Systèmes probabilistes en physique et leur définition au moyen d'appareils de mesure

JEAN DAYANTIS

Laboratoire des Verres, Université Montpellier II
Place Eugène Bataillon, 34095 Montpellier Cedex, France

RÉSUMÉ. On discute d'abord les différentes propriétés auxquels doivent satisfaire les appareils de mesure en général, y compris l'appareil de mesure constitué de l'observateur se servant de ses sens pour procéder à l'estimation de la valeur prise par une variable aléatoire. Ceci permet d'introduire la notion d'appareil de mesure idéal et non-idéal. On classifie ensuite les différentes catégories de systèmes probabilistes que l'on rencontre, non seulement lors de l'activité scientifique, mais aussi dans la vie quotidienne. Cette classification permet d'étendre, au moins en principe, le champ d'application du calcul des probabilités usuel à des systèmes probabilistes qui ne sont pas en général pris en considération. On rappelle ensuite la notion d'équivocation, introduite par Shannon en vue d'étudier les propriétés des canaux bruités, et on l'applique aux appareils de mesure idéaux et non-idéaux, ce qui permet de quantifier la différence d'information obtenue lorsque l'on corrige les résultats obtenus par un appareil de mesure non-idéal grâce à l'utilisation subséquente d'un appareil de mesure idéal. Dans un paragraphe final, on analyse le problème posé par Tribus et on montre qu'il s'agit d'un paradoxe apparent, et que la probabilité peut y être exprimée en tant que fréquence en accord avec la démarche de von Mises.

ABSTRACT. The properties of measuring devices, including human observers making use of their senses to estimate the value taken by a random variable are first discussed. This allows one to introduce the concept of ideal and non-ideal measuring devices. Then the probabilistic systems, which are encountered, not only during scientific research, but also in ordinary life, are described and classified in distinct categories. This allows to extend the field of application of Probability Theory, at least in principle, to systems which were not taken into account by this theory. Recalling the concept of equivocation, introduced by Shannon in order to address channels with noise in Communication Theory, we apply it to non-ideal measuring devices, and calculate the excess information obtained when the output of a probabilistic system is measured first using a non-ideal measuring device, and then an ideal one. After shortly examining improper probabilistic propositions, we conclude the article by showing that in the Tribus experiment, the probability also complies to its usual definition as a ratio of frequencies, in accordance with the von Mises approach.

1 Préliminaires : propriétés rattachées aux appareils de mesure

L'exposé qui suit sera uniquement consacré aux systèmes physiques macroscopiques, de sorte qu'une mesure de grandeur sur l'un quelconque de ces systèmes ne pourra jamais modifier l'état du système. Ceci exclue d'emblée tout système microscopique faisant entrée d'une manière ou d'une autre des considérations de mécanique quantique. Soit alors un système probabiliste (SP) à une variable aléatoire (v.a.) X, pouvant prendre N états (issues) distincts i, à l'état i correspondant la mesure x_i, et soit p_i la probabilité associée à la réalisation de l'état i, de sorte que l'on peut écrire

$$\sum_i p_{\mathrm{i}} = 1 \tag{1}$$

La distribution de probabilités définie par les p_i sera représentée par P. Par ailleurs, la généralisation au cas de plusieurs v.a. X_k, certaines ou toutes d'entre elles pouvant être continues, ne pose pas de problème, de sorte que pour l'objet du présent article nous nous abstiendrons de prendre en compte ces généralisations.

Supposons que les valeurs x_i prises par la v.a. X nécessitent en vue de leur détermination l'utilisation d'un appareil de mesure (a.d.m.) adéquat. Remarquons cependant de suite que tous les SP ne nécessitent pas en vue de la détermination du résultat d'une épreuve l'utilisation d'un a.d.m. adapté ; ceci est le cas par exemple du jeu du pile ou face ou encore le jet d'un dé. Toutefois, le cas où il faut utiliser un a.d.m. sont fréquents et importants en pratique, aussi faut-il d'abord définir les propriétés qui permettent de caractériser tout a.d.m. Elles sont au nombre de trois, respectivement la stationnarité [1], la fiabilité et la précision. Un a.d.m. est dit stationnaire si la distributions de probabilité P qu'il détermine sur le SP sous considération ne varie pas en fonction du temps. (On suppose que les SP lui-même est aussi stationnaire, à distinguer de la stationnarité de l'a.d.m.). Il est dit fiable, si, compte tenu d'une marge d'erreur supposée définie par avance, il donne des valeurs x_i qui correspondent à celles qui sont objectivement celles du SP [2]. Par exemple, un thermomètre mal calibré n'est pas un a.d.m. fiable, bien qu'il puisse être stationnaire, à savoir conduire, compte tenu de la marge d'erreur permise, au même résultat pour la valeur x_i associée à l'issue i de la v.a. X. Enfin, la précision d'un a.d.m. correspond à l'écart moyen entre plusieurs déterminations x_i de l'issue i, en supposant que l'a.d.m. est stationnaire ; mathématiquement, elle peut être définie avec exactitude à partir des notions de variance et d'écart type. Un a.d.m. fiable est nécessairement stationnaire, mais la réciproque n'est pas vraie ; par ailleurs, un appareil précis peut ne pas être fiable, comme ce serais le cas d'un thermomètre d'une bonne précision mais mal calibré, et la réciproque est aussi vrai, un appareil fiable peut ne pas être précis. Il peut aussi être précis sans être ni fiable ni stationnaire. De manière facile à comprendre, la notion de fiabilité est reliée à celle de précision, cette dernière étant une notion essentiellement arbitraire, et déterminée au cas par cas selon les besoins de l'expérimentateur ou de l'observateur. Le cas idéal est bien sûr celui où l'a.d.m. utilisé est à la fois stationnaire, fiable et précis, la notion de fiabilité comme nous venons de le dire étant reliée à celle de la précision requise par l'observateur. Un a.d.m. à la fois stationnaire, fiable et précis sera dit un a.d.m. idéal. La distribution de probabilités P, satisfaisant à (1) et mesurée au moyen d'un a.d.m. idéal sera dite la distribution de probabilités objective du SP sous examen. Nous supposerons désormais que les p_i qui interviennent dans (1) correspondent à une distribution P objective. Dans ces conditions, la relation (1) nous permettra de définir l'information I obtenue, au sens de Shannon [3,4,5], lorsque l'observateur détermine la distribution P et ensuite le résultat d'une épreuve particulière au moyen d'un a.d.m. idéal.

