CHAPITRE I

LE MODELE PROBABILISTE

Les dieux nous créent bien des surprises :
l'attendu ne s'accomplit pas,
et à l'inattendu un dieu ouvre la voie
Euripide

1 - INTRODUCTION

Il paraît naturel de débuter un discours sur le calcul des probabilités en définissant ce qu'est une probabilité, et avant cela ce qui caractérise un phénomène aléatoire. Or, cette tâche n'est pas forcément des plus simples car la plupart du temps, considérer un système comme étant aléatoire est un choix de modélisation résultant de la méconnaissance de certains facteurs. Par exemple, le radical du mot aléatoire, l'*alea* signifie en latin *jeu de dés*. Or rien n'est plus déterministe que le lancer d'un dé, car la dynamique et les conditions initiales déterminent complètement le mouvement du dé, et donc la position sur laquelle il s'arrêtera in fine. Pour autant, il ne vient à l'idée de personne de considérer ce système comme déterministe pour la simple raison que de très faibles variations des conditions initiales peuvent mener à des résultats différents. Ainsi dans le cas d'un dé, la mesure de ces conditions est beaucoup trop difficile pour un expérimentateur moyennement équipé, sans parler du client du café qui vient jouer au 421… Cette idée est d'ailleurs celle de Laplace, telle qu'exposée dans son important traité *Théorie analytique des probabilités* dont la première édition date de 1812. Pour s'en convaincre, le plus simple est évidemment de le citer. *Nous devons donc envisager l'état présent de l'univers comme l'effet de son état antérieur et comme la cause de ce qui va suivre. Une intelligence qui […] connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent […] embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle […]. La courbe décrite par une simple molécule d'air […] est réglée de manière aussi certaine que les orbites planétaires : il n'y a de différence entre elles que celle qu'y met notre ignorance.* Le développement des sciences nous a depuis montré que cette vision est un peu restrictive. On pense bien sûr par exemple à la physique quantique où les objets manipulés sont intrinsèquement probabilistes quoique le formalisme soit un peu particulier[1]. Pour autant, cette vision non figée de la distinction entre phénomène déterministe et aléatoire (ou stochastique) nous paraît importante dans une démarche de modélisation, et largement suffisante pour un ouvrage d'initiation comme celui-ci.

Ces précautions sur la caractérisation d'un phénomène aléatoire étant prises, il reste à définir la notion de probabilité. Mais le travail à faire reste d'envergure comme le montrent les trois exemples qui suivent :

[1] Il s'agit du domaine des probabilités quantiques.

Exemple 1 - contrôle de qualité

Une machine produit des pièces en acier d'une certaine masse et on désire qualifier la qualité de la production. A cet effet, on cherche à évaluer la probabilité que la masse d'une pièce produite n'appartienne pas à un intervalle $[m_0, m_1]$, appelé intervalle de tolérance, hors duquel la pièce sera rejetée car non utilisable dans la suite du processus de production.

L'évaluation de cette probabilité paraît à première vue assez simple : il suffit d'observer un grand nombre de pièces produites et de faire le rapport du nombre de pièces hors normes sur le nombre de pièces observées. On utilise ainsi une définition fréquentiste de la probabilité qui suppose que le phénomène sous-jacent soit indéfiniment, ou tout au moins un grand nombre de fois, répétable. C'est la vision de Laplace, qui suppose comme il l'explique très clairement que les pièces aient le même comportement. Un peu de réflexion montre que la répétabilité du phénomène n'est pas du tout évidente. Dans le cas présent, de nombreux facteurs peuvent influencer l'observation, comme l'usure de la machine, les changements d'équipes, la qualité de la matière première…

Exemple 2 - prospection pétrolière

A la suite d'une campagne de prospection, un expert d'une compagnie pétrolière estime que la probabilité qu'un gisement donné contienne au moins 50 millions de barils de pétrole exploitable[2] est de 60 %. Quel sens peut avoir une telle phrase ?

Plusieurs interprétations sont possibles ; on peut tenter de se rapprocher de la définition fréquentiste précédente en affirmant que 60 % des gisements connus précédemment ayant des caractéristiques proches de celui considéré ici contenaient du pétrole sous une forme exploitable. On imagine cependant, compte tenu des variations de forme et de nature de roche d'un champ pétrolier, qu'il doit être difficile dans un tel contexte de trouver suffisamment de gisements de caractéristiques proches de celui étudié. Ce qui rend en même temps la définition de la probabilité difficile, mais aussi - ce qui est tout aussi grave - son évaluation impossible en pratique. On peut de plus imaginer de nombreuses situations où un tel historique n'est pas du tout disponible (penser par exemple à la probabilité qu'une centrale nucléaire « explose », par ailleurs évaluée par EDF !).

Deux interprétations sont encore possibles : soit il faut traduire la phrase précédente de façon plus qualitative, en disant que l'expert considère plus probable qu'il y ait du pétrole plutôt qu'il n'y en ait pas mais que la différence entre les deux hypothèses n'est pas très contrastée. On aboutit ainsi à une probabilité **subjective** de 60 %. On peut penser à première vue qu'une telle « définition » est sans intérêt car dépendante de l'individu qui l'énonce sur des bases non formalisées. Et une telle opinion est licite la plupart du temps. Pour autant, cette évaluation peut alors être considérée comme une première estimation formalisant la connaissance de l'expert, devant être raffinée. Dans un tel cas, nous sommes revenus à notre point de départ pour améliorer l'estimation. On notera pour finir sur cette question que la notion de probabilité dépendant d'un individu est à la base de la théorie de la décision économique ou financière où chaque acteur possède son propre système de valeurs[3].

La dernière interprétation, qui est la bonne dans ce contexte d'exploration et la plus courante dans les sciences de l'ingénieur, consiste à forger un **modèle probabiliste** du gisement et à évaluer ensuite, à l'aide de ce modèle, la probabilité que ce gisement possède du pétrole sous une forme exploitable. Sans détailler cet exemple très complexe, nous

---

[2] Seuil au delà duquel la compagnie pétrolière concernée lancera la production.

[3] Ou sa propre « fonction d'utilité », pour reprendre le terme utilisé par Bernoulli.

pouvons donner ici le principe d'une telle modélisation. Il s'agit de se ramener à une expérience *supposée* répétable, en ayant modélisé les influences, la plupart du temps déterministes, qui rendaient le phénomène non répétable.

Terminons cette analyse préalable par un troisième exemple qui nous permettra d'expliciter cette démarche de modélisation dans un cas simple mais réel.

Exemple 3 – prévision de la consommation électrique

Pour la semaine du nouvel an[4] (nous sommes alors dimanche 26 décembre 1999), un ingénieur en charge de la production d'électricité en France se demande s'il doit décider de la mise en route (très coûteuse) d'une centrale thermique supplémentaire pour répondre aux besoins futurs de la consommation. Il sait que ses capacités de production ne peuvent pas en l'état dépasser 9650 GWh pour la semaine à venir et s'interroge donc sur la consommation électrique à laquelle il va devoir faire face. Il regroupe à cet effet les consommations observées depuis janvier 1996 et trace la figure qui suit.

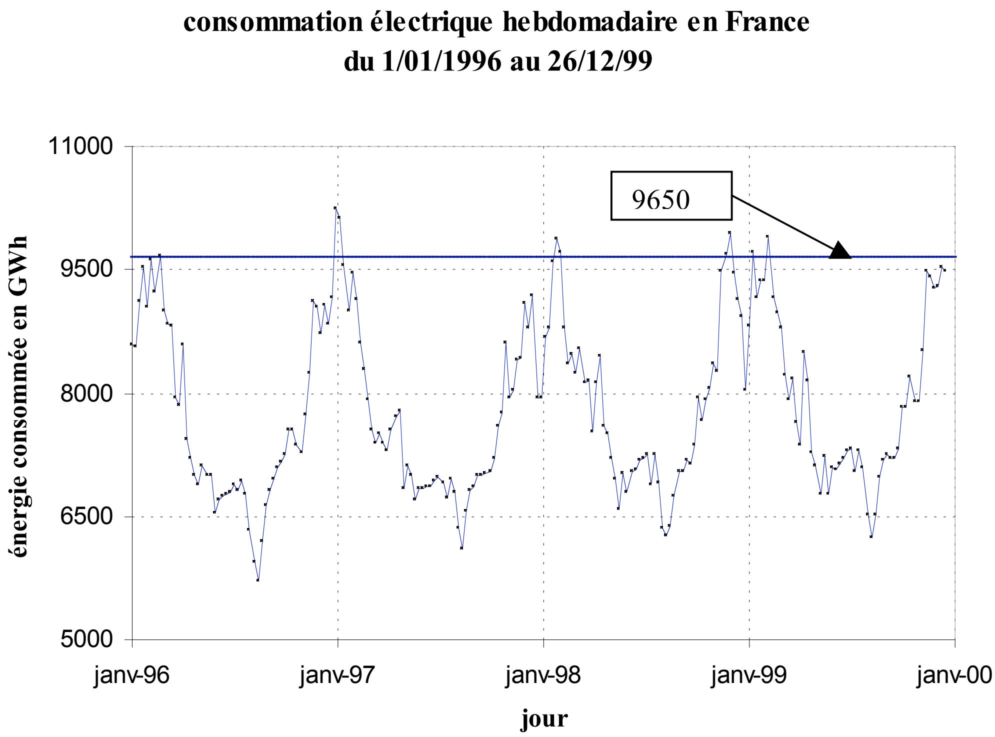

Figure 1. Source : RTE (www.rte-france.com)

On observe que la consommation de la semaine écoulée n'a été que de 9480 GWh et a été assez stable depuis plus d'un mois. Mais la courbe de consommation montre une grande variabilité avec des sauts d'amplitude forte en particulier en décembre et janvier. Par exemple, l'observation de la fin de l'année 1997 montre qu'il est tout à fait possible que la consommation dépasse 9650 GWh après des semaines de consommation inférieure à 9000 GWh. L'ingénieur a donc besoin d'une information complémentaire pour l'aider

[4] Dans la réalité, ces questions se posent à un horizon beaucoup plus court, comme le jour ou l'heure.

dans sa décision, comme la connaissance de la probabilité que la consommation de la prochaine semaine dépasse 9650 GWh.

