

Simulation de variables aléatoires - Estimation ponctuelle ou par intervalle de confiance - Chaîne de Markov

Formation Académique 19–21 octobre 2026 – LYMED · Tétouan

KANBER Ahmed

LAAROUSSI Saâdeddine

19 Octobre 2026

Estimation ponctuelle

La simulation permet de générer des observations suivant une loi de probabilité donnée.

En pratique, on rencontre souvent le problème inverse :

- on dispose d'observations issues d'une expérience aléatoire ;
- la loi de probabilité est connue, mais certains de ses paramètres sont inconnus ;
- on souhaite utiliser les observations pour estimer ces paramètres.

On observe le résultat de n lancers indépendants d'une pièce. La probabilité p d'obtenir pile est inconnue.

Comment estimer p à partir des résultats observés ?

Construire des méthodes permettant d'estimer un paramètre inconnu à partir d'un échantillon.

Échantillon et estimateur

Soit X une variable aléatoire dont la loi dépend d'un paramètre inconnu θ .

On observe un échantillon

$$X_1, X_2, \dots, X_n$$

de variables aléatoires indépendantes et de même loi que X .

Définition

Un estimateur de θ est une statistique, c'est-à-dire une fonction des observations :

$$\hat{\theta}_n = T(X_1, \dots, X_n).$$

Estimateur = variable aléatoire

Estimation = valeur observée

La moyenne empirique

Un estimateur fondamental est la **moyenne empirique** :

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Si

$$\mu = \mathbb{E}(X),$$

alors $\overline{X}_n$ est naturellement utilisée pour estimer μ .

Principe

On remplace une quantité théorique par sa moyenne calculée sur les observations.

Cette idée sera au cœur de nombreuses méthodes de **Monte-Carlo**.

Exemple : estimer une probabilité

Soit A un événement.

On introduit sa fonction indicatrice :

$$\mathbf{1}_A = \begin{cases} 1 & \text{si } A \text{ est réalisé,} \\ 0 & \text{sinon.} \end{cases}$$

On a :

$$\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A).$$

Si $X_1, \dots, X_n$ sont des réalisations indépendantes, alors :

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_A(X_i).$$

La probabilité est estimée par la fréquence empirique de réalisation de l'événement.

Exemple : loi de Bernoulli

Soit $X \sim \mathcal{B}(p)$.

On observe : $X_1, \dots, X_n$.

Comme

$$\mathbb{E}(X) = p,$$

on utilise :

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Par exemple, pour

$$(X_1, \dots, X_{10}) = (1, 0, 0, 1, 0, 1, 0, 0, 0, 0),$$

on obtient :

$$\hat{p}_{10} = \frac{3}{10} = 0.3.$$

Comment juger un estimateur ?

Un estimateur peut être étudié à travers plusieurs propriétés notamment :

- Biais
- Variance

Biais d'un estimateur

Définition

Soit $\hat{\theta}$ un estimateur d'un paramètre inconnu θ . Le **biais** de $\hat{\theta}$ est défini par :

$$B(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$$

Le biais permet de mesurer l'écart entre l'espérance de l'estimateur et la vraie valeur du paramètre.

- Si $B(\hat{\theta}) = 0$, l'estimateur est dit **sans biais**.
- Si $B(\hat{\theta}) > 0$, l'estimateur a tendance à surestimer θ .
- Si $B(\hat{\theta}) < 0$, l'estimateur a tendance à sous-estimer θ .

Variance d'un estimateur

Définition

La **variance** d'un estimateur $\hat{\theta}$ est définie par :

$$\text{Var}(\hat{\theta}) = \mathbb{E} \left[\left(\hat{\theta} - \mathbb{E}[\hat{\theta}] \right)^2 \right] = \mathbb{E}[\hat{\theta}^2] - \left(\mathbb{E}[\hat{\theta}] \right)^2$$

La variance permet de mesurer la dispersion des valeurs de l'estimateur autour de son espérance.

- Une variance faible indique que les valeurs de l'estimateur sont peu dispersées.
- Une variance élevée indique que les valeurs de l'estimateur sont plus dispersées.

Biais et variance : exemple de Bernoulli

Pour

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad X_i \sim \mathcal{B}(p),$$

on a :

$$\mathbb{E}(\hat{p}_n) = p.$$

L'estimateur est donc sans biais.

De plus :

$$\text{Var}(\hat{p}_n) = \frac{p(1-p)}{n}.$$

Ainsi, lorsque n augmente :

$$\boxed{\text{Var}(\hat{p}_n) \longrightarrow 0}$$

Erreur quadratique moyenne

Pour évaluer la qualité d'un estimateur $\hat{\theta}$ de θ , on peut considérer son erreur quadratique moyenne :

$$\text{EQM}(\hat{\theta}) = \mathbb{E} \left[(\hat{\theta} - \theta)^2 \right]$$

$$\text{EQM}(\hat{\theta}) = \text{Var}(\hat{\theta}) + B(\hat{\theta})^2$$

- Le biais mesure l'écart entre l'espérance de l'estimateur et le paramètre.
- La variance mesure la dispersion de l'estimateur.

Un estimateur de bonne qualité possède une erreur quadratique moyenne faible.

Du calcul théorique à Monte-Carlo

L'idée de Monte-Carlo est très générale.

Supposons que la quantité recherchée puisse s'écrire :

$$\theta = \mathbb{E}(g(X)).$$

On simule :

$$X_1, \dots, X_n$$

puis on construit :

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n g(X_i)$$

Principe de Monte-Carlo

Remplacer une quantité théorique par une moyenne calculée sur un grand nombre de simulations.

Monte-Carlo : estimer une probabilité

Considérons deux dés équilibrés.

On souhaite estimer :

$$p = \mathbb{P}(D_1 + D_2 = 7).$$

Théoriquement :

$$p = \frac{6}{36} = \frac{1}{6}.$$

On peut simuler un grand nombre de lancers et compter la fréquence d'apparition de la somme 7.

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{D_{1,i} + D_{2,i} = 7\}}.$$

En Python, l'idée essentielle est simplement de calculer la `mean` des indicatrices.

