X. LES FONDEMENTS DE LA STATISTIQUE.

X.1. Le principe des tests statistiques.

Nous avions vu que si on jette par exemple mille fois une pièce de monnaie, le nombre N de pile est une variable aléatoire pouvant prendre toutes les valeurs entre zéro et mille ; la plus probable est 500 et la probabilité d'avoir $500 + j$ pile est environ

$$\frac{1}{\sqrt{500\,\pi}}\ \exp\left\{-\frac{j^2}{1000}\right\}$$

en supposant que pile ou face a à chaque fois exactement une chance sur deux de sortir. On peut constater ceci : pour $j = 10$, $\exp\{-j^2/1000\} = e^{-0.1} \simeq 0.99$, c'est-à-dire que la probabilité d'obtenir 510 (ou 505, ou 507) est pratiquement la même que celle d'obtenir 500. Par contre pour $j = 100$, $\exp\{-j^2/1000\} = e^{-10} \simeq 0.000045$, c'est-à-dire que la probabilité d'obtenir 600 est 22000 fois plus petite que celle d'obtenir 500. Si on fait l'expérience avec une pièce correctement équilibrée, il n'y a pratiquement pas plus de chances d'obtenir *exactement* 500 pile plutôt que 501 ou 499 ; la probabilité d'obtenir 500 (ou 501 ou 499) est à peu près $1/\sqrt{500\pi} \simeq 1/40$), et il y a à peine moins de chances d'obtenir 507, 485, ou 511. Par contre il est extrêmement peu probable d'obtenir 600 (une chance sur 873000) ; la probabilité d'obtenir pile 600 fois *ou plus* est la somme des probabilités d'obtenir k pile, de $k = 600$ à $k = 1000$, et est environ 1/250000.

Par conséquent, si après avoir fait cette expérience on obtient 600 pile, on est fondé à en déduire que la pièce est mal équilibrée et que pile avait plus de chances de sortir que face (que par exemple la probabilité d'avoir pile n'est pas 0.5, mais plutôt 0.6). Par contre si on obtient 511 pile, on n'est absolument pas fondé à en déduire que la pièce est mal équilibrée (que par exemple la probabilité d'avoir pile serait 0.511 plutôt que 0.500), car avec une pièce parfaitement équilibrée on avait à peu près autant de chances de faire 511 que 500.

Un des problèmes de la statistique est de pouvoir déterminer ainsi, par des tests *quantitatifs*, si on est fondé ou non à considérer un écart par rapport à la valeur attendue d'un paramètre comme une fluctuation normale due au hasard, ou si au contraire cet écart est extrêmement peu probable, ce qui conduit alors à penser que la valeur attendue avait été incorrectement déterminée. Dans l'analyse ci-dessus du jeu de pile ou face, on a pu calculer la probabilité pour qu'un écart donné soit dû à une fluctuation aléatoire, *parce qu*'on connaissait la loi de probabilité (binômiale) de ces fluctuations.

Le jeu de pile ou face est ici un exemple scolaire, c'est-à-dire un exemple choisi pour son extrême simplicité, qui permet de tout calculer et de montrer ainsi comment la théorie fonctionne. Dans la pratique (du moins celle où justement la statistique est le plus nécessaire), les situations ne sont pas aussi nettes.

Si par exemple on teste un nouveau médicament contre l'hypertension, on compare la diminution de tension à celle qui est attendue en l'absence de médicament. Il est très rare que les effets d'un médicament soient clairs au point d'amener une amélioration qui n'aurait qu'une chance sur un million de se produire par hasard en l'absence de médicament (mais on peut considérer que les *véritables* médicaments sont précisément ceux-là). Le plus souvent on admet que le médicament a un effet si l'amélioration constatée n'avait qu'une chance sur vingt ou sur cent de se produire par hasard. Les tests de nouveaux médicaments sont généralement effectués selon le principe suivant : on administre le médicament à un groupe de malades (disons de cent personnes pour fixer les idées) pendant une durée déterminée. Il n'y a aucun moyen de connaître a priori la probabilité des améliorations spontanées de l'état des malades ; c'est pourquoi on choisit un autre groupe de cent personnes ayant la même maladie (en l'occurrence de l'hypertension), et qui ne recevront pas le médicament (ce second groupe est appelé le groupe témoin). Dans le groupe témoin on peut également observer des améliorations de l'état des malades, qui sont soit spontanées, soit dues aux soins normaux (par exemple régime sans sel). La différence entre la valeur moyenne de la tension dans le groupe témoin et le groupe traité est alors considérée comme une estimation de l'espérance mathématique de la baisse de tension induite par le médicament que l'on veut tester.

Afin d'éliminer tout effet subjectif, au lieu de simplement s'abstenir d'administrer un médicament aux personnes du groupe témoin, on leur distribue un placebo sous une forme exactement identique au médicament (par exemple le médicament est préparé sous forme de comprimé contenant la substance active et un excipient ; le placebo est alors un comprimé d'apparence exactement identique, fabriqué avec le même excipient, mais sans la substance active), et en veillant à ce que les malades ignorent si le comprimé qu'ils reçoivent est actif ou inactif. On dit alors que l'expérience a lieu en (simple) aveugle. Afin d'éviter également un effet subjectif par l'intermédiaire du personnel soignant, on distribue les boîtes de comprimés de façon indifférenciée, le placebo ne se distinguant du médicament que par un code que les expérimentateurs sont seuls à connaître. On dit alors que l'expérience a lieu en double aveugle. Tel est du moins le *principe* de l'expérimentation, car la pratique peut s'en écarter. Ainsi, une enquête de l'épidémiologiste Kenneth Schulz révèle que sur quatre cents médecins, la moitié déclare connaître des collègues qui n'ont pas respecté le protocole et obtenu illicitement les codes qui devaient être tenus secrets (*New Scientist*, vol **2008**, 1996, page 10).

Si on constate davantage d'améliorations dans le groupe traité que dans le groupe témoin, on ne peut pas en déduire que le médicament est efficace, car la différence entre les deux groupes peut être due au seul hasard : soit que par hasard il y a eu moins d'améliorations spontanées que la moyenne dans le groupe témoin, soit que par hasard il y en ait eu plus que la moyenne dans le groupe testé, le médicament lui-même étant totalement inefficace. Toutefois, si on peut montrer que la probabilité pour que la différence constatée entre les deux groupes soit due au hasard est très faible (par exemple 0.01), on peut tenir le raisonnement suivant : il n'y avait qu'une chance sur cent pour que les résultats constatés se produisent par hasard ; il n'y a donc "qu'une chance sur cent" de se tromper en attribuant la différence à une cause ; or on a veillé, en rendant toutes les conditions identiques pour les deux groupes, à ce que la seule cause possible soit le médicament. On tient alors un raisonnement analogue à celui qui avait été tenu pour le jeu de pile ou face. Mais tout le problème est justement de savoir *comment on calcule* cette probabilité : dans le jeu de pile ou face, on connaissait la loi de probabilité ; dans le cas de l'action d'un médicament sur un groupe de personnes, on ne sait rien a priori. Généralement, on postule que la loi de probabilité inconnue a une densité gaussienne, car on admet que, l'action du médicament étant par nature complexe, des causes innombrables et indépendantes les unes des autres se superposent, de sorte que d'après la propriété établie au chapitre **VII**, la loi de probabilité des fluctuations devrait avoir une densité gaussienne, bien qu'on ignore tout de leur cause. Dans certains cas rares, où on devine mieux les causes, on peut toutefois être conduit à postuler des lois non gaussiennes. La règle générale établie au chapitre **VII** est qu'on obtient des lois gaussiennes quand un grand nombre de perturbations aléatoires *indépendantes* se superposent ; mais on peut obtenir des lois non gaussiennes, si les différentes perturbations qui agissent s'influencent mutuellement : un exemple de cela est fourni par les processus en cascade (chapitre **VIII**), où la variable aléatoire Z_n n'est pas la somme de variables indépendantes.

Une fois admis que la densité de la loi de probabilité est gaussienne, il suffit de connaître la moyenne et l'écart-type pour la déterminer entièrement (et connaître la densité de la loi suffit pour calculer approximativement les probabilités). Ces deux paramètres peuvent être devinés à partir des résultats observés, et on peut alors estimer quantitativement la probabilité pour que les valeurs devinées soient fausses, selon un raisonnement du même type que celui qui a été tenu plus haut, à propos du jeu de pile ou face, pour estimer à partir des résultats observés la probabilité que la pièce soit pipée.

Bien entendu ce type de raisonnement statistique n'est pas rigoureux : dans le cas de la pièce de monnaie, il était rigoureux ; mais pour des mécanismes complexes dont une grande part reste inconnue, on fait des hypothèses qui sont peut-être tout à fait gratuites. Rien ne garantit que les probabilités d'amélioration spontanée de l'état d'un malade soient réellement évaluables ; par exemple, que faut-il penser si au lieu de prendre un seul groupe témoin, on en prend beaucoup et qu'on constate une très forte variabilité dans la proportion de guérisons spontanées ? En outre, comme on va le voir, la question de savoir de quoi au juste on mesure ainsi la probabilité n'est elle-même pas très clairement posée.

Il se trouve que les réponses à ces questions ne sont pas simples si on veut les traiter honnêtement. Le fond de l'affaire est de savoir où intervient le hasard. Nous disions en effet “si on peut montrer que la probabilité pour que la différence constatée entre les deux groupes *soit due au hasard* est très faible, alors ...”. Mais comment intervient le hasard dans cette affaire ? Dans les premiers chapitres nous avions déjà montré que le hasard n'agit pas toujours de la même façon (par exemple pour des boules ou pour des particules de Bose). Pour pouvoir dire qu'un problème de probabilité ou de statistique est traité scientifiquement, il ne suffit pas d'appliquer des formules mathématiques, il faut aussi que les conclusions qu'on en tire soient *réellement logiques*, et pour cela il faut savoir comment le hasard a agi. Le problème des tests statistiques de médicaments est brouillé par la complexité des processus biologiques ; c'est pourquoi, avant de l'aborder, nous avons étudié un exemple très simple, celui du jeu de pile ou face. Un autre exemple simple qui joue un rôle essentiel par rapport aux principes de la statistique est le *sondage par échantillons*, que nous allons étudier maintenant. La théorie mathématique du sondage par échantillon sera traitée dans la section suivante (**X.2.**), tout comme la théorie du jeu de pile ou face a été traitée (sous la forme de la marche aléatoire) au chapitre **III**. Toutefois, on peut discuter de son sens avant d'aborder la théorie mathématique. Nous reprendrons ensuite la discussion sur les tests de médicaments, à la lumière des ces deux exemples simples.