Naturellement d'un point de vue pragmatique, la question est de savoir comment calculer cette probabilité ; toutefois, une autre question non moins essentielle est de savoir quel sens lui donner. Laissons momentanément la deuxième et attachons-nous pour l'heure au « comment ». Nous verrons en fait que la démarche suivie pour évaluer la probabilité cherchée répondra au passage à la question du « quoi » à travers un modèle probabiliste. Comme dans les exemples précédents, on peut être tenté en premier lieu d'utiliser une approche fréquentiste et de chercher à évaluer le nombre de fois dans le passé où la consommation a dépassé le seuil fatidique. Ainsi, en remarquant que sur les 208 semaines de données, seules 9 valeurs dépassent 9650 GWh, une façon quelque peu expéditive d'évaluer la probabilité cherchée consiste à proposer le rapport 9/208, soit 4,3 %. Bien sûr, cette approche ne tient pas compte de la cyclicité de la consommation, et il serait sans doute judicieux de se restreindre aux mois d'hiver pour faire ce calcul, mais encore faut-il définir précisément les mois à considérer. On obtiendrait alors un quotient au numérateur inchangé mais dont le dénominateur diminue, d'où une probabilité augmentée. Poussée à l'extrême, cette démarche amènerait à rechercher des « scénarios » identiques dans l'historique de consommation[5]. Ici par exemple, on peut considérer les seules dernières semaines de l'année : la croissance de consommation entre ces semaines et leur suivante est de –118 en 1996, 730 en 1997, et 778 en 1998. En réutilisant ces chiffres pour la dernière semaine de 1999 qui nous préoccupe, on observe que 2 accroissements conduisent au dépassement du seuil de 9650 et un à son non-dépassement. Doit-on en déduire que la probabilité cherchée est finalement de 2/3 ? On voit là combien nos estimations varient en fonction du raisonnement appliqué !

Les arguments précédents sont pour certains d'entre eux pour le moins discutables, et ont deux principaux défauts. Le premier est de ne pas expliciter, et donc a fortiori de ne pas vérifier, les hypothèses faites (répétabilité par exemple). Le second est de ne pas se baser sur une *compréhension du phénomène*. Or, il est au moins un facteur dont dépend fortement la consommation électrique : la température. A titre d'illustration, on a représenté sur la figure ci-dessous la température hebdomadaire moyenne mesurée à Paris sur la période considérée. On peut observer en particulier que le record de consommation électrique de la semaine du nouvel an 1997 correspond à un record de froid dans cette semaine avec une température moyenne de –5°C. Les dates des autres pics de consommation coïncident également avec les pics de température hivernale.

[5] On utilise alors une méthodologie analogue à la démarche « chartiste » d'analyse des courbes bien connue dans le domaine de la finance. Dans ce cadre, on peut trouver une « justification » a posteriori de cette démarche car le mouvement de la bourse va être influencé par les décisions des acheteurs et vendeurs : si tous sont chartistes, le mouvement prédit peut alors être effectivement vérifié… et la démarche a priori discutable devient non discutée puisque vérifiée !

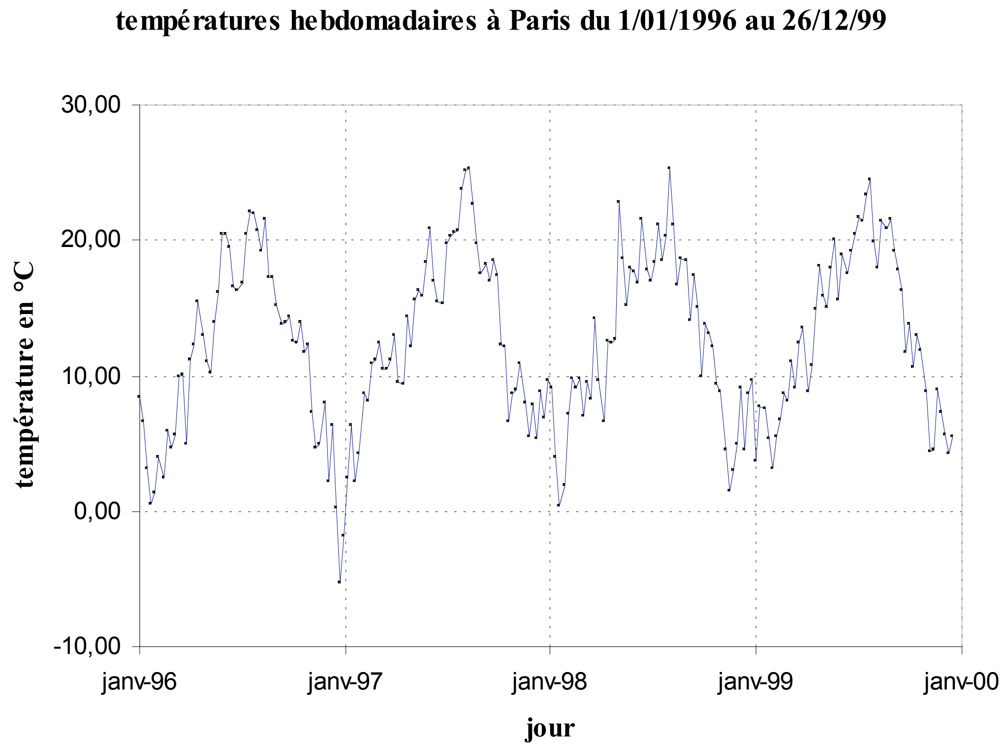

Figure 2. Source : Météo-France

Pour se convaincre du lien entre ces deux variables, il est d'ailleurs utile de tracer l'une en fonction de l'autre :

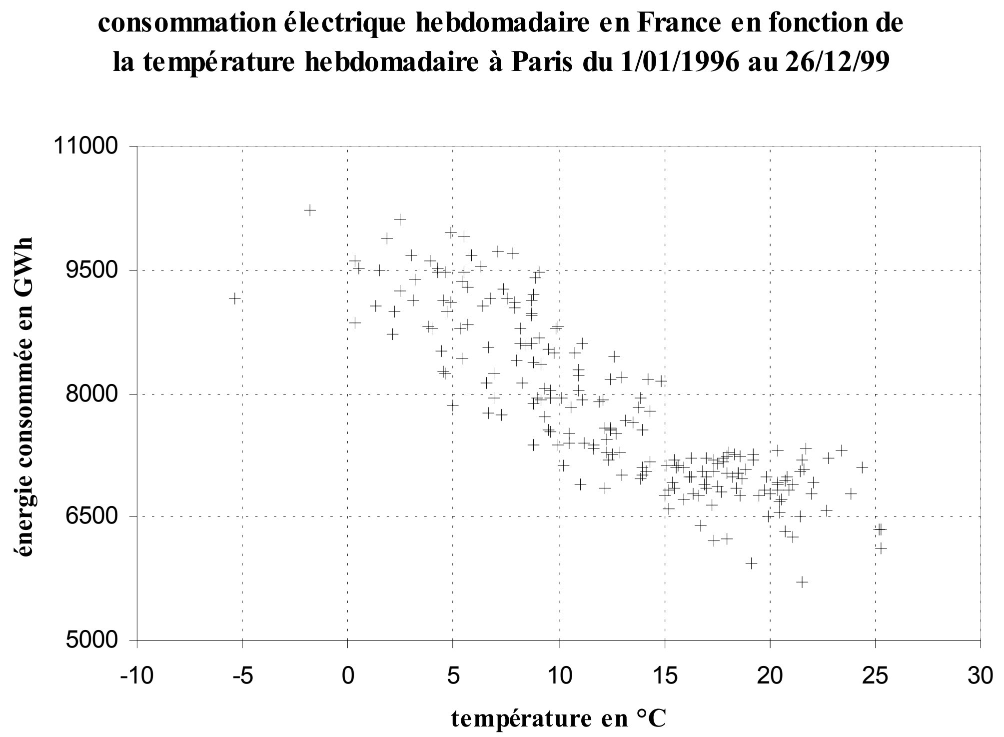

Figure 3.

En observant un tel graphique, on prendra bien garde à ne pas tirer des conclusions hâtives. Et nous allons donc marquer une courte pause dans notre objectif d'évaluation de la probabilité pour mettre en garde le lecteur sur un travers courant. Il semble en effet que le graphique précédent montre que la température a un effet sur la consommation. Mais que dirait le lecteur si l'on renversait les axes : a-t-on une influence de la consommation électrique sur la température ? Et si l'on remplaçait la consommation électrique par celle de foie gras, la figure aurait sans doute un aspect proche de la précédente, avec un pic pour les températures faibles et une consommation moindre pour de fortes températures. Nous aurions pourtant du mal à convaincre notre interlocuteur que la diminution de la température de 1°C augmente la consommation de foie gras d'un certain pourcentage ! Ici, des informations non contenues dans les données (effet de la température sur le chauffage par exemple) indiquent que la température influence la consommation.