Monte-Carlo : estimer une espérance

Supposons :

$$X \sim \mathcal{U}([0, 1]).$$

On sait théoriquement que :

$$\mathbb{E}(X) = \frac{1}{2}.$$

Après simulation de n réalisations :

$$X_1, \dots, X_n,$$

on estime cette espérance par :

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Le principe général de Monte-Carlo

De nombreux problèmes peuvent être ramenés à une espérance :

$$\theta = \mathbb{E}(g(X))$$

On procède alors en trois étapes :

- 1 Simuler $X_1, \dots, X_n$;
- 2 calculer $g(X_1), \dots, g(X_n)$;
- 3 prendre la moyenne.

On obtient :

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n g(X_i)$$

Estimer une fonction de répartition

Soit X une variable aléatoire de fonction de répartition :

$$F_X(x) = \mathbb{P}(X \leq x).$$

En utilisant une indicatrice :

$$F_X(x) = \mathbb{E} \left[\mathbf{1}_{\{X \leq x\}} \right].$$

On obtient naturellement l'estimateur :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}$$

$\hat{F}_n(x)$ est la proportion des observations inférieures ou égales à x .

Fonction de répartition empirique

Exemple

Pour

$$(0.5, 1.2, 0.8, 1.7, 1.0),$$

on a :

$$\hat{F}_5(1) = \frac{3}{5} = 0.6.$$

Deux estimateurs d'un même paramètre

Revenons à : $X \sim \mathcal{B}(p)$.

On peut considérer deux estimateurs de p .

Estimateur 1 : moyenne empirique

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Estimateur 2 : première observation

$$\tilde{p} = X_1.$$

Les deux sont sans biais :

$$\mathbb{E}(\hat{p}_n) = \mathbb{E}(\tilde{p}) = p.$$

Si les deux estimateurs sont sans biais, lequel choisir ?

Comparer les deux estimateurs

Pour la moyenne empirique :

$$\text{Var}(\hat{p}_n) = \frac{p(1-p)}{n}.$$

Pour le second estimateur :

$$\text{Var}(\tilde{p}) = p(1-p).$$

Ainsi :

$$\frac{\text{Var}(\tilde{p})}{\text{Var}(\hat{p}_n)} = n.$$

Conclusion

La moyenne empirique exploite l'ensemble des observations et devient beaucoup moins fluctuante lorsque n augmente.

Comparer par simulation

La simulation permet également de comparer concrètement les estimateurs.

Pour une même valeur de p et un même nombre d'observations :

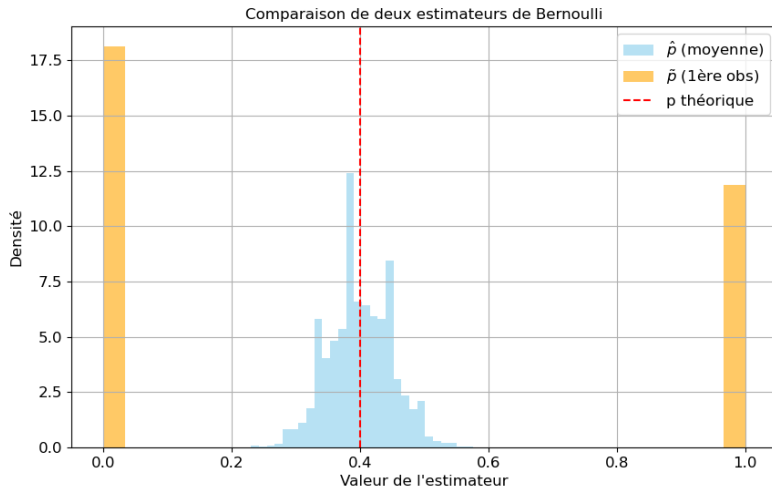
- on répète l'expérience un grand nombre de fois ;
- on calcule $\hat{p}_n$ et $\tilde{p}$;
- on observe leur dispersion.

On s'attend à observer :

$$\text{Var}(\hat{p}_n) = \frac{p(1-p)}{n} < p(1-p) = \text{Var}(\tilde{p}).$$

La simulation permet d'illustrer les propriétés théoriques des estimateurs.

Comparer par simulation



Loi des grands nombres et convergence des estimateurs

De l'estimation à la convergence

On cherche à estimer un paramètre inconnu θ à partir d'un échantillon $X_1, \dots, X_n$.

Un estimateur de θ est une variable aléatoire de la forme

$$\hat{\theta}_n = g(X_1, \dots, X_n).$$

Lorsque la taille de l'échantillon augmente, on espère que l'estimateur se rapproche de la vraie valeur θ .

Question

Comment traduire mathématiquement cette idée de rapprochement ?

Convergence en probabilité

Définition

Une suite de variables aléatoires (X_n) converge en probabilité vers une variable aléatoire X si, pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

On note alors

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X.$$

Interprétation : lorsque n devient grand, la probabilité que X_n s'écarte de X de plus de ε devient arbitrairement petite.

Application aux estimateurs

Pour un estimateur $\hat{\theta}_n$ d'un paramètre θ , la convergence en probabilité s'écrit

$$\forall \varepsilon > 0, \quad \mathbb{P} \left(|\hat{\theta}_n - \theta| > \varepsilon \right) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Cela signifie que, lorsque la taille de l'échantillon augmente, l'estimateur a une probabilité de plus en plus faible de s'écarter sensiblement de la vraie valeur du paramètre.

Objectif

Trouver des résultats qui permettent de garantir cette convergence.

La moyenne empirique

Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et de même loi que X , d'espérance μ .

La moyenne empirique est définie par

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Elle permet notamment d'estimer l'espérance μ de la loi.

Lorsque n augmente, la moyenne empirique $\bar{X}_n$ se rapproche-t-elle de μ ?

