# Statistiques

# 3 Calcul des probabilités

## 3.1 Histoire

Cette section est développée et illustrée dans le diaporama « Histoire ».

### 3.1.1 Le probable et son histoire ancienne

On trouve déjà chez les Grecs une opposition entre la connaissance certaine (epistémê) parce que démontrée (comme en mathématique et philosophie) et la connaissance incertaine ou opinion (doxa) attachée aux faits que l'on ne peut connaître que par des témoignages de nos sens ou de nos semblables. Les sceptiques en ont tiré une raison de douter de toute chose, et les chrétiens ont au contraire inventé la révélation comme troisième forme de certitude concernant l'existence divine.

La scolastique, introduite par Saint Thomas d'Aquin et développée par les Jésuites à partir du XVIème siècle, consistait en l'étude des cas à travers la littérature consacrée. Elle a défini la connaissance probable comme opinion digne de foi dès lors qu'elle est émise par une autorité religieuse. Ce qui déchaîna une certaine colère des jansénistes.

A la Renaissance et à l'époque des Lumières, les hommes sont conviés à abandonner la méthode scolastique qui avait envahi les universités et à se fier davantage à l'étude du grand Livre de la Nature c'est à dire l'observation du Monde. Ce tournant marque la naissance des Sciences, mais pose de nouveau la question du type de connaissance que procure cette observation directe de la nature, même contrôlée expérimentalement. Les conclusions de Locke et Hume ont plutôt tendance à faire de la physique une connaissance probable.

### 3.1.2 Les contrats aléatoires

Comme l'a montré un bel article de l'historien des sciences Ernest Coumet, le calcul des probabilités n'est pas né par hasard. Il est le fruit d'une réflexion autour de nombreuses situations concrètes vécues dans les différents domaines du commerce, du droit, de la médecine, des pratiques actuarielles qui toutes visent à chercher le juste prix d'un contrat aléatoire. Un contrat aléatoire est ce qui règle un échange entre une loterie (c'est à dire une suite d'issues incertaines dont on sait évaluer les probabilités et les conséquences) et un montant fixe qui lui serait équivalent du double point de vue de la justice et de la justesse du calcul. Le prototype en est la prime d'assurance qui serait le prix d'un achat des différents bénéfices et pertes que peut escompter l'armateur d'un navire partant faire du commerce dans des contrées lointaines, ou encore le prix à payer pour une rente viagère jusqu'à ma mort, dont la date est bien sûr incertaine.

Si l'on a prétendu que le calcul des probabilités était né de la pratique et de la théorisation des jeux de hasard, c'est parce que les premiers mathématiciens à s'intéresser à ce calcul (Pascal, Fermat, Montmort, Huygens…) en ont établi les règles en résolvant des questions posées à propos de jeux très simples (type lancer de dés) pour lesquels la détermination des probabilités initiales est non problématique. Mais leur motivation et celles de leurs concitoyens était bien davantage de répondre à une demande sociale liée à ces différentes situations où un gain dépend d'un futur contingent (c'est à dire d'un futur qui semble advenir par hasard et pas par nécessité).

### 3.1.3 La règle de l'espérance

C'est à l'occasion d'un échange de correspondance avec Fermat en 1654 et de recherches simultanées sur le triangle des nombres binomiaux (cf infra) que Pascal met en place la règle de l'espérance mathématique, du mot *expecatio* – ce à quoi on s'attend - que l'on peut traduire aussi bien par « espérance » que par « anticipation ». La règle est établie à l'occasion du problème des partis c'est à dire du juste partage qu'il faut faire d'une mise lorsque le jeu de hasard s'interrompt à un moment où les deux joueurs n'ont pas forcément des chances égales de l'emporter. C'est une règle qui consiste à calculer par récurrence arrière, depuis les issues du jeu, la somme anticipée ou espérée comme une moyenne des gains pondérée par les probabilités. Historiquement l'idée de juste moyenne est donc antérieure et fondatrice de la notion de probabilité. Dans la version axiomatique enseignée aujourd'hui c'est l'inverse. Cette règle de l'espérance a parfois été mise en cause, comme dans le paradoxe de Saint Petersbourg :