De façon générale, ce que révèle un graphique de ce type est une corrélation entre les deux variables et non un lien de causalité. Ce qui signifie en pratique qu'il existe un troisième facteur qui influence les deux variables observées, ce qui est le cas évidemment avec notre exemple de foie gras où le facteur « fêtes de fin d'année » est prépondérant. Pour *démontrer* une causalité, il faut que l'on observe une corrélation, **toutes choses égales par ailleurs**. Ce qui signifie qu'il faut être capable de fixer le niveau des facteurs alternatifs et éventuellement influents[6]. Pour nos questions de consommation électrique, on peut par exemple observer les semaines de novembre (on évite décembre et les fêtes qui perturbent l'activité et donc la consommation électrique). On espère ainsi s'affranchir d'un éventuel effet de saison, ce qui donne le graphique qui suit.

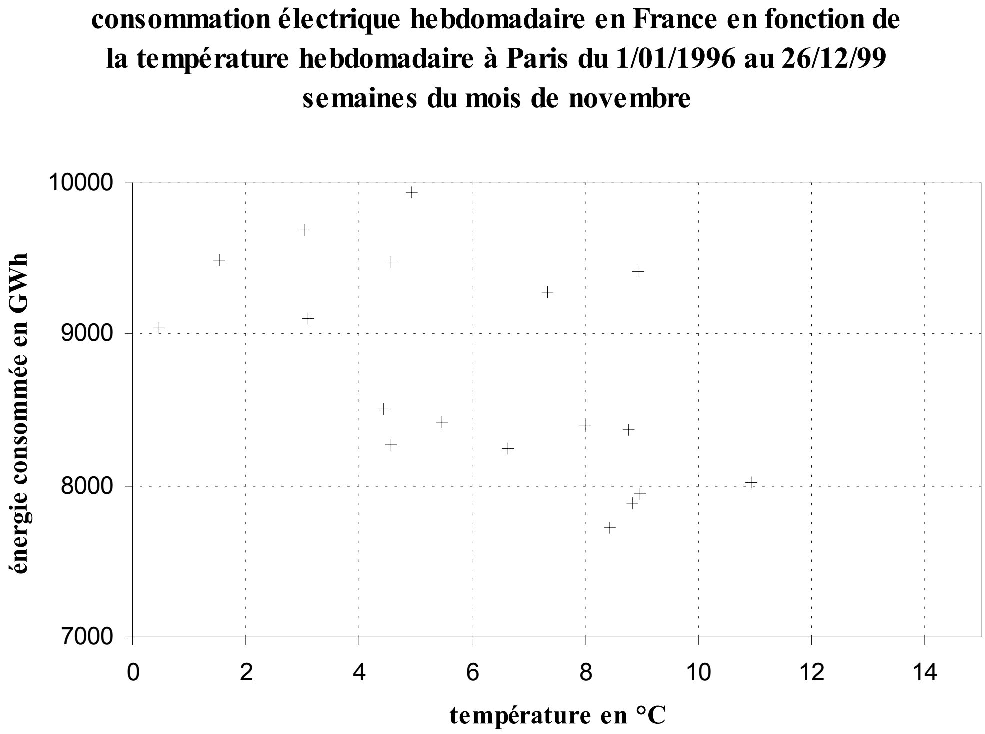

Figure 4.

[6] Lorsque les variables peuvent être fixées par l'expérimentateur, des procédures existent pour réaliser ce programme, ceci constitue une partie de la théorie des plans d'expériences.

On continue d'observer une corrélation, qui conforte notre idée d'influence et de causalité. On remarque néanmoins que le lien est moins net, et que – comme attendu – l'échantillon observé s'est grandement réduit. Idéalement, pour être assuré que tous les facteurs pouvant influencer la consommation sont fixés, il faudrait se limiter à un échantillon de taille 1, d'une portée pratique très limitée pour examiner l'effet d'une variable sur une autre !

Compte tenu de ce qui précède, nous allons tenter de proposer un modèle permettant d'expliquer la consommation. Pour cela, nous désignerons par s la semaine et n l'année. Un tel modèle peut par exemple exprimer la consommation de la semaine future C(s+1,n) en fonction de la température de la semaine courante T(s,n), de la prévision de température pour la semaine à venir PT(s+1,n), et - de façon à tenir compte de la spécificité de la semaine du nouvel an – de la consommation moyenne observée dans le passé lors de la semaine du nouvel an CM(s+1,.). Il aurait donc la forme suivante :

$$C(s+1,n) = F\left[T(s,n), PT(s+1,n), CM(s+1,.)\right] + \varepsilon(s,n)$$

où F est une fonction déterministe et ε(s,n) représente l'erreur d'approximation entre la consommation donnée par le modèle et la consommation réellement observée. Ce dernier terme peut alors être considéré comme un terme purement aléatoire : c'est son existence qui distingue un *modèle probabiliste* d'un modèle déterministe.

Pour que ce modèle soit utilisable, on doit faire des hypothèses sur ε(s,n), sur la *distibution* ou *loi* de ε(s,n) (voir chapitre 3) et la *dépendance* (ou non-dépendance) dans le temps entre ε(s,n), ε(s-1,n), … Sans expliciter précisément ces hypothèses, retenons simplement qu'elles définissent de façon rigoureuse la façon dont le hasard se manifeste, et qu'elles consistent en substance à se ramener à l'approche fréquentiste en isolant dans le phénomène observé, à l'aide d'une modélisation adéquate, un événement aléatoire, ou une variable, qui se répète.

Pour finir, disons un mot de la manière dont on utiliserait le modèle de consommation pour calculer la probabilité de dépasser le seuil de 9650 GWh. Après avoir contrôlé que les hypothèses sur les erreurs aléatoires ε(s,n), ε(s-1,n),… sont satisfaites, on peut alors, et seulement alors, obtenir des prévisions plausibles de la consommation en procédant à des tirages au hasard de ε(s,n) (de la même façon, en quelque sorte, qu'on lance un jeu de dés) en nombre aussi important que l'on veut ; il suffit alors de calculer parmi ces prévisions la proportion de celles qui dépassent le seuil connu.

On observe cette fois que l'on a donné, en plus du calcul, un sens précis à la notion de probabilité, car la « façon de tirer au hasard » a été explicitée. Ce point est essentiel, comme nous allons le voir avec le paradoxe de Bertrand.

Paradoxe de Bertrand

Joseph Bertrand publie en 1888 un ouvrage intitulé *Calcul des Probabilités* dans lequel il rappelle la définition de la probabilité, donnée déjà par Laplace :

$$\text{probabilité} = \frac{\text{nombre de résultats favorables}}{\text{nombre de résultats possibles}}$$

Il fait suivre cette définition d'un certain nombre de critiques, comme la difficulté d'évaluer numérateur et dénominateur lorsque l'on s'intéresse à des probabilités *géométriques* où les résultats possibles forment un continuum. Il exhibe notamment un exemple particulièrement simple, repris et popularisé par H. Poincaré dans son ouvrage homonyme paru en 1912 ; cet exemple est connu depuis sous le nom de paradoxe de Bertrand.

On trace au hasard une corde dans un cercle. Quelle est la probabilité pour qu'elle soit plus petite que le côté du triangle équilatéral inscrit ?

Notons R le rayon du cercle, L la longueur du côté d'un triangle équilatéral inscrit et p la probabilité cherchée. La géométrie élémentaire nous montre que le rayon du cercle inscrit dans le triangle équilatéral est de R/2. Nous allons conduire plusieurs raisonnements pour répondre à la question.

- Premier raisonnement :

Une corde est repérée par son milieu car la corde est orthogonale au rayon passant par ce milieu. Donc choisir la corde revient à choisir ce point milieu. Or, les cordes dont la longueur dépasse L sont celles dont le point milieu est à l'intérieur du cercle concentrique de rayon R/2. Les points correspondants sont indiqués en grisé sur la figure. Un exemple de corde satisfaisant la condition est reportée en pointillés.

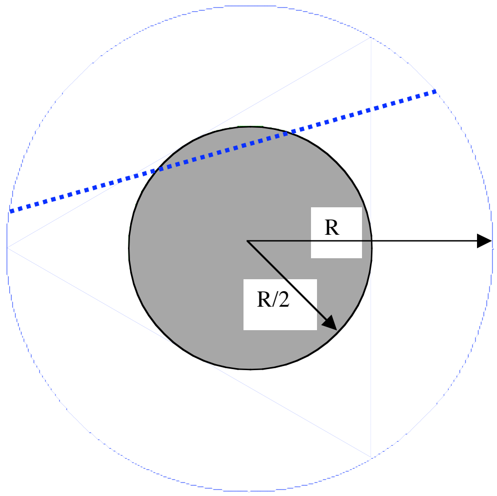

Figure 5. Premier raisonnement : la corde est repérée par son milieu

A partir de la définition intuitive de la probabilité, et en notant D_ρ le disque de rayon ρ, on aboutit à l'estimation suivante pour p :

$$p = \frac{\text{aire}(D_{R/2})}{\text{aire}(D_R)} = \frac{\pi\,(R/2)^2}{\pi\,R^2} = \frac{1}{4}$$

- Deuxième raisonnement :

On reprend le raisonnement précédent en remarquant que le problème est invariant par rotation. Par suite, le point milieu de la corde peut être choisi sur un rayon du cercle. Les points du rayon correspondants aux cas favorables sont indiqués en grisé sur la figure et une corde satisfaisant la condition est reproduite en pointillés.