Après ces considérations préliminaires mais essentielles, supposons qu'un observateur ne dispose que de ses sens en vue de déterminer la valeur x_i prise par une v.a. X quelconque. On peut considérer que les sens d'un observateur constituent un appareil de mesure comme tout autre ; cependant, ce genre d'a.d.m. n'est au départ ni stationnaire ni fiable. Cela est à peu près évident, la stationnarité n'est pas acquise car elle peut varier en fonction des circonstances : si par exemple on demande à une personne d'estimer la température d'une eau on y plongeant les mains, son estimation pourra varier de manière considérable si auparavant il avait plongé ses mains dans de l'eau froide ou dans de l'eau chaude. La stationnarité n'étant pas acquise, la fiabilité ne peut pas non plus l'être. La précision (écart entre mesures faite à peu près au même moment) peut aussi varier avec le temps. L'a.d.m. constitué par les sens de l'observateur n'est donc à priori ni stationnaire, ni fiable, et d'une précision qui peut être satisfaisante ou non selon les exigences de l'observateur. Pour autant, les estimations d'un observateur au moyen de ses seuls sens ne sont pas sans valeur, elles peuvent apporter une information non-négligeable. De manière analogue à la distribution objective P satisfaisant à

(1), nous allons donc introduire une distribution de probabilités subjective P' définie par des p'_i et la relation normée

$$\sum_i p_i^{'} = 1 \qquad (1')$$

Alors qu'il existe une seule distribution objective P définie par les p_i, il existe autant de distributions subjectives P' définies par les p'_i que d'observateurs.

Le problème que nous nous posons, et qui constitue en définitive l'objet du présent article, est celui de l'information obtenue, au sens de Shannon, lorsque l'on passe d'une distribution subjective P' à la distribution objective P ; en d'autres termes, quelle est l'information supplémentaire obtenue lorsqu'un observateur procède d'abord à l'estimation du résultat d'une épreuve sur une grandeur aléatoire X en se servant de ses sens, ensuite à sa mesure exacte au moyen d'un a.d.m. idéal. Pour pouvoir cependant poursuivre dans cette voie, il convient d'abord de développer quelques considérations sur les différentes catégories de SP que l'on peut distinguer dans l'univers de la physique, ou encore, sur les différentes catégories qu'induit le mot commun de "probabilité", tel qu'on l'utilise dans le langage quotidien.

2 Considérations sur les catégories de propositions probabilistes en physique

D'emblée, on peut distinguer les catégories suivantes de propositions probabilistes conduisant à des SP : a.- Propositions probabilistes propres directes; b.- Propositions probabilistes propres inverses (ou problèmes de probabilités des causes) ; et c.- Propositions probabilistes impropres. Les propositions des catégories a. et b. sont les propositions probabilistes classiques, objet de tout traité de calcul des probabilités. Ce n'est pas le cas des propositions de la catégorie c., où le mot "probable" intervient souvent au sens courant du langage ordinaire. Nous allons dans ce qui suit examiner d'abord les propositions probabilistes de la catégorie a., en essayant d'en étendre l'application à des cas qui ne sont pas en général envisagés dans les traités de calcul des probabilités usuels. Ensuite, nous traiterons brièvement des propositions probabilistes de la catégorie c.