Le sondage par échantillons est un procédé très connu du grand public, puisque les media en font un usage fréquent. Pour connaître par exemple les intentions de vote des électeurs français (dont le nombre est de trente millions environ), on mène une enquête auprès d'un échantillon de mille à deux mille d'entre eux, choisis selon des critères que nous étudierons à la section suivante. On part du principe (qui sera justifié théoriquement à la section **X.2.**) que les intentions de vote pour les différents partis se distribuent dans l'échantillon selon des proportions proches des proportions exactes, celles de la population totale. Or il se trouve que le hasard n'intervient pas du tout de la même façon dans le sondage que dans le jeu de pile ou face.

Dans le jeu de pile ou face, le hasard intervient dans le mouvement de la pièce de monnaie, car ce mouvement est chaotique, tout comme celui de la bille de billard ou de roulette analysé au chapitre **I** : en principe la position et les vitesses (de translation et de rotation) initiales déterminent le côté sur lequel la pièce tombera, mais un changement absolument infime dans ces données initiales suffit à changer ce résultat. D'un lancer à l'autre, ces paramètres initiaux sont variables. Dans le sondage, le hasard intervient dans le choix de l'échantillon.

Il n'est pas nécessaire de connaître exactement les paramètres initiaux du mouvement ni de calculer le mouvement exact, pour pouvoir dire que, si la pièce est parfaitement symétrique, elle ne peut tomber plus souvent sur pile que sur face car cela contredirait les symétries spatiales auxquelles le mouvement, tout chaotique qu'il soit, est soumis ; c'est pourquoi on admet a priori que les deux côtés sont équiprobables.

Par contre si par exemple la pièce a une forme symétrique, mais est plus lourde du côté pile (le centre de gravité est alors plus près du côté pile que du côté face), ou bien si le métal est homogène, mais que la tranche est légèrement biseautée (de telle sorte que le côté face ait un diamètre légèrement plus grand que le côté pile), il se produira le phénomène suivant : lorsque la pièce sera encore sur la tranche et roulera sur la table, le centre de gravité ne sera pas à la verticale de l'aire de sustentation et la pesanteur fera pencher la pièce du côté où se trouve le centre de gravité, favorisant ainsi le côté face. Ces considérations sur le mouvement montrent donc que le hasard agit *à travers le mouvement de la pièce*. Si la pièce est bien équilibrée, il y a une raison *a priori* pour que la probabilité d'avoir face soit égale à $\frac{1}{2}$, même si on ne lance la pièce qu'une seule fois. Cette probabilité est déterminée par les lois de la Physique avant même qu'on ait procédé au premier lancer. Ce sera vrai aussi si la pièce n'est pas équilibrée et que le côté face est plus probable : il y aura une probabilité inconnue $x > \frac{1}{2}$ d'avoir face et une probabilité $1 - x < \frac{1}{2}$ d'avoir pile, mais bien qu'inconnue (et impossible ou du moins très difficile à calculer à partir des équations du mouvement et de la forme de la pièce) cette probabilité x sera néanmoins déterminée a priori par les lois de la Physique.

Dans le sondage les choses sont très différentes. Il y a (à un moment donné de la campagne électorale) une proportion x_1 d'électeurs qui se prononceraient pour le parti du Progrès si on leur posait la question, une proportion x_2 qui se prononceraient pour le parti de la Démocratie, une proportion x_3 qui se prononceraient pour le parti de la Liberté, etc. Cela veut dire que si on posait la question à chacun des N électeurs, il y en aurait p_1 qui se prononceraient pour le parti du Progrès, p_2 qui se prononceraient pour le parti de la Démocratie, p_3 qui se prononceraient pour le parti de la Liberté, et les nombres $x_1,\ x_2,\ x_3$ sont simplement les rapports $p_1/N\ ,\ p_2/N\ ,\ p_3/N$.

Ces rapports *ne sont pas* la probabilité a priori pour qu'un électeur particulier vote pour tel ou tel parti. On peut d'ailleurs se demander ici ce que seraient de telles probabilités a priori ; si on croit au réductionnisme mécaniste absolu, par exemple, on dira que la réponse de l'électeur Jean Dupond à la question du sondeur est déterminée par les lois de la Physique, étant donné que tous les mécanismes neuronaux actionnés dans le cerveau de Jean Dupond, et aboutissant à l'état dans lequel Jean Dupond éprouve la sensation que le parti du Progrès est le meilleur, sont des mouvements d'atomes et de molécules qui obéissent aux lois de la Physique. De même que pour la pièce de monnaie, la roulette, ou l'agitation thermique, on peut alors considérer ces mouvements moléculaires de neurotransmetteurs comme chaotiques, qu'ils créent donc du hasard et déterminent une probabilité a priori.

Il importe peu qu'on croie au déterminisme mécaniste absolu ou qu'on n'y croie pas ; le problème débattu ici est que, s'il y a un sens à parler de probabilité a priori, ce sera la probabilité, déterminée par des causes particulières internes à Jean Dupond, pour que celui-ci déclare au sondeur son intention de voter pour le parti du Progrès. Il n'y a alors aucune raison pour que la probabilité a priori de Pierre Martin, ou de Paul Duval soit la même ; il n'y a aucune raison non plus pour qu'elle soit égale à x_1, ni même pour que la moyenne de ces probabilités a priori sur l'ensemble des électeurs soit égale à x_1. Au lieu de prendre l'exemple des intentions de vote, pour lequel le raisonnement mécaniste est confus et très discutable, on pourrait prendre celui de la mucoviscidose (voir sections **III. 6** et **IV. 4**). Si on prend au hasard un couple marié dans la population française, il y a une chance sur deux mille qu'il procrée un enfant ayant cette maladie ; mais la probabilité pour un couple donné est soit 0, soit 0.25 (ce pourrait même être 1 si les parents étaient tous deux atteints, c'est-à-dire tous deux homozygotes, mais ce cas ne se produit jamais). En prenant un couple au hasard dans la population, on fait intervenir le hasard dans le choix de ce couple ; en considérant un couple particulier donné, le hasard intervient au moment de la fécondation, dans la combinaison chromosomique. Le hasard ne joue pas du tout le même rôle dans les deux cas.

Dans le jeu de pile ou face, la notion de probabilité a priori a un sens simple et clair (elle se comprend d'après la Physique et le chaos, comme cela a été expliqué au chapitre **I**), et cette probabilité a priori est celle que nous retenons pour traiter le problème par le Calcul des probabilités. C'est ce que nous avions fait pour les problèmes de boules qu'on jette dans des boîtes, de particules de Bose qu'on place dans des états quantiques, la bille de roulette, etc. Dans le cas des sondages sur les intentions de vote, il y a aussi des probabilités a priori, mais qui n'ont rien à voir avec la probabilité a priori pour que tel ou tel électeur particulier donne telle ou telle réponse : ce sont les probabilités pour que, un électeur étant choisi au hasard, ce soit un partisan de l'un ou l'autre des candidats. Quant à la probabilité a priori pour qu'un individu particulier vote comme ceci ou comme cela, on ne voit guère comment lui donner un sens concret. La *signification* de la probabilité est donnée par la connaissance du niveau où intervient le hasard.

Il en va de même pour l'action d'un médicament. Un médicament agit par voie chimique : il y a donc réellement un mécanisme moléculaire par lequel le médicament fait baisser la tension artérielle (point n'est besoin de se réclamer du déterminisme mécaniste absolu pour cela). L'effet du médicament sur l'organisme est donc causal, tout comme le mouvement de la pièce de monnaie qui induisait la probabilité a priori x d'avoir face. Il n'est pas dépourvu de sens de parler de la probabilité a priori pour que le médicament, administré à Jean Dupond, induise dans les cinq heures qui suivent une baisse de tension nettement supérieure aux fluctuations *spontanées* de la tension. La calculer ou même la mesurer empiriquement est une autre histoire, mais il est également rationnel d'admettre que les fluctuations spontanées de la tension suivent une loi de probabilité approximativement gaussienne. On peut tenir pour fantaisiste l'explication mécaniste des intentions de vote, mais le mécanisme moléculaire de l'action d'un médicament est un fait soigneusement établi.

Les deux notions de probabilité se distinguent par le lieu où intervient le hasard :

— lorsqu'une personne est choisie au hasard dans un groupe et qu'on demande la probabilité pour que le médicament ait agi (ou pour que la personne choisie préfère le candidat du Progrès), le hasard intervient dans le choix, par le sondeur, de cette personne ;

— lorsqu'il n'y a pas de groupe, mais une seule personne présente, et qu'on demande la probabilité pour que, si on administre le médicament *à cette personne*, celui-ci agisse sur elle, le hasard intervient dans le chaos moléculaire du métabolisme.

Appelons la première *probabilité statistique* et la seconde *probabilité biochimique*. Il n'y a évidemment aucune raison pour que ces probabilités coïncident.