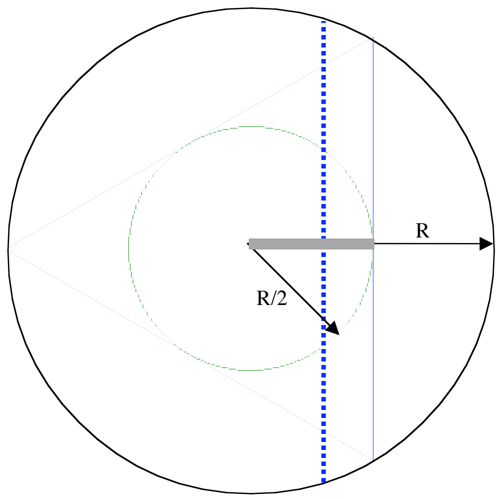

Figure 6. Deuxième raisonnement : le milieu de la corde est choisi sur un rayon

On aboutit ainsi à l'estimation suivante pour p :

$$p = \frac{\text{longueur}([0,R/2])}{\text{longueur}([0,R])} = \frac{R/2}{R} = \frac{1}{2}$$

- Troisième raisonnement :

Le problème posé étant invariant par rotation, on peut supposer qu'une extrémité de la corde est fixée, par exemple à un sommet du triangle équilatéral inscrit. Par suite, choisir la corde revient à choisir la seconde extrémité sur le cercle C de rayon R. Les cordes dont la longueur dépasse L sont celles pour lesquelles la seconde extrémité est située dans le tiers de cercle présenté sur la figure ci-dessous en grisé. On a là encore reproduit un exemple de corde satisfaisant les propriétés requises en pointillés.

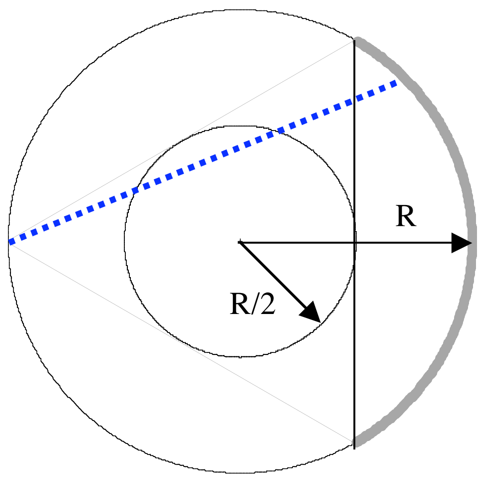

Figure 7. Troisième raisonnement : une extrémité de la corde est fixée

En désignant par C le cercle de rayon R et C_1 le tiers de cercle regroupant les cas favorables, on obtient comme estimation de la probabilité cherchée :

$$p = \frac{\text{longueur}(C_1)}{\text{longueur}(C)} = \frac{2\pi R/3}{2\pi R} = \frac{1}{3}$$

Nous conseillons au lecteur de rechercher la faille dans les raisonnements élémentaires qui précèdent. Il n'en trouvera malheureusement pas. Un bon réflexe serait de se ramener à la définition expérimentale, fréquentiste, de la probabilité. Il s'agirait, pour distinguer le *bon* raisonnement, de réaliser l'expérience consistant à lancer la corde un grand nombre de fois, et d'examiner la fréquence d'apparition obtenue. Cet essai bute très rapidement sur la définition des conditions expérimentales et on voit très vite que leur détermination nécessite de choisir la façon dont on lance la corde , et donc que l'on tourne en rond.

La résolution de cet apparent paradoxe est pourtant très simple, et Bertrand en donne évidemment la raison :

Entre ces trois réponses, quelle est la véritable ? Aucune des trois n'est fausse, aucune n'est exacte, la question est mal posée.

En effet, l'énoncé n'est pas suffisamment précis pour déterminer le modèle probabiliste sous-jacent, i.e. les conditions expérimentales et il y a de nombreuses façons de lancer la corde, et les trois figures précédentes en donnent trois exemples.

- Dans le premier cas, on choisit le milieu de la corde uniformément dans le disque ; la définition rigoureuse d'un tirage uniforme sera bientôt donné, mais on peut déjà dire qu'il s'agit de choisir un point dans le disque sans privilégier une zone par rapport à une autre.
- Dans le deuxième cas, on choisit le milieu de la corde en choisissant de façon indépendante (là aussi, une définition rigoureuse sera donnée par la suite) un angle uniformément entre 0 et 360°, et une distance au centre uniformément entre 0 et R. Un brin de réflexion suffit à se convaincre que ce procédé, lorsqu'on le répète, place un plus grand nombre de points vers le centre du cercle que vers la circonférence : il n'est donc pas du tout équivalent au premier.
- Enfin, dans le troisième cas, on choisit de façon indépendante les deux extrémités de la corde uniformément sur la circonférence.

En fait, il existe bien d'autres réponses à la question posée par Bertrand, autant qu'on peut trouver de procédés différents de simulation… c'est-à-dire une infinité[7] !

Fonctionnement général d'un modèle

Comme nous l'a montré le paradoxe de Bertrand, toute évaluation probabiliste doit être précédée par une modélisation précise. Le fonctionnement d'une telle démarche, valable bien au-delà de notre contexte probabiliste, est reproduit dans le schéma ci-dessous.

[7] Par exemple dans le premier cas, il y a une infinité de façons de simuler le milieu de la corde dans un disque, autres qu'uniformément.

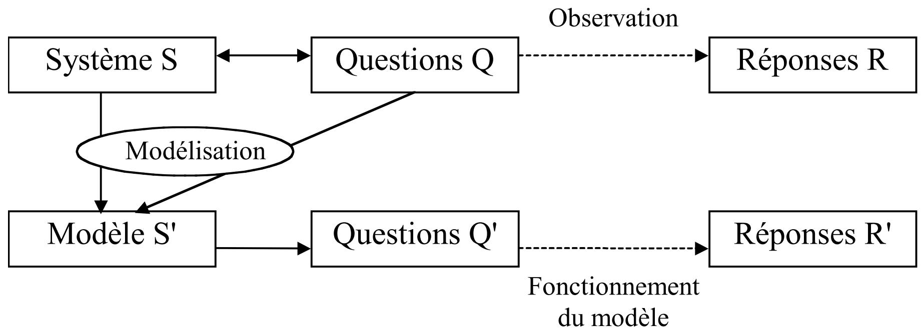

Figure 8.

La lecture du graphique précédent se fait de haut en bas et de gauche à droite. Tout d'abord, il convient de bien noter que la considération du seul système S n'est pas suffisante pour entamer une démarche de modélisation et qu'il faut toujours considérer le **couple** (système S, questions Q). Par exemple dans le contexte de l'exemple 1, on ne construira pas le même modèle si l'on veut seulement connaître la proportion de pièces satisfaisant les tolérances (i.e. masse $\in [m_0, m_1]$) , ou si l'on désire contrôler précisément leur masse et identifier les causes possibles de variation. Cette remarque assez évidente peut paraître un peu superflue ; pourtant il arrive souvent dans la pratique que les questions soient un peu oubliées et que les défauts induits soient perçus lors de l'étape - finale - de validation. On s'attachera donc toujours à bien définir les problèmes à résoudre avant de modéliser ou d'utiliser une méthode de résolution. Si ce conseil de bon sens était suivi par tous les ingénieurs, des gains de temps substantiels seraient obtenus ! Une fois le modèle S' construit, les questions posées sont traduites dans cette modélisation en questions Q'. La validation du modèle est ensuite effectuée en comparant réponses prévues R' et réponses observées R.

Notons que dans le cadre du calcul des probabilités, la validation est effectuée à l'aide des statistiques, mais ce n'est pas, loin de là, la seule utilisation de l'outil statistique. Une fois la confiance dans le modèle acquise, tout au moins dans un certain domaine de validité, de nouvelles questions Q_1 fourniront des prévisions R'_1 que l'on espère évidemment proches des véritables réponses R_1.

2 - DEFINITIONS

A la lumière de ce qui précède, on peut tenter une première formalisation : une expérience probabiliste est formée d'un couple (Ω, P) défini ci-après. Déjà, nous ne savons pas prévoir le résultat de l'expérience, c'est pourquoi nous considérons l'ensemble Ω, appelé **espace des épreuves**, dont chaque élément ω de Ω représente un résultat possible de l'expérience réalisée. Les parties de Ω, appelées **événements**, sont les parties dont nous allons vouloir calculer la probabilité. Pour revenir sur l'exemple 1, pour une production à venir de 500 pièces, on va vouloir calculer la probabilité que plus de 10 pièces ne satisfassent pas les tolérances. Si nous considérons que cette probabilité est trop forte, nous pourrons vouloir intervenir sur le processus de production pour l'améliorer.

