

Notion de probabilité

L'approche fréquentiste

Qu'est-ce qu'une probabilité ? Henri Poincaré résume la situation en 1856 : on n'a pas de " définition satisfaisante ". La gêne exprimée par Poincaré vient du fait que la plupart des probabilités manipulées sont "connues" par l'intermédiaire de la "loi des grands nombres", laquelle ne saurait "définir" la probabilité puisque son énoncé fait appel à la notion de probabilité.

En fait, le cercle vicieux d'une probabilité définie par elle-même a été brisé en 1933 par Kolmogorov. Sa théorie axiomatique des probabilités, purement mathématique, est basée sur la théorie de la mesure de Borel (1897), l'intégrale de Lebesgue (1901) et la mesure abstraite de Radon (1913). Si cette théorie apporte une réponse théorique rassurante sur la validité du calcul des probabilités, elle ignore volontairement l'interprétation de la probabilité et l'utilisation qui peut en être faite. Le calcul des probabilités fournit des « modèles », mais il ne dit pas quelles lois choisir. L'évaluation des probabilités « physiques », celles de la vie sociale, pose la question de la modélisation et de l'adéquation du modèle à la réalité, éventuellement avec essais et rectifications.

Pratiquement dès l'origine du calcul des probabilités, à la fin du XVII^e siècle, émergent deux aspects de la probabilité.

- L'un, plus théorique, se fonde sur le rapport des cas favorables aux cas possibles dans les situations d'équiprobabilité. Issu de l'étude des jeux de hasard, il a peu d'applications concrètes. Les probabilités tirées "du nombre des causes", c'est-à-dire, dans le cadre de l'équiprobabilité, du rapport des cas favorables aux cas possibles, ne sont calculables que dans un cadre très limité, grosso modo, celui des jeux de hasard et des modèles d'urnes, dont les règles sont déterminées et la modélisation assez simple.
- L'autre approche, "fondée sur l'expérience du passé", c'est-à-dire sur l'observation des fréquences lorsque l'on répète un grand nombre de fois l'expérience, ouvre davantage de perspectives, au moins dans les cas d'expériences répétables. Cette approche fréquentiste des probabilités est fondée sur la loi des grands nombres dont Jacques Bernoulli est à l'origine.

Un énoncé "vulgarisé" de cette loi des grands nombres peut être le suivant :

La fréquence f d'un événement, dans une suite de n expériences aléatoires identiques et indépendantes, tend à se rapprocher de sa probabilité p quand n augmente indéfiniment.

La simulation

On peut faire remonter l'histoire de la simulation statistique jusqu'au milieu du XVIII^e siècle, où le comte de Buffon fait lancer à un enfant 2048 séries de pile ou face, jusqu'à obtenir face sur chaque série (de manière à déterminer une valeur empirique du gain moyen dans le problème de Saint-Pétersbourg)

À la fin du XIX^e siècle encore, Karl Pearson (1857-1936), l'inventeur du critère du "khi-deux", eut recours à des lancements de pièces ou de dés, embauchant pour ce faire amis et élèves, pour mettre ses théories à l'épreuve du hasard.

L'intérêt pédagogique de la simulation réside dans la nature expérimentale qu'elle donne à l'enseignement de la statistique et des probabilités, donnant davantage de sens aux concepts et motivant les élèves par l'aspect novateur de cette approche et de l'utilisation de l'ordinateur.

L'aspect de modélisation par les probabilités n'étant abordé pour l'instant qu'en première, on adopte en seconde un point de vue pragmatique :

"Dans le cadre du programme de seconde, simuler une expérience consistera à produire une liste de résultats que l'on pourra assimiler à un échantillon de cette expérience."

Document d'accompagnement du programme 2000 de seconde.

On peut dire que, pour nos élèves de seconde, simuler une expérience aléatoire consiste à produire "virtuellement" des résultats analogues à ceux que l'on aurait obtenus en réalisant "physiquement" l'expérience aléatoire.

La simulation permet de mettre à l'épreuve de l'expérience certains résultats. Elle favorise le débat scientifique, obligeant les élèves à confronter leurs observations et à les analyser. On peut également, en sens inverse, expérimenter d'abord, pour émettre des conjectures ou introduire une notion.

Comment simuler le hasard ? Quelques repères historiques

Ce n'est pas facile de simuler le hasard et les premières "tables de nombres au hasard" ont été construites en recueillant le hasard "physique". Les premières tables ont été obtenues dans les années 1925 à partir des numéros gagnants de la loterie.



La méthode de Monte-Carlo...

En 1955, la « Rand Corporation » édita une table, A Million Random Digits, obtenue à partir de bruits de fond électroniques (fluctuations du débit de tubes électroniques). Il s'agit alors d'un générateur aléatoire physique.