L'effet biochimique du médicament sur l'organisme est différent chez chaque personne : le médicament peut provoquer avec certitude une chute de tension chez Jean Dupond, mais rien du tout (si ce n'est des effets secondaires) chez Pierre Martin. Cette différence est due à une cause, par exemple des métabolismes différents : il se peut que l'hypertension de Jean Dupond soit due à une mauvaise élimination du sel (insuffisance rénale), tandis que celle de Pierre Martin serait due à une autre cause (par exemple des états de stress peuvent induire la production d'hormones provoquant une vasoconstriction, c'est-à-dire un resserrement des artères) ; si le médicament accélère l'élimination du sel, il agit sur l'hypertension de Jean Dupond, mais pas sur celle de Pierre Martin. Dans les deux cas la probabilité biochimique a priori est différente : celle de Jean Dupond est proche de 1, celle de Pierre Martin proche de 0. Le mécanisme biochimique détermine une probabilité a priori pour que le médicament entraîne une chute de tension, tout comme le mouvement chaotique de la pièce de monnaie déterminait une probabilité a priori d'avoir pile ou face.

Le test clinique du médicament par contre consiste à faire un sondage, en prélevant sur l'ensemble de la population des personnes hypertendues un groupe de cinquante, cinq cents, ou cinq mille personnes, et à noter combien réagissent positivement au traitement. Si le médicament agit sur 80% des personnes du groupe, on pourra dire qu'environ 80% de la population générale des personnes hypertendues répond au traitement. Mais cela ne signifie pas que chaque personne a une probabilité biochimique 0.8 d'être guérie, cela signifierait plutôt que 80% des personnes hypertendues élimine mal le sel, l'hypertension ayant alors chez les 20% restants une cause différente. La signification de cette probabilité statistique est très prosaïque : si vous allez consulter pour hypertension, alors, pour le médecin qui prescrira le médicament en ne connaissant que le symptôme (l'hypertension), mais non la cause (mauvaise élimination du sel), il y aura 80% de chances qu'il s'applique à vous ; par contre pour vous, si votre hypertension a une autre cause, il y aura 0% de chances qu'il agisse.

Si on veut *mesurer* la probabilité biochimique, on peut procéder comme suit (de tels tests de médicament sont réellement pratiqués en clinique) : une séquence comportant des placebos et des substances actives (sous une présentation identique) est prescrite ; la succession des médicaments (M) et placebos (P) n'est connue que de l'expérimentateur ; on choisit délibérément des séquences très irrégulières, de type aléatoire (randomisées), par exemple $PPMPMMPMPPPMPMPPMPMMPPMP$.

Le test se présente sous la forme de comprimés identiques, mais numérotés afin que l'ordre puisse être respecté ; la personne testée doit prendre un comprimé chaque jour à la même heure. On se rapproche ainsi de l'idéal d'une expérience reproductible, quoique la variabilité inhérente à tout organisme vivant ne peut être éliminée. Un tel test effectué sur une seule et même personne mesure alors (même si dans le principe la précision est très médiocre) la probabilité a priori pour une personne donnée. Si par exemple ce test est effectué sur Jean Dupond, qui élimine mal le sel, on s'attend en théorie à ce qu'un médicament favorisant l'élimination du sel agisse efficacement, tandis qu'un médicament qui s'attaque aux causes hormonales reste sans action. Pour Pierre Martin, dont l'hypertension a des causes hormonales, ce devrait être l'inverse.

En réalité, ces expériences montrent, lorsqu'elles sont effectuées, que l'action des médicaments hypotenseurs n'est jamais aussi nettement spécifique. La raison à cela est ce qu'on appelle la *multifactorialité* (ici, de l'hypertension). Il faut comprendre que l'organisme vivant est un tout inséparable, dans lequel la séparation rigoureuse des différentes causes, qui est le principe des sciences expérimentales, est impossible. Par exemple, on peut avoir diagnostiqué une mauvaise élimination du sel chez Jean Dupond ; mais cela suffit-il à affirmer qu'elle est la cause (ou la cause unique) de l'hypertension ? Celle-ci pourrait être due à la conjuguaison de deux, trois facteurs, ou plus, l'un connu (la mauvaise élimination du sel) et les autres inconnus ; en outre la mauvaise élimination du sel, ou bien l'hypertension qui en résulte, peut créer un état de stress qui a son tour agira comme seconde cause.

C'est ici qu'on rencontre les limites de la méthode statistique : pour que des tests statistiques puissent apporter une véritable information, il faut pouvoir séparer les différentes causes qui agissent (lorsque cette séparation est possible, les techniques statistiques qui permettent de les séparer constituent la *régression* (voir chapitre **XII**) ; mais pour que celle-ci puisse avoir une signification rigoureuse, il faut que les causes soient toutes connues, et que leurs effets puissent être séparés par un choix convenable des protocoles expérimentaux). Une simple corrélation statistique, sans isoler une cause, n'a pas de signification utilisable en pratique : par exemple, on peut avoir établi qu'il y a trois fois plus de cancer de l'oesophage dans la ville X que dans la ville Y ; mais cela ne dit rien sur les causes, et en particulier on ne peut pas en déduire que ces causes agiraient sur une personne quittant Y pour venir s'installer à X (par exemple si la cause était l'alcoolisme endémique des habitants de X, elle ne pourrait agir sur un résident qui ne s'adonne pas à ce vice).

La notion de probabilité biochimique est donc, comme on voit, peu opératoire. Elle le deviendrait cependant s'il était possible de découvrir dans le métabolisme des invariances conduisant à des épreuves équiprobables comme c'est le cas en génétique. Là ce n'est pas le cas, et c'est pourquoi cette notion de probabilité biochimique n'est pas utilisée ; elle est simplement passée sous silence ou ignorée. Mais justement à cause de cette omission, l'idée qu'il y aurait une différence essentielle entre la probabilité statistique et la probabilité biochimique n'est jamais expressément discutée, ce qui conduit à la confusion courante sur la signification d'un résultat statistique. C'est pour empêcher cette confusion que nous avons introduit ici la notion techniquement inopérante de probabilité biochimique. Elle permet de se rendre compte que la simple mesure statistique de probabilités empiriques ne suffit pas pour comprendre un phénomène, **il faut aussi savoir où intervient le hasard**. Par exemple, une enquête peut établir que un Français sur deux mille est frappé par la mucoviscidose ; mais la *compréhension* du phénomène résulte de la découverte du niveau où est intervenu le hasard (la combinaison des chromosomes). Le simple résultat de l'enquête n'apporte aucune compréhension. Il en est de même pour comprendre l'action d'un médicament : on la comprendrait si on découvrait où intervient le hasard dont le test n'a montré que l'effet statistique.

Bien entendu, pour que le sondage soit correct, il faut que l'échantillon soit choisi selon un protocole *stochastiquement indépendant* de la variable qu'on cherche à mesurer (l'effet statistique du médicament dans la population ou le nombre d'électeurs de tel ou tel parti). Un tel protocole est appelé une *randomisation*.

Il n'existe en général pas de critère théorique de randomisation, c'est-à-dire qu'il n'existe pas, pour choisir n personnes sur une population totale de N, de critère dont on puisse démontrer théoriquement a priori qu'il est stochastiquement indépendant du paramètre à mesurer ; ces critères sont soit basés sur des «évidences», soit établis empiriquement par essais et erreurs. Ainsi on peut adopter le protocole suivant : la population totale de N personnes est enregistrée dans un fichier séquentiel, et une fonction random choisit n fois un nombre compris entre 1 et N ; on prend alors l'échantillon formé par les n personnes dont les numéros d'ordre dans le fichier sont les nombres fournis par la fonction random ; dans ce cas, on tient pour «évident» qu'un tel protocole est stochastiquement indépendant de la variable étudiée (la fonction random «ignore» la couleur des yeux des personnes). C'est ce type de protocole qui est utilisé pour des enquêtes médicales à grande échelle (voir encadré ci-contre).

Toutefois pour un sondage sur une population de 30 000 000 de personnes, la gestion d'un fichier qui les contient toutes avec leurs noms et adresses pose des problèmes techniques considérables (coût énorme pour la création et l'entretien du fichier, quantité de mémoire et temps d'accès), et c'est pourquoi des méthodes plus souples sont employées par les instituts de sondage, mais bien sûr ces méthodes doivent être validées empiriquement, c'est-à-dire que leur indépendance stochastique (par rapport aux intentions de vote par exemple) doit être vérifiée et maintenue par des corrections constantes (consulter des ouvrages spécialisés pour en savoir plus).

Ce qui vient d'être dit laisse de côté le rôle joué par le groupe témoin. Celui-ci sert, comme son nom l'indique, à écarter les effets de clinique ;

streptokinase + aspirine et héparine	tPA + aspirine et héparine	APSAC + aspirine et héparine
streptokinase + aspirine seule	tPA + aspirine seule	APSAC + aspirine seule

une enquête médicale à grande échelle.

Le tableau ci-dessus est extrait d'un article publié dans *The Lancet* (édition française) de Septembre 1992, pages 7 – 25 . Il s'agit d'une étude de la survie après infarctus du myocarde, portant sur 41 299 patients, répartis dans 914 hôpitaux situés dans 20 pays. L'ensemble des 41 299 patients a été réuni sur une base volontaire, mais ensuite leur répartition entre les six groupes indiqués dans le tableau a été effectuée par randomisation. Cela signifie que les coordonnées des 41 299 malades ont été enregistrées séquentiellement dans un fichier, puis que les six groupes ont été choisis par une fonction random opérant sur le fichier. Le but de l'étude était d'étudier l'effet de l'héparine (un médicament fibrinolytique) sur le nombre de récidives d'infarctus ; ce médicament n'étant pas destiné à être administré seul, mais toujours accompagné d'aspirine et d'une troisième substance qui peut être selon les cas, la streptokinase, l'APSAC, ou le tPA, on voit que tout repose sur la comparaison d'un groupe traité à l'héparine (en haut) et d'un groupe témoin (en bas) ; afin que la comparaison entre un groupe traité et le groupe témoin correspondant soit significative, il était essentiel que ces deux groupes ne diffèrent que par l'administration d'héparine : c'est pourquoi les deux autres substances actives sont administrées à l'identique dans les deux groupes. La randomisation a pour but de garantir l'indépendance stochastique entre le choix de l'échantillon et les caractéristiques des soins particulières à chaque hôpital ; en effet, si par exemple dans le groupe traité avec APSAC les hôpitaux européens étaient favorisés, tandis que dans le groupe traité avec tPA les hôpitaux américains étaient favorisés, on ne pourrait plus écarter l'influence de facteurs inconnus caractérisant les soins hospitaliers de part et d'autre de l'Atlantique et cela fausserait l'étude.