P est une probabilité, c'est-à-dire une application définie sur l'ensemble des événements de Ω, à valeurs dans $[0,1]$, vérifiant les propriétés « évidentes » suivantes :

(i) $P(\emptyset) = 0$

(ii) $P(\Omega) = 1$

(iii) $\forall$ A, B ; A et B disjoints $\Rightarrow P(A \cup B) = P(A) + P(B)$ **(additivité)**

Notons que le point (iii) s'étend alors facilement par récurrence en :

(iii') $\forall A_1, A_2, ..., A_p$; $A_m \cap A_n = \varnothing$ pour $m \neq n \Rightarrow P(\bigcup_{n=1}^{p} A_n) = \sum_{n=1}^{p} P(A_n)$

(iv) $\forall$ A, B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Nous ne cherchons absolument pas à donner ici des axiomes minimaux : par exemple, (ii) et (iii) impliquent (i), (ii) entraîne (iv), etc… Après tous ces préliminaires, nous allons donner de manière brutale les définitions nécessaires avant de retrouver commentaires et exemples.

2.1 - TRIBU

Une famille $\mathcal{F}$ formée de parties de Ω est une **tribu** (en anglais σ-field) si :

(i) $\varnothing \in \mathcal{F}$

(ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (A^c est le complémentaire de la partie A)

(iii) $A_n \in \mathcal{F}$ pour tout $n \in \mathbb{N} \Rightarrow \bigcup_{n \geq 0} A_n \in \mathcal{F}$

Dans ce contexte, les éléments de la tribu $\mathcal{F}$ sont appelés événements , ce sont eux dont nous allons chercher à évaluer la probabilité.

Remarque

On retiendra qu'une tribu est stable par toutes les opérations ensemblistes (dénombrables) : passage au complémentaire, réunion $\cup$, intersection $\cap$, différence \, différence symétrique Δ.

2.2 - PROBABILITE

- Soit $\mathcal{F}$ une tribu sur Ω. On appelle **probabilité** sur $\mathcal{F}$ toute application $P : \mathcal{F} \to [0, 1]$ telle que :

(i) $P(\Omega) = 1$

(ii) Pour toute suite dénombrable $(A_n)_{n\geq 1}$ d'éléments de $\mathcal{F}$, on a :

$A_m \cap A_n = \varnothing$, pour $m \neq n \Rightarrow P(\bigcup_n A_n) = \sum_n P(A_n)$ **(σ-additivité)**

- Un **espace probabilisé**, ou **modèle probabiliste**, est un triplet $(\Omega, \mathcal{F}, P)$

Remarque

L'apparition de la notion de tribu peut dérouter le lecteur ; elle vient du fait que l'on ne peut pas en pratique calculer la probabilité de tous les événements mais seulement des plus simples (ce qui suffit largement aux besoins). En outre, elle est utilisée de façon très courante pour modéliser l'information disponible dans un contexte donné. Nous n'insisterons pas ici sur cette difficile notion, pourtant constitutive du calcul des probabilités.

En ce qui concerne l'axiomatique définissant une probabilité, on aura noté une apparition du dénombrable qui peut également surprendre, alors que l'on définit souvent dans les textes élémentaires une probabilité comme satisfaisant la propriété (i) et la propriété d'additivité :

$$A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$$

Cette dernière propriété, cas particulier de (ii), est suffisante lorsque l'on travaille sur des espaces des épreuves finis. Et certains font remarquer à juste titre que le continu n'est qu'une abstraction, et que l'on devrait se contenter des ensembles finis. Nous n'irons pas aussi loin dans cette voie car elle mène à l'extrême à s'interroger sur le calcul infinitésimal !

Pour étayer cette hypothèse de σ-additivité, disons seulement pour l'instant que de très nombreux calculs de probabilité nécessitent de dépasser le seuil d'opérations ensemblistes finies pour accéder à des concepts limites, au sens mathématique du terme. Prenons l'exemple du jeu de pile ou face, et imaginons que nous souhaitions démontrer la véracité de la loi empirique des grands nombres, à savoir que la fréquence d'apparition de *pile* converge vers ½. Pour ce faire, nous devrons considérer une infinité de lancers, et donc nous devrons évaluer la probabilité d'événements se déduisant de manière ensembliste à partir d'une infinité dénombrable d'événements élémentaires. Dans ce contexte, les notions de tribu et de probabilité telles qu'introduites ici nous seront précieuses.

2.3 - EXERCICE

(i) Montrer que $P(A^c) = 1 - P(A)$

(ii) Montrer que dans le cas où A et B sont deux éléments de $\mathcal{F}$, on a:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(iii) Généraliser la formule précédente au cas de la réunion de trois parties (faire une figure).

2.4 - PROPRIETES

Soit P une probabilité sur une tribu $\mathcal{F}$; alors[8] :

(i) $A \subset B \Rightarrow P(A) \leq P(B)$

(ii) $A \subset B \Rightarrow P(B \setminus A) = P(B) - P(A)$

(iii) $A \subset \bigcup_n A_n \Rightarrow P(A) \leq \sum_n P(A_n)$

(iv) $A_n \uparrow A \Rightarrow P(A_n) \uparrow P(A)$

(v) $A_n \downarrow A \Rightarrow P(A_n) \downarrow P(A)$

Notation

Il convient de préciser les notations employées dans l'énoncé de la proposition : pour une suite réelle $(\alpha_n)_n$, la notation $\alpha_n \uparrow \alpha$ signifie que la suite $(\alpha_n)_n$ est croissante, de limite α. D'une manière analogue, lorsque l'on écrit $A_n \uparrow A$, il faut comprendre que la suite $(A_n)_n$ est croissante pour l'inclusion et que $A = \bigcup_n A_n$; il en est de même de la notation $A_n \downarrow A$ qui signifie que la suite est décroissante, avec $A = \bigcap_n A_n$.

Indications de preuve

Pour établir (i) et(ii), il suffit de remarquer que B est réunion disjointe de B\A et A.

[8] Tous les événements A cités sont évidemment supposés appartenir à F.

Commençons par démontrer la propriété (iv). Pour cela, il suffit de transformer la suite croissante $(A_n)_n$ d'événements en une suite disjointe $(B_n)_n$, selon :

$$B_0 = A_0$$
$$B_n = A_n \backslash A_{n-1}, \text{ pour } n \geq 1$$

On en déduit (iv) immédiatement.
Pour obtenir (iii), on commence par l'établir pour deux événements (exercice) :
si $A \subset A_1 \cup A_2$, $P(A) \leq P(A_1) + P(A_2)$
On en déduit par récurrence immédiate la propriété pour n événements.
Pour conclure, on part de l'égalité :

$$A = \bigcup_n (A_n \cap A),$$

puis on transforme la suite $(A_n \cap A)_n$ en une suite croissante en cumulant les parties. On en déduit que :

$$P(A) = \lim_n P\Big(\bigcup_{i=1}^{n} (A_i \cap A)\Big) \leq \lim_n \sum_{i=1}^{n} P(A_i \cap A) \leq \sum_n P(A_n \cap A) \leq \sum_n P(A_n).$$

La dernière propriété (v) s'établit très simplement en transformant la suite décroissante $(A_n)_n$ d'événements en une suite croissante $(B_n)_n$ en posant par exemple :

$$B_n = A_0 \setminus A_n. \square$$

Avant d'aller plus loin, donnons deux exemples de tribus, l'ensemble de toutes les parties de Ω et la tribu borélienne :

Notation

L'ensemble de toutes les parties de Ω est une tribu ; on la note $\mathcal{P}(\Omega)$.

2.5 - TRIBU BORELIENNE

L'intersection des tribus contenant la famille $\{]-\infty, x[, x \in \mathbb{R}\}$ est une tribu : c'est la plus petite tribu contenant cette famille. On appelle cette tribu la **tribu borélienne** de $\mathbb{R}$; elle est notée $\mathcal{B}(\mathbb{R})$ et ses éléments sont les **boréliens** de $\mathbb{R}$.

On définirait de même la tribu borélienne de $\mathbb{R}^2$ (ou même de $\mathbb{R}^d$) en remplaçant les intervalles $]-\infty, x[$ par les quarts de plan $]-\infty, x[\times]-\infty, y[$; elle est notée $\mathcal{B}(\mathbb{R}^2)$.

Pourquoi de telles complications ? On peut le comprendre à l'aide d'une expérience aléatoire très simple : le tirage d'un nombre au hasard entre 0 et 1. L'espace des épreuves Ω est alors $[0, 1]$ et la probabilité P d'une partie de $[0, 1]$ est simplement sa longueur. Or on ne peut pas calculer la longueur de n'importe quel ensemble de même qu'on ne peut pas calculer l'intégrale de n'importe quelle fonction : il faut ajouter quelques conditions de régularité pour pouvoir le faire[9]. On peut en fait montrer que les boréliens sont les parties dont on peut mesurer la longueur, et donc la probabilité ; ce qui justifie leur introduction dans notre contexte probabiliste.

Donnons dès à présent quelques exemples élémentaires de probabilités.

[9] Voir l'exercice 1 pour une preuve.

