



janvier, 2022

Information Algorithmique & Intelligence Artificielle

Micro-étude

teaching.dessalles.fr/

Nom : Alexandre LE BRIS

Évolution naturelle et probabilité algorithmique

Résumé

L'évolution naturelle est conçue de nos jours comme un processus graduel, alimenté génération après génération par de légers changements dans le patrimoine génétique des populations (on pense par exemple aux mutations ou aux recombinaisons). Ces changements sont généralement modélisés selon une distribution probabiliste uniforme classique. Toutefois, de nombreuses observations, comme des épisodes d'explosion de la diversité ou d'extinction massive dans l'histoire naturelle, ou encore la récurrence de certaines structures dans le vivant, peuvent être en partie justifiée par l'utilisation des probabilités algorithmiques.

Problématique

À la suite des travaux de Darwin, la thèse de la sélection naturelle pour expliquer l'évolution des formes vivantes s'est peu à peu imposée pour devenir aujourd'hui un quasi-paradigme, intégrant avec succès les avancées modernes de la génétique à travers les notions de mutations et de recombinaisons pour expliquer les changements graduels observés au cours du temps dans les populations. Toutefois, certains phénomènes semblent aller à l'encontre de la vision darwinienne.

Dès le début du 20^e siècle, certains biologistes s'étonnaient de la récurrence de certaines formes organisées sur des branches de l'évolution pourtant trop éloignées pour que ces formes puissent ne serait-ce qu'être ébauchées sur une branche éloignée commune. Bergson s'en était fait l'écho dans son livre « L'évolution créatrice » :

Voici, à côté de l'œil d'un Vertébré, celui d'un Mollusque tel que le Peigne. Ce sont, dans l'un et dans l'autre, les mêmes parties essentielles, composées d'éléments analogues. [...] à

quelque opinion qu'on se rallie, on accordera que Mollusques et Vertébrés se sont séparés de leur tronc commun bien avant l'apparition d'un œil aussi complexe que celui du Peigne. D'où vient alors l'analogie de structure ? (Bergson 1907)

Cette question indirectement interroge également le rôle du hasard dans la production de ces structures :

Une accumulation de variations accidentelles, comme il en faut pour produire une structure compliquée, exige le concours d'un nombre pour ainsi dire infini de causes infinitésimales. Comment ces causes, toutes accidentelles, réapparaîtraient-elles les mêmes, et dans le même ordre, sur des points différents de l'espace et du temps ? (Bergson 1907)

De plus, un hasard « tout accidentel » aurait également bien du mal à expliquer l'explosion naturelle qu'a représentée le Cambrien, avec l'apparition en quelques dizaines de millions d'années de la quasi-totalité des grands plans d'organisation connus actuellement¹. En somme, « le hasard ferait trop bien les choses pour servir le goût des explications mécanistes » (Barreau 2007).

Cet aléatoire aveugle pourrait se traduire de nos jours, et en termes extrêmement simplifiés, comme une chance équiprobable de mutation à quelque endroit donné du matériel génétique. Toutefois, les notions essentielles de ce débat, à savoir celles de structure et de complexité, nous poussent à nous intéresser à une toute autre forme de hasard, pensée dans le cadre de la complexité algorithmique (ou complexité de Kolmogorov) et développée notamment par Solomonoff à travers les concepts de probabilité algorithmique et de distribution universelle (Solomonoff 2003).

Nous reproduisons ici une expérience menée par (Hernández-Orozco, Zenil et Kiani 2017) pour comparer les vitesses d'évolution, selon que la sélection parmi les mutations s'opère selon une distribution algorithmique (hypothèse 1) ou bien une distribution uniforme classique (hypothèse 0).

Méthode

La figure 1 détaille le protocole mis en place dans le cadre de notre expérience :

1. On définit une matrice cible² que l'on souhaite atteindre, on génère aléatoirement une matrice de départ que l'on prend pour référence.
2. A partir de la matrice de référence, on génère l'ensemble des mutations possibles (on se limitera ici à 1 ou 2 mutations). Pour chacune des mutations, on calculera sa probabilité de sélection, à savoir :
 - a. Hypothèse 0 : une probabilité égale pour toutes les matrices,
 - b. Hypothèse 1 : la probabilité algorithmique, à savoir $2^{-K(s)}$ où $K(s)$ est la complexité (avec préfixe) de Kolmogorov et s le plus court programme qui génère la matrice en question. Nous avons utilisé à cet effet l'algorithme BDM (Block Decomposition Method) (Zenil, et al. 2018), dont le code Python est disponible sur internet³.
3. On réalise un tirage aléatoire (sans remise) parmi les matrices mutées selon la distribution calculée.