D'une manière tout à fait générale, une proposition probabiliste ne présente d'intérêt d'un point de vue strictement scientifique que si elle génère un SP dont la probabilité des issues s'expriment par des chiffres bien définis, toujours inférieurs à un et satisfaisant aux relations (1) ou (1'). Le chiffre associé à toute issue se définit de la manière usuelle, en tant que rapport des épreuves favorables sur le nombre total d'épreuves. Ainsi, dans un dé sans défaut, la probabilité d'obtenir le six est 1/6, valeur que l'on obtient à la limite d'un nombre d'épreuves tendant vers l'infini. Par ailleurs, on peut subdiviser les SP de la catégorie a., en SP à épreuves spontanées, en SP à épreuves non-spontanées mais réalisables, enfin en SP à épreuves non-spontanées et non-réalisables. Cette dernière sous-catégorie n'est pas à notre connaissance prise en compte dans les traités de probabilités usuels.

2.1 SP à épreuves spontanées

On peut définir les SP à épreuves spontanées comme ceux dont les épreuves se produisent de manière spontanée dans la nature, en d'autres termes sans intervention humaine. Les probabilités de transition en mécanique quantique sont un exemple de SP à épreuves spontanées. Un autre exemple est la désintégration d'un corps radioactif. On sait depuis Rutherford que dans des périodes de temps courtes par rapport à la durée de vie du corps radioactif, l'émission de particules obéit à une distribution de Poisson. Un troisième exemple peut être tiré de la météorologie : quel temps fera-t-il le 15 janvier 2005 à un endroit donné du globe ? On suppose qu'il existe pour l'endroit des données météorologiques suffisantes pour permettre une projection sur l'avenir. Pour simplifier, on peut supposer qu'il n'existe que deux types de temps, le beau temps réalisé en moyenne A fois et le mauvais temps réalisé B fois. Les probabilités respectives sont A/(A+B) et B/(A+B). L'épreuve se fera d'elle même le 15 janvier 2005 et les observateurs humains n'auront qu'à enregistrer le résultat.

2.2 SP à épreuves non-spontanées mais réalisables. Degrés de liberté agissant sur la loi de distribution de la variable aléatoire

L'exemple le plus simple d'un tel SP est le jeu du pile ou face. L'épreuve n'est pas spontanée, car une intervention, en principe humaine, est nécessaire pour lancer la pièce en l'air. Mais les épreuves sont facilement réalisables, car il n'y a aucune difficulté pour lancer la pièce en l'air un nombre de fois statistiquement significatif. Ce problème (et autres problèmes analogues) sont parfaitement étudiés, et ont donné naissance au chapitre du calcul des probabilités traitant de la distribution binomiale. Un autre exemple de SP à épreuves non-spontanées peut-être tiré du sport : un tireur au but et un gardien de but étant choisis, il s'agit de savoir combien de buts en moyenne le tireur marque au gardien de but à partir du point de penalty au bout de cinq tirs. Il convient de définir la distribution de probabilités P pour marquer 0, 1, 2, 3, 4, ou 5 buts. On voit dans cet exemple surgir la notion de degrés de liberté agissant sur ce type de SP. En effet, on a un degré de liberté pour la définition du gardien de but, et un autre pour celle du tireur au but. On élimine un degré de liberté en choisissant de toujours garder le même gardien de but, alors que le tireur lui change à chaque tir. On en éliminerait un deuxième en fixant également le tireur au but. On sait que ce procédé n'est pas sans importance lors de parties du football au résultat nul après la fin du temps réglementaire, afin de désigner un vainqueur. Ici aussi on peut considérer que l'on a un SP à épreuves non-spontanées mais réalisables, cette dernière notion étant quelque peu arbitraire. En effet, et éliminant pour simplifier les degrés de liberté (ce qui revient à conserver toujours le même gardien de but et le même tireur au but), il est concevable que l'on puisse répéter les épreuves un nombre suffisant de fois pour avoir des résultats statistiquement significatifs, à savoir connaître avec une précision raisonnable la distribution de probabilité P pour le nombre de buts marqués au bout de cinq essais.

2.3 SP probabilistes à épreuves non-spontanées et non-réalisables en pratique

Tel ne serait plus le cas si au lieu de considérer cinq tirs au but, on prenait pour SP des parties de football entières et leur résultat, par exemple les rencontres entre les équipes de France et d'Angleterre. Théoriquement, il ne devrait pas exister de SP à épreuves non-spontanées et non-réalisables un nombre de fois statistiquement significatif, car, si l'on sait réaliser une épreuve une fois, on devrait savoir la réaliser un nombre arbitraire de fois. Toutefois, en pratique, il est clair que les choses ne sont pas ainsi. On n'imagine pas en effet que l'on puisse répéter les rencontres France-Angleterre plusieurs centaines de fois de suite et à courts intervalles de temps, afin d'avoir un décompte significatif des victoires, défaites, et match nuls de l'une des équipes par rapport à l'autre. En outre, dans le cas des équipes entières de football, les degrés de liberté s'accroissent considérablement (vu le nombre de joueurs différents qui peuvent jouer lors des différentes rencontres), et chaque degré de liberté ajouté modifie quelque peu davantage la distribution (inconnue) initiale P, jusqu'à rendre l'existence même de celle-ci incertaine (absence de stationnarité, condition indispensable à la définition sans équivoque de P). On peut évidemment réduire considérablement le nombre de degrés de liberté en assumant que les deux équipes conservent la même composition lors des différentes rencontres, que le terrain reste le même, que les conditions climatiques et les autres paramètres (par exemple, le soutient du public) sont aussi peu différents que possible d'une rencontre à la suivante. On peut ainsi assumer l'existence d'une distribution P objective, mais en général, contrairement au jeu du pile ou face, on ne saura pas répéter les épreuves un nombre de fois statistiquement significatif, afin de connaître effectivement cette distribution. Mais le fait que nous ne pouvons pas dans la pratique la connaître, ne signifie pas qu'elle n'existe pas. Dans son essence même le problème est identique à celui des tirs au but, c'est un problème de calculs des probabilités objectif, même si en pratique on ne sait pas le résoudre de manière satisfaisante. Pour citer un analogue, un problème d'analyse peut comporter une solution, bien que cette dernière ne nous soit pas connue, alors que, par exemple, la quadrature du cercle au moyen de le règle et du compas est un problème qui ne comporte pas de solution (aucune solution n'existe). Les estimations (pronostics) qui interviennent dans les courses de chevaux (tiercé, etc...) relèvent exactement de la même problématique.