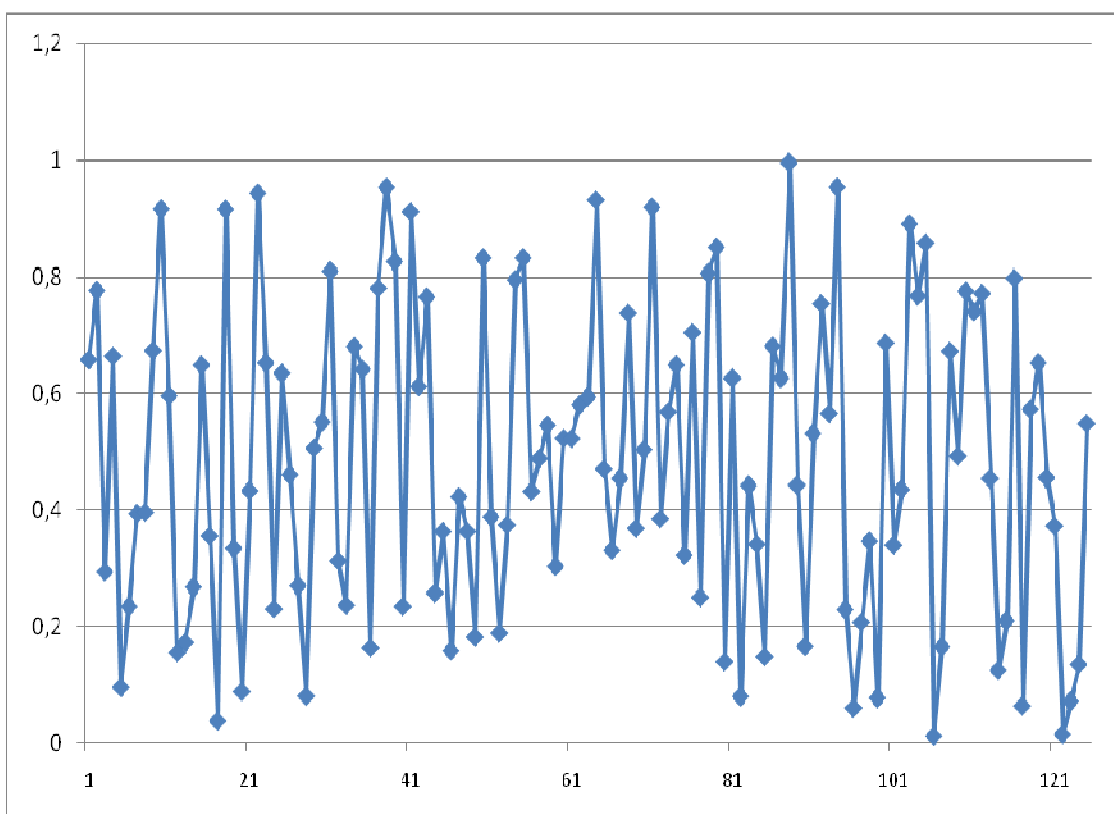
Avec l'apparition des ordinateurs, on chercha à générer des nombres aléatoires, à l'aide d'algorithmes produisant une suite de nombres telle qu'un statisticien en l'analysant, ne soit pas capable de détecter si elle a été produite par un procédé mathématique ou un réel phénomène aléatoire physique : qu'il lui soit impossible, par exemple, sur une suite assez grande de 0 et de 1, de savoir s'ils ont été générés par un ordinateur ou en lançant une pièce de monnaie bien équilibrée. Une telle suite est pseudo-aléatoire. Il ne s'agit plus de hasard physique mais d'un « hasard calculé ».

Sur un modèle IBN des années 1980, on trouvait par exemple le générateur de nombres pseudoaléatoires suivant. On utilise le nombre premier $p = 2^{31} - 1$ (c'est un nombre de Mersenne), puis, à partir de la valeur $r_0 = 5$ on calcule la suite $r_{n+1} = 7^5 \times r_n$ modulo p (c'est-à-dire que r_{n+1} est le reste de la division euclidienne de $7^5 \times r_n$ par p).

On constate que la suite (x_n) définie par $x_n = \frac{r_n}{p}$ fournit des nombres pseudo-aléatoires de l'intervalle $[0;1[$.

On donne ci-dessous les premières valeurs de (x_n) , calculées par le tableur, le graphique montre leur caractère aléatoire qui resterait à contrôler par des tests statistiques.

0,657688941
0,778026611
0,29325066
0,663836187
0,094795932
0,235223081
0,394323584
0,396482029
0,67346448
0,917510387
0,59708186
0,154826731
0,172860553
0,267308175
0,648500967
0,35574692
0,038490931
0,917078254
0,334211188
0,087429872
0,433863562



source

L'induction statistique au lycée,
Philippe Dutarte, IREM Paris-Nord.