La loi des grands nombres

Théorème : loi faible des grands nombres

Soient $(X_n)_{n \geq 1}$ des variables aléatoires indépendantes et de même loi, telles que

$$\mathbb{E}(|X_1|) < +\infty.$$

Alors

$$\overline{X}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mathbb{E}(X_1).$$

Théorème : loi forte des grands nombres

Sous les mêmes hypothèses,

$$\overline{X}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}(X_1).$$

Interprétation de la loi des grands nombres

Considérons une suite de variables de Bernoulli indépendantes de paramètre p :

$$X_i = \begin{cases} 1 & \text{si le succès se produit,} \\ 0 & \text{sinon.} \end{cases}$$

On a $\mathbb{E}(X_i) = p$.

La moyenne empirique

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

représente la fréquence observée des succès.

D'après la loi des grands nombres,

$$\boxed{\overline{X}_n \xrightarrow{\mathbb{P}} p.}$$

Lorsque le nombre d'expériences augmente, la fréquence observée se rapproche de la probabilité du succès.

Application : estimation par Monte-Carlo

On souhaite estimer

$$\theta = \mathbb{E}[g(X)].$$

Soient $X_1, \dots, X_n$ des simulations indépendantes de même loi que X . On considère l'estimateur

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n g(X_i).$$

Si $\mathbb{E}[|g(X)|] < +\infty$, la loi des grands nombres donne

$$\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mathbb{E}[g(X)] = \theta.$$

Application : estimation par Monte-Carlo

Exemple : si $X \sim \mathcal{U}([0, 1])$, alors

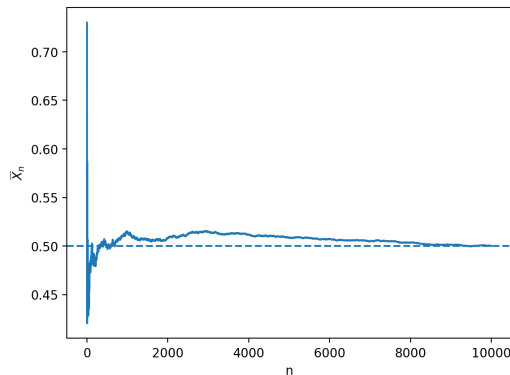
$$\mathbb{E}[X] = \frac{1}{2}, \quad \overline{X}_n \xrightarrow{\mathbb{P}} \frac{1}{2}.$$

On peut donc estimer $\frac{1}{2}$ en simulant un grand nombre de variables uniformes et en calculant leur moyenne.

Illustration numérique de la convergence

On simule des variables aléatoires indépendantes $X_i \sim \mathcal{U}([0, 1])$ et on calcule les moyennes successives $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

La figure représente l'évolution de $\bar{X}_n$ en fonction de n .



Un résultat simple avec la variance

Supposons que $X_1, \dots, X_n$ soient indépendantes, de même loi, d'espérance μ et de variance finie σ^2 .

Pour la moyenne empirique

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

on a

$$\mathbb{E}[\bar{X}_n] = \mu$$

et, par indépendance,

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}.$$

Ainsi,

$$\boxed{\text{Var}(\bar{X}_n) \xrightarrow{n \rightarrow +\infty} 0.}$$

Inégalité de Bienaymé–Tchebychev

Rappel

Pour toute variable aléatoire Y d'espérance $\mathbb{E}[Y]$ et de variance finie,

$$\mathbb{P}(|Y - \mathbb{E}[Y]| \geq \varepsilon) \leq \frac{\text{Var}(Y)}{\varepsilon^2}, \quad \varepsilon > 0.$$

Appliquons cette inégalité à la moyenne empirique :

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2}.$$

Or, $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$.

Par conséquent,

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0.$$

De la loi des grands nombres à la consistance

Définition

Un estimateur $\hat{\theta}_n$ d'un paramètre θ est dit **convergent** ou **consistant** si

$$\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta.$$

La loi des grands nombres permet notamment d'établir la consistance de la moyenne empirique utilisée pour estimer une espérance.

En effet, si $\mathbb{E}[|X|] < +\infty$, alors

$$\bar{X}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mathbb{E}[X].$$

La moyenne empirique est donc un estimateur consistant de $\mathbb{E}[X]$.

Mais la loi des grands nombres a des hypothèses

La loi des grands nombres ne s'applique pas à toutes les lois de probabilité.

Pour la version présentée ici, on suppose notamment que

$$\mathbb{E}[|X|] < +\infty.$$

Question

Que se passe-t-il si une variable aléatoire n'admet pas d'espérance finie ?

Pour étudier cette situation, considérons la loi de Cauchy, dont les queues sont particulièrement lourdes.

La loi de Cauchy

Une variable aléatoire X suit une loi de Cauchy centrée si elle admet pour densité

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

Cette loi présente deux particularités :

- son espérance n'existe pas ;
- sa variance n'est pas finie.

En particulier,

$$\mathbb{E}[|X|] = +\infty.$$

La condition d'intégrabilité nécessaire à la loi des grands nombres présentée précédemment n'est donc pas vérifiée.

La moyenne empirique d'une loi de Cauchy

Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes suivant une loi de Cauchy centrée.

On considère leur moyenne empirique :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

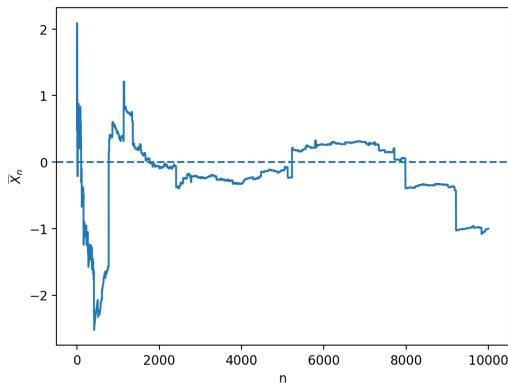
Question

La moyenne empirique se rapproche-t-elle d'une valeur fixe lorsque n augmente ?

Pour répondre à cette question, nous allons observer le comportement de cette moyenne par simulation.

Simulation : moyenne d'une loi de Cauchy

On simule des variables indépendantes suivant une loi de Cauchy centrée, puis on calcule les moyennes successives $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.