Toutefois une telle étude a forcément les limites suivantes :

a) les 914 hôpitaux n'ont pas été choisis par randomisation ; s'ils l'avaient été, d'ailleurs, l'intérêt médical de l'enquête n'en aurait pas été augmenté ;

b) si on considère les proportions de récidive d'infarctus (ou d'hémorragie cérébrale consécutive au traitement) dans les groupes traités comme la mesure par sondage d'une probabilité, il s'agit de la probabilité pour qu'une personne venant de subir une crise, prise au hasard dans la population des 20 pays de l'enquête, et soignée dans l'un des 914 hôpitaux, soit victime d'une récidive (ou d'une hémorragie cérébrale consécutive au traitement). La probabilité du même événement dans un hôpital misérable d'une région deshéritée du tiers-monde peut être très différente. Autrement dit, il s'agit de la probabilité *statistique* pour une personne prise au hasard dans la population de l'enquête, et non de la probabilité *biochimique* a priori d'action du médicament (résultant des mécanismes moléculaires, au demeurant inconnus) pour une personne particulière et bien déterminée. Une telle étude ne peut donc fournir aucune information sur les mécanismes d'action de l'héparine. Son ambition est surtout de décider si statistiquement le nouveau médicament vaut la peine d'être commercialisé ; c'est pourquoi l'enquête est restreinte aux hopitaux des régions solvables.

en effet les patients du groupe traité ne subissent pas seulement les effets du médicament testé, mais aussi ceux des soins généraux consécutifs à l'hospitalisation, ne serait-ce que le régime alimentaire. Si on veut utiliser le sondage pour connaître les effets sur une population *non* hospitalisée, il faut pouvoir soustraire les effets dûs à l'hospitalisation : la constatation que dans le groupe témoin il y a dix améliorations (alors que l'écart-type n'est que $3,46$) indique une assez forte probabilité pour qu'un tel effet de clinique se soit produit, et il faudrait donc corriger en baisse la proportion de $3/5$. Pour effectuer de telles soustractions, il existe des méthodes statistiques mathématiquement fondées, déjà évoquées plus haut, appelées *régression* (voir chapitre **XII**).

X.2. La théorie des échantillons de Bernoulli.

Dans l'exemple du test de médicament, il a fallu distinguer entre les probabilités biochimiques d'une action du médicament sur le métabolisme de chaque personne particulière, et les probabilités statistiques liées au choix d'un échantillon de population. Les deux sortes de probabilités proviennent de phénomènes aléatoires, mais les premières proviennent des échanges aléatoires de molécules dans le métabolisme, tandis que les secondes proviennent du choix aléatoire d'un groupe de population.

Pour que les choses soient plus claires on peut prendre un exemple où le hasard n'intervient que dans le choix de l'échantillon statistique et n'interfère pas du tout avec les mécanismes biologiques. Supposons qu'on veuille déterminer la proportion de Français qui ont les yeux bleus. Ce caractère ne varie pratiquement pas au cours du temps (il faut plusieurs générations pour que se produise une variation notable), de sorte que les mécanismes biologiques sous-jacents peuvent être complètement négligés. Faire un recensement de toute la population pour connaître le nombre de personnes dont les yeux sont bleus permettrait certes de connaître ce nombre à l'unité près, mais aurait un coût disproportionné par rapport au bénéfice attendu de l'information : il est en effet bien rare qu'une telle précision soit utile. La méthode qu'on utilise alors pour déterminer cette proportion est celle du sondage : on choisit *au hasard* un échantillon de mille ou deux mille personnes dans l'ensemble de la population, et on compte la proportion de personnes aux yeux bleus dans cet échantillon : on admet communément que plus l'échantillon est grand, plus cette proportion sera proche de la proportion exacte, ce qui est à nouveau l'expression de la *loi des grands nombres.* Il s'agit là d'une évidence commune tant qu'on se contente d'appréciations qualitatives. Mais dans cette section nous allons soumettre cette évidence à l'analyse *quantitative.* Au lieu de se contenter de dire "*plus l'échantillon est grand, plus cette proportion sera proche de* *la proportion exacte*", on peut établir une relation mathématique entre la taille de l'échantillon et la précision du sondage. Cette théorie mathématique du sondage a été créée par Jakob (Jacques) Bernoulli et publiée dans un livre posthume en 1713, huit ans après sa mort : l'*Ars conjectandi*, c'est-à-dire "l'art de conjecturer"). Le terme même de "loi des grands nombres" n'apparaîtra que bien plus tard. Voici comment Jacques Bernoulli lui-même pose le problème :

> Je suppose que dans une urne, à ton insu [le tutoiement du lecteur est un style d'exposition latin] soient placées trois mille pierres blanches et deux mille pierres noires ; je suppose que pour connaître leurs nombres par expérience, tu tires une pierre après l'autre (en replaçant cependant à chaque fois la pierre que tu as tirée avant de choisir la suivante, pour que le nombre des pierres ne diminue pas dans l'urne) ; tu observes combien de fois sort une pierre blanche et combien de fois une noire. On demande si tu peux le faire tant de fois qu'il devienne dix fois, cent fois, mille fois, etc. plus probable (c'est-à-dire qu'il devienne moralement certain) que le nombre de fois où tu choisis une pierre blanche et le nombre de fois où tu choisis une pierre noire soient dans ce même rapport sesquialtère [1, 5 fois] où se complaisent à être entre eux les nombres de pierres ou de cas, plutôt que dans tout autre rapport différent de celui-ci. (...)
>
> Mais pour que cela ne soit pas compris autrement qu'il ne convient, il faut bien noter ce qui suit ; je voudrais que le rapport entre les nombres que nous entreprenons de déterminer expérimentalement, ne fût pas pris de façon nette et sans partage (car ainsi c'est tout le contraire qui arriverait et il deviendrait d'autant moins probable de découvrir le vrai rapport qu'on ferait de plus nombreuses observations), mais je voudrais que le rapport fût admis avec une certaine latitude, c'est-à-dire compris entre une paire de limites, pouvant être prises aussi rapprochées qu'on voudra. Assurément, si dans l'exemple des pierres proposé plus haut nous prenons les deux rapports $\frac{301}{200}$ et $\frac{299}{200}$, ou $\frac{3001}{2000}$ et $\frac{2999}{2000}$, etc. dont le sesquialtère est très près et du plus grand et du plus petit, on montrera que l'on peut arriver à ce que le rapport trouvé grâce à des expériences recommencées de nombreuses fois tombe entre ces limites du rapport sesquialtère plus probablement, de toute probabilité donnée, qu'en dehors.
>
> (Trad. Norbert Meusnier, document IREM de Besançon, 1989)

La *nécessité logique* de la loi des grands nombres signifie que celle-ci peut être *démontrée*, et non simplement connue comme une loi de la nature dont Dieu seul connaît la justification. Il ne faut cependant pas oublier que démontrer consiste à déduire une vérité à partir d'une autre, et qu'on ne démontre rien à partir de rien. Jacques Bernoulli en est parfaitement conscient :

> On en est ainsi venu à ce point que pour former selon les règles des conjectures sur n'importe quelle chose il est seulement requis d'une part que les nombres de cas soient soigneusement déterminés, et d'autre part que soit défini combien les uns peuvent arriver plus facilement que les autres. (...) Mais cela peut se voir à peine dans quelques très rares cas et ne se produit presque pas en dehors des jeux de hasard (...) Ainsi les cas "d'égale facilité" sont-ils connus pour les dés, pour une urne contenant des bulletins noirs et blancs, mais que dire du nombre des maladies qui peuvent engendrer la mort, des changements qui peuvent affecter le climat, (...) Mais à la vérité ici s'offre à nous un autre

> chemin pour obtenir ce que nous cherchons. Ce qu'il n'est pas donné d'obtenir a priori l'est du moins a posteriori, c'est-à-dire qu'il sera possible de l'extraire en observant l'issue de nombreux exemples semblables ; car on doit présumer que, par la suite, chaque fait peut arriver et ne pas arriver dans le même nombre de cas qu'il avait été constaté auparavant, dans un état de choses semblables, qu'il arrivait ou n'arrivait pas.

Autrement dit, on démontre la loi des grands nombres par la logique, *à partir de l'hypothèse d'une invariance*. Lorsque les séries d'événements ne se reproduisent pas à l'identique, il n'est plus question de démontrer quoi que ce soit ; en revanche, on peut mesurer par la statistique.

C'est à l'exposé de cette "théorie des échantillons" que cette section est consacrée ; très précisément, il s'agit d'établir la relation mathématique entre la taille de l'échantillon et la précision ou la certitude du résultat.