2.6 - EXEMPLES

a) Probabilité (ou masse) **de Dirac** au point $a \in \Omega$. C'est la probabilité δ_a définie sur $\mathcal{P}(\Omega)$ par :

$$\delta_a(A) = \begin{cases} 1 \text{ si } a \in A \\ 0 \text{ si } a \notin A \end{cases}$$

b) Probabilité discrète :

On fixe une suite de points $(a_n)_n$ de Ω et une suite de réels positifs $(\alpha_n)_n$ de somme unité (i.e. $\sum_n \alpha_n = 1$). On montre que l'application $P : \mathcal{P}(\Omega) \to [0, 1]$, définie par :

$$P(A) = \sum_{a_n \in A} \alpha_n = \sum_n \alpha_n \, \delta_{a_n}(A)$$

est une probabilité, notée $P = \sum_n \alpha_n \, \delta_{a_n}$

c) Probabilité sur un ensemble fini ou dénombrable :

On suppose que Ω est <u>fini ou dénombrable</u>. Toute probabilité P sur $\mathcal{P}(\Omega)$ est discrète et est donc déterminée par ses valeurs sur les singletons $\{\omega\}$.

En effet, si $A \in \mathcal{P}(\Omega)$, $A = \bigcup_{\omega \in A} \{\omega\}$ (et la réunion est au plus dénombrable et disjointe), donc :

$$P(A) = \sum_{\omega \in A} P(\{\omega\}).$$

Ainsi, se donner une probabilité P sur un ensemble fini ou dénombrable revient à se donner une suite $(p_\omega)_\omega$ de réels positifs, de somme 1, et on a :

$$P(A) = \sum_{\omega \in A} p_\omega \, \delta_\omega.$$

Par exemple, le tirage d'un dé à six faces équilibré est modélisé par la probabilité

$$P = \frac{1}{6}(\delta_1+\delta_2+\delta_3+\delta_4+\delta_5+\delta_6).$$

💣 On se retiendra d'effectuer un raisonnement analogue dans le cas d'un ensemble ni fini, ni dénombrable ; par exemple, dans le cas du tirage d'un nombre au hasard entre 0 et 1, il mènerait à :

$$1 = P([0,1]) = \sum_{x \in [0,1]} P(\{x\}) = \sum_{x \in [0,1]} 0 = 0 \;!!!$$

d) Modèle probabiliste en contrôle qualité

On reprend ici l'exemple 1 présenté en introduction.

- Observation d'une seule pièce :

Notons « + » si la masse de la pièce considérée appartient à l'intervalle de tolérance $[m_0, m_1]$ et « – »sinon.

Dans ce cas, l'espace des épreuves $\Omega = \{-, +\}$ est fini et la probabilité P est définie par :

$P(\{-\}) = p$ (p est la probabilité de produire une pièce défectueuse)

et par suite :

$P(\{+\}) = 1-p$

- Observation de deux pièces :

Ici, l'espace des épreuves est formé de couples :

$$\Omega = \{(-,-), (-,+), (+,-), (+,+)\}$$

Mais la définition de la probabilité P n'est pas immédiate, ni évidente.
Si l'on suppose que le phénomène est répétable, c'est-à-dire si les deux pièces ont le même comportement, on doit avoir avec des notations évidentes :

$$P(\text{« } 1^{\text{ère}} \text{ pièce défectueuse »}) = P(\{(-,-), (-,+)\}) = p_{-,-} + p_{-,+} = p$$

$$P(\text{« } 2^{\text{ème}} \text{ pièce défectueuse »}) = P(\{(-,-), (+,-)\}) = p_{-,-} + p_{+,-} = p$$

A ces conditions s'ajoute évidemment la condition de normalisation de la probabilité :

$$p_{-,-} + p_{-,+} + p_{+,-} + p_{+,+} = 1$$

En supposant p connu, on voit que nous obtenons 3 équations pour 4 inconnues et il reste donc un degré de liberté pour définir le modèle probabiliste. Une hypothèse supplémentaire d'indépendance entre les comportements des 2 pièces, qu'il s'agira de vérifier, donne une possibilité de fixer notre modèle. Nous verrons plus tard que cette seule hypothèse permettra de définir complètement le modèle dans le cas de l'observation de n pièces. Ceci explique d'ailleurs les raisons de l'utilisation intensive, pas toujours justifiée, de ces notions d'indépendance.

e) Probabilité uniforme sur un ensemble fini :
Ω est ici un ensemble fini, de cardinal $\#(\Omega)$ et aucun ω n'est a priori privilégié, donc

$$P(\{\omega\}) = \frac{1}{\#(\Omega)}, \text{ pour tout } \omega \in \Omega \text{ et par suite,}$$

$$P(A) = \frac{\#(A)}{\#(\Omega)}, \text{ si } A \in \mathcal{P}(\Omega).$$

On retrouve alors la définition « naïve » de la probabilité : nombre de cas favorables divisé par nombre de cas possibles.

3 - PROBABILITES CONDITIONNELLES ET INDEPENDANCE

On aborde là les concepts sans doute les plus importants du calcul des probabilités qui cherchent à formaliser la notion vague d'influence ou de non influence d'un événement sur un autre. La plupart des erreurs de raisonnement faites par des non-spécialistes ont trait à ces deux notions dont les définitions très simples cachent la complexité.

Commençons par la notion de probabilité conditionnelle. Il s'agit ici de voir de quelle manière on peut réévaluer la probabilité d'un événement lorsque l'on dispose d'une information supplémentaire. Prenons un exemple élémentaire pour examiner la démarche :

Imaginons un enquêteur qui désire savoir avec quelle probabilité une personne choisie « au hasard » dans la population a un revenu annuel supérieur à 30 kiloeuros (k€). La réponse est théoriquement aisée.

Si Ω représente l'ensemble de la population, avec $\#\Omega = N$, et si A désigne l'événement « la personne a un revenu annuel supérieur à 30 k€ », on a :

$P(A) = \frac{\#A}{\#\Omega}$, car la probabilité P est supposée uniforme sur Ω[10].

Supposons maintenant que l'enquêteur ne se pose la question que pour les femmes ; il est bien clair que son estimation va être modifiée[11]. Si B désigne l'événement « la personne est une femme », il sera tenté de proposer pour probabilité :

$$P(A/B) = \frac{\#A \cap B}{\#B} = \frac{P(A \cap B)}{P(B)}$$

En d'autres termes, l'enquêteur a décidé de ne considérer que des aléas appartenant à l'événement B. Il a en quelque sorte réalisé un zoom de la probabilité P sur le seul événement B.

Dans la suite de la section, on fixe un espace probabilisé (Ω, $\mathcal{F}$, P).

3.1 - PROBABILITE CONDITIONNELLE

Soit B un événement de probabilité non nulle.
La **probabilité conditionnelle** de l'événement A sachant que l'événement B s'est réalisé, notée P(A/B), est définie par :

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

On la note également $P_B(A)$

3.2 - PROPRIETE

Soit B un événement tel que P(B) > 0. L'application P_B est une probabilité sur $\mathcal{F}$.

Preuve

Il suffit de vérifier les axiomes d'une probabilité, ce qui ne pose aucune difficulté. ❐

L'exercice commenté qui suit montre que cette simple définition peut mener, si l'on ne réfléchit pas assez, à d'apparents paradoxes :

EXERCICE COMMENTE (BELOTE)

Une personne joue à la belote (on rappelle que ce jeu nécessite 32 cartes et que chaque joueur en reçoit 8). On considère les événements suivants :

- A_1 : « le joueur reçoit au moins un as »
- A_2 : « le joueur reçoit au moins deux as »
- B : « le joueur reçoit l'as de cœur »

On demande si les probabilités $P(A_2/A_1)$ et $P(A_2/B)$ sont égales comme on pourrait le penser vu que les as sont équiprobables.

Plutôt que se perdre en conjectures, on débute en précisant le modèle probabiliste : Ω est l'ensemble des tirages possibles de 8 cartes parmi 32, $\mathcal{F} = \mathcal{P}(\Omega)$ et P est la probabilité uniforme sur Ω ($\#(\Omega) = C_{32}^{8}$).

Notons E_i l'événement « le joueur reçoit exactement i as » et $p_i = P(E_i)$. Comme $A_2 = \Omega \setminus (E_0 \cup E_1)$, on a :

$$P(A_2/A_1) = 1 - P(E_0/A_1) - P(E_1/A_1) = 1 - P(E_1/A_1),$$

et comme $E_1 \subset A_1$, on a donc :

---

[10] Evidemment, l'enquêteur n'interrogera pas toute la population mais seulement un échantillon; c'est tout le problème de l'inférence statistique.

[11] Malheureusement toujours à la baisse…

$$P(A_2/A_1) = 1 - \frac{P(E_1)}{P(A_1)} = 1 - \frac{p_1}{1 - p_0}$$

Par ailleurs, on obtient comme précédemment l'identité :

$$P(A_2/B) = 1 - P(E_1/B), \text{ d'où :}$$

$$P(A_2/B) = 1 - \frac{P(E_1 \cap B)}{P(B)} = 1 - \frac{C_{28}^7/C_{32}^8}{C_{31}^7/C_{32}^8} = 1 - 4\frac{C_{28}^7}{C_{32}^8}$$

Enfin, comme il est facile de voir que l'on a :

$$p_1 = 4\frac{C_{28}^7}{C_{32}^8}, \text{ on en déduit que } P(A_2/B) = 1 - p_1$$

et donc finalement que $P(A_2/A_1) < P(A_2/B)$, contrairement à ce qu'indiquerait la première impression. La raison intuitive en est qu'il est plus difficile d'obtenir l'as de cœur qu'un as quelconque, et donc que conditionner par rapport à cet événement donne de plus grandes chances d'obtenir d'autres as.

