

Suites, statistiques et entropie

Trois autres situations concrètes de mathématiques dans l'IA

À savoir avant de commencer

Ce document complète la série avec trois situations indépendantes les unes des autres : la décroissance du taux d'erreur pendant l'entraînement (suites), la détection d'anomalies (statistiques), et la mesure de l'incertitude d'une prédiction (logarithme et entropie). Chaque partie peut être lue séparément.

Niveau : Première → Terminale (chaque partie indique son niveau)

Introduction

Les documents précédents ont montré comment une IA minimise une erreur (documents 0, 4, 5), comment elle représente et compare des mots (document 2), et comment elle raisonne en probabilités (document 1, document 3). Ce document explore trois autres situations, plus ponctuelles mais tout aussi réelles : comment le taux d'erreur d'un modèle évolue au fil de l'entraînement, comment une IA peut repérer une donnée suspecte, et comment on mesure mathématiquement le degré d'incertitude d'une prédiction.

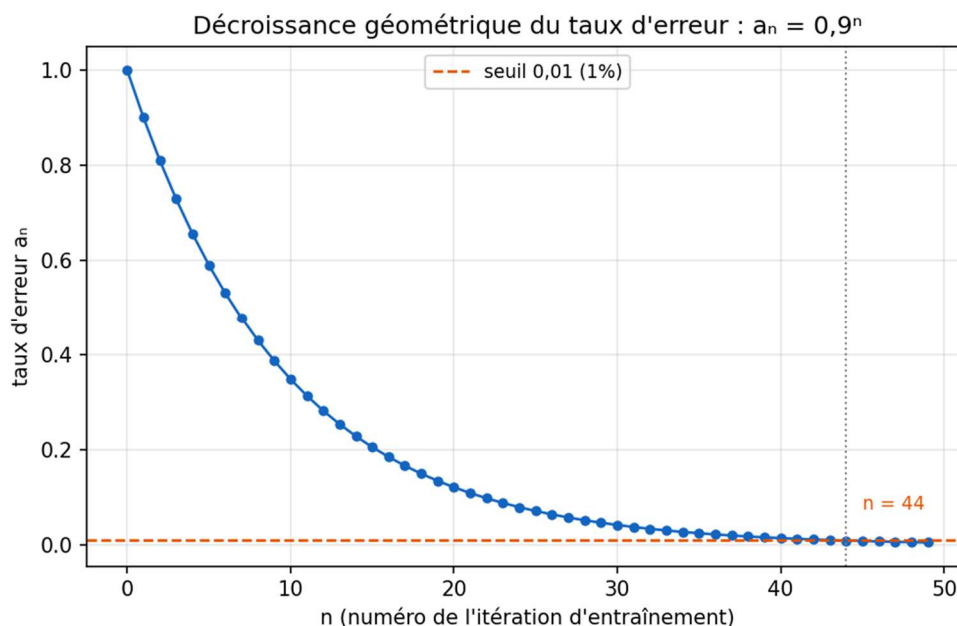
Partie A — Suites : l'IA qui apprend de ses erreurs

Niveau : Première / Terminale

Pendant l'entraînement d'une IA, le taux d'erreur diminue progressivement à chaque itération. On modélise ce taux d'erreur par une suite (a_n) , où n représente le numéro de l'itération d'entraînement (on parle parfois d'« époque »).

On suppose qu'à chaque itération, l'erreur est réduite de 10 % : $a_0 = 1$ (100 % d'erreur au départ, avant tout entraînement), et $a_{n+1} = 0,9 \times a_n$.

1. Quelle est la nature de la suite (a_n) ? Donne son expression explicite en fonction de n .
2. Calcule a_1 , a_2 , a_3 , a_4 et a_5 .
3. Démontre par récurrence que, pour tout entier naturel n , $a_n > 0$.
4. Montre que la suite (a_n) est strictement décroissante.
5. Quelle est la limite de (a_n) quand n tend vers $+\infty$? Interprète ce résultat : le taux d'erreur peut-il un jour atteindre exactement 0 ?



6. Résous l'inéquation $0,9^n < 0,01$ (utilise le logarithme népérien). Au bout de combien d'itérations le taux d'erreur passe-t-il, dans ce modèle, sous la barre des 1 % ? Vérifie ta réponse sur le graphique.