« On suppose que A joue contre B à croix ou pile, à cette condition que s'il amène pile au premier coup il donnera une pièce à B, deux s'il l'amène au second coup, Quatre s'il l'amène au troisième, huit s'il l'amène au quatrième, et ainsi de suite; et on demande quelle somme B doit donner à A avant le jeu pour que leur sort soit égal. »

La réponse par la règle de l'espérance aboutit à un résultat inadmissible :

$1.1/2 + 2.1/4 + 4.1/8 +8.1/16 +.... = 1/2+1/2+1/2+1/2+... = \infty$

La réponse de Daniel Bernoulli par la règle dite de l'espérance morale consiste alors à remarquer q'un gain de 1 n'a pas la même utilité pour un riche que pour un pauvre, et il vaut mieux appliquer cette même règle de l'espérance à l'utilité $u$ du gain $x$ que l'on suppose croissante et plus précisément de la forme $u = k.\log x - \log a$ donc $du = kdx/x$ : l'utilité marginale est proportionnelle au gain relatif (à la fortune).

### 3.1.4 Les règles du calcul de Huygens à Laplace

A la fin du XVIIème, les règles élémentaires du calcul des probabilités sont établies. La probabilité d'un événement se définit comme le rapport du nombre des cas favorables aux nombre total des cas possibles. C'est ce qu'on appelle les probabilités totales. Mais comme le fera remarquer Laplace ce calcul ne convient que lorsque les cas sont également possibles, c'est à dire également probables. On tourne donc en rond ; la probabilité se définit par…la probabilité. Et Laplace ne sortira pas de ce cercle vicieux dans son œuvre magistrale qui est *la théorie analytique des probabilités* (1812). Il faut attendre pour cela la refondation axiomatique au début du XXème siècle par Borel, von Mises, et finalement Kolmogorov (1930). Par contre une fois définies, ces probabilités initiales permettent sans trop de difficultés d'évaluer par addition les probabilités totales d'une réunion de cas, et par multiplication les probabilités composées d'une conjonction d'épreuves (cf 3.5)

### 3.1.5 L'Ars conjectandi de Bernoulli

Cet ouvrage posthume de Jacques Bernoulli (1713) est souvent considéré comme la première pierre de la statistique mathématique, c'est à dire du lien statistique-probabilité. La première partie est une reprise du petit traité de Huygens sur le calcul des probabilités et l'espérance mathématique. La seconde partie est un petit traité des combinaisons et la troisième traite des jeux de hasard. La quatrième contient d'abord une suite de définitions de la certitude, du hasard, de la contingence, et de l'art de la conjecture : où l'on voit que la statistique est d'abord un outil utile à la décision. Une sorte de calcul des arguments et de la probation y est d'ailleurs esquissé. Mais le morceau de bravoure qui a rendu cette 4ème partie si célèbre est le théorème de Bernoulli ou loi des grands nombres qui y est présenté et démontré. Nous en verrons l'essentiel un peu plus loin (1.3). Mais il est important d'en comprendre le principe qui est de fournir un moyen d'estimer les probabilités inconnues des phénomènes par un nombre suffisant d'observations statistiques (cf. 1.2.5)

### 3.1.6 Le théorème de Bayes-Laplace

Laplace a proposé de « *ramener a deux classes de problèmes tous ceux qui dépendent de la Théorie des hasards* » selon que c'est la cause ou l'événement qui sont inconnus. Il a traité d'abord en 1772 du cas direct de la détermination des événements par des causes connues grâce à la résolution des équations aux différences partielles. En 1774, dans un « Mémoire sur la probabilité des causes par les évènements », c'est au problème inverse qu'il s'attaque, c'est à dire « *à déterminer la probabilité des causes par les événements, matière neuve à bien des égards et qui mérite d'autant plus d'être cultivée que c'est principalement sous ce point de vue que la science des hasards peut être utile dans la vie civile.* » Bien que les travaux de Thomas Bayes publiés par Price après sa mort en 1765 sous le titre « An Essay towards solving a problem in the doctrine of chance » ne soient pas connus des géomètres français avant 1780, et qu'ils se distinguent par quelques éléments sémantiques de l'interprétation laplacienne et moderne, le théorème de Bayes dans le cas particulier d'une distribution a priori uniforme et le principe de Laplace peuvent être considérés comme formellement identiques. Dans les deux cas, le principe est, connaissant les probabilités conditionnelles d'un événement pour chacune de ses causes possibles, de calculer la probabilité (inverse) de chacune des causes sachant qu'il s'est produit.