¹ https://fr.wikipedia.org/wiki/Explosion_cambrienne

² Par souci de simplicité, on se limitera à des matrices binaires (composées de 0 et de 1).

³ <https://github.com/sztal/pybdm>

4. On passe la matrice tirée au comparateur f et on la compare à la matrice cible (on prendra la distance de Hamming, i.e. le nombre d'éléments différents).
5. Si, cette distance est égale ou plus grande que celle entre la matrice de référence et la matrice cible, la matrice sélectionnée est rejetée et une autre matrice mutée est tirée au sort ; si cette distance améliore la distance de référence à la matrice cible, alors la matrice mutée devient la nouvelle matrice de référence. Cette boucle est répétée jusqu'à ce que la matrice cible soit atteinte⁴. Le résultat est le nombre de pas S nécessaires pour réaliser cette évolution, un pas étant considéré comme un passage par le comparateur f .

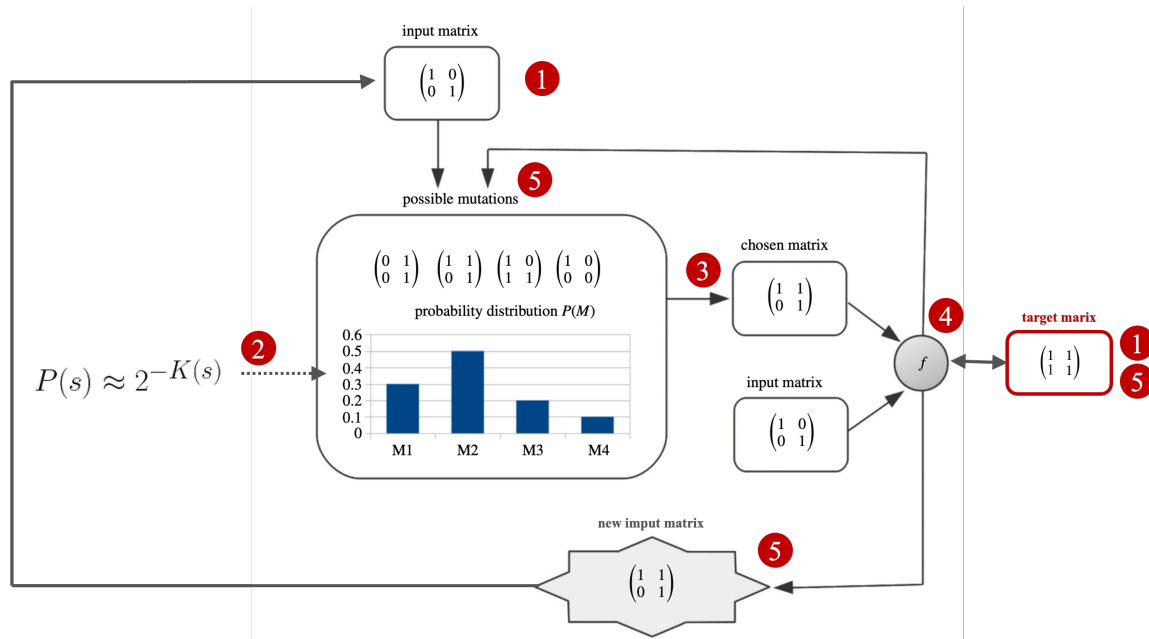


Figure 1 - Schéma du protocole expérimental, étape par étape

Tout ce protocole a été codé en python à l'occasion de cette micro-étude (à travers la classe `Evolution`). Le code de la simulation est disponible à cette adresse :

https://github.com/Dubris/AIT_Natural_Selection

Résultats

Nous définissons le speed-up quotient δ comme le rapport entre le nombre de pas nécessaires S_u dans l'hypothèse 0 d'une distribution uniforme classique, et le nombre de pas nécessaires S_{BDM} dans l'hypothèse 1 d'une distribution de probabilité algorithmique (calculée selon la méthode BDM) : $\delta = S_u/S_{BDM}$. Un rapport supérieur à 1 montre une accélération de l'évolution quand la probabilité algorithmique est utilisée.