Partie B — Statistiques : détecter une anomalie

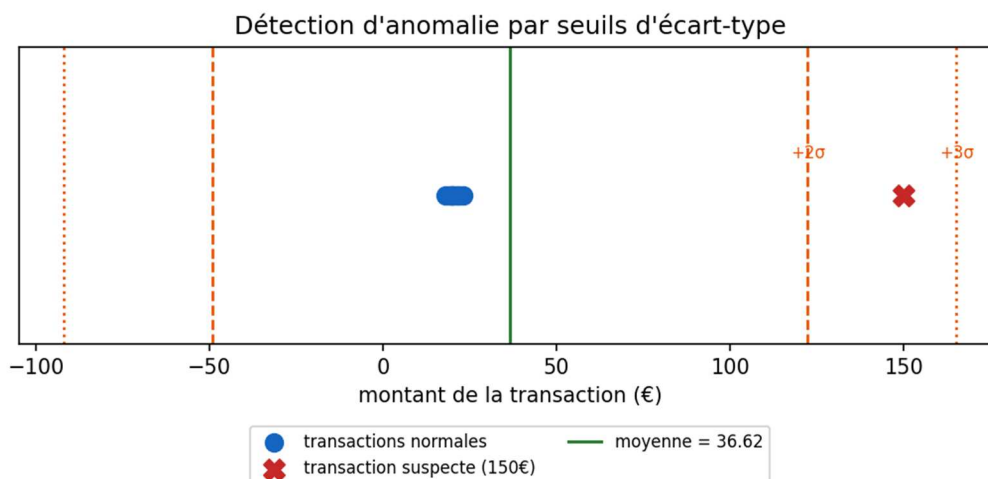
Niveau : Terminale

Une IA de détection de fraude bancaire compare chaque nouvelle transaction à l'ensemble des transactions habituelles d'un client, et signale celles qui s'écartent anormalement de la moyenne. Une méthode simple consiste à calculer la moyenne $\bar{x}$ et l'écart-type σ d'un ensemble de transactions, puis à signaler comme suspecte toute transaction en dehors de l'intervalle $[\bar{x} - k\sigma ; \bar{x} + k\sigma]$, pour une valeur de k choisie (souvent $k=2$ ou $k=3$).

Voici les 8 dernières transactions d'un client (en euros) :

Transaction	1	2	3	4	5	6	7	8
Montant (€)	20	22	19	21	20	23	18	150

- Calcule la moyenne $\bar{x}$ de ces 8 montants.
- Calcule l'écart-type σ de cette série (arrondis à 10^{-2} près).



- Détermine l'intervalle $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$. La transaction de 150 € est-elle signalée comme suspecte avec ce seuil ?
- Détermine l'intervalle $[\bar{x} - 3\sigma ; \bar{x} + 3\sigma]$. La transaction de 150 € est-elle encore signalée comme suspecte avec ce seuil, plus strict ?
- Que remarques-tu en comparant les réponses aux deux questions précédentes ? Explique pourquoi la présence même de la transaction suspecte dans le calcul de σ peut rendre sa détection plus difficile avec un seuil trop strict.

Partie C — Logarithme : mesurer l'incertitude d'une prédiction

Niveau : Terminale

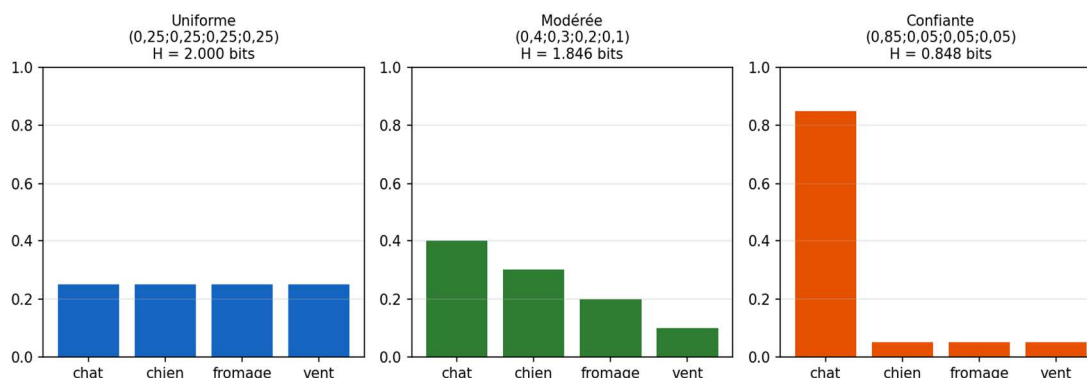
Dans le document 1, une IA prédisait une distribution de probabilité sur le mot suivant. On peut mesurer à quel point cette distribution est « incertaine » (l'IA hésite) ou au contraire « confiante » (l'IA est presque sûre de son choix), à l'aide de l'entropie de Shannon :