### 3.1.7 La théorie analytique de Laplace

Le grand savant français est sans doute celui qui a poussé le plus loin le calcul des probabilités. Il a surtout le mérite d'avoir utilisé toutes les ressources de l'analyse – cette branche des mathématiques qui s'est dotée grâce à Newton et quelques autres, d'un calcul infinitésimal (dérivées et intégrales) – pour répondre à la plupart des difficultés posées par le calcul des probabilités. Le couronnement de son travail commencé dans les années 1770 est la *Théorie analytique des probabilités* (1812) considérée pendant plus d'un siècle comme un sommet des mathématiques. Une des choses les plus remarquables est que Laplace a su présenter les principes, les résultats, et les applications de cette théorie dans un petit livre sans aucune formule, encore tout à fait lisible aujourd'hui (et réédité en 1986) : *Essai philosophique sur les probabilités*, d'après la 5ème édition (1825), Paris, C. Bourgois. Comme principes, on retrouvera ceux du calcul élaborés par Pascal, Huygens et Bernoulli, mais aussi sa philosophie du hasard comme « expression de notre ignorance » si bien traduite par la parabole du génie omniscient (cf. infra). Comme résultats on retrouvera ceux de cette leçon et de la suivante sur la théorie des erreurs et le théorème limite, qui apparaît comme central pour la discipline, et la loi de Laplace-Gauss, ainsi que des résultats de la théorie de l'estimation repris un siècle plus tard par Fischer. Quant aux applications, Laplace les a décrites dans l'*Essai philosophique*, et en a jeté les bases : il s'agit principalement des applications astronomiques et géodésiques, et de celles qui seront faites à la démographie. Mais d'autres sont esquissées. « *La considération des probabilités peut servir à démêler les petites inégalités des mouvements célestes, enveloppées dans les erreurs des observations, et à remonter à la cause des anomalies observées dans ces mouvements ».* Et Laplace cite les inégalités lunaires, de Saturne et Jupiter, la reconnaissance des satellites de Jupiter, le problème des marées, les fluctuations barométriques etc... Le calcul des probabilités est finalement à la base des sciences naturelles et des sciences de l'homme.

### 3.1.8 La reconstruction mathématique et les applications du XXème siècle

La théorie laplacienne s'est diffusée non sans contestation et sans améliorations aussi tout au long du XIXème siècle. Le début du XXème siècle a été cependant l'occasion d'une double révision du calcul des probabilités : la théorie de la mesure et celle de la convergence de sommes de variables aléatoires ont subi des avancées de la part des mathématiciens français Lebesgue, Borel, Fréchet et Lévy entre 1906 et 1937; tandis que les fondements du calcul étaient repris dans une théorie axiomatique par Kolmogorov. Dans le cadre de la recherche militaire développée à l'occasion de la seconde guerre mondiale, les mathématiques de la décision fournissaient de nouveau un cadre très large de problèmes traités par le calcul des probabilités (processus, stocks, files d'attentes, contrôle optimal…).

## 3.2 Significations et détermination d'une probabilité

### 3.2.1 Epistémique ou Ontique

L'interprétation classique des probabilités jusqu'au milieu du XIXème est le plus souvent de type *épistémique*, c'est à dire liée à la connaissance des choses du monde. Le hasard n'est pas dans les choses mais dans la connaissance que nous en avons. Le monde est totalement régi par la nécessité, c'est à dire qu'il y a toujours une cause ou une raison à toute chose. Mais nous n'avons pas toujours accès à ces causes. Il en résulte que le probabilisme n'est pas opposé au déterminisme : il lui est totalement complémentaire. C'est le sens de la célèbre tirade de Laplace qui définit cette articulation :

*« …Une chose ne peut commencer d'être sans une cause qui la produise. Cet axiome, connu sous le nom de principe de la raison suffisante, s'étend aux actions mêmes que l'on juge indifférentes (…). Nous devons donc envisager l'état présent de l'univers comme l'effet de son état antérieur et comme la cause de celui qui va suivre. Une intelligence supérieure qui pour un instant donné connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, et l'avenir, comme le passé, serait présent à ses yeux (…). La courbe décrite par une simple molécule d'air ou de vapeurs est réglée d'une manière aussi certaine que les orbites planétaires : il n'y a de différence entre elles que celle qu'y met notre ignorance. La probabilité est relative en partie à cette ignorance, en partie à nos connaissances… »*