3.4 - FORMULE DE BAYES

Soit $(B_1, B_2, ..., B_n)$ une partition de Ω par des événements de probabilité non nulle et soit A un événement. Alors :

• **Formules des probabilités totales et composées :**

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A/B_i)\, P(B_i)$$

• **Formule de la probabilité des causes (dite formule de Bayes) :**

Soit $j \in \{1, 2, ..., n\}$. Si $P(A) > 0$, on a :

$$P(B_j/A) = \frac{P(A/B_j)\, P(B_j)}{P(A)} = \frac{P(A/B_j)\, P(B_j)}{\sum_{i=1}^{n} P(A/B_i)\, P(B_i)}$$

Preuve

La preuve de la formule des probabilités totales et composées est évidente. La formule de Bayes en découle, si l'on écrit que l'on a à la fois les identités :

$$P(A \cap B_j) = P(A/B_j)\, P(B_j), \text{ et}$$
$$P(A \cap B_j) = P(B_j/A)\, P(A). \ \square$$

Remarque

La formule de Bayes a fait couler beaucoup d'encre parmi la communauté statistique et continue de provoquer de sérieux débats scientifiques. On comprendra facilement que ce qui est en jeu dans ces débats n'est aucunement lié au fait que cette formule est correcte, car la démonstration en est immédiate. C'est plutôt l'utilisation faite par certains statisticiens, qualifiés de bayésiens, qui enflamme depuis fort longtemps la communauté. Nous y reviendrons succinctement lors du chapitre portant sur l'inférence statistique.

3.5 - EXERCICE COMMENTE (ALCOOLEMIE ET ACCIDENTS)

On sait que la consommation d'alcool a une influence sur la probabilité d'avoir un accident de voiture. Par ailleurs, des contrôles effectués dans le passé ont permis également d'évaluer la fréquence d'alcoolémie. Tous ces renseignements sont consignés dans les tableaux ci-dessous[12].

taux d'alcoolémie t (en g/l)	$t < 0{,}1$	$0{,}1 \le t < 0{,}3$	$0{,}3 \le t < 0{,}5$	$0{,}5 \le t$
fréquence d'apparition	0,78	0,13	0,08	0,01

taux d'alcoolémie t (en g/l)	$t < 0{,}1$	$0{,}1 \le t < 0{,}3$	$0{,}3 \le t < 0{,}5$	$0{,}5 \le t$
probabilité d'accident	0,00032	0,00054	0,0009	0,0017

Supposons qu'un accident de voiture vienne de se produire et que nous souhaitions savoir, avant toute analyse, avec quelle probabilité l'individu impliqué à un taux d'alcoolémie donné.

Désignons par Ω l'ensemble des individus de la population considérée, muni de la probabilité uniforme. A représente les individus qui ont un accident au moment donné, et B_1, B_2, B_3, B_4 sont respectivement les événements « l'individu a un taux t inférieur à 0,1 », « l'individu a un taux t compris entre 0,1 et 0,3 », « l'individu a un taux t compris entre 0,3 et 0,5 », « l'individu a un taux t supérieur à 0,5 ».

Faisons le raisonnement pour B_1. Il suffit d'appliquer la formule de Bayes, en utilisant au passage la probabilité d'avoir un accident :

$$\begin{aligned} P(A) &= P(A/B_1)\, P(B_1) + P(A/B_2)\, P(B_2) + P(A/B_3)\, P(B_3) + P(A/B_4)\, P(B_4) \\ &= 0{,}00032 \times 0{,}78 + 0{,}00054 \times 0{,}13 + 0{,}0009 \times 0{,}08 + 0{,}0017 \times 0{,}01 \\ &= 0{,}000409 \end{aligned}$$

Puis,

$$P(B_1/A) = \frac{P(A/B_1)\, P(B_1)}{P(A)} = \frac{0{,}00032 \times 0{,}78}{0{,}000409} = 0{,}61$$

On évalue de même les autres probabilités :

$$P(B_2/A) = 0{,}17$$
$$P(B_3/A) = 0{,}18$$
$$P(B_4/A) = 0{,}04$$

et on observe que la probabilité de l'événement B_1 a diminué, contrairement aux probabilités des événements B_2, B_3, B_4.

Terminons en notant que cet exemple montre bien les raisons qui mènent à qualifier la formule de Bayes de formule de probabilité des causes. En effet, ce contexte correspond au cas où des causes $B_1, B_2, \ldots, B_n$ mènent à l'apparition d'un phénomène A, avec des probabilités associées $P(A/B_1), \ldots, P(A/B_n)$. L'observation des conséquences, ici un accident, permet alors de remonter à la probabilité des causes, ici le taux d'alcoolémie.

Venons-en maintenant à la notion d'indépendance. De façon intuitive, un événement A est indépendant d'un événement B si la réalisation (ou la non réalisation) de B n'influe pas sur la probabilité de A ; ce qui signifie que $P(A/B) = P(A)$. De façon plus symétrique, on posera la définition suivante.

3.6 - INDEPENDANCE DE DEUX EVENEMENTS

> Les événements A et B sont dits **indépendants** si $P(A \cap B) = P(A)\,.\,P(B)$

[12] Ces données sont fictives.

On notera que si A et B sont de probabilités non nulles, on a équivalence entre les deux notions :

- A n'influe par sur B, i.e. $P(A/B) = P(A)$
- B n'influe par sur A, i.e. $P(B/A) = P(B)$

Ceci conforte notre définition mais montre également que les notions présentées ici impliquent que la non influence d'un événement sur un autre est une notion symétrique (contrairement à une relation de cause à effet). En d'autres termes, ce formalisme donne une traduction au phénomène d'influence qui peut devenir restrictive dans certains cas. C'est le prix à payer pour toute activité de modélisation[13].

Comme premier exemple, considérons le lancer de deux dés. Le modèle le plus naturel a priori est $\Omega = \{1,2,3,4,5,6\}^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ et P est la probabilité uniforme sur Ω ($\#(\Omega) = 36$). Il est alors facile de voir que les résultats du premier et du deuxième dé sont indépendants ; par exemple, si A est l'événement « 3 pour le premier dé » et B : « 2 pour le deuxième dé », on a évidemment :

$$P(A) = 6/36 = 1/6,\ P(B) = 6/36 = 1/6 \text{ et } P(A \cap B) = 1/36$$

Par conséquent, les événements A et B sont indépendants.

Cependant, les choses ne sont dans la pratique pas toujours aussi simples et un coup d'oeil à l'exercice commenté consacré à la belote permet de prendre un peu mieux la mesure de cette délicate notion. En effet, le lecteur attentif aura remarqué que l'on a démontré au passage $P(E_1/B) = P(E_1)$; ce qui signifie que les événements « recevoir l'as de cœur » et « recevoir exactement un as » sont indépendants...

L'indépendance de plus de deux événements est moins facile à définir.

3.7 - INDEPENDANCE DE PLUSIEURS EVENEMENTS

Les événements $A_1, A_2, \ldots, A_n$ sont dits **indépendants** (dans leur ensemble, ou mutuellement) si :

$$\forall J \subset \{1, 2, \ldots, n\},\ P\Big(\bigcap_{j \in J} A_j\Big) = \prod_{j \in J} P(A_j).$$

Pour bien comprendre cette définition, on peut commencer par examiner le cas de 3 événements. La définition précédente donne comme conditions d'indépendance les identités[14] :

$$P(A_1 \cap A_2) = P(A_1)\, P(A_2) \qquad \text{cas } J=\{1,2\}$$
$$P(A_1 \cap A_3) = P(A_1)\, P(A_3) \qquad \text{cas } J=\{1,3\}$$
$$P(A_2 \cap A_3) = P(A_2)\, P(A_3) \qquad \text{cas } J=\{2,3\}$$
$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\, P(A_2)\, P(A_3) \qquad \text{cas } J=\{1,2,3\}$$

[13] Le lecteur attentif nous reprochera à juste titre d'avoir en fait introduit l'indépendance entre les deux événements avec notre modèle probabiliste. Nous y reviendrons, car rien n'est plus néfaste que d'introduire des hypothèses de modélisation sans le savoir.

[14] Le cas où J est un singleton n'est pas mentionné puisque l'identité est alors toujours satisfaite.

4 – EXERCICES - ENONCES

I.1 – tout ne peut pas être probabilisé

On va montrer ici que la probabilité uniforme P sur [0, 1] ne peut pas être définie sur toutes les parties de [0, 1] et on raisonne pour cela par l'absurde.
On considère alors l'application f définie sur l'intervalle [0, ½] par :

$f(x) = \{y \in [0, \tfrac{1}{2}],\ x - y \in \mathbb{Q}\}$

puis on choisit un système de représentants de ces parties f(x) (ou classes d'équivalence); ce qui donne un ensemble A contenu dans [0, ½] qui vérifie :
$\forall x \in [0, \tfrac{1}{2}],\ \exists\ a \in A,\ x \in f(a)$
On note enfin A+x l'ensemble $\{a+x,\ a \in A\}$

1) Montrer que $[0, \tfrac{1}{2}] \subset \bigcup_{x \in \mathbb{Q} \cap [0,1/2]} (A + x) \subset [0, 1]$ et que la réunion est disjointe.

2) En remarquant que P(A+x) = P(A) (probabilité uniforme), en déduire que P(A) ne peut être ni nul ni non nul !