Pourquoi la loi de Cauchy est-elle différente ?

La loi de Cauchy possède des queues lourdes : des valeurs extrêmes peuvent apparaître avec une probabilité non négligeable.

Ces valeurs peuvent fortement influencer la moyenne empirique.

Contrairement au cas d'une variable intégrable, on ne peut pas appliquer ici la loi des grands nombres présentée précédemment, car

$$\mathbb{E}[|X|] = +\infty.$$

Plus précisément, la moyenne de variables indépendantes de loi de Cauchy centrée suit encore une loi de Cauchy centrée :

$$\overline{X}_n \sim \mathcal{C}(0, 1).$$

Sa distribution ne se concentre donc pas autour d'une constante lorsque n augmente.

TCL, fluctuations et intervalles de confiance

De la convergence aux fluctuations

Dans la partie précédente, nous avons établi que, sous certaines hypothèses,

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta.$$

La loi des grands nombres nous renseigne donc sur la **limite** de l'estimateur.

Cependant, elle ne décrit pas précisément les fluctuations de $\hat{\theta}_n$ autour de θ .

Pour un échantillon de taille finie n , quelle est l'amplitude de l'erreur

$$\hat{\theta}_n - \theta ?$$

À quelle vitesse l'erreur diminue-t-elle ?

Considérons la moyenne empirique

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

où les variables X_i sont indépendantes, de même loi, d'espérance μ et de variance finie σ^2 .

On a

$$\mathbb{E}(\bar{X}_n) = \mu \quad \text{et} \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Son écart-type vaut donc

$$\boxed{\frac{\sigma}{\sqrt{n}}}.$$

Les fluctuations de la moyenne empirique sont de l'ordre de $\frac{1}{\sqrt{n}}$.

Pourquoi normaliser par $\sqrt{n}$?

L'erreur de la moyenne empirique est $\bar{X}_n - \mu$.

Son ordre de grandeur est $\frac{\sigma}{\sqrt{n}}$.

Pour obtenir une quantité dont les fluctuations ne s'annulent pas lorsque n augmente, on multiplie par $\sqrt{n}$:

$$\sqrt{n}(\bar{X}_n - \mu).$$

Puis on normalise par σ , supposée strictement positive :

$$\boxed{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}}.$$

La normalisation par $\sqrt{n}$ permet d'étudier les fluctuations de la moyenne à une échelle non dégénérée.

Le théorème central limite

Soient $X_1, X_2, \dots$ des variables aléatoires indépendantes et de même loi, telles que

$$\mathbb{E}(X_1) = \mu, \quad \text{Var}(X_1) = \sigma^2, \quad 0 < \sigma^2 < +\infty.$$

Le théorème central limite affirme que

$$\boxed{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)}$$

où

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Pour n suffisamment grand, la distribution de la moyenne empirique, une fois centrée et normalisée, est approximativement une loi normale centrée réduite.

Conséquence du TCL pour la moyenne

Pour n suffisamment grand,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \approx Z, \quad Z \sim \mathcal{N}(0, 1).$$

On en déduit l'approximation

$$\bar{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Ainsi, l'erreur typique de la moyenne empirique est de l'ordre de

$$\frac{\sigma}{\sqrt{n}}.$$

Cette approximation permet de construire des intervalles de confiance pour l'espérance μ .

Illustration numérique du TCL

Considérons des variables aléatoires indépendantes

$$X_i \sim \mathcal{U}([0, 1]).$$

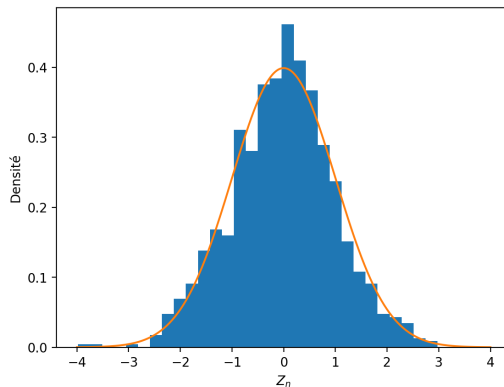
On a

$$\mu = \frac{1}{2}, \quad \sigma^2 = \frac{1}{12}.$$

Pour chaque répétition de l'expérience, on calcule la moyenne empirique puis la variable normalisée

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}.$$

Illustration numérique du TCL



Lorsque n est suffisamment grand, l'histogramme des valeurs de Z_n se rapproche de la densité de la loi normale centrée réduite.

De la loi normale à une probabilité

Soit

$$Z \sim \mathcal{N}(0, 1).$$

On sait que

$$\mathbb{P}(-1.96 \leq Z \leq 1.96) \approx 0.95.$$

D'après le TCL, pour n suffisamment grand,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \approx Z.$$

Par conséquent,

$$\mathbb{P}\left(-1.96 \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq 1.96\right) \approx 0.95.$$

Cette approximation constitue le point de départ de la construction d'un intervalle de confiance pour μ .

Construire un intervalle de confiance

À partir de

$$-1.96 \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq 1.96,$$

on obtient

$$-1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq 1.96 \frac{\sigma}{\sqrt{n}}.$$

En isolant μ , on trouve

$$\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}.$$

On obtient ainsi l'intervalle de confiance asymptotique à 95 % :

$$IC_{95\%} = \left[\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right].$$

Cette formule suppose notamment que σ est connue.

Comment interpréter un intervalle de confiance ?

Un intervalle de confiance à 95 % ne signifie pas qu'après avoir calculé l'intervalle, le paramètre μ a une probabilité de 95 % d'y appartenir.

Interprétation correcte

Le paramètre μ est fixe, tandis que l'intervalle dépend de l'échantillon.

Si l'on répétait un grand nombre de fois la procédure de construction de l'intervalle, environ 95 % des intervalles obtenus contiendraient μ .

Avant l'observation des données, l'intervalle est aléatoire. Après observation, il devient un intervalle numérique.

Application : loi de Bernoulli

Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes de loi de Bernoulli de paramètre p : $X_i \sim \mathcal{B}(p)$.

