

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{1}{a^2} \operatorname{var}(X)$$

Markov : $\mathbb{P}(X \geq a) \leq \frac{1}{a} \mathbb{E}[|X|]$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) = \mathbb{P}(|X - \mathbb{E}[X]|^2 \geq a^2)$$
$$= \mathbb{P}((X - \mathbb{E}[X])^2 \geq a^2)$$
$$\underset{\text{per Markov}}{\leq} \frac{1}{a^2} \mathbb{E}[|(X - \mathbb{E}[X])^2|]$$
$$= \frac{1}{a^2} \mathbb{E}[(X - \mathbb{E}[X])^2]$$
$$= \frac{1}{a^2} \operatorname{var}(X) .$$

$\blacksquare$

Application: loi des grands nombres

Soit $X_1, X_2, \dots$ une suite de variable aléatoire indépendantes et de même loi et à valeurs dans $\mathcal{X}$ fini. Soit $S_n = \frac{1}{n}\sum_{k=1}^{n} X_k$. Alors: ← moyenne des n premières valeurs.

$$\forall \epsilon > 0 : \qquad \lim_{n\to\infty} \mathbb{P}\left[|S_n - \mathbb{E}[X]| \geq \varepsilon\right] = 0.$$

Preuve : $\mathrm{var}(S_n) = \mathrm{var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)$

$$= \frac{1}{n^2}\,\mathrm{var}\left(\sum_{k=1}^{n} X_k\right)$$

$$= \frac{1}{n^2}\left(\mathrm{var}(X_1) + \mathrm{var}(X_2) + \dots + \mathrm{var}(X_n)\right)$$ car les X_i sont indépendants.

$$= \frac{1}{n^2}\, n\,\mathrm{var}(X) = \frac{\mathrm{var}(X)}{n}.$$

$$\mathbb{E}[S_n] = \frac{1}{n}\left(\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]\right) = \mathbb{E}[X]$$

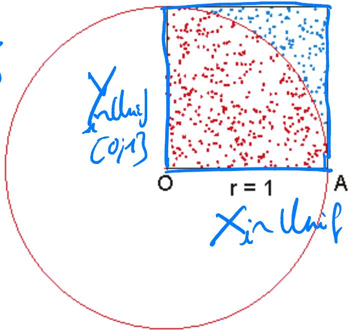
Chebychev: $\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}\,\mathrm{var}(S_n) = \frac{\mathrm{var}(X)}{n\,\varepsilon^2} \xrightarrow{n\to\infty} 0$

Illustration: Estimation de π.

$$Z_i = \begin{cases} 1 & \text{si } (X_i, Y_i) \text{ est dans le cercle } \underline{\text{ssi}} \ X_i^2 + Y_i^2 \leq 1 \\ 0 & \text{sinon} \end{cases}$$

$$\frac{1}{n}\sum_{i=1}^{n} Z_i \approx \mathbb{E}[Z]$$

quand n est grand



$$\mathbb{E}[Z] = \frac{\pi}{4}$$

$$\pi \approx 4 \cdot \frac{1}{n} \sum_{i=1}^{n} Z_i \, .$$

Confiance?

On dispose d'une pièce biaisée $\mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0] = p$.

- On cherche à estimer p.

Combien de tirages sont nécessaire?

Confiance?

On dispose d'une pièce biaisée $\mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0] = p$.

- On cherche à estimer p.

Combien de tirages sont nécessaire?

$X = (0,$

Confiance?

On dispose d'une pièce biaisée $\mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0] = p$.

- On cherche à estimer p.

Combien de tirages sont nécessaire?

$X = ($0, 1, 0, 1, 0, 1,

Confiance?

On dispose d'une pièce biaisée $\mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0] = p$.

- On cherche à estimer p.

Combien de tirages sont nécessaire?

$X = ($0, 1, 0, 1, 0, 1,1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0,

Confiance?

On dispose d'une pièce biaisée $\mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0] = p$.

- On cherche à estimer p.

Combien de tirages sont nécessaire?