Définition (entropie de Shannon)

$$H = - \sum_i p_i \times \log_2(p_i)$$

H se mesure en bits. Plus H est grand, plus la distribution est incertaine (proche d'un choix « au hasard ») ; plus H est petit, plus l'IA est confiante dans un mot en particulier.

On considère trois distributions différentes que pourrait prédire une IA pour le mot suivant, parmi 4 candidats (chat, chien, fromage, vent) :



- Calcule l'entropie H de la distribution uniforme $(0,25 ; 0,25 ; 0,25 ; 0,25)$. (Indice : $\log_2(0,25) = \log_2(1/4) = -2$.)
- Calcule l'entropie de la distribution « modérée » $(0,4 ; 0,3 ; 0,2 ; 0,1)$ (arrondis à 10^{-3} près). Tu pourras utiliser $\log_2(x) = \ln(x)/\ln(2)$.
- Calcule l'entropie de la distribution « confiante » $(0,85 ; 0,05 ; 0,05 ; 0,05)$ (arrondis à 10^{-3} près).
- Classe ces trois entropies dans l'ordre croissant. Ce classement est-il cohérent avec l'idée intuitive de « confiance » de l'IA dans sa prédiction ?
- Pourquoi la distribution uniforme donne-t-elle l'entropie la plus grande possible parmi toutes les distributions sur 4 issues ? (Tu peux répondre intuitivement, sans démonstration.)

Partie D — Ouverture : les nombres complexes pour le son et l'image

Niveau : Terminale (option mathématiques expertes)

Certaines IA traitent des signaux audio (reconnaissance vocale, génération de musique). Un son peut être représenté comme une somme de signaux sinusoidaux de fréquences différentes, ce qu'on appelle une transformée de Fourier — un outil qui repose entièrement sur les nombres complexes, via la formule d'Euler $e^{i\theta} = \cos(\theta) + i \sin(\theta)$.

- Rappelle la forme algébrique du nombre complexe $e^{i\pi/2}$, puis calcule son module.

Ce sujet dépasse largement le programme de terminale, mais il illustre que même les nombres complexes — souvent perçus comme une notion « abstraite » — ont une application concrète dans certaines IA de traitement du signal.

Ce que ces situations simplifient, et ce que fait réellement une IA comme Claude

Ce que ce modèle simplifie	Ce que fait réellement un modèle comme Claude
Un taux d'erreur qui décroît de façon parfaitement régulière (toujours $\times 0,9$). Une détection d'anomalie fondée uniquement sur la moyenne et l'écart-type d'une seule variable. Une entropie calculée sur seulement 4 issues possibles.	Une décroissance de l'erreur réelle, mesurée par la fonction de coût (documents 0, 4, 5), qui est beaucoup plus irrégulière et dépend de très nombreux facteurs (données, architecture, pas d'apprentissage...). Des systèmes de détection d'anomalies réels combinant de nombreuses variables à la fois, souvent eux-mêmes des modèles d'IA entraînés (et pas seulement une formule statistique simple). Une entropie calculée, à chaque étape de génération, sur un vocabulaire de dizaines de milliers de tokens possibles — le même calcul, mais à une tout autre échelle.

Ces trois situations restent de bons points d'entrée pour comprendre des notions réellement utilisées (vitesse de convergence, détection de valeurs aberrantes, mesure d'incertitude) — à l'échelle d'un exercice de lycée plutôt qu'à celle d'un centre de données.