L'estimateur de p est la fréquence empirique $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

On a

$$\mathbb{E}(X_i) = p, \quad \text{Var}(X_i) = p(1 - p).$$

Le TCL donne

$$\frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{p(1 - p)}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Difficulté

La variance dépend de p , qui est précisément le paramètre inconnu que nous cherchons à estimer.

Intervalle de confiance pour une proportion

Pour n suffisamment grand, on utilise l'approximation

$$\hat{p}_n \approx \mathcal{N} \left(p, \frac{p(1-p)}{n} \right).$$

Comme p est inconnu, on remplace p par son estimateur $\hat{p}_n$ dans la formule de l'écart-type.

On obtient l'intervalle de Wald à 95 % :

$$IC_{95\%} = \left[\hat{p}_n - 1.96 \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + 1.96 \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \right].$$

Cet intervalle est une approximation asymptotique. Il peut être imprécis lorsque l'échantillon est petit ou lorsque $\hat{p}_n$ est proche de 0 ou de 1.

Exemple numérique

Supposons que l'on observe $n = 100$ essais de Bernoulli et que la fréquence observée soit

$$\hat{p} = 0.40.$$

L'erreur-type estimée vaut

$$\sqrt{\frac{0.4(1 - 0.4)}{100}} \approx 0.049.$$

L'intervalle de Wald à 95 % est donc approximativement

$$0.40 \pm 1.96 \times 0.049.$$

Ainsi,

$$IC_{95\%} \approx [0.304; 0.496].$$

Cet intervalle est obtenu à partir de la fréquence observée pour estimer la proportion inconnue.

Effet de la taille de l'échantillon

Pour la moyenne, la marge d'erreur de l'intervalle de confiance à 95 % est

$$1.96 \frac{\sigma}{\sqrt{n}}.$$

Ainsi,

$$n \uparrow \implies \text{marge d'erreur} \downarrow .$$

Si l'on multiplie la taille de l'échantillon par 4, alors

$$\sqrt{4n} = 2\sqrt{n}.$$

La marge d'erreur est donc divisée par 2.

Pour diviser approximativement la marge d'erreur par 2, il faut multiplier la taille de l'échantillon par 4.

Simulation d'intervalles de confiance

On considère une population dans laquelle la proportion recherchée vaut $p = 0.4$.

- ① On prélève un échantillon de taille $n = 50$: $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{B}(p)$.
- ② On calcule la fréquence observée :

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i.$$

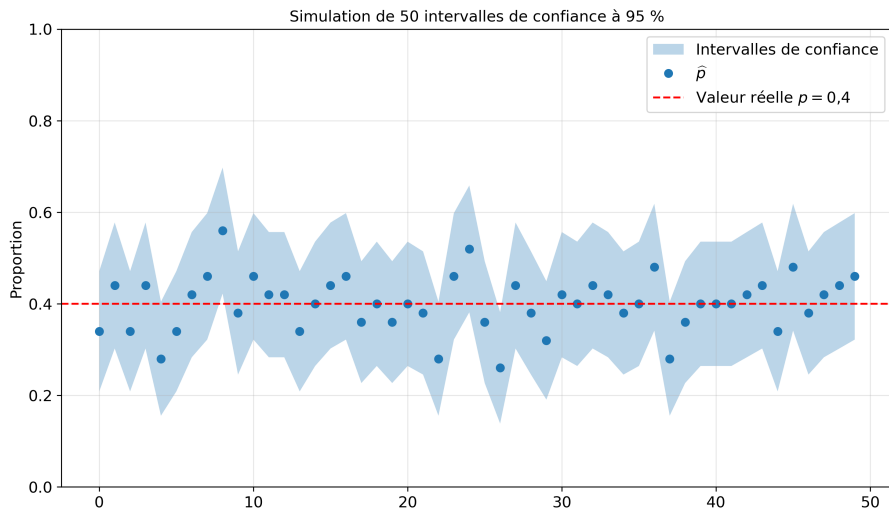
- ③ On construit l'intervalle de Wald à 95 % :

$$IC = \left[\hat{p} - 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right].$$

- ④ On répète la procédure $N_{\text{sim}} = 1000$ fois.

Quelle proportion des intervalles obtenus contient-elle la vraie valeur $p = 0.4$?

Visualisation des intervalles de confiance



Comment interpréter le graphique ?

Pour $N_{\text{sim}} = 1000$ simulations, supposons que 932 intervalles contiennent la vraie proportion $p = 0.4$.

Le taux de couverture observé est alors

$$\frac{932}{1000} = 0.932 = 93.2 \, \%.$$

Un intervalle qui ne contient pas la ligne rouge ne couvre pas la vraie valeur p .

Le taux de couverture observé peut différer du niveau théorique de 95 % en raison de la variabilité des simulations et du caractère approximatif de l'intervalle de Wald.

Loi de Pareto : exemple

La loi de Pareto est une loi à densité utilisée pour modéliser des phénomènes présentant des valeurs extrêmes, comme certaines fortunes, tailles de sinistres ou durées.

On dit que X suit une loi de Pareto de paramètres $x_m > 0$ et $\alpha > 0$ si sa densité est

$$f_X(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}}, & x \geq x_m, \\ 0, & x < x_m. \end{cases}$$

Sa fonction de répartition est

$$F_X(x) = \begin{cases} 0, & x < x_m, \\ 1 - \left(\frac{x_m}{x}\right)^\alpha, & x \geq x_m. \end{cases}$$

Espérance et variance de la loi de Pareto

Pour $X \sim \mathcal{P}(x_m, \alpha)$, on a

$$\mathbb{E}[X] = \begin{cases} \frac{\alpha x_m}{\alpha - 1}, & \alpha > 1, \\ +\infty, & 0 < \alpha \leq 1. \end{cases}$$

Lorsque $\alpha > 1$, la variance vaut

$$\text{Var}(X) = \begin{cases} \frac{\alpha x_m^2}{(\alpha - 1)^2(\alpha - 2)}, & \alpha > 2, \\ +\infty, & 1 < \alpha \leq 2. \end{cases}$$

Le paramètre α détermine l'existence des moments. En particulier, une espérance finie n'implique pas nécessairement une variance finie.