⁴ Dans cette micro-étude, nous ne prendrons pas en compte la notion d'extinction, formalisée dans l'étude (Hernández-Orozco, Zenil et Kiani 2017) comme le nombre limite de comparaisons faites sans convergence vers la matrice cible (définie dans l'article à 2500). Ce paramètre est toutefois disponible dans le code mis à disposition.

Nous nous intéressons ici à la variation du speed-up quotient selon la complexité de la matrice cible (plus la complexité est faible, plus la matrice est considérée comme structurée). La simulation a été lancée dans 3 configurations différentes :

1. Matrices binaires $8 * 8$ et mutations limitées à 1 changement (64 mutations possibles),
2. Matrices binaires $8 * 8$ et mutations limitées à maximum 2 changement (2080 mutations possibles),
3. Matrices binaires $16 * 16$ et mutations limitées à 1 changement (256 mutations possibles).

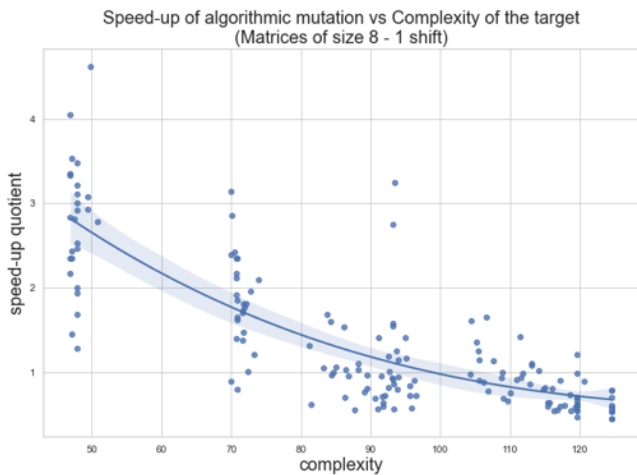


Figure 2 - Configuration 1

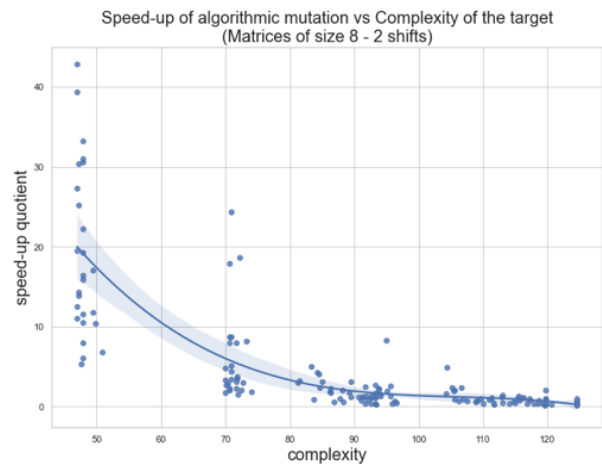


Figure 3 - Configuration 2

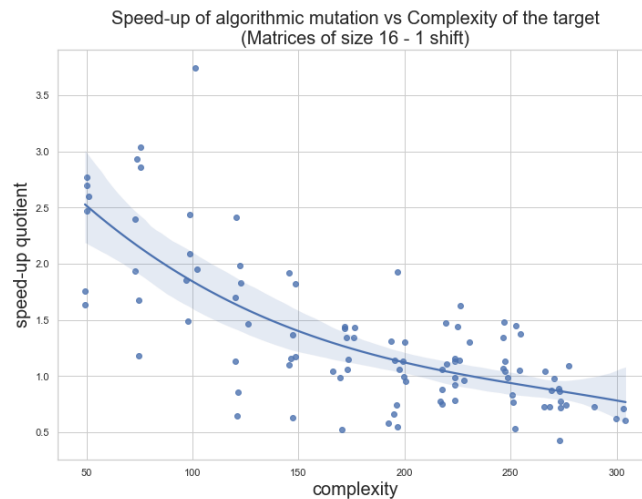


Figure 4 - Configuration 3

A travers ces résultats, on observe très clairement qu'une sélection de mutation sur la base de probabilités algorithmiques accélère l'évolution vers une cible suffisamment structurée. Cette évolution est d'autant plus rapide que la complexité à atteindre est basse et que l'espace des mutations est importante (figure 3).