Exercices d'entraînement

Exercice 1 — Une autre suite d'erreur

Niveau : Première / Terminale

On modélise le taux d'erreur d'un autre modèle par $b_0 = 0,8$ et $b_{n+1} = 0,95 \times b_n$.

18. Donne l'expression explicite de b_n en fonction de n .
19. Calcule b_1 , b_2 et b_3 (arrondis à 10^{-3} près).
20. Résous l'inéquation $0,95^n < 0,1$ pour déterminer à partir de quelle itération b_n passe sous 0,08 (soit 10 % de b_0).

Exercice 2 — Anomalie dans un autre jeu de données

Niveau : Terminale

Un capteur relève les températures suivantes (en °C) sur une semaine : 18, 19, 17, 20, 18, 19, 40.

21. Calcule la moyenne et l'écart-type de cette série (arrondis à 10^{-2} près).
22. La valeur 40°C est-elle en dehors de l'intervalle $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$?

Exercice 3 — Entropie d'une pièce truquée

Niveau : Terminale

On lance une pièce truquée : elle donne « pile » avec une probabilité de 0,9, et « face » avec une probabilité de 0,1.

23. Calcule l'entropie $H = -[0,9 \times \log_2(0,9) + 0,1 \times \log_2(0,1)]$ de cette expérience (arrondis à 10^{-3} près).
24. Compare avec l'entropie d'une pièce équilibrée (0,5 ; 0,5), qui vaut exactement 1 bit. Ton résultat est-il cohérent avec l'idée qu'une pièce truquée est plus « prévisible » qu'une pièce équilibrée ?

Corrigé

Partie A

- A.1. (a_n) est une suite géométrique de raison 0,9 et de premier terme $a_0=1$. Donc $a_n = 0,9^n$.
- A.2. $a_1=0,9$; $a_2=0,81$; $a_3=0,729$; $a_4=0,6561$; $a_5=0,59049$.
- A.3. Initialisation : $a_0=1>0$, vrai. Hérité : si $a_n>0$, alors $a_{n+1}=0,9 \times a_n>0$ (produit de deux nombres strictement positifs). Par principe de récurrence, $a_n>0$ pour tout n .
- A.4. $a_{n+1}-a_n = 0,9a_n - a_n = -0,1a_n$. Comme $a_n>0$ (question précédente), on a $a_{n+1}-a_n<0$: la suite est strictement décroissante.
- A.5. Comme $0<0,9<1$, on a $\lim(0,9^n)=0$ quand $n \rightarrow +\infty$. Le taux d'erreur se rapproche indéfiniment de 0 sans jamais l'atteindre exactement dans ce modèle continu — il devient seulement de plus en plus petit.
- A.6. $0,9^n < 0,01 \Leftrightarrow n \times \ln(0,9) < \ln(0,01) \Leftrightarrow n > \ln(0,01)/\ln(0,9)$ (l'inégalité change de sens car $\ln(0,9)<0$) $\Leftrightarrow n > (-4,605)/(-0,105) \approx 43,7$. Comme n est un entier, il faut $n \geq 44$: le taux d'erreur passe sous 1 % à partir de la 44^e itération.

Partie B

- B.1. $\bar{x} = (20+22+19+21+20+23+18+150)/8 = 293/8 = 36,625$.
- B.2. Écart-type (population) : on calcule la moyenne des carrés des écarts à la moyenne. On trouve $\sigma \approx 42,88$.
- B.3. $[\bar{x}-2\sigma ; \bar{x}+2\sigma] \approx [36,625-85,76 ; 36,625+85,76] \approx [-49,13 ; 122,38]$. Comme $150 > 122,38$, la transaction de 150 € est bien signalée comme suspecte avec ce seuil.
- B.4. $[\bar{x}-3\sigma ; \bar{x}+3\sigma] \approx [36,625-128,63 ; 36,625+128,63] \approx [-92 ; 165,26]$. Comme $150 < 165,26$, la transaction de 150 € n'est plus signalée comme suspecte avec ce seuil plus strict !
- B.5. La transaction suspecte elle-même a fortement augmenté l'écart-type (elle est très éloignée des autres valeurs), ce qui élargit l'intervalle « normal » et peut donc, paradoxalement, masquer sa propre anomalie avec un seuil trop strict (3σ). C'est une limite connue de cette méthode simple : une anomalie peut fausser le calcul même de ce qui sert à la détecter.