Un nouveau paramètre doit cependant être pris en compte lorsqu'intervient le facteur humain, celui de l'indépendance des épreuves. Il est clair que le jet d'une pièce de monnaie ou d'un dé n'a pas de mémoire, de sorte que le résultat de la $(n+1)^{\text{ième}}$ épreuve ne saurait être influencé par celui de la $n^{\text{ième}}$. Toutefois, dans le cas des rencontres de football ou d'autres problèmes équivalents, des facteurs de psychologie humaine peuvent également intervenir : une équipe B constamment battue par une équipe A, risque d'aborder la $(n+1)^{\text{ième}}$ rencontre entre les deux équipes dans un esprit défaitiste. La psychologie, pour utiliser la terminologie introduite dans la partie A, risque donc de rendre la distribution P non-stationnaire, alors que la stationnarité est un présupposé nécessaire en vue de raisonner en termes de probabilités objectives.

Si l'on néglige cette difficulté supplémentaire, on peut représenter par un chiffre la probabilité qu'a l'équipe A de l'emporter sur l'équipe B. Si i est la différence de buts a l'avantage de l'équipe A, cette probabilité est donnée par

$$\sum_{i=0}^{+\infty} p_i \Big/ \sum_{i=-\infty}^{+\infty} p_i \qquad (2)$$

alors que la probabilité d'un résultat nul est

$$p_o \Big/ \sum_{i=-\infty}^{+\infty} p_i \qquad (3)$$

Si le rapport (2) vaut par exemple 0.75, l'équipe A aura 3 chances sur 4 de l'emporter sur l'équipe B, résultat objectif, mais que l'on ne peut connaître dans la pratique pour les raisons ci-dessus développées. Si un observateur estime que l'équipe A a deux chances sur trois de l'emporter sur l'équipe B, il fera là une estimation subjective, dans le sens correct certes, mais en-deçà de la réalité objective. Les SP où la distribution de probabilités P existe mais ne peut être effectivement déterminée seront appelés SP gibbsiens, du nom du grand thermodynamicien Gibbs[6] qui a introduit la notion d'expériences réalisées par la pensée (thought experiments).

Ce qui précède nous conduit à la deuxième partie du problème que nous aimerions traiter, à savoir quantifier l'information, au sens de Shannon [3,4,5] gagnée par l'observateur qui a risqué une estimation, lorsque le résultat de l'épreuve est acquis et mesuré si nécessaire au moyen d'un appareil de mesure idéal. Il nous faut donc calculer la différence entre deux informations, celle apportée par l'estimation subjective de l'observateur (qui, nous le répétons, n'est pas forcément sans valeur), et celle acquise par la mesure du résultat objectif de l'épreuve, en particulier grâce à l'utilisation éventuelle d'un a.d.m. supposé idéal. Pour cela il nous faut d'abord faire appel à la notion d'équivocation, notion introduite par Shannon[3].

3 Equivocation d'un observateur assimilé à un appareil de mesure

Nous nous plaçons d'emblée, comme indiqué dans la partie 1, dans le cas où le SP ne comporte qu'une suite discrète d'issues possibles. Si le SP n'est pas de nature discrète par essence, il est souvent discrétisé pour les besoins de la pratique. C'est ce qui se passe par exemple en statistique des populations. Du reste, la généralisation au cas continu ne présente pas de difficultés. Voici, pour fixer les idées, deux exemples de tels SP discretisés :

1) On dispose d'une longue série de pieux de longueur variable. On sait, grâce à une règle graduée, mesurer la longueur de chaque pieu au centimètre près ; pour chaque longueur l_i on aura une probabilité associée p_i.

2) Au bord d'une autoroute, on sait, grâce à un appareil de mesure adéquat (radar) mesurer la vitesse de tout véhicule au kilomètre près. On aura ainsi, par une statistique suffisante, la probabilité p_i pour qu'un véhicule choisi au hasard ait la vitesse v_i.