$X = ($0, 1, 0, 1, 0, 1,1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0,0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1,$\ldots)$

Confiance?

On dispose d'une pièce biaisée $\mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0] = p$.

- On cherche à estimer p.

Combien de tirages sont nécessaire?

$X = (0, 1, 0, 1, 0, 1,1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1,$
$1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0,$
$1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0,0, 0, 1, 1, 1, 1, 0, 0,$
$1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1,$
$0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1,$
$0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1,\dots)$

n	1	3	5				
$\frac{1}{n}\sum_{i=1}^{n} X_i$	0.0	0.333...	.4				

Confiance?

On dispose d'une pièce biaisée $\mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0] = p$.

- On cherche à estimer p.

Combien de tirages sont nécessaire?

$X = ($0, 1, 0, 1, 0, 1,1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0,0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1,$\ldots)$

	1	3	5	10	20	100	
$\frac{1}{n}\sum_{i=1}^{n} X_i$	0.0	0.333...	.4	0.7	0.7	0.57	

Confiance?

On dispose d'une pièce biaisée $\mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0] = p$.

- On cherche à estimer p.

Combien de tirages sont nécessaire?

$X = ($0, 1, 0, 1, 0, 1,1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0,0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1,$\dots)$

	1	3	5	10	20	100	1000
$\frac{1}{n}\sum_{i=1}^{n} X_i$	0.0	0.333...	.4	0.7	0.7	0.57	0.606

$0.606 \approx p$?

Confiance?

On dispose d'une pièce biaisée $\mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0] = p$.

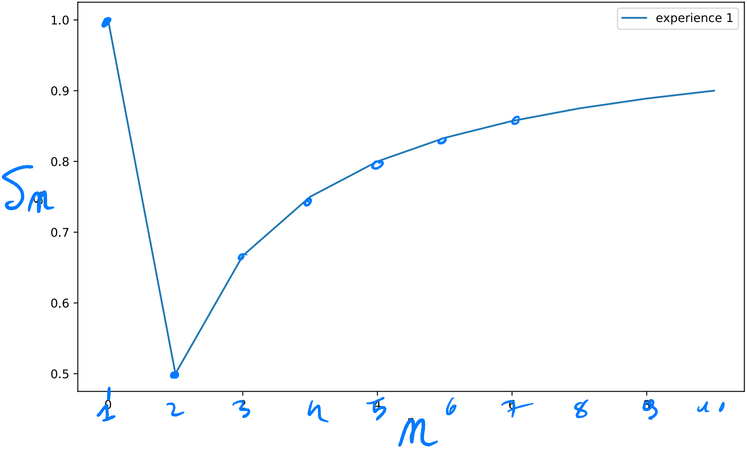
- On cherche à estimer p.

Combien de tirages sont nécessaire?

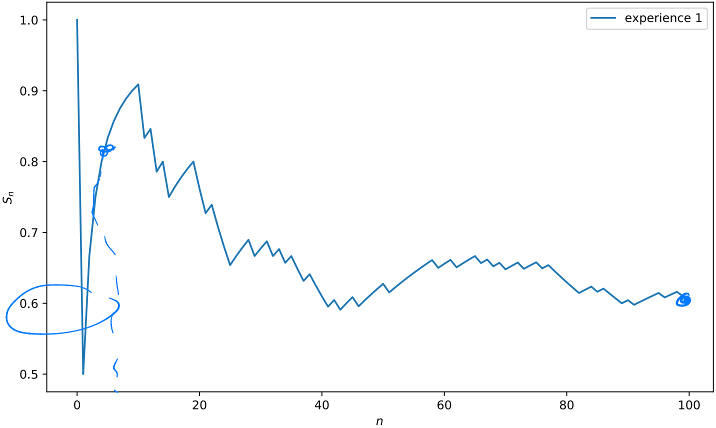
$X = ($0, 1, 0, 1, 0, 1,1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0,0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1,$\dots)$