Soit donc N le nombre d'individus dans la population totale, et n la taille de l'échantillon. On appellera p le nombre d'individus ayant les yeux bleus dans la population totale. Dans cet exemple le hasard n'intervient pas au niveau des mécanismes moléculaires, mais dans le choix (fait par l'expérimentateur) de l'échantillon ; on considère donc que par principe aucun échantillon n'a été privilégié. L'espace Ω est alors l'ensemble de tous les échantillons possibles de taille n : son cardinal est $\#\Omega = \binom{N}{n}$ (voir chapitre **II**, section **3**)

Pour un entier k compris entre 0 et n, le nombre d'échantillons contenant exactement k personnes aux yeux bleus est $\binom{p}{k} \cdot \binom{N-p}{n-k}$: en effet on peut considérer qu'un tel échantillon s'obtient en prenant d'abord un échantillon de k personnes parmi les p qui ont les yeux bleus (ce qui fait $\binom{p}{k}$ possibilités d'après **II.3.**), puis en prenant $n-k$ personnes parmi les $N-p$ qui n'ont pas les yeux bleus (ce qui fait $\binom{N-p}{n-k}$ possibilités pour *chacun* des choix précédents). Par conséquent la probabilité d'obtenir un échantillon contenant exactement k personnes aux yeux bleus est

$$p_k \;=\; \frac{\binom{p}{k} \cdot \binom{N-p}{n-k}}{\binom{N}{n}} \;=\; \frac{\dfrac{p!}{k!\,(p-k)!} \cdot \dfrac{(N-p)!}{(n-k)!\,(N-p-n+k)!}}{\dfrac{N!}{n!\,(N-n)!}} \qquad (X.1.)$$

En ne faisant rien d'autre que regrouper les factorielles dans un autre ordre, on voit que cela est aussi égal à

$$p_k \;=\; \frac{\dfrac{n!}{k!\,(n-k)!} \cdot \dfrac{(N-n)!}{(p-k)!\,(N-p-n+k)!}}{\dfrac{N!}{p!\,(N-p)!}} \;=\; \frac{\binom{n}{k} \cdot \binom{N-n}{p-k}}{\binom{N}{p}}$$

C'est la loi *hypergéométrique* (cf. **IX.5**). Mais si n est beaucoup plus petit que N, on peut faire pour ces factorielles les approximations suivantes :

$$\begin{aligned}
(N-n)! &= \frac{N!}{N(N-1)(N-2)\cdots(N-n+1)} \simeq \frac{N!}{N^n} \\
(p-k)! &= \frac{p!}{p(p-1)(p-2)\cdots(p-k+1)} \simeq \frac{p!}{p^k} \\
(N-p-n+k)! &= \frac{(N-p)!}{(N-p)(N-p-1)\cdots(N-p-n+k+1)} \\
&\simeq \frac{(N-p)!}{(N-p)^{n-k}}
\end{aligned}$$

(ces approximations supposent évidemment que $n \ll N$, $k \ll p$, et $n-k \ll N-p$). Par conséquent, si on pose $\alpha = p/N$ et $\beta = (N-p)/N = 1-\alpha$, on aura :

$$p_k = \binom{n}{k} \cdot \frac{\binom{N-n}{p-k}}{\binom{N}{p}} \simeq \binom{n}{k} \alpha^k \beta^{n-k} \qquad (X.2.)$$

Cette loi de probabilité *asymptotique* (valable pour N grand) est appelée *loi de Bernoulli*. Si on choisit au hasard une personne parmi les N de la population totale, α est la probabilité (exacte) pour que cette personne ait les yeux bleus (et β la probabilité pour qu'elle ne les ait pas bleus) : cela est évident a priori, puisque dans la population totale il y a exactement p personnes aux yeux bleus sur N en tout. La loi de Bernoulli dit alors que si on choisit au hasard un échantillon de plusieurs (n) personnes, la probabilité d'en avoir k aux yeux bleus dans l'échantillon sera environ $\binom{n}{k} \alpha^k \beta^{n-k}$.

L'idée (conforme au sens commun) que dans un échantillon de 1000 ou 2000 personnes on doit retrouver à peu près la même proportion de personnes aux yeux bleus que dans la population totale, doit alors se traduire par le fait que la probabilité que $k \simeq n\alpha$ doit être proche de 1, tandis que la probabilité pour que k soit nettement différent de $n\alpha$ doit être faible. Il n'est en effet pas exclu que l'on puisse tomber sur un échantillon sans aucune personne aux yeux bleus, de même qu'en jouant à pile ou face mille fois il n'est pas absolument exclu de n'obtenir aucune fois face ; cela est simplement très peu probable. Nous allons donc étudier de plus près cette loi de Bernoulli.

Commençons par chercher le maximum de p_k. Il est clair que lorsque p_k est maximum, on doit avoir à la fois $p_{k+1} \leq p_k$ et $p_{k-1} \leq p_k$. On trouvera donc les maxima de p_k en cherchant les valeurs de k pour lesquelles ces deux inégalités sont vérifiées à la fois. Or

$$\begin{aligned}\binom{n}{k+1} &= \binom{n}{k} \cdot \frac{n-k}{k+1} \\ \binom{n}{k-1} &= \binom{n}{k} \cdot \frac{k}{n-k+1}\end{aligned}$$

d'où on déduit

$$\begin{aligned}p_{k+1} &= p_k \cdot \frac{n-k}{k+1} \cdot \frac{\alpha}{\beta} \\ p_{k-1} &= p_k \cdot \frac{k}{n-k+1} \cdot \frac{\beta}{\alpha}\end{aligned}$$

Les deux inégalités caractérisant le maximum seront donc

$$\begin{aligned}\frac{n-k}{k+1} \cdot \frac{\alpha}{\beta} \leq 1 \qquad &\Longleftrightarrow \qquad k \geq n\alpha - \beta \\ \frac{k}{n-k+1} \cdot \frac{\beta}{\alpha} \leq 1 \qquad &\Longleftrightarrow \qquad k \leq n\alpha + \alpha\end{aligned}$$

On voit que k doit être compris entre $n\alpha - \beta$ et $n\alpha + \alpha$. Mais la différence entre ces deux nombres est $n\alpha + \alpha - n\alpha + \beta = \alpha + \beta = 1$, or entre deux nombres réels qui diffèrent de 1 il y a exactement un entier (à la rigueur deux dans le cas exceptionnel où $n\alpha + \alpha$ est lui-même entier) ; soit k_0 cet entier (il est donc très proche de $n\alpha$). Ainsi p_k est maximum pour $k = k_0 \simeq n\alpha$. Cela confirme le sens commun, qui voulait que dans un échantillon pris au hasard on retrouve à peu près la même proportion de personnes aux yeux bleus que dans l'ensemble de la population. Ici nous voyons par le calcul que la probabilité de trouver la même proportion est la probabilité maximum. Reste à voir comment elle décroît autour du maximum.

Pour cela on va faire comme déjà à plusieurs reprises dans ce cours : posons $k = k_0 + j$; j sera ainsi l'écart (positif ou négatif) de k par rapport à k_0. On peut écrire pour $j > 0$:

$$\begin{aligned}\binom{n}{k_0+j} &= \binom{n}{k_0} \cdot \frac{(n-k_0)^j}{k_0^j} \cdot \frac{(1-\frac{1}{n-k_0})(1-\frac{2}{n-k_0})\cdots(1-\frac{j-1}{n-k_0})}{(1+\frac{1}{k_0})(1+\frac{2}{k_0})\cdots(1+\frac{j}{k_0})} \\ &\simeq \binom{n}{k_0} \cdot \frac{\beta^j}{\alpha^j} \cdot \frac{\exp\{-j^2/2\,(n-k_0)\}}{\exp\{j^2/2\ k_0\}}\end{aligned}$$

et de la même façon

$$\binom{n}{k_0 - j} = \binom{n}{k_0} \cdot \frac{k_0^j}{(n-k_0)^j} \cdot \frac{(1-\frac{1}{k_0})(1-\frac{2}{k_0})\cdots(1-\frac{j-1}{k_0})}{(1+\frac{1}{n-k_0})(1+\frac{2}{n-k_0})\cdots(1+\frac{j}{n-k_0})}$$

$$\simeq \binom{n}{k_0} \cdot \frac{\alpha^j}{\beta^j} \cdot \frac{\exp\{-j^2/2\,(n-k_0)\}}{\exp\{j^2/2\ k_0\}}$$

Il est facile de voir que

$$\frac{\exp\{-j^2/2\,k_0\}}{\exp\{j^2/2\,(n-k_0)\}} = \exp\Big\{-\frac{j^2}{2}\Big[\frac{1}{k_0} + \frac{1}{n-k_0}\Big]\Big\} \simeq \exp\Big\{-\frac{j^2}{2n\alpha\beta}\Big\}$$

ce qui montre que, aussi bien pour $j > 0$ que pour $j < 0$ on aura

$$p_{k_0+j} \simeq p_{k_0} \cdot \exp\Big\{-\frac{j^2}{2n\alpha\beta}\Big\} \qquad (X.3.)$$

On constate une fois de plus que la probabilité diminue autour du maximum selon une loi gaussienne. Le calcul détaillé que nous venons d'effectuer montre surtout que l'écart-type de la gaussienne est $\sqrt{n\alpha\beta}$. En regardant une table de la loi normale, on voit que la probabilité pour que j soit supérieur à trois fois l'écart-type est égale à 0.0026. Donc la probabilité pour que j soit supérieur à $3\sqrt{n\alpha\beta}$ est 0.0026. Par exemple si $\alpha = 1/3$, $\beta = 2/3$, et $n = 1800$, on a $k_0 = 600$ et la probabilité pour que $|j| > 60$ est 0.0026. Autrement dit, la probabilité pour que la proportion de personnes aux yeux bleus *dans l'échantillon* soit supérieure à $\frac{660}{1800} = 36.7\%$ ou inférieure à $\frac{540}{1800} = 30.0\%$, est 0.0026. Cela signifie qu'on a seulement une chance sur 385 ($0.0026 \simeq \frac{1}{385}$) de se tromper en affirmant qu'à trois points de pourcentage près, la proportion de personnes aux yeux bleus dans l'échantillon est la même que dans la population totale. On peut alors inverser le raisonnement : si α est inconnu et qu'on trouve 31% dans un échantillon pris au hasard, on pourra dire qu'il n'y avait que une chance sur 385 pour que la proportion dans l'échantillon s'écarte de la proportion réelle α de plus de trois points de pourcentage, et que donc α est, avec probabilité $1 - \frac{1}{385}$, compris entre 28% et 34%.

L'*inversion du raisonnement* est une évidence empirique en ce sens que la vie de tous les jours se fonde sur elle : cette évidence empirique est à l'origine de ce qu'on nomme l'*habitude*. Pourtant ce n'est pas une évidence logique, cela ne se voit pas *immédiatement* à partir des principes. En outre l'évidence empirique n'est pas quantitative. C'est pourquoi nous allons soumettre cette *inversion du raisonnement* à l'analyse.