Choisissons maintenant un observateur démuni d'a.d.m. et à qui on demande d'estimer, en se servant uniquement de ses sens (en l'occurrence la vue), la longueur des pieux au centimètre près et la vitesse des véhicules sur l'autoroute au kilomètre près. Pour chaque longueur objective l_i des pieux et pour chaque vitesse objective v_k des véhicules, l'observateur donnera, du fait de la déficience de ses sens, des distributions plus ou moins étalées de longueurs et de vitesses (qui peuvent ou non contenir la longueur réelle l_i ou la vitesse réelle v_k). Appelons q_{ij} la probabilité conditionnelle pour que l'observateur estime la longueur d'un pieu donné égale à l_j alors que la longueur objective est l_i (même démarche pour les vitesses des véhicules) ; les q_{ij} définissent une matrice carrée Q d'ordre N, N étant le nombre d'issues possibles du SP. (S'il y a des estimations l_j qui sortent du cadre des issues objectives possibles, on augmentera le nombre de ces dernières pour y inclure les valeurs subjectives sortant de ce cadre, mais en leur attribuant des probabilités objectives p_i nulles). Tous les q_{ij} sont des quantités positives, supérieures ou égales à zéro et inférieures à un, satisfaisant aux conditions

$$\sum_i q_{ij} = 1 \qquad 0 < q_{ij} < 1 \tag{4a}$$

et

$$\sum_i p_i q_{ij} = p'_j \tag{4b}$$

soit un total de 2N conditions. Q est une matrice stochastique. Par ailleurs, les probabilités objectives p_i définissent un vecteur N-dimensionnel $\vec{P}$, vecteur stochastique, alors que les probabilités estimées (probabilités subjectives) p'_k définissent un autre vecteur stochastique N-dimensionnel $\vec{P}'$; certaines des composantes aussi bien de $\vec{P}$ que de $\vec{P}'$ peuvent éventuellement être nulles. Compte tenu de (4a), on établit facilement les relations

$$\vec{P} = {}^t Q \vec{P}' \tag{5a}$$

et

$$\vec{P}^{©} = \left[{}^t Q \right]^{-1} \vec{P} \tag{5b}$$

où ${}^tQ$ est la matrice transposée de Q et $[{}^tQ]^{-1}$ est l'inverse de la transposée. Q doit donc être régulière, donc à déterminant non nul [7]. La matrice inverse Q^{-1} aura des composantes Q_{ij} déterminés selon les règles du calcul matriciel à partir des composantes q_{ij} de la matrice Q, et l'on aura encore

$$\sum_j Q_{ji} = 1 \tag{6a}$$

et $$\sum_j p'_j Q_{ji} = p_i \tag{6b}$$

Q_{ji} est la probabilité conditionnelle pour que la hauteur du pieu soit l_i lorsque l'estimation aura été l_j. Les éléments Q_{ji} introduisent la notion d'équivocation pour les estimations de l'observateur, notion introduite au départ par Shannon en vue de quantifier le défaut d'information suite aux erreurs de transmission. Lorsque l'observateur estime la longueur d'un pieu égale à l_j, la longueur objective (celle que donne un a.d.m. idéal) pourra être l_i, l_j, l_k... avec des probabilités conditionnelles respectives Q_{ji}, Q_{jj}, Q_{jk}..., on est donc incertains sur la longueur objective (= réelle) du pieu, mais cette incertitude n'est pas quelconque, elle obéit à la distribution de probabilités conditionnelles définies par les Q_{ji} de la matrice Q^{-1}. Cette incertitude de la valeur objective de la variable par rapport à l'estimation de l'observateur définit comme nous venons de le dire l'équivocation de l'observateur et, plus généralement, d'un appareil de mesure quelconque non-idéal. (Pour un a.d.m. idéal il n'y a pas d'équivocation, ce qui se traduit mathématiquement par le fait que la matrice Q et son inverse Q^{-1} ont tous leurs éléments non-diagonaux nuls et ceux de la diagonale égaux à 1, de sorte que $Q = Q^{-1} = {}^tQ$ et que le vecteur $\vec{P}$ est identique au vecteur $\vec{P}'$).

Pour un appareil de mesure non-idéal usuel, ont peut admettre que les composantes q_{ij} de Q sont des constantes indépendantes du temps, et de même pour les composantes Q_{ij} ; cependant, pour un observateur humain, s'ajoutent les difficultés supplémentaires provenant de la non-stationnarité éventuelle de sa distribution estimée P' et de la non-indépendance éventuelle de ses estimations successives, notions qui sont corrélées sans être identiques. L'hypothèse de stationnarité signifie que les probabilités conditionnelles q_{ij} ne dépendent pas de conditions externes ou internes à l'observateur (température, luminosité, heure du jour, saison de l'année, état psychologique de l'observateur, et ainsi de suite). La réalisation de cette propriété, s'agissant d'êtres humains, est donc plus que douteuse. L'indépendance des estimations signifie quant à elle que l'observateur n'est pas influencé par ses précédentes estimations, qu'il les a en quelque sorte oubliées. Cette propriété est également fort douteuse. Elle est manifestement fausse dans un exemple précédemment cité, où l'on demande à un observateur d'estimer la température d'un bain d'eau, en lui faisant au préalable plonger les mains dans de l'eau chaude puis lors l'épreuve suivante dans de l'eau froide. En revanche, pour les a.d.m. "mécaniques", et compte tenu de leur utilisation correcte, on peut raisonnablement admettre qu'elles sont sensiblement satisfaites. Elles sont évidemment satisfaites pour un a.d.m. idéal, sans quoi il ne pourrait être idéal.