	1	3	5	10	20	100	1000
$\frac{1}{n}\sum_{i=1}^{n} X_i$	0.0	0.333...	.4	0.7	0.7	0.57	0.606
	0.0	0.666...	0.8	0.7	0.65	0.71	0.62
	1.0	0.666...	0.6	0.4	0.55	0.6	0.597
	0.0	0.0	0.2	0.3	0.45	0.53	0.574
	0.0	0.333...	0.2	0.5	0.6	0.59	0.592
	0.0	0.333...	0.4	0.6	0.55	0.6	0.623

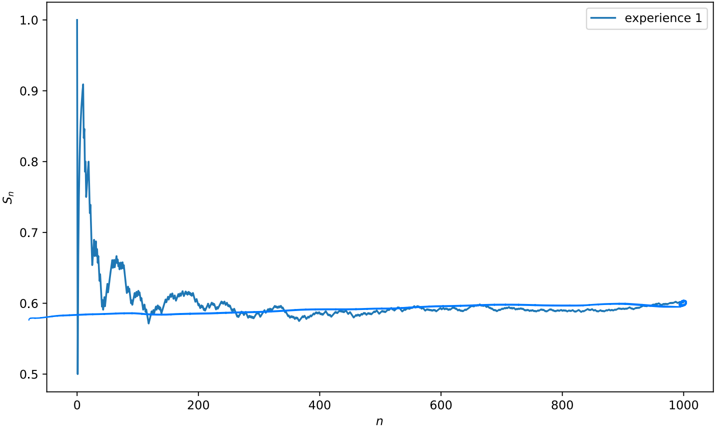
Étudions S_n



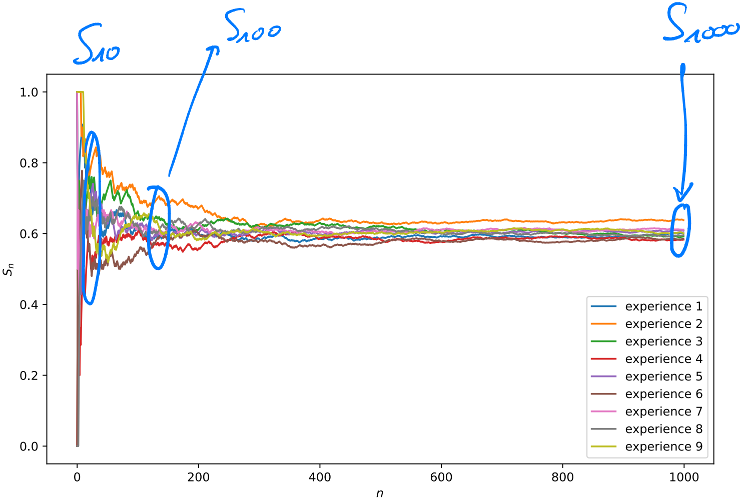
Études S_n



Études S_n

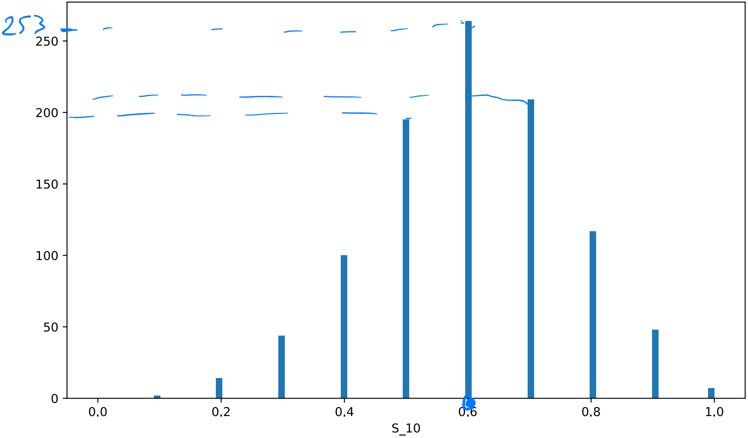