Objectif de l'expérience

On souhaite étudier par simulation le comportement de la moyenne empirique

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

dans trois situations différentes.

On choisit $x_m = 1$ et trois valeurs du paramètre α :

Modèle	α	Espérance	Variance
1	0.8	Infinie	Infinie
2	1.5	3	Infinie
3	3	1.5	0.75

Protocole expérimental

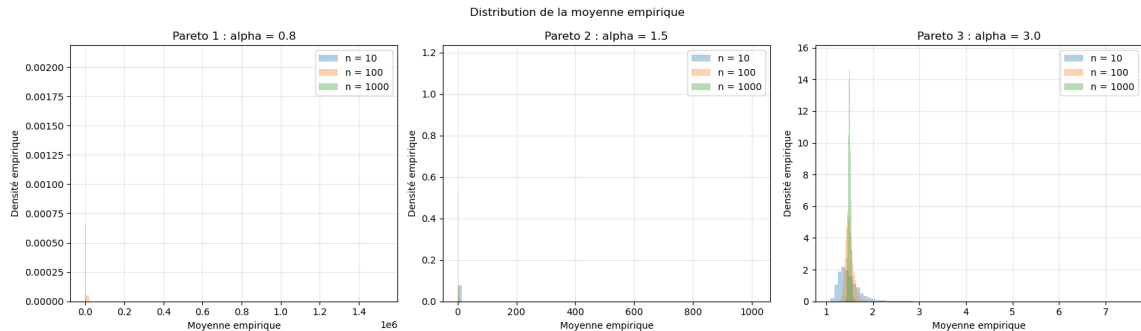
Pour chaque modèle, on effectue les opérations suivantes.

- ① Générer une suite de variables aléatoires indépendantes $X_1, \dots, X_N$ suivant la loi de Pareto choisie.
- ② Calculer les moyennes empiriques successives

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

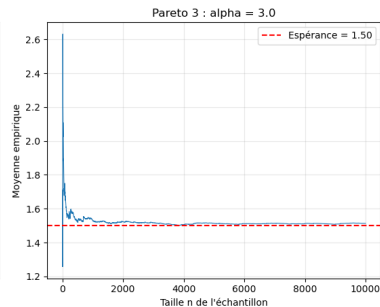
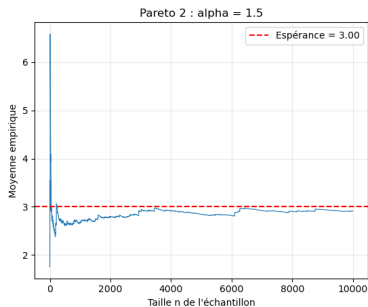
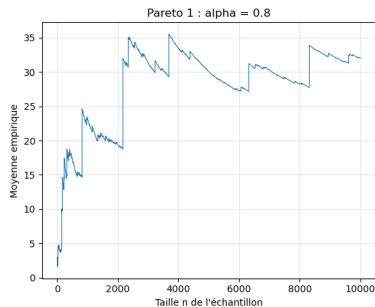
- ③ Répéter l'expérience M fois, indépendamment, pour plusieurs tailles d'échantillon n .
- ④ Étudier les distributions des moyennes empiriques, puis calculer les indicateurs empiriques de performance.

Distribution des valeurs

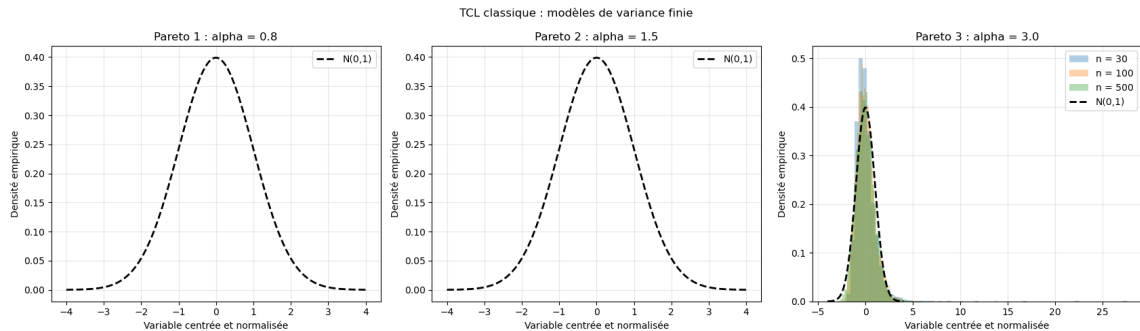


Loi des grands nombres

Loi des grands nombres : moyennes empiriques



TCL



Biais, variance et EQM : résultats pour $n = 1000$

Modèle 1 : $\alpha = 0.8$

	Théorique
Espérance	$+\infty$
Variance	$+\infty$
Biais	Non défini
EQM	Non défini

Valeurs empiriques

Variance	8.8303×10^6
Biais	Non calculé
EQM	Non calculé

Modèle 2 : $\alpha = 1.5$

	Théorique
Espérance	3
Variance	$+\infty$
Biais	0
EQM	$+\infty$

Valeurs empiriques

Biais	-0.0035006
Variance	0.6019
EQM	0.60179

Modèle 3 : $\alpha = 3$

	Théorique
Espérance	1.5
Variance	0.75
Biais	0
EQM	0.00075

Valeurs empiriques

Biais	-0.00027322
Variance	0.00075146
EQM	0.00075138

Chaînes de Markov

Pourquoi des chaînes de Markov ?

Dans de nombreux systèmes, l'état futur dépend de l'état actuel.

Exemples :

- météo : beau / pluie ;
- état d'une machine : normal / dégradé / panne ;
- file d'attente : nombre de clients ;
- évolution d'un stock ;
- position d'un joueur.

L'idée :

Le futur dépend du présent, et non de tout le passé.

Un exemple : la météo

On considère deux états :

$$E = \{\text{Beau}, \text{Pluie}\}.$$

On suppose que la météo de demain dépend uniquement de celle d'aujourd'hui.