Le problème du sondage se posait ainsi : un ensemble de N éléments était donné, et on y choisissait au hasard une partie de n éléments ; l'espace des épreuves était donc l'ensemble de ces parties, dont le cardinal était $\binom{N}{n}$. Dans le problème inverse, on ne peut plus dire que l'ensemble de N éléments est donné puisqu'on veut considérer toutes les possibilités pour les nombres p et q ; ces possibilités sont au nombre de $N+1$:

$$\begin{cases} p=0 \\ q=N \end{cases} \quad \begin{cases} p=1 \\ q=N-1 \end{cases} \quad \begin{cases} p=2 \\ q=N-2 \end{cases} \quad \cdots \quad \begin{cases} p=N \\ q=0 \end{cases}$$

Il s'agit donc de trouver la loi de probabilité de p, sachant que le résultat du sondage est k (au lieu de trouver la loi de k pour un p donné). Comme toujours depuis le début, il faut chercher ce qui est équiprobable. Il est clair que si $k=k_0$ est donné, les différentes possibilités pour p ne sont pas équiprobables (il sera plus probable de trouver un p tel que $p/N \simeq k_0/n$ qu'un p tel que p/N soit très différent de k_0/n, c'est justement ce que nous dit l'évidence empirique). Il faut donc imaginer l'expérience de pensée suivante : on prépare $N+1$ ensembles, correspondant à toutes les valeurs possibles de p et q. Puis on effectue un très grand nombre de sondages dans ces $N+1$ ensembles sans en favoriser aucun ; ces sondages donneront toutes les valeurs possibles pour k, mais on ne retiendra que le sous-échantillon de ceux qui donnent la valeur k_0. Puis on comptera *dans ce sous-échantillon* la répartition des p. En termes d'espace des épreuves, cela signifie qu'on prend pour Ω la réunion de $N+1$ répliques de l'ensemble des parties à n éléments de la population à N éléments, dont le cardinal sera $(N+1)\times\binom{N}{n}$. Ces répliques sont les Ω_p, tous de même cardinal $\binom{N}{n}$: chacun est l'ensemble des parties à n éléments de la population à N éléments, et ne se distingue des autres répliques que par la répartition en p et q. Les épreuves de l'espace Ω sont alors équiprobables. Quant à l'espace des épreuves de notre problème inverse, ce sera le sous-espace Ω_{k_0} des épreuves $\omega \in \Omega$ pour lesquelles le sondage aura donné $k=k_0$.

Il s'agit donc d'un problème de probabilités conditionnelles : on se restreint à un sous-ensemble de l'espace Ω (voir fin de la section **IV.3**). En reprenant la notation introduite par $IV.1$, cela s'écrit

$$\mathcal{P}\,(p=p_0 \mid k=k_0) \quad = \quad \frac{\mathcal{P}\,(p=p_0 \ \text{ et } \ k=k_0)}{\mathcal{P}\,(k=k_0)} \qquad (X.4.)$$

Le numérateur peut également s'écrire

$$\mathcal{P}\,(p=p_0 \ \text{ et } \ k=k_0) \quad = \quad \mathcal{P}\,(k=k_0 \mid p=p_0) \times \mathcal{P}\,(p=p_0) \qquad (X.5.)$$

Conformément à ce qui a été dit à la fin de la section **IV.3** sur la nature des probabilités conditionnelles, les probabilités absolues $\mathcal{P}\,(p=p_0$ et $k=k_0)$, $\mathcal{P}\,(p=p_0)$, et $\mathcal{P}\,(k=k_0)$, concernent l'espace des épreuves Ω introduit ci-dessus, tandis que les probabilités conditionnelles $\mathcal{P}\,(p=p_0 \mid k=k_0)$

et $\mathcal{P}\,(k = k_0 \mid p = p_0)$ sont les probabilités sur un sous-espace de Ω, respectivement Ω_{k_0} (l'ensemble des épreuves de Ω où le résultat du sondage est k_0) et Ω_{p_0} déjà introduit (l'ensemble des épreuves de Ω où $p = p_0$).

Il est facile de voir que $\mathcal{P}\,(p = p_0) = 1/(N+1)$: en effet, Ω est la réunion des Ω_{p_0}, qui ont tous le même cardinal $\binom{N}{n}$. Par contre il n'y a aucune évidence a priori pour la valeur de $\mathcal{P}\,(k = k_0)$; mais on peut calculer : l'espace Ω étant la réunion des Ω_{p_0}, ces derniers forment donc une *famille exhaustive d'événements*, de sorte que

$$\begin{aligned} \mathcal{P}\,(k = k_0) &= \sum_{p_0=0}^{N} \mathcal{P}\,(k = k_0 \mid p = p_0) \times \mathcal{P}\,(p = p_0) \\ &= \frac{1}{N+1} \sum_{p_0=0}^{N} \mathcal{P}\,(k = k_0 \mid p = p_0) \end{aligned} \tag{X.6.}$$

Remarque. Les relations $X.4$, $X.5$ et $X.6$ qui nous ont permis de calculer $\mathcal{P}\,(p = p_0 \mid k = k_0)$ à partir des $\mathcal{P}\,(k = k_0 \mid p)$ sont connues dans la littérature sous le nom de relation de Bayes. Si les A_i sont une famille exhaustive d'événements, on a en effet :

$$\mathcal{P}\,(A_{i_0} \mid B) = \frac{\mathcal{P}\,(B \mid A_{i_0}) \cdot \mathcal{P}\,(A_{i_0})}{\sum_i \mathcal{P}\,(B \mid A_i) \cdot \mathcal{P}\,(A_i)}$$

En général, cette relation de Bayes sert, comme c'est le cas ici, à inverser le conditionnement. Mais elle n'est qu'une combinaison immédiate des relations $IV.2$ et $IV.4$; c'est pourquoi dans les traités modernes on perd peu à peu l'habitude d'en faire un théorème séparé.

En reprenant l'ensemble de ces décompositions, on va arriver au résultat cherché ; en effet :

— $\mathcal{P}\,(k = k_0 \mid p = p_0)$ est déjà connu puisque c'est la loi du problème direct donnée par $X.1$:

$$\mathcal{P}\,(k = k_0 \mid p = p_0) \;=\; \binom{p_0}{k_0} \cdot \binom{N - p_0}{n - k_0} \Big/ \binom{N}{n}$$

— alors d'après $X.6$:

$$\mathcal{P}\,(k = k_0) = \frac{1}{N+1} \sum_{p_0=0}^{N} \binom{p_0}{k_0} \cdot \binom{N - p_0}{n - k_0} \Big/ \binom{N}{n}$$

— de sorte que $X.4$ se traduit par

$$\mathcal{P}\left(p = p_0 \mid k = k_0\right) \quad = \quad \frac{\dfrac{1}{N+1}\;\dfrac{\binom{p_0}{k_0}\binom{N-p_0}{n-k_0}}{\binom{N}{n}}}{\dfrac{1}{N+1}\displaystyle\sum_{p_0=0}^{N}\dfrac{\binom{p_0}{k_0}\binom{N-p_0}{n-k_0}}{\binom{N}{n}}} \quad = \quad \frac{\dbinom{p_0}{k_0}\dbinom{N-p_0}{n-k_0}}{\displaystyle\sum_{p=0}^{N}\dbinom{p}{k}\dbinom{N-p}{n-k_0}} \tag{X.7.}$$

La loi de probabilité de p est donc $C^{\text{te}} \cdot \binom{p}{k_0} \cdot \binom{N-p}{n-k_0}$. Comme toujours cette probabilité «exacte» est sans intérêt et il nous faut chercher sa densité asymptotique (cf. chapitre **IX**). Pour cela on peut suivre la méthode habituelle, qui a déjà fait ses preuves pour $II.7$, $II.8$, $II.10$, et encore ci-dessus pour $X.3$: d'abord chercher la probabilité maximum ; le maximum se produit pour $p = p_{\max} \simeq Nk_0/n$ (très exactement $p_{\max}$ est la partie entière de $[N+1]\,k_0/n$). Puis chercher la variation autour de $p_{\max}$. Cela conduirait à

$$\binom{p}{k_0} \cdot \binom{N-p}{n-k_0} \quad \simeq \quad \max \times \exp\left\{-\,n\,\frac{x^2}{2\alpha(1-\alpha)}\right\}$$

avec $\alpha = k_0/n$ et $x = (p - p_{\max})/N$. On reconnaît la même densité macroscopique que dans $X.3$. (la variable macroscopique qui dans $X.3$ correspond à $x = (p - p_{\max})/N$, est évidemment $y = j/n = (k - k_{\max})/n$)

Dans le cas présent on peut cependant, à partir des lois exactes, retrouver cela plus simplement. En effet, le dénominateur de $X.7$ peut être calculé très facilement ; il suffit de considérer les deux séries entières

$$\sum_{p=k}^{\infty}\binom{p}{k}z^p \;=\; \frac{z^k}{(1-z)^{k+1}} \qquad \text{et} \qquad \sum_{p=n-k}^{\infty}\binom{p}{n-k}z^p \;=\; \frac{z^{n-k}}{(1-z)^{n-k+1}}$$

Pour la commodité de l'écriture convient (comme déjà dans $X.7$) que les coefficients binômiaux $\binom{p}{k}$ sont nuls pour $p < k$; en faisant le produit de ces deux séries on obtient

$$\sum_{p=0}^{\infty}\binom{p}{k}z^p \times \sum_{p=0}^{\infty}\binom{p}{n-k}z^p \quad = \quad \sum_{N=0}^{\infty}A_N\,z^N \quad = \quad \frac{z^n}{(1-z)^{n+2}}$$

avec $A_N = \displaystyle\sum_{p=0}^{N}\binom{p}{k}\binom{N-p}{n-k}$. Or on a aussi

$$\frac{z^n}{(1-z)^{n+2}} \quad = \quad \frac{1}{z}\sum_{N=n+1}^{\infty}\binom{N}{n+1}z^N \quad = \quad \sum_{N=n}^{\infty}\binom{N+1}{n+1}z^N$$

Par conséquent, en identifiant les coefficients,

$$A_N \quad = \quad \sum_{p=0}^{N} \binom{p}{k} \binom{N-p}{n-k} \quad = \quad \binom{N+1}{n+1}$$

Ainsi $X.7$ devient

$$\mathcal{P}\left(p = p_0 \mid k = k_0\right) \quad = \quad \frac{\binom{p_0}{k_0}\binom{N-p_0}{n-k_0}}{\binom{N+1}{n+1}} \quad = \quad \frac{n+1}{N+1} \cdot \frac{\binom{p_0}{k_0}\binom{N-p_0}{n-k_0}}{\binom{N}{n}}$$

On reconnaît ci-dessus l'expression $X.1$ de la probabilité $\mathcal{P}\left(k = k_0 \mid p = p_0\right)$, de sorte que finalement

$$(N+1) \cdot \mathcal{P}\left(p = p_0 \mid k = k_0\right) \quad = \quad (n+1) \cdot \mathcal{P}\left(k = k_0 \mid p = p_0\right) \qquad (X.8.)$$

Cette relation montre qu'à un coefficient de normalisation près (ce qui est logique, puisque p peut prendre $N+1$ valeurs, alors que k en prend $n+1$), les lois $\mathcal{P}\left(k = k_0 \mid p = p_0\right)$ et $\mathcal{P}\left(p = p_0 \mid k = k_0\right)$ sont identiques.

