Statistique des écarts

de la vie quotidienne à la conjecture de Riemann

Michel BAUER et Philippe Di FRANCESCO, CEA

De nombreux phénomènes font intervenir des séries de nombres rangés par ordre croissant et dont la répartition des écarts est souvent révélatrice.

Nous allons en donner quelques exemples.

Les horaires d'autobus

Même si les horaires de l'autobus qui s'arrête près de chez nous sont programmés au long d'une journée à intervalles réguliers, par exemple toutes les 15 minutes, les heures de passage réelles peuvent être différentes. Nous avons tous un jour attendu bien plus d'une demi-heure. Il arrive aussi que deux autobus se suivent à moins d'une minute. Il est possible de quantifier ces fluctuations en faisant un histogramme qui compte combien de fois l'écart entre deux bus successifs a été, par exemple, de moins de 5 minutes, de 5 à 10 minutes, de 10 à 15 minutes, etc. Si T_i est l'heure de passage du i+1ème (à T_0 passe le premier bus, le suivant à T_1, etc), on compte pour combien de valeurs de i, $T_{i+1}-T_i$ est plus petit que 5 minutes, compris entre 5 et 10 minutes, 10 et 15 minutes, etc.

En l'absence de tout aléa sur le parcours, on s'attend à ce que presque tous les écarts soient proches de 15 minutes, donc que seules les tranches entre 10 et 20 minutes soient représentées. En revanche, si de nombreux petits incidents viennent perturber chaque trajet mais qu'ils sont assez brefs pour n'avoir une influence notable que sur un des bus, le résultat peut ressembler à celui décrit dans la figure 1, dont l'interprétation, surprenante au premier abord, est que de nombreux bus se suivent de près, mais que ceci est compensé par de rares écarts très importants ; phénomène effectivement observé dans de grandes villes.

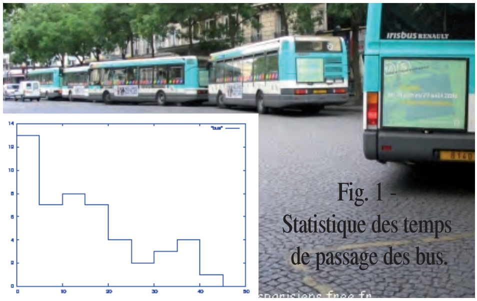

Fig. 1 - Statistique des temps de passage des bus.

Les cageots de fruits

De nos jours, les fruits de l'étal du marchand sont calibrés, mais malgré cela, deux pommes du même cageot n'ont pas exactement le même poids. Muni d'une balance de précision, on peut trier les pommes du cageot par poids croissant, P_0 est le poids de la plus légère, P_1 le poids suivant, etc, jusqu'au poids de la pomme la plus lourde, P_n si le cageot contient n + 1 pommes. Pour un cageot réel, n serait typiquement de l'ordre de la cinquantaine. L'écart de poids moyen Δ est simplement la différence entre le poids de la plus légère et de la plus lourde (P_n-P_0) divisé par n. On peut faire le même type d'histogramme que pour les temps d'attente entre deux bus : pour combien de valeurs de i la différence P_{i+1} - P_i est-elle de moins d'1/5 de l'écart moyen Δ, ou comprise

Statistique des écarts

entre 1/5 et 2/5 de Δ , etc ? Et on observe un résultat analogue : à nouveau, de nombreuses pommes ont des poids qui ne diffèrent que d'une faible fraction de l'écart moyen, et ceci est compensé par quelques écarts bien plus grands que la moyenne. Dans le cas présent, on dispose d'un modèle probabiliste très simple pour expliquer le résultat. Ce modèle prédit que si le nombre de pommes est très grand, dans une échelle appropriée, l'histogramme s'approche d'une exponentielle, comme sur la figure 2. La statistique des écarts des bus ne suit que qualitativement cette loi limite car le nombre d'échantillons est faible.