4 Quantification de l'information acquise lors de l'estimation d'une variable aléatoire par un appareil de mesure non-idéal

Si la valeur prise par la v.a. était exactement mesurée par l'intermédiaire d'un a.d.m. idéal, il n'y aurait pas d'équivocation, et l'information obtenue serait donnée par la relation de Shannon,

$$I = -\sum_i p_i \ln p_i \tag{7}$$

En fait, la mesure sera le plus souvent biaisée, ce sera en pratique toujours le cas lorsque l'a.d.m. est tout simplement constitué des sens de l'observateur. Pour simplifier, nous supposerons ici que quel que soit l'a.d.m. non-idéal, il satisfait aux conditions de stationnarité et d'indépendance. Shannon a montré[8] que si une source produit de l'information au taux donné par la relation (7) (information produite par unité de temps), et si cette information est transmise par l'intermédiaire d'un canal bruité, l'information I' obtenue à la sortie du canal est moindre que I et donnée par

$$I' = -\sum_i p_i \ln p_i + \sum_i \sum_j p'_j Q_{ji} \ln Q_{ji} \tag{8}$$

où les Q_{ij} ont la même signification que ci-dessus. (Q_{ji} est la probabilité conditionnelle pour que le signal envoyé en tant que i soit reçu en tant que j). La différence d'information I produite à la source et reçu à la sortie du canal sera donc la différence entre (8) et (7), soit

$$\Delta I = I - I' = -\sum_i \sum_j p'_j Q_{ji} \ln Q_{ji} \tag{9}$$

La théorie de Shannon pour le taux de transmission de l'information se transpose à notre problème sans modifications, en éliminant tout simplement la variable temps. L'information obtenue par un a.d.m. non-idéal (en particulier, l'observateur humain) est donnée par (8) et l'information supplémentaire obtenue lorsque l'on corrige le résultat de l'a.d.m. non-idéal par celui donné par un a.d.m. idéal est donnée par (9) changée de signe. On remarque que le calcul de cette information complémentaire fait intervenir les vecteurs P et P' ainsi que la matrice Q. Si Q ainsi que l'un de ces vecteurs sont connus, le deuxième s'en déduit à partir des relations (5) ou (6) ; en revanche, si $\vec{P}$ et $\vec{P}'$

sont connus mais non pas Q, cette dernière matrice ne pourra pas en général être complètement déterminée, les composantes Q_{ji} étant en général plus nombreuses que les conditions auxquelles elles sont astreintes. En plus de P et P', une connaissance complète de la matrice Q^{-1} (ou Q) ne sera en général possible que si un certain nombre de ses composantes Q_{ji} auront aussi été déterminées expérimentalement.

A partir de la relation (8) on peut définir deux cas extrêmes : premièrement, celui où $Q_{ij} = \delta_{ij}$, où δ_{ij} est le symbole de Kronecker, auquel cas l'estimation sera toujours confondue avec la valeur objective (exacte) de la v.a., en d'autres termes, l'appareil de mesure utilisé sera un a.d.m. idéal ; l'information I obtenue alors aura sa valeur maximale donnée par la formule (7) de Shannon. L'autre cas limite est celui où il n'y a aucune corrélation entre l'estimation j et la valeur objective i prise par la v.a. Dans ce cas on peut poser $Q_{ji} = p_i$, ce qui veut dire que quelle que soit la valeur de l'estimation j, la probabilité conditionnelle pour que la valeur objective soit i est égale à la probabilité non conditionnelle p_i, en accord avec la distribution objective P. En reportant cette valeur des Q_{ji} dans (8), on voit aisément que les deux termes de droite s'annulent, de sorte que l'information obtenue lors d'une estimation est nulle, résultat attendu.

La relation (8) résout en théorie la question posée en 2,3, à savoir l'information I' obtenue lorsqu'un observateur avance un pronostic sur le résultat d'une rencontre de football ou bien au sujet du tiercé. En théorie seulement, car en pratique la distribution objective gibbsienne P restera le plus souvent inconnue, ainsi que la matrice Q. Elle les résout non seulement en théorie mais aussi en pratique dans les cas où la distribution n'est pas gibbsienne, c'est-à-dire que la détermination de la distribution P n'induit pas des difficultés très grandes, voire insurmontables. Cela est le cas pour les exemples donnés en début de section 3, la distribution de la longueur de pieux et la distribution des vitesses des automobiles sur l'autoroute.