Études S_n

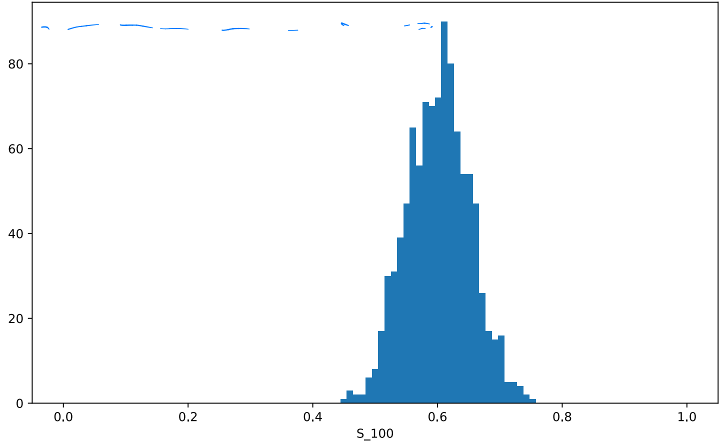


S_{10}
S_{100}
S_{1000}
1.0
0.8
0.6
0.4
0.2
0.0
S_n
0
200
400
600
800
1000
n
experience 1
experience 2
experience 3
experience 4
experience 5
experience 6
experience 7
experience 8
experience 9

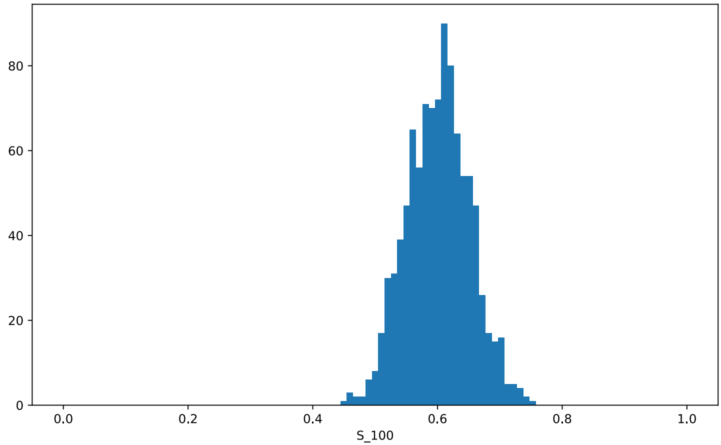
Distribution de S_n pour $n \in \{10, 100, 1000\}$



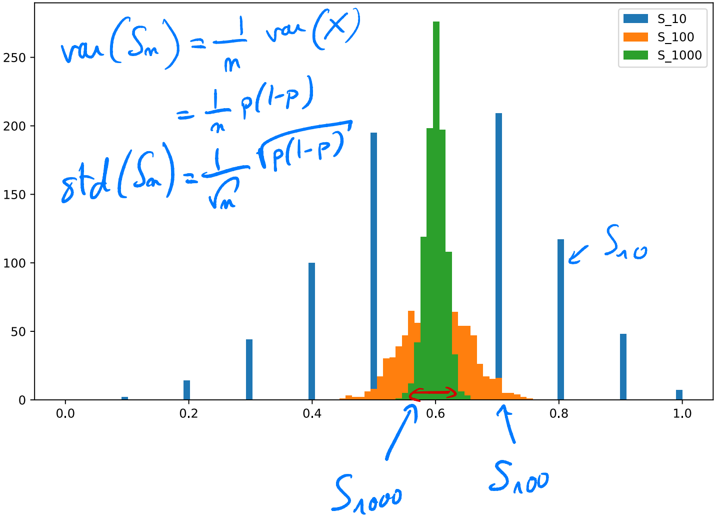
Distribution de S_n pour $n \in \{10, 100, 1000\}$