En conclusion, ces deux interprétations du sondage sont équivalentes :

— si dans la population étudiée, on sait qu'il y a 1/3 de personnes aux yeux bleus, la probabilité pour qu'un sondage effectué sur 1800 personnes prises au hasard en donne une proportion inférieure à 30.0% ou supérieure à 36.7%, est 0.0026 ;

— si on ne sait rien sur la population, mais que dans un sondage effectué sur 1800 personnes prises au hasard il y en ait 1/3, soit 600, qui ont les yeux bleus, la probabilité pour que dans la population totale la proportion soit inférieure à 30.0% ou supérieure à 36.7%, est 0.0026.

Ces deux interprétations sont, d'après $X.8$, exactement équivalentes, à condition évidemment qu'on donne au terme "probabilité" le sens qui a présidé à nos calculs :

— dans la première interprétation, que les échantillons de 1800 personnes sont tous équiprobables ;

— dans la deuxième interprétation, que les échantillons pris sur les $N+1$ populations virtuelles de l'expérience de pensée envisagée plus haut soient équiprobables.

C'est ainsi qu'est fondée la méthode du sondage. La précision d'un sondage est donc mesurée par *deux* paramètres :

a) le *seuil* de certitude (qui est ici 0.0026) est ce qu'on accepte comme probabilité de se tromper ;

b) la *marge d'erreur* ou *fourchette* ou encore *intervalle de confiance* (qui est ici $\pm 3.33\%$) est la différence maximum qu'on accepte entre la valeur réelle de α et la valeur d'échantillon.

Lorsque la taille de l'échantillon est fixée, on ne peut bien sûr réduire la marge d'erreur qu'en augmentant les chances de se tromper ; ou inversement on ne peut réduire les chances de se tromper qu'en élargissant la fourchette. Si on veut à la fois resserrer l'intervalle de confiance et les chances de se tromper, il faut augmenter la taille de l'échantillon. Mathématiquement cela se traduit de la façon suivante : introduisons la fonction

$$\mathcal{N}(x) \quad = \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{t^2}{2}\right\} dt$$

cette fonction est parfois désignée aussi par $\mathbf{erf}\,(x)$ (**er**ror **f**unction). On a la relation suivante entre le seuil de certitude ρ et la fourchette $\pm\varepsilon$:

$$\rho \;\simeq\; 2\,\mathcal{N}\left(-\sqrt{\frac{n}{\alpha\beta}}\,\varepsilon\right) \tag{X.9.}$$

En effet, le seuil de certitude est la probabilité pour que $|\frac{k}{n} - \alpha| > \varepsilon$; or $\alpha \simeq \frac{k_0}{n}$, donc

$$\left|\frac{k}{n} - \alpha\right| > \varepsilon \qquad \Longleftrightarrow \qquad |j| > n\varepsilon \qquad \Longleftrightarrow \qquad \frac{|j|}{\sqrt{n\alpha\beta}} > \sqrt{\frac{n}{\alpha\beta}}\,\varepsilon$$

Si alors on assimile la loi discrète à la densité gaussienne, on aura d'après $(VII.6.)$

$$\begin{aligned} \mathcal{P}\left\{\frac{|j|}{\sqrt{n\alpha\beta}} > \sqrt{\frac{n}{\alpha\beta}}\,\varepsilon\right\} \;&=\; \mathcal{P}\left\{|x| > \sqrt{\frac{n}{\alpha\beta}}\,\varepsilon\right\} \;=\; 2\,\mathcal{P}\left\{x < -\sqrt{\frac{n}{\alpha\beta}}\,\varepsilon\right\} \\ &\simeq\; 2\mathcal{N}\left(-\sqrt{\frac{n}{\alpha\beta}}\,\varepsilon\right) \end{aligned}$$

La relation $X.9.$ relie entre eux le seuil ρ, la fourchette ε, et la taille de l'échantillon n. Comme nous l'avons vu plus haut, si $\rho = 0.0026$ et $n = 1800$, on aura $\varepsilon = 0.033$. Si on abaisse le seuil de certitude à $\rho = 0.072$ (soit une chance sur quatorze de se tromper), la fourchette se réduit à $\pm 2\%$; par contre si on veut augmenter la certitude à $\rho = 0.0001$ (une chance sur dix mille) la fourchette se relâche à $\pm 4.2\%$. Si on veut avoir une fourchette de $\pm 1\%$ avec le seuil de certitude de 0.26%, il faut augmenter la taille de l'échantillon : on voit d'après $X.9$ que ρ ne dépend en réalité que de $\sqrt{n}\,\varepsilon$; si ε passe de 3.33% à 1%, ρ restera inchangé si $\sqrt{n}$ est conjointement multiplié par 3.33, ce qui signifie que l'échantillon devra être 11 fois plus nombreux

$(1800 \times 11 = 19800)$: pour un seuil de certitude fixé, l'intervalle de confiance est inversement proportionnel à la *racine carrée* de la taille de l'échantillon.

La relation $X.9$ exprime la limite théorique de la mesure par échantillon. Il n'existe aucun moyen (par exemple un traitement mathématique sophistiqué) permettant d'augmenter la certitude ou de resserrer la fourchette sans augmenter conjointement la taille de l'échantillon. Par contre on s'expose facilement à des causes d'erreur additionnelles. En effet, la formule $X.9$ a été obtenue en supposant que l'échantillon est choisi *au hasard* parmi $\binom{N}{n}$ échantillons équiprobables. Cela suppose qu'on n'a pas favorisé certains échantillons par rapport aux autres. Or, si par exemple les expérimentateurs choisissent leur échantillon dans une région particulière où *à leur insu* les yeux bleus sont plus rares, ils favorisent sans le savoir les échantillons où k est inférieur à la moyenne ; il n'y alors plus équiprobabilité et la formule $X.9$ n'est plus valide. La qualité ou la difficulté des sondages ne provient donc pas du traitement mathématique des résultats, mais du protocole d'expérience : celui-ci doit déterminer un procédé de choix de l'échantillon qui soit stochastiquement indépendant de la couleur des yeux. Si par exemple on possède un fichier où toute la population est enregistrée avec un numéro d'ordre, on peut choisir l'échantillon avec une fonction *random* qui parcourt le fichier au hasard : par ce procédé l'équiprobabilité est garantie. Mais pour des populations considérables (par exemple l'électorat français) cette méthode n'est pas pratique et on a recours à des procédés de sélection souvent sophistiqués, mais plus économiques.

Dans de tels procédés l'indépendance stochastique est essentielle. Supposons par exemple que l'échantillon soit choisi selon un procédé déterminé (et déterministe : n'oublions pas que les fonctions *random* sont elles aussi toujours des algorithmes déterministes). On peut toujours ramener logiquement un tel procédé à un classement des échantillons dans un certain ordre croissant ou décroissant ; par exemple on organise un concours dans la population et les n premiers au classement constituent l'échantillon. Cela revient à considérer une variable aléatoire X : dans notre exemple, pour un échantillon (c'est-à-dire une épreuve) ω, $X(\omega)$ pourrait être la moyenne des notes obtenues au concours par les n personnes de l'échantillon ; l'échantillon formé par les n premiers au classement est alors celui qui a la meilleure moyenne, pour lequel X est maximum. Le contenu du concours ou les critères d'évaluation importent peu, il suffit qu'ils soient stochastiquement indépendants de la couleur des yeux : le classement peut être effectué par une fonction *random*. On considérera aussi une autre variable aléatoire Y : $Y(\omega)$ sera le nombre de personnes aux yeux bleus dans l'échantillon ω.

Si par exemple les échantillons sont prélevés dans la population française, et que la note donnée à chaque personne au concours est le code postal de son domicile, il n'y aura pas d'indépendance stochastique entre X et Y. Ainsi les échantillons prélevés entièrement dans les départements d'outremer donneront à X des valeurs élevées ($\sim$ 97000), tandis que les échantillons entièrement prélevés dans l'Ain donneront à X les valeurs les plus faibles ($\sim$ 1000). L'indépendance stochastique implique que les probabilités conditionnelles $\mathcal{P}\left(X(\omega) \leq 1900 \mid Y(\omega) = k\right)$ ou $\mathcal{P}\left(X(\omega) \geq 97400 \mid Y(\omega) = k\right)$ sont indépendantes de k. Cela signifierait (si on prend par exemple $k = 18$) que la probabilité de prélever des échantillons de 1800 personnes dont seulement 18 ont les yeux bleus est la même à l'île de la Réunion que dans le département de l'Ain. L'inégale répartition géographique des yeux bleus a ainsi pour conséquence que la variable aléatoire X, qui reflète l'origine géographique de l'échantillon, n'est pas stochastiquement indépendante de Y. Le sondage ne sera donc correct que si le procédé de sélection de l'échantillon garantit cette indépendance stochastique, et c'est pourquoi les procédés de sélection sont l'objet du plus grand soin.