Partie C

- C.1. $H = -4 \times 0,25 \times \log_2(0,25) = -4 \times 0,25 \times (-2) = 2$ bits.
- C.2. $\log_2(0,4) = \ln(0,4)/\ln(2) \approx -0,916/0,693 \approx -1,322$; $\log_2(0,3) \approx -1,737$; $\log_2(0,2) \approx -2,322$; $\log_2(0,1) \approx -3,322$. $H \approx -(0,4 \times (-1,322) + 0,3 \times (-1,737) + 0,2 \times (-2,322) + 0,1 \times (-3,322)) \approx -(-0,529 - 0,521 - 0,464 - 0,332) \approx 1,846$ bits.
- C.3. $\log_2(0,85) \approx -0,234$; $\log_2(0,05) \approx -4,322$. $H \approx -(0,85 \times (-0,234) + 3 \times 0,05 \times (-4,322)) \approx -(-0,199 - 0,648) \approx 0,848$ bits.
- C.4. Ordre croissant : confiante (0,848) < modérée (1,846) < uniforme (2). C'est cohérent : plus l'IA est confiante dans un mot précis, plus l'entropie (l'incertitude) est faible.

C.5. La distribution uniforme correspond à une absence totale de préférence entre les issues : c'est la situation où l'incertitude est maximale, puisqu'aucune issue n'est privilégiée par rapport aux autres — on ne peut faire aucune prédiction meilleure qu'une autre.

Partie D

D.1. $e^{i\pi/2} = \cos(\pi/2) + i \sin(\pi/2) = 0 + i \times 1 = i$. Son module est $|i| = 1$.

Exercice 1

1. $b_n = 0,8 \times 0,95^n$.

2. $b_1 = 0,8 \times 0,95 = 0,76$; $b_2 = 0,8 \times 0,95^2 = 0,722$; $b_3 = 0,8 \times 0,95^3 \approx 0,686$.

3. On cherche $b_n < 0,08$, soit $0,8 \times 0,95^n < 0,08$, soit $0,95^n < 0,1$. On résout : $n \times \ln(0,95) < \ln(0,1) \Leftrightarrow n > \ln(0,1) / \ln(0,95) \approx (-2,303) / (-0,0513) \approx 44,9$, donc $n \geq 45$.

Exercice 2

1. Moyenne = $(18+19+17+20+18+19+40)/7 = 151/7 \approx 21,57$. Écart-type $\approx 7,63$ (calcul similaire à la partie B).

2. $[\bar{x}-2\sigma ; \bar{x}+2\sigma] \approx [21,57-15,26 ; 21,57+15,26] \approx [6,31 ; 36,83]$. Comme $40 > 36,83$, la valeur 40°C est bien en dehors de l'intervalle : elle est signalée comme anomalie.

Exercice 3

1. $\log_2(0,9) = \ln(0,9) / \ln(2) \approx -0,152$; $\log_2(0,1) \approx -3,322$. $H \approx -(0,9 \times (-0,152) + 0,1 \times (-3,322)) \approx -(-0,137 - 0,332) \approx 0,469$ bits.

2. $0,469$ bits < 1 bit (pièce équilibrée) : c'est cohérent, une pièce truquée (qui donne très souvent le même résultat) est plus prévisible qu'une pièce équilibrée, donc son entropie (incertitude) est plus faible.

Mathématiques du programme mobilisées dans ce document

Première (spécialité mathématiques)

Suites géométriques, expression explicite, sens de variation.

Démonstration par récurrence.

Terminale (spécialité mathématiques)

Limite d'une suite géométrique, résolution d'inéquations avec le logarithme népérien.

Écart-type et moyenne d'une série statistique, intervalles à k écarts-types.

Fonction logarithme (base 2 et népérien), changement de base.

Terminale (option mathématiques expertes) / ouverture

Nombres complexes : forme exponentielle, formule d'Euler, module.