Distribution de S_n pour $n \in \{10, 100, 1000\}$



Distribution de S_n pour $n \in \{10, 100, 1000\}$



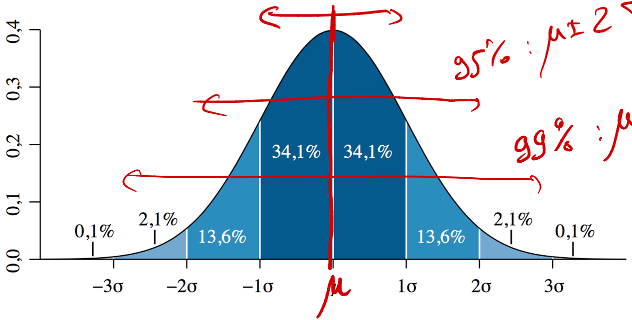
Loi normale (μ, σ)

Densité: $f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

70% : $\mu \pm 1\sigma$

95% : $\mu \pm 2\sigma$

99% : $\mu \pm 3\sigma$



Théorème central limite

indépendantes et identiquement distribuées.

Soit $X_1, X_2, X_3, \dots$ une suite de variable aléatoires *i.i.d.* de ~~moyenne~~ d'espérance μ et de variance σ^2. Alors:

$$\underbrace{\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu)}\to N(0,1) \qquad \text{en loi}$$

Historique de la démonstration: de Moivre 1733 (pour une pièce non biaisée), Laplace 1812 (pour les pièces biaisées), Pólya 1920 (cas général)

$$S_n = \frac{1}{n}\sum_{k=1}^{n} X_k \approx \boxed{\mu + \frac{\sigma}{\sqrt{n}}\mathcal{N}(0;1)}$$

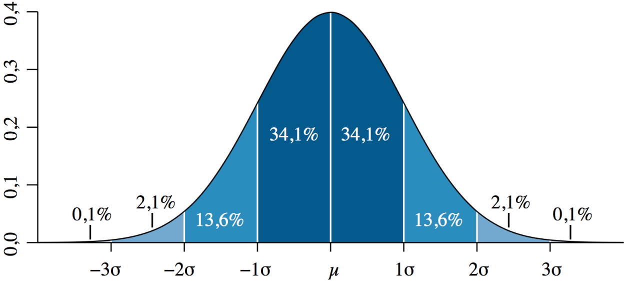
$$\frac{\sqrt{n}}{\sigma}(S_n - \mu) \to \mathcal{N}(0;1)$$

Intervalle de confiance

$$\frac{1}{n}\sum_{k=1}^{n} X_k \approx \mu + \frac{1}{\sqrt{n}}\sigma.$$

En particulier:

$$\mathbb{P}\left[\frac{1}{n}\sum_{k=1}^{n} X_k \in [\mu - \frac{2}{\sqrt{n}}\sigma; \mu + \frac{2}{\sqrt{n}}\sigma]\right] \geq 95\%.$$



Quel σ utiliser? Cas des variables sont de Bernoulli

- $\sigma = \sqrt{p(1-p)} \leq 1/2$. Application: sondage et confiance à 3 points.

$$S_n = \frac{1}{n}\sum_{k=1}^{n} X_k = \mu + \frac{\sigma}{\sqrt{n}} \mathcal{N}(0;1)$$

$$\mathbb{P}(|S_n - \mu| \geq x) \simeq \mathbb{P}\left(\frac{\sigma}{\sqrt{n}}|\mathcal{N}(0;1)| \geq x\right)$$

$$\mathbb{P}\left(|S_n - \mu| \geq \frac{2\sigma}{\sqrt{n}}\right) \simeq 95\%$$

Population $\simeq$ 70M hab.

Sondage sur 1000 personnes

$X_i \in \{0;1\}$ vote de la personne i pour le candidat qui nous intéresse.

si une fraction "p" de la population vote pour ce candidat

$X_i \sim$ Bernoulli de param p. Les X_i sont à peu près indépendants car $1000 < 70\cdot 10^6$.

$$\mathbb{P}\left(|S_n - p| \geq \frac{2\sqrt{p(1-p)}}{\sqrt{1000}}\right) \simeq 95\%$$

$$\frac{2\sqrt{p(1-p)}}{\sqrt{1000}} \leq \frac{1}{\sqrt{1000}} \simeq 0.03 = 3\%$$