Par exemple :

	Beau demain	Pluie demain
Beau aujourd'hui	0.8	0.2
Pluie aujourd'hui	0.4	0.6

Cette information décrit les **transitions** entre les états.

Propriété de Markov

Une suite de variables aléatoires $(X_n)_{n \geq 0}$ est une chaîne de Markov si, pour tout $n \geq 0$,

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, \dots, X_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

Autrement dit :

Connaître le présent suffit pour déterminer les probabilités du futur.

Espace des états

On note

$$E = \{1, \dots, m\}$$

l'ensemble des états possibles.

À chaque instant n , le système se trouve dans un état :

$$X_n \in E.$$

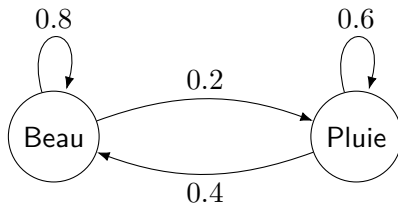
Dans l'exemple de la météo :

$$E = \{\text{Beau}, \text{Pluie}\}.$$

Représentation graphique des transitions

On peut représenter les transitions entre les états par un graphe orienté.

Pour notre exemple :



Chaque flèche indique une transition possible, et son étiquette donne la probabilité correspondante.

La matrice de transition

Les probabilités de transition sont regroupées dans une matrice :

$$P = (p_{ij})$$

où

$$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i).$$

Pour notre exemple :

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix}$$

La ligne i correspond à l'état de départ et la colonne j à l'état d'arrivée.

Propriétés d'une matrice de transition

Chaque coefficient vérifie :

$$p_{ij} \geq 0.$$

De plus, depuis un état i , le système doit nécessairement aller vers l'un des états possibles.

Ainsi, pour tout i :

$$\sum_{j=1}^m p_{ij} = 1.$$

Chaque ligne de P est donc une distribution de probabilités.

Graphe ou matrice ?

Les deux représentations contiennent la même information.

- Le **graphe** donne une vision intuitive des transitions entre les états.
- La **matrice** permet d'effectuer les calculs et de déterminer l'évolution de la distribution.

Pour une chaîne à m états, la matrice de transition s'écrit :

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix}.$$

Simuler une trajectoire

Une trajectoire est une réalisation de la suite :

$$X_0, X_1, X_2, \dots$$

Par exemple :

Beau $\rightarrow$ Beau $\rightarrow$ Pluie $\rightarrow$ Pluie $\rightarrow$ Beau $\rightarrow \dots$

À chaque étape, on choisit le prochain état selon les probabilités de la ligne correspondante de P .

En Python, on peut utiliser :

```
np.random.choice(..., p=...)
```

La distribution du système

Une trajectoire représente une réalisation particulière du système.

On peut aussi s'intéresser aux probabilités des différents états à un instant donné.

À l'instant n , on définit la distribution :

$$\mu_n = (\mathbb{P}(X_n = 1), \dots, \mathbb{P}(X_n = m)).$$

Elle vérifie :

$$\mathbb{P}(X_n = i) \geq 0$$

et

$$\sum_{i=1}^m \mathbb{P}(X_n = i) = 1.$$

Ainsi, μ_n est un vecteur de probabilités.

Évolution de la distribution

La matrice de transition permet de calculer la distribution à l'instant suivant.

$$\mu_{n+1} = \mu_n P$$

Par conséquent :

$$\mu_1 = \mu_0 P,$$

$$\mu_2 = \mu_0 P^2,$$

et, plus généralement :

$$\mu_n = \mu_0 P^n.$$

Exemple : évolution de la météo

Supposons qu'aujourd'hui il fasse certainement beau :

$$\mu_0 = (1, 0).$$

Avec

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix},$$

on obtient :

$$\mu_1 = \mu_0 P = (0.8, 0.2).$$

Puis :

$$\mu_2 = \mu_1 P = (0.72, 0.28).$$

La distribution évolue donc au cours du temps.

Vers un régime stable

On peut alors poser la question :

Que devient la distribution lorsque n devient grand ?

Dans certains systèmes, il existe une distribution π telle que

$$\mu_n \longrightarrow \pi.$$

Une telle limite, lorsqu'elle existe, est nécessairement une distribution stationnaire.

Une distribution stationnaire est une distribution qui reste inchangée après une transition.

Distribution stationnaire

Une distribution $\pi = (\pi_1, \dots, \pi_m)$ est **stationnaire** si

$$\pi P = \pi.$$

Elle doit également être une distribution de probabilités :

$$\pi_i \geq 0 \quad \text{et} \quad \sum_{i=1}^m \pi_i = 1.$$

On cherche donc π en résolvant :

$$\begin{cases} \pi P = \pi, \\ \pi_i \geq 0, \quad i = 1, \dots, m, \\ \sum_{i=1}^m \pi_i = 1. \end{cases}$$

Calcul de la distribution stationnaire

Pour la météo :

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix}.$$

On cherche :

$$\pi = (\pi_1, \pi_2).$$

La condition $\pi P = \pi$ donne :

$$\begin{cases} 0.8\pi_1 + 0.4\pi_2 = \pi_1, \\ 0.2\pi_1 + 0.6\pi_2 = \pi_2. \end{cases}$$

On ajoute la condition de normalisation :

$$\pi_1 + \pi_2 = 1.$$

Résultat pour la météo

En résolvant le système précédent, on obtient :

$$\pi = \left(\frac{2}{3}, \frac{1}{3} \right).$$

Ainsi, dans le régime stationnaire :

$$\mathbb{P}(\text{Beau}) = \frac{2}{3},$$

$$\mathbb{P}(\text{Pluie}) = \frac{1}{3}.$$

La distribution stationnaire attribue une probabilité de $\frac{2}{3}$ au beau temps.

Distribution à un instant ou à long terme ?

Il faut distinguer deux notions.

La distribution à l'instant n :

Elle décrit les probabilités des différents états à cet instant. Elle dépend de la distribution initiale et du nombre de transitions.