5 Propositions probabilistes impropres

Le long de ce travail nous avons adopté la définition de von Mises[9] selon laquelle une probabilité est une fréquence dans un ensemble et peut donc toujours être représentée par un chiffre compris entre zéro et un. En fait, nous voudrions montrer ici que si l'on ne sait pas faire correspondre un chiffre au mots "probable", "probabilité", ou tout autre mot ou locution équivalente, la dite proposition conduit à un problème probabiliste impropre, en ce sens qu'elle ne comporte pas d'alternative mais a une issue unique. Considérons en effet les propositions suivantes :

a. "Paul est probablement un homme heureux."

b. "Les courbes de tension de vapeur au voisinage du point critique ne sont probablement pas analytiques" [10].

c. "la dernière conjecture de Fermat est probablement vraie" [11].

Dans toutes ces propositions interviennent les mots "probable", "probablement." Cependant ces propositions ne peuvent conduire à un SP aux épreuves répétitives conduisant à des issues différentes, à chacune desquelles l'on associe une probabilité. Examinons d'abord la première proposition : il faut d'abord savoir définir exactement ce qu'est un homme heureux, ce qui sur le plan conceptuel peut présenter certaines difficultés. Admettons que ces difficultés aient été surmontées, il est alors clair que toute épreuve (= enquête bien menée) conduira à la conclusion unique que Paul est un homme heureux ou non heureux. Ceci s'oppose donc complètement au jeu du pile ou face, ou à la succession des épreuves d'un SP dans lequel plusieurs issues sont possibles. Dans le jeu du pile ou face ordinaire, il y a deux issues possibles, qui se produisent toutes les deux lors de la succession des épreuves. Ici, on serait plutôt dans le cas d'une pièce de monnaie ayant deux "pile" ou deux "face", l'observateur étant au départ dans l'ignorance de la structure que la pièce de monnaie pourrait avoir. De la sorte, quelle que soit l'épreuve, l'issue serait toujours la même. La question est cependant délicate, de sorte qu'il est utile d'analyser brièvement le problème correspondant à une distribution $p(l_i)$ de pieux de longueur l_i. Dans ce cas, l'observateur ayant choisi un pieu, il dira que "ce pieu est probablement de longueur l supérieure à l_0." Toutes les mesures successives que l'on pourra faire avec un a.d.m. idéal sur la longueur de ce pieu, conduirons toujours à l'une des deux conclusions, soit qu'effectivement la longueur du pieu est supérieure à l_0, soit à la conclusion inverse (on néglige le cas où on aurait exactement $l = l_0$.) Il n'y a pas différentes issues possibles, chacune avec sa probabilité propre. On voit sur cet exemple la subtile différence qui distingue le SP propre "ayant choisi au hasard un pieu dans la population définie par les $p(l_i)$, quelle est la probabilité pour que sa longueur soit supérieure à l_0 ?", du système impropre, "ayant choisi un pieu au hasard dans la population définie par les $p(l_i)$, un observateur estime probable que sa longueur est supérieure à l_0". Dans le premier cas le problème en tant que mesure de probabilité se chiffre correctement à l'aide d'une expression de type de la relation (2), c'est-à-dire à partir de la distribution objective des $p(l_i)$. Dans le deuxième cas, si $P'(l_i)$ est la distribution estimée par l'observateur correspondant à la longueur objective l_i, on peut certes toujours poser, en tant que probabilité subjective propre à l'observateur, le rapport

$$R = \sum_{ij} P'_{ij}(l_j \rangle l_0) \Big/ \sum_{ij} P'_{ij}(l_j \rangle 0) = \sum_{ij} P'_{ij}(l_j \rangle l_0) \qquad (10)$$

Mais ce rapport a une valeur variable selon l'observateur et ne peut être rattaché à une quelconque fréquence objective. Ainsi, son intérêt est fort réduit. Des considérations analogues peuvent également être développées pour les propositions b. et c. ci-dessus, et il en ressort en conclusion, que l'on ne peut relier les propositions à issue unique, propositions impropres, à une quelconque probabilité objective, traduite par un nombre positif inférieur à un et bien défini.

Les propositions probabilistes propres inverses ou problèmes de probabilités des causes (voir début de la section B) obéissent à la même problématique que les propositions probabilistes propres directes, mais ne seront pas davantage examinées ici, pour ne pas trop allonger le présent exposé.