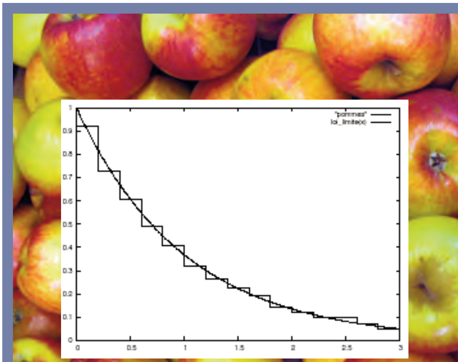

Fig. 2
Statistique des différences de poids des pommes vs la loi limite.

Les vibrations en construction automobile

L'exemple suivant est inspiré de l'automobile. Il n'y a pas si longtemps, il n'était pas rare que, pour certains régimes moteur, la carrosserie d'un véhicule vibre tout à fait perceptiblement. Le phénomène est analogue à celui d'une personne qui pousse un enfant sur une balançoire. Si elle choisit la bonne fréquence, il suffit de donner des pichenettes pour que l'amplitude augmente fortement. Les vibrations d'un véhicule sont une nuisance et l'utilisation d'ordinateurs pour la conception a permis de supprimer quasiment le phénomène. On peut voir la carrosserie comme un ensemble de pièces $p_0 , \dots , p_n$ liées plus ou moins fortement les unes aux autres et on peut quantifier ceci dans une tableau de nombres m_{ij}, $0 < i, j < n$ d'autant plus grands que la liaison entre les pièces numérotées i et j est forte. Un tel tableau de nombres s'appelle une matrice, et il existe des algorithmes mathématiques pour déduire de ce tableau de nombres les fréquences d'excitation qui feront vibrer fortement la carrosserie, il y en a en général n+1, que l'on peut ordonner de f_0 (la plus petite) à f_n, la plus grande. On peut alors s'intéresser à la statistique des écarts, comme pour les pommes et les bus. Nous ne connaissons pas l'allure de l' histogramme des écarts de fréquence pour la carrosserie d'une voiture, mais les mêmes objets mathématiques décrivent des systèmes physiques plus "fondamentaux", avec des résultats remarquables.

Généralisons à tout système physique

Si l'on ne craint pas d'être caricatural, on peut dire que le principe fondamental de la mécanique quantique est que tout système physique est décrit par un tableau de nombres (une matrice appelée *hamiltonien*), même si l'interprétation des nombres m_{ij} dans ce cas est délicate, et assez différente de celle de l'exemple d'une carrosserie. Il reste néanmoins vrai que la première chose à comprendre est l'ensemble des fréquences de vibrations correspondantes, que l'on appelle dans ce cas les niveaux d'énergie, et qui forment le spectre du système considéré. Les raies d'émission (ou raies spectrales) des atomes sont par exemple liées à des écarts entre niveaux d'énergie. Dans certains systèmes, mais rares, on peut calculer exactement les niveaux d'énergie. L'atome d'hydrogène est un exemple très important. Mais le spectre des noyaux même les plus simples, échappe à une approche analytique. En fait le *hamiltonien* des noyaux lui même est encore très mal compris aujourd'hui sur le plan fondamental. Les mesures expérimentales du spectre donnent des résultats qui semblent chaotiques. C'est ce qui a amené au début des années 1950 le physicien Eugène Wigner à poser la question suivante : peut-on comprendre certains aspects du spectre en remplaçant les coefficients m_{ij} mal connus par des nombres aléatoires ? Il n'y a pas d'espoir que les niveaux d'énergie individuels soient reproduits par un système aléatoire, mais la statistique des écarts de niveaux par exemple est un meilleur candidat car c'est une quantité moyenne sur le spectre. De plus, expérimentalement, les écarts entre les niveaux d'énergie ont une statistique simple et grosso modo indépendante du noyau considéré. Cette statistique est bien différente de celles des écarts entre les temps de passage des autobus ou entre les poids des pommes comme le montre la figure 3.