La distribution stationnaire :

Elle décrit une distribution qui reste inchangée après une transition :

$$\pi P = \pi.$$

La distribution du système peut converger vers une distribution stationnaire, mais cette convergence n'est pas automatique.

Fréquence des visites

On peut également observer une seule trajectoire :

$$X_0, X_1, \dots, X_N.$$

La fréquence de visite de l'état i est définie par :

$$f_N(i) = \frac{1}{N} \sum_{k=1}^N \mathbf{1}_{\{X_k=i\}}.$$

où

$$\mathbf{1}_{\{X_k=i\}} = \begin{cases} 1, & X_k = i, \\ 0, & X_k \neq i. \end{cases}$$

Fréquences et comportement à long terme

Pour une chaîne de Markov finie et irréductible, les fréquences de visite convergent presque sûrement vers les probabilités de la distribution stationnaire :

$$f_N(i) \longrightarrow \pi_i \quad \text{presque sûrement.}$$

Ainsi, π_i représente la proportion de temps passée dans l'état i à long terme.

Dans notre exemple :

$$f_N(\text{Beau}) \longrightarrow \frac{2}{3},$$

$$f_N(\text{Pluie}) \longrightarrow \frac{1}{3}.$$

Deux notions à ne pas confondre

Probabilité à un instant donné

$$\mathbb{P}(X_n = i)$$

C'est la probabilité que le système soit dans l'état i à l'instant n .

Fréquence de visite

$$f_N(i) = \frac{1}{N} \sum_{k=1}^N \mathbf{1}_{\{X_k=i\}}$$

C'est la proportion de visites dans l'état i au cours des N premières étapes.

La première décrit une probabilité à un instant donné ; la seconde décrit le comportement observé sur une longue trajectoire.

Un exemple de quantité observée

On associe une température à chaque état :

Beau $\longrightarrow 25^{\circ}\text{C}$,

Pluie $\longrightarrow 15^{\circ}\text{C}$.

Sur une trajectoire, on peut calculer la température moyenne :

$$\frac{1}{N} \sum_{k=1}^N T(X_k).$$

Si les fréquences de visites convergent, cette moyenne converge presque sûrement vers :

$$\frac{2}{3}(25) + \frac{1}{3}(15).$$

Donc :

$$\frac{65}{3} \approx 21.67^{\circ}\text{C}.$$

Quand la convergence peut-elle échouer ?

La convergence de la distribution vers une distribution stationnaire n'est pas automatique.

Considérons deux états :

$$E = \{0, 1\}$$

avec la matrice de transition :

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Le système change obligatoirement d'état :

$$0 \rightarrow 1 \rightarrow 0 \rightarrow 1 \rightarrow \dots$$

Un système périodique

Si $X_0 = 0$, alors :

$$X_0 = 0, \quad X_1 = 1, \quad X_2 = 0, \quad X_3 = 1, \dots$$

La probabilité d'être dans l'état 0 à l'instant n vaut :

$$\mathbb{P}(X_n = 0) = \begin{cases} 1, & n \text{ pair,} \\ 0, & n \text{ impair.} \end{cases}$$

La distribution μ_n ne converge donc pas.

Elle oscille indéfiniment.

Pourtant, une distribution stationnaire existe

Cherchons une distribution stationnaire :

$$\pi = (\pi_1, \pi_2)$$

vérifiant :

$$\pi P = \pi \quad \text{et} \quad \pi_1 + \pi_2 = 1.$$

On obtient :

$$\pi = \left(\frac{1}{2}, \frac{1}{2} \right).$$

Pourtant, lorsque $X_0 = 0$:

Existence d'une distribution stationnaire

Pour toute chaîne de Markov à espace d'états fini, il existe au moins une distribution stationnaire.

Cependant, cette existence ne garantit pas que la distribution du système converge vers cette distribution stationnaire.

La convergence

$$\mu_n \longrightarrow \pi$$

est garantie, quelle que soit la distribution initiale, si la chaîne est **irréductible et apériodique**.

Irréductibilité

Une chaîne est **irréductible** si, à partir de n'importe quel état, on peut atteindre n'importe quel autre état en un nombre fini de transitions avec une probabilité strictement positive.

Autrement dit, pour tous $i, j \in E$, il existe $n \geq 0$ tel que

$$(P^n)_{ij} > 0.$$

Il n'existe donc pas de sous-ensemble d'états complètement isolé du reste du système.

Période d'un état

Pour un état i , on considère les temps $n \geq 1$ pour lesquels il est possible de revenir à i :

$$(P^n)_{ii} > 0.$$

La période de i est définie par :

$$d(i) = \text{pgcd}\{n \geq 1 : (P^n)_{ii} > 0\}.$$

- Si $d(i) = 1$, l'état est **apériodique**.
- Si $d(i) > 1$, l'état est **périodique**.

Exemple : une chaîne périodique

Considérons :

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

À partir de l'état 0, on ne peut revenir à 0 qu'après un nombre pair de transitions :

$$2, 4, 6, \dots$$

Ainsi :

$$d(0) = \text{pgcd}(2, 4, 6, \dots) = 2.$$

La chaîne est donc **périodique**.

Apériodicité

Une chaîne irréductible est dite **apériodique** si sa période vaut 1.

$$d = 1$$

Intuitivement, les retours vers un état ne sont pas contraints à se produire selon un rythme périodique.

Par exemple, si un état possède une boucle

$$i \longrightarrow i$$

avec une probabilité strictement positive, alors sa période vaut 1.

Convergence vers la distribution stationnaire

Pour une chaîne de Markov finie, si elle est **irréductible et apériodique**, alors elle possède une unique distribution stationnaire :

$$\pi = (\pi_1, \dots, \pi_m).$$

Cette distribution vérifie :

$$\pi P = \pi, \quad \pi_i \geq 0, \quad \sum_{i=1}^m \pi_i = 1.$$

De plus, quelle que soit la distribution initiale μ_0 :

$$\mu_n = \mu_0 P^n \longrightarrow \pi.$$