Pour terminer, montrons comment un problème cité par Tribus,[12] en tant qu'exemple de SP n'obéissant pas à la définition de la probabilité à partir des fréquences selon von Mises, rentre en fait dans nos catégories précédentes des systèmes probabilistes propres : Tribus cite le cas d'un dé en sucre, qui au départ a une probabilité de 1/6 pour chaque face, mais qui se trouve rapidement détérioré par les jets successifs, de sorte que l'on ne peut effectuer avec une suite d'épreuves de Bernoulli. Bien que la probabilité soit au départ chiffrable, nous dit Tribus, sur l'ensemble des jets on ne peut la ramener à un rapport de fréquences. Dans le langage utilisé dans le présent article, nous dirons qu'ici la distribution objective P n'est pas stationnaire, ou plus exactement que le SP lui-même n'est pas stationnaire, alors que notre souci jusqu'à présent aura été plutôt de savoir si ce ne serait pas la distribution P' (correspondant à l'estimation subjective d'un observateur lors d'une mesure nécessitant un a.d.m.) qui pourrait ne pas être stationnaire. Il est cependant facile de ramener ce problème de Tribus de SP à distribution P non-stationnaire à des probabilités représentées par le rapport d'épreuves favorables sur le nombre total d'épreuves. Il suffit pour cela, après le jet j, de fabriquer un dé en matière résistante exactement identique au dé de sucre détérioré par les j jets précédents. Avant de jeter le dé pour une $(j+1)^{\text{ième}}$ fois, on déterminera la distribution objective P(j) correspondant au $j^{\text{ième}}$ jet et on aura ainsi la probabilité du résultat de chaque épreuve à l'étape j en tant que rapport de fréquences. De plus, en utilisant la relation (7) de Shannon, on aura aussi l'information (variable) obtenue à chaque jet du dé : le paradoxe de Tribus n'est qu'une apparence. Par ailleurs, il est inutile d'insister sur le fait qu'en pratique, il sera plus que recommandé d'opérer toujours avec des distributions objectives P stationnaires.

6 CONCLUSION

La classification du présent article a conduit à la notion de distribution de probabilités gibbsiennes, qui permet d'étendre à une large classe de systèmes probabilistes la définition de la probabilité en tant que rapport des épreuves favorables au nombre total d'épreuves, ou encore, en accord avec von Mises, la fréquence avec laquelle les diverses issues possibles apparaissent lors des épreuves successives. Si on exclue les systèmes probabilistes impropres, à issue unique, il est conceptuellement possible de faire correspondre à toute proposition contenant le mot "probable" ou une locution équivalente, un chiffre compris entre 0 et 1 et représentant sa probabilité. Ce chiffre ne sera cependant pas effectivement accessible pour les systèmes probabilistes gibbsien, en raison des efforts prohibitifs nécessités pour la détermination de la distribution P, mais ceci n'infirme en rien son existence objective.

Admettons par ailleurs que même pour les appareils de mesure non-idéaux (dont les observateurs humains utilisant uniquement leurs sens en vue de leurs estimations), les deux conditions de la stationnarité de la distribution des probabilités et de l'indépendance des épreuves sont satisfaites. Alors, toutes les fois que le vecteur objectif $\vec{P}$ et la matrice Q seront connus, on saura quantifier au moyen de la relation (8) l'information I' obtenue lors d'une mesure avec un appareil de mesure non-idéal et en particulier lors de l'estimation d'une grandeur quelconque par un observateur humain. Remarquons enfin, pour terminer, que si, du point de vue conceptuel, les deux distributions P et P' peuvent être admises de nature différente (la première étant objective et obtenue grâce à un appareil de mesure idéal, la deuxième étant subjective à un observateur donné et pouvant avoir peu ou pas de relations avec la réalité), du point de vue strictement mathématique, elles jouent des rôles parfaitement symétriques.

Références et notes

[1] Le mot "stationnarité" est un barbarisme, mais son usage est beaucoup trop commode pour que nous décidions de nous en passer.

[2] Nous n'entrerons pas ici dans une discussion philosophique sur l'existence d'un monde extérieur indépendant de l'observateur. Ce débat pourrait n'être jamais clos, et en pratique il n'y a pas d'intérêt de faire de la physique si l'on adopte le

point de vu de ceux qui disent qu'aucune réalité n'existe en l'absence de la conscience d'un observateur.

[3] C.E. Shannon and W. Weaver, The Mathematical Theory of Communication,The University of Illinois Press, Urbana 1949.

[4] A.L. Khinchin, Mathematical Foundations of Information Theory, Dover, N.Y. 1957.

[5] L. Brillouin, Science and Information Theory, Academic Press, N.Y. 1962.

[6] Josiah Willard Gibbs (1839-1903), savant américain, principal père fondateur de ce que nous appelons aujourd'hui la physique (ou mécanique) statistique.

[7] Ce n'est pas ici la place pour détailler les propriétés auxquelles doivent satisfaire les composantes q_{ij} de la matrice Q (ou Q_{ji} de la matrice Q^{-1}) ainsi que les possibilités de leur détermination éventuelle hors expériences adéquates. Des indications en ce sens sont données dans l'article de l'auteur, Mesures Physiques et Théorie de l'Information, Ann. Fond. L. de Broglie, 6(1), 40-71(1981).

[8] Voir la réf. 3 ci-dessus, chap.II, ainsi que la réf. 7.

[9] R. von Mises, Probability, Statistics and Truth, MacMillan, N.Y. 1939.

[10] J.S. Rowlinson, Nature, 225,1196(1970).

[11] J. Itard, Arithmétique et Théorie des Nombres, P.U.F., Collection "Que sais-je", Paris 1967. Depuis la parution de ce petit livre, la dernière conjecture de Fermat a été résolue par l'affirmative.

[12] M. Tribus, Décisions rationnelles dans l'incertain, Masson, Paris 1972

(Manuscrit reçu le 28 septembre 2000)