D'autres résultats, comme la distribution de nombre de pas nécessaires selon l'hypothèse de sélection, sont disponibles sur le répertoire GitHub.

Discussion

Nous venons de voir que plus l'espace d'évolution est riche, plus l'accélération est importante : **les probabilités algorithmiques ont cet effet de réduire considérablement l'espace des mutations possiblement intéressantes**. Il est donc légitime de s'attendre à des phénomènes d'accélération substantielle dans un environnement naturel : à titre d'exemple, 700 mégabits sont nécessaires pour coder le génome humain, alors que notre simulation se limite à des matrices de 256 bits, avec des résultats déjà intéressants. Ceci pourrait fournir une explication plausible au « Big-Bang » zoologique observé lors du Cambrien.

L'article (Hernández-Orozco, Zenil et Kiani 2017) évoque d'autres aspects de l'évolution que les probabilités algorithmiques seraient plus à même d'expliquer, comme **certaines extinctions massives d'origine naturelle, la mémoire génétique ou encore une forme de modularité structurelle**, telle que celle que mentionnée en introduction pour l'œil.

Un point que l'on pourrait objecter à ce protocole est la **définition *a priori* d'une cible, forme de but prédéterminé que notre processus d'évolution doit atteindre** : ce serait adopter ici de façon implicite un point de vue que l'on pourrait qualifier de finaliste, pour ne pas dire d'interventionniste (la main de Dieu). Toutefois, cette hypothèse est loin d'être aussi farfelue qu'il y paraît à première vue, notamment si l'on adopte la vision de Bergson, qui concevait l'évolution comme la marque d'un élan vital tentant de se frayer un chemin dans un milieu naturel soumis avant tout à l'inertie de la matière qui le compose. Celle-ci constitue alors pour l'élan vital comme autant d'obstacles à contourner, plutôt que comme des moyens qu'il userait pour se déployer. Aussi, les formes à venir deviennent inscrites en creux dans la matière, selon la configuration prise par celle-ci en un lieu et en un temps donné (l'environnement), et l'élan vital, dans son effort de manifestation, adoptera alors plus rapidement les formes les plus structurées - ou les moins complexes algorithmiquement parlant (ce qui n'exclut pas une forme de complexité si l'on considère le résultat et si l'on tente de le décrire, comme serait l'anatomie d'un œil « vu » par un humain).

Un axe de travail intéressant, pour creuser encore l'intuition de Bergson, serait de mettre en œuvre sa notion d'évolution créatrice, « *l'imprévisibilité des formes que la vie crée de toutes pièces, par des sauts discontinus* », au moyen de la théorie de la simplicité et de la notion centrale de *Unexpectedness* (Saillenfest, Dessalles et Auber 2016).

Bibliographie

- Barreau, Hervé. 2007. «Bergson et le darwinisme : une critique pertinente.» *HAL*.
- Bergson, Henri. 1907. *L'évolution créatrice*. Paris: F. Alcan.
- Hernández-Orozco, Santiago, Hector Zenil, et Narsis A. Kiani. 2017. «Algorithmically probable mutations reproduce aspects of evolution such as convergence rate, genetic memory, modularity, diversity explosions, and mass extinction.» *CoRR* abs/1709.00268.
- Saillenfest, Antoine, Jean-Louis Dessalles, et Olivier Auber. 2016. «Role of Simplicity in Creative Behaviour: The Case of the Poietic Generator.» *ICCC* 33-41.
- Solomonoff, Ray J. 2003. «The Kolmogorov Lecture - The Universal Distribution and Machine Learning.» *Comput. J.* 46: 598-601.
- Zenil, Hector, Santiago Hernández-Orozco, Narsis Aftab Kiani, Fernando Soler-Toscano, Antonio Rueda-Toicen, et Jesper Tegnér. 2018. «A Decomposition Method for Global Evaluation of Shannon Entropy and Local Estimations of Algorithmic Complexity.» *Entropy* 20: 605.