X.3. Conclusion.

Dans ce chapitre nous avons montré que dans toute situation où interviennent des probabilités, il ne suffisait pas de les calculer ou de les mesurer, mais il fallait également savoir sur quoi portent les choix du hasard. Selon les cas, la méthode du sondage peut permettre de mesurer une probabilité a priori qu'on ne peut pas calculer par la théorie (dans ce cas on fabriquait une population en reproduisant un processus), ou bien de mesurer approximativement des données statistiques non aléatoires sur une population déjà existante, et il importe de ne pas confondre ces deux situations très différentes. La première concerne les expériences reproductibles, mais donnant un résultat aléatoire : nous avions évoqué à titre d'exemple le lancer d'une pièce de monnaie ; mais n'importe quelle expérience de Mécanique quantique tombe dans cette rubrique. Il s'agit d'expériences reproductibles dont le résultat est aléatoirement variable. Ce qui reste invariable dans ces expériences, n'est pas la valeur numérique du résultat, mais sa loi de probabilité : lorsqu'on relance une pièce de monnaie, elle ne retombe pas forcément du même côté, mais la probabilité de chaque côté reste la même à chaque lancer ; de même dans une expérience de Stern et Gerlach, l'atome a toujours une chance sur deux d'être dans un état de spin, et si on recommence l'expérience, la déviation de l'atome varie aléatoirement, mais la loi de probabilité reste la même. La *reproductibilité* de l'expérience consiste en ce que

a) la loi de probabilité du résultat reste la même ;

b) les expériences reproduites ne s'influencent pas mutuellement (indépendance stochastique).

Dans le cas du sondage sur une population on ne reproduit pas une expérience ; le sondage est lui-même une expérience, et on ne la fait qu'une fois (sauf exception). On se fonde sur la relation (X.4.) pour tirer d'une observation faite sur l'échantillon, une conclusion statistique sur la population. Nous avions insisté sur le fait que si un sondage permet d'établir qu'un médicament sera efficace pour quatre personnes sur cinq dans la population, cela ne signifie pas que chaque personne a quatre chances sur cinq pour que le médicament agisse sur elle. La différence entre ces deux conclusions provient d'une différence dans l'intervention du hasard : lorsqu'on dit que le médicament sera efficace pour quatre personnes sur cinq, on sous-entend que le hasard intervient dans le choix d'une personne parmi une population ; lorsqu'on dit que pour une personne donnée, le médicament a une probabilité 3/5 d'être efficace, on sous-entend que le hasard intervient dans le métabolisme de la personne.

Il se trouve cependant que ces deux modes d'action du hasard, quoique complètement différents, se confondent objectivement lorsqu'on a affaire a des phénomènes reproductibles ; dans le cas du médicament, ils ne se confondaient pas objectivement, car passer d'une personne à une autre n'est pas la reproduction d'une expérience.

Reprenons l'exemple de la pièce de monnaie (on aurait tout aussi bien pu prendre une quelconque expérience de Mécanique quantique). Si on la lance trente millions de fois, on disposera d'une population de résultats, par exemple 15 003 147 face et 14 996 853 pile. Si au lieu de relever tous les trente millions de résultats, on n'en relève que 1 800, on aura une estimation moins précise de la loi de probabilité $(\frac{1}{2}, \frac{1}{2})$. La précision est, comme on l'a vu au §2, de l'ordre de l'écart-type $\sqrt{n\alpha\beta}$, donc de ± 42 sur les 1 800. L'incertitude moyenne sur la *mesure* de la probabilité est donc $42/1800 \simeq 0.023$. Avec la population complète de 30 000 000 de résultats, 15 003 147 face et 14 996 853 pile, on estimerait les probabilités à 0, 500 1049 pour face et 0.499 8951 pour pile, soit une erreur de 0.000 1049 par rapport à la valeur théorique 0.5. Cette erreur peut tout aussi bien provenir des fluctuations aléatoires que des imperfections de la pièce utilisée pour l'expérience. L'incertitude a priori sur la mesure (écart-type des fluctuations) est $5\,477/30\,000\,000 \simeq 0.000\,182$, et l'erreur est bien de cet ordre, de sorte qu'avec le résultat effectivement obtenu on n'est pas fondé à penser que la pièce est mal équilibrée.

Faisons semblant de croire que la pièce est parfaite. Alors on peut ne pas faire réellement l'expérience et se contenter d'une expérience de pensée, dont le résultat est prévisible : si on la lancait dix milliards de fois, l'écart quadratique moyen des fluctuations serait de l'ordre de 70 000 sur dix milliards, soit 0.000 007 (faire cette expérience réellement demanderait un travail considérable). Le grand avantage d'une expérience de pensée est de ne pas être pas limitée par des contingences matérielles ; on peut donc supposer que la pièce est lancée un nombre N aussi grand qu'on veut de fois. Ainsi, au lieu d'avoir une population déterminée (par exemple les électeurs français) dont le nombre N ne peut pas être augmenté à volonté, on dispose d'une population potentiellement infinie, c'est-à-dire qu'on peut faire tendre N vers l'infini. Lorsque N tend vers l'infini, non seulement le nombre de pile ou de face reste proche de $N/2$, mais en outre l'incertitude liée aux fluctuations tend vers zéro, ce qui permet de "mesurer" (en quelque sorte) la probabilité avec une précision arbitrairement grande.

Dans ces conditions, la proportion de pile ou de face dans la population infinie de résultats est *égale* à la probabilité a priori. On a démontré mathématiquement (cela résulte du chapitre **III**) que si la probabilité a priori d'obtenir face est 0.5, alors, sur une population potentiellement infinie de lancers, il y aurait une proportion de pile ou de face exactement égale à 0.5 : la probabilité pour que cette proportion s'en écarte, même d'une quantité infinitésimale, serait nulle. Il en serait de même si la probabilité a priori n'était pas 0.5, mais un nombre quelconque x. La vérité mathématique de cette affirmation résulte, comme cela a déjà été dit, de deux conditions :

a) que les lancers soient reproductibles à l'identique ;

b) qu'ils soient stochastiquement indépendants.

À cause de cela il n'y a pas de différence entre la probabilité a priori dûe à un choix aléatoire dans les paramètres rapidement fluctuants du mouvement de la pièce de monnaie, et la probabilité a posteriori qui se révèle dans la population de N résultats lorsque N croît indéfiniment. Pour dire les choses autrement : si on effectue l'expérience consistant à lancer la pièce 1 800 fois, on peut indifféremment considérer que

a) on a répété 1 800 fois un lancer (à chaque fois de manière identique et stochastiquement indépendante des lancers précédents) avec une probabilité 0.5 *a priori*, c'est-à-dire une probabilité décidée *avant* la manifestation du résultat, provenant de choix faits par le hasard pur sur les *causes* mécaniques du résultat ;

ou bien que

b) on a prélevé *au hasard* un échantillon de 1 800 résultats dans une population potentiellement infinie, de sorte que le choix du hasard n'a pas porté 1 800 fois sur les causes du mouvement d'une pièce de monnaie, mais une seule fois sur le choix d'un échantillon de 1 800 mouvements, parmi tous les mouvements possibles.

Si on présuppose l'existence d'une "population" formée d'un nombre énorme de résultats possibles (mais comportant bien sûr une moitié de pile et une moitié de face) et qu'on prélève au hasard un échantillon de 1 800 résultats dans cette population, on obtiendra en appliquant la théorie des échantillons développée ci-dessus exactement les mêmes lois de probabilité qu'en appliquant la théorie de la marche aléatoire à la répétition de 1 800 lancers. Aucune expérience, même de pensée, ne pourrait mettre en évidence une différence entre les deux situations, et par conséquent permettre de trancher en faveur de l'une des deux interprétations. Il en est ainsi parce que les deux interprétations sont mathématiquement équivalentes : on a *démontré*, par la logique et non par l'expérience, que l'une entraîne l'autre, de sorte que le choix pour l'une plutôt que pour l'autre n'est qu'un choix de forme : l'interprétation b) peut donner le sentiment que les résultats sont prélevés dans l'ensemble, supposé déjà inscrit dans le Grand Livre, des résultats possibles, tandis que l'interprétation a) laisse croire que le hasard fait son choix *avant* que la pièce ne s'immobilise. Mais ces différences ne sont que des illusions dues à des manières différentes de dire la même chose.

Par contre la différence que nous avions signalée auparavant, entre le choix au hasard d'un échantillon sur une population existante, et l'action du hasard dans les mécanismes moléculaires du métabolisme, est une différence réelle et non une illusion. On peut imaginer une expérience concrète qui distingue les deux sortes d'actions du hasard : il suffit de mesurer sur une seule personne l'écart-type des fluctuations de tension artérielle dues au second hasard (celui qui agit dans les mécanismes moléculaires du métabolisme), et de constater qu'il n'est pas le même que celui des fluctuations de la tension d'une personne à l'autre. Même sans songer à cette expérience, il apparaît tout de suite à l'entendement qu'il n'y a aucun lien de cause à effet entre ce qui se passe dans le métabolisme d'une personne particulière, et les différences de constitution que d'autres personnes d'une population peuvent avoir entre elles ; par exemple, que Pierre Martin soit obèse n'a aucune influence sur le fait que Jean Dupond élimine mal le sel, surtout s'ils ne se sont jamais rencontrés et ne se connaissent pas. C'est que dans ce cas la population qu'on sonde n'est pas constituée de clones de Jean Dupond. La population de pile et de face imaginée ci-dessus a été fabriquée artificiellement par reproduction à l'infini d'un même processus et il est donc logique qu'elle ne contienne pas plus d'information qu'il n'y en avait au départ dans le processus, tandis que la population humaine existe indépendamment du métabolisme de chacun de ses membres, et apporte une information qui ne peut pas être extrapolée à partir de ce qu'on sait d'un individu particulier.