I.2 – sur le paradoxe de Bertrand

Soit une urne contenant 2 boules blanches et une boule noire. On désire calculer la probabilité p de tirer au hasard 2 boules de couleurs différentes. On note A l'événement correspondant.

1) On effectue un premier tirage puis, après remise et mélange de la boule dans l'urne, on procède à un deuxième tirage. Donner un modèle probabiliste de cette expérience. Identifier A puis calculer p.

2) On considère maintenant le cas d'un tirage sans remise : on choisit au hasard une boule puis, sans la replacer dans l'urne, on effectue un deuxième tirage. Donner à nouveau un modèle probabiliste de cette expérience. Identifier A et calculer p.

3) Que peut-on conclure des questions 1 et 2 ?

I.3 – modèle probabiliste et dénombrement

n étudiants sont actuellement présents dans la salle (remplacer n par sa valeur) ; et on désire calculer la probabilité qu'aucun d'entre eux ne soit né le même jour. Donner un modèle probabiliste qui va permettre de répondre à cette question (on négligera les années bissextiles). Calculer alors la probabilité cherchée.

I.4 – axiomes probabilistes

Dans une école maternelle, les élèves pendent leurs vêtements aux portemanteaux situés à l'extérieur de la salle de classe. A l'heure de la récréation, il sortent et choisissent au hasard un vêtement. On désire évaluer la probabilité p qu'au moins l'un d'entre eux soit vêtu de son propre habit.

1) Donner un modèle probabiliste de cette expérience et exprimer dans ce modèle les événements '*au moins un élève a le bon habit*' et '*l'élève numéro i a le bon habit*'. Quelles relations ensemblistes peut-on établir entre ces événements ?

2) Etablir la formule, dite formule de Poincaré, donnant la probabilité de la réunion de n événements pas forcément deux à deux disjoints.

3) Utiliser le résultat précédent pour calculer p et examiner le comportement de p lorsque la taille de la classe tend vers l'infini.

I.5 – indépendance d'événements

1) On lance une pièce de monnaie 2 fois. On considère le cas général où la pièce peut être pipée : la probabilité d'obtenir 'pile' vaut p ($p \in [0, 1]$). On s'intéresse aux événements suivants :

A : '*le premier lancer donne pile*'
B : '*le deuxième lancer donne pile*'
C : '*les deux lancers donnent le même résultat*'

Proposer un modèle probabiliste, puis étudier l'indépendance des événements A, B et C (deux à deux, puis mutuelle). Interpréter.

2) On tire trois élèves E_1, E_2, E_3 dans la promotion et on s'interroge sur leur date de naissance ; on choisira donc comme modèle $\Omega = \{1,2,...,365\}^3$, $\mathcal{F} = \mathcal{P}(\Omega)$ et P la probabilité uniforme (on négligera ici les années bissextiles). On considère alors les événements A : « E_1 et E_2 ont le même anniversaire », B : « E_1 et E_3 ont le même anniversaire » et C : « E_2 et E_3 ont le même anniversaire ». Montrer que les événements A, B et C sont deux à deux indépendants mais pas dans leur ensemble.

3) Montrer que les événements A_1, ..., A_n sont indépendants si et seulement si les événements A_1^c , ..., A_n^c le sont.

I.6 – probabilités conditionnelles

Lors d'une fête entre amis, deux d'entre eux, Xavier et Anca, parlent de leur progéniture. Vous arrivez au moment où Xavier apprend à Anca qu'il a deux enfants dont un prénommé Ronan.
réponse de Anca :*Tu as donc également une fille et disant cela, je n'ai qu'une chance sur 3 de me tromper.*
Xavier : *Sais-tu également que Ronan est mon aîné ?*
Anca : *Alors là, je ne peux plus rien dire.*

Pouvez-vous expliquer le contenu de cette conversation ?

I.7 – les pièges du conditionnement

Un test de dépistage d'une maladie est mis au point par un laboratoire pharmaceutique qui se propose d'en déterminer l'efficacité. Il estime pour cela la probabilité α (resp. β) que le test soit positif (resp.négatif) alors que la personne considérée est malade (resp. bien portante) en testant un échantillon de personnes malades (resp. bien portantes) au sein d'une population donnée.

On suppose α et β proches de 1 (par exemple $\alpha = \beta = 0.99$). Proposer des indicateurs d'efficacité du test ; les évaluer.

I.8 – suicides et prisons

Un récent article du journal Le Monde était consacré au suicide dans les prisons. On y apprenait notamment que le taux de suicide y est d'environ 1/1000 et que parmi les détenus s'étant suicidés, on trouvait une population ayant déjà fait une tentative de suicide 15 fois plus importante que de détenus n'ayant jamais attenté à leur vie. Le Monde concluait à la nécessité de surveiller de près les détenus ayant déjà fait une tentative de suicide.

1) Donner un modèle probabiliste prenant en compte les éventualités précisées plus haut (tentative de suicide, suicide). Exprimer dans ce modèle la conclusion du quotidien.

2) Proposer un moyen d'évaluer – à l'aide des renseignements fournis - la justesse de la conclusion précédente, puis faire cette évaluation.

I.9 – probabilités conditionnelles

On considère un ensemble de familles ayant 0,1 ou 2 enfants. La probabilité pour une famille d'avoir 0,1 ou 2 enfants est supposée égale à 1/4, 1/2, 1/4. Vous rencontrez le petit Pierre dans la rue, qui appartient à une de ces familles. Calculez la probabilité qu'il soit fils unique lorsqu'il vous dit :

1) qu'il n'a pas de frère.

2) qu'il n'a pas de soeur.

5 – EXERCICES – INDICATIONS DE SOLUTIONS

Exercice I.1

1) Pour voir que la réunion est disjointe, montrer dans un premier temps que

$a + x = a' + x'$ avec a, a' dans A et x, x' dans $\mathbb{Q}\cap[0, 1/2[$ entraîne $a = a'$

2) Utiliser les propriétés d'une probabilité et remarquer que $\mathbb{Q}\cap[0, 1/2[$ est infini dénombrable…

Exercice I.2

1) et 2) Choisir chaque fois un ensemble Ω le plus simple qui soit !

3) Le fait de trouver p = 4/9 puis p = 2/3 est-il paradoxal ?

Exercice I.3

On peut prendre le modèle suivant :

$\Omega = \{1, \dots , 365\}^n$, de cardinal 365^n, et P est la probabilité uniforme sur Ω.

Exercice I.4

1) On peut choisir le modèle :
- $\Omega = \{ \omega$ application de $\{1, \dots , \}$ dans $\{1, \dots , n\}$ *bijective* $\}$ de cardinal n!
- P probabilité uniforme sur Ω

2) La formule demandée est la suivante :

$$P(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{j_1<j_2<\ldots<j_k} P(A_{j_1} \cap \ldots \cap A_{j_k})$$

L'établir pour n = 2 et n = 3 puis par récurrence sur n…

3) Remarquer que $\#(A_{j_1} \cap \ldots \cap A_{j_k}) = (n-k)!$, cardinal de l'ensemble des bijections de $\{1, \ldots, n\}/\{j_1, \ldots, j_k\}$ sur lui-même…

Exercice I.5

1) Considérer $\Omega = \{P,F\}\times\{P,F\}$ et P probabilité entièrement définie par :

$P(\{(P,F)\}) = p.(1-p)$; $P(\{(P,P)\}) = p^2$

$P(\{(F,P)\}) = (1-p).p$; $P(\{(P,P)\}) = (1-p)^2$

2) Le modèle étant donné, il suffit de faire les calculs…

3) Commencer par n = 2 et s'y ramener dans le cas général.

Exercice I.6

On peut choisir le modèle probabiliste suivant :

$\Omega = \{G,F\}\times\{G,F\}$; P probabilité uniforme

en convenant pour $\omega = (\omega_1, \omega_2)$ dans Ω que ω_1 est le sexe (Garçon ou Fille) de l'aîné.

Exercice I.7

Considérer les événements T^+ « test positif », T^- « test négatif », M « malade », $M^c = \Omega/M$ « non malade » et remarquer que

$\alpha = P(T^+/M)$; $\beta = P(T^-/M^c)$

S'intéresser alors aux deux indicateurs suivants

$I_1 = P(M/T^+)$; $I_2 = P(M^c/T^-)$

Sont-ils différents ? Si oui, en quoi ? Faire une application numérique pour $\alpha = \beta = 99\%$ et $p = P(M) = 1\%$.

Exercice I.8

Considérer Ω l'ensemble de la population carcérale et P la probabilité uniforme sur Ω. En notant S l'événement associé à un suicide et TS à une tentative de suicide, remarquer que les données du Monde se traduisent par :

$P(S) = 1/1000$; $P(TS/S) = 15\times P(TS^c/S)$

La conclusion du Monde revient à penser que P(S/TS) est importante, en tout cas bien plus que P(S). Faire une discussion en fonction de la valeur non fournie de la probabilité P(TS) (on prendra garde au fait que TS peut correspondre à une tentative de suicide en prison).

Exercice I.9

Prendre $\Omega = \{0, 1F, 1G, 2F, 2G, 2GF\}$ avec 0 pour « sans enfant », 1F pour « une fille », etc. Puis, expliquer pourquoi on considère sur Ω la probabilité définie par :

$P(\{0\}) = 1/4$
$P(\{1G\}) = P(\{1F\}) = 1/4$
$P(\{2G\}) = P(\{2F\}) = 1/16$ et $P(\{2GF\}) = 1/8$