Wigner et les matrices aléatoires

En particulier, les petits écarts sont peu nombreux : on dit que les niveaux d'énergie se repoussent. Le calcul de Wigner est un des grands succès de la

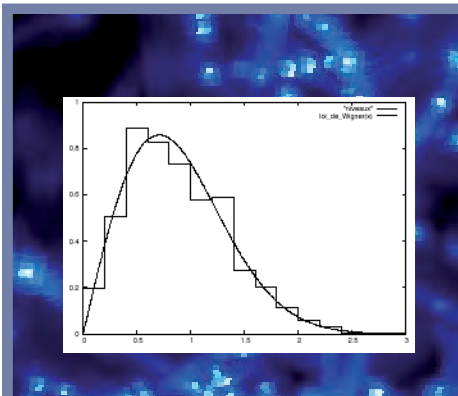

Fig. 3 - Statistique des écarts de niveau des noyaux vs la loi de Wigner.

physique théorique, car ses prédictions sont très proches des résultats expérimentaux. Son approche originale, qui a donné naissance à la vaste théorie des matrices aléatoires, ne cesse de trouver des applications dans tous les domaines de la science ou presque. (voir l'article d'Alice Guionnet sur les Grandes Matrices Aléatoires qui évoque le vaste champ d'application des permutations aléatoires).

Riemann et nombres premiers

Un exemple plus surprenant encore est donné par la théorie des nombres. Nous avons appris au collège que tout nombre entier s'écrit sans ambiguité comme un produit de nombres premiers (premier car divisible que par 1 et lui même). La suite des nombres premiers (qui est infinie, on le sait depuis Euclide) commence ainsi : 2, 3, 5, 7, 11, 13, ... Leur répartition reste aujourd'hui encore mystérieuse. Il est possible de coder tous les nombres premiers implicitement dans une fonction appelée fonction Zeta. Cette fonction qui lie une somme portant sur tous les entiers et un produit où interviennent les nombres premiers permet de définir une droite spéciale D, dite droite critique, sur laquelle s'alignent tous les zéros de la fonction. C'est sur cette fonction que

Statistique des écarts

Riemann formula une fameuse conjecture abstraite (voir article de Benoît Rittaud).

Le théorème dit *des nombres premiers*, un résultat difficile, peut se formuler intuitivement comme suit : si n est un entier très grand, les nombres premiers proches de n représentent une fraction $1/\ln n$ de la totalité des entiers proches de n. Par exemple, la densité de nombres premiers autour de 10^{10} est en gros double de la densité des nombres premiers autour de 10^{20}. Le théorème des nombres premiers est un résultat asymptotique : plus n est grand et plus la densité des nombres premiers proches de n tend vers $1/\ln n$. On aimerait connaître l'erreur typique, et la conjecture de Riemann est équivalente au fait que cette erreur est d'ordre de l'inverse de racine carrée de n. On pourrait penser que cette formulation concrète se prête mieux à démonstration que la conjecture sur la fonction Zeta, mais pour l'heure c'est l'approche abstraite qui est jugée comme la plus prometteuse.

visualisation artistique
de la fonction Zeta de Riemann
Jean François Colonna

Et les statistiques d'écarts ?

On pourrait essayer d'appliquer les idées précédentes sur la statistique des écarts aux nombres premiers, mais le fait qu'il soient justement entiers est un obstacle. En revanche, la famille des zéros critiques situés sur le droite critique D, forme un ensemble de points isolés et ordonnés. A l'aide d'ordinateurs on peut en calculer numériquement des millions et le résultat est que la statistique des écarts semble être exactement celle des écarts d'énergie des noyaux ou des matrices aléatoires correspondantes !

Au prix de nombreuses simplifications que les experts pourront juger outrancières, nous avons illustré deux statistiques d'écarts dans les lignes précédentes. La théorie des matrices aléatoires a permis d'en mettre en évidence un petit nombre d'autres qui semblent universelles. Comprendre leur ubiquité, et plus ambitieusement encore démontrer qu'en un sens elles sont les seules statistiques d'écarts possibles est un défi scientifique d'actualité. Ces questions sont vraiment à la frontière de la physique, des mathématiques voir de la biologie.

Une anecdote pour conclure.

En certains endroits d'Amérique latine, la statistique des écarts entre passages ressemble plus au cas des niveaux d'énergie des noyaux qu'à celui des poids des pommes d'un cageot, en particulier les bus ne se suivent jamais de près ; une explication est que dans ces endroits les bus sont la propriété des chauffeurs, qui sont donc rémunérés en proportion directe du nombre de passagers qu'ils transportent et n'ont aucun intérêt à passer juste après leur prédécesseur, induisant ainsi un phénomène de répulsion tout comme pour les niveaux d'énergie des noyaux.