Vers 1860 apparaît une autre interprétation de la probabilité que l'on peut dire *ontique* parce qu'elle affirme que le hasard est constitutif du monde, il est dans les choses elles-mêmes. On trouve une première version de cette thèse dans les travaux du physicien Maxwell sur la théorie statistique des gaz parfaits qui imagine ceux ci constitués de molécules animées d'un mouvement aléatoire continuel. La physique statistique puis la physique quantique prolongeront au XXème siècle cette conception d'un monde régi par le hasard. On trouve une seconde version de la conception ontique des probabilités dans la théorie de l'évolution qui prétend, à la suite de l'ouvrage de Darwin sur l'évolution des espèces, que celle-ci s'explique entièrement par le jeu de la variation au hasard des descendants d'un couple lorsqu'il est couplé avec la règle de sélection des mieux adaptés. Ici encore l'idée que la vie est régie par une combinaison du hasard et de la nécessité sera le credo des biologistes du XXème siècle de Watson – le découvreur de la structure de l'ADN à Monod. Lire à ce sujet l'ouvrage de Lestienne.

### 3.2.2 Subjective / Objective

Dans la même conception classique (XVIIème et XVIIIème) des probabilités la détermination des probabilités repose sur une conception mixte, à la fois objective et subjective. Prenons l'exemple du lancer d'un dé à six faces. Qu'est-ce qui fait que j'attribue une probabilité 1/6 à chaque face :

- <u>une considération objective a priori</u> (liée fortement à l'objet et mesurable) : le dé étant homogène et symétrique, chaque face a la même chance de sortir. Cette considération peut être renforcée par le *principe d'indifférence* ou encore de *raison insuffisante* : il n'y a pas de raison qu'une face ait plus de probabilité qu'une autre donc donnons-leur la même probabilité.

- une considération subjective a priori : compte tenu de l'information que j'ai, je *crois* que le dé est régulier et que les faces sont équiprobables ou au contraire je crois qu'il est truqué. Dans d'autres exemples, en particulier d'épreuves non répétables, c'est bien souvent le seul moyen que j'aie de chiffrer une probabilité : j'estime la probabilité qu'il va pleuvoir dans la journée à ¾, je pense que la probabilité de se faire écraser si on traverse en dehors des clous est de …Condorcet a plusieurs fois défini la probabilité comme une mesure de la *raison de croire* en un événement.

- une considération objective a posteriori : le fréquentisme. Le théorème de Bernoulli ou *loi des grands nombres* affirme (cf infra) que si l'on répète une même épreuve aléatoire un nombre de fois suffisamment grand, la fréquence d'un certain événement est en probabilité assez proche de sa probabilité, ce qui fait de cette fréquence une bonne estimation de la probabilité inconnue. Cette méthode, aussi précise que l'on veut, offre une réponse dite fréquentiste à la question de savoir comment déterminer une probabilité. Toutefois la méthode n'est applicable que si l'épreuve est répétable. Condorcet en a d'ailleurs déduit qu'il n'existait plus de différence entre la certitude totale des sciences déductives, et l'incertitude des sciences d'observation fondées sur les témoignages, puisque la répétition des observations nous fournit la même *raison de croire* à un événement : les probabilités subjectives et objectives sont reliées par le théorème de Bernoulli. Le mathématicien et philosophe Augustin Cournot fut le grand défenseur d'une conception objective des probabilités.

- une considération subjective a posteriori : l'évaluation subjective a priori d'une probabilité (des causes) peut être révisée grâce à la formule de Bayes en tenant compte d'observations nouvelles, qui renforcent ou affaiblissent la première évaluation. Les probabilités obtenues a posteriori sont alors un mixte de considérations subjectives et fréquentistes. C'est l'approche privilégiée aujourd'hui par l'école bayésienne en statistique mathématique, plutôt minoritaire dans l'ensemble des statisticiens, mais très répandue chez les économistes, et les décideurs.

## 3.3 La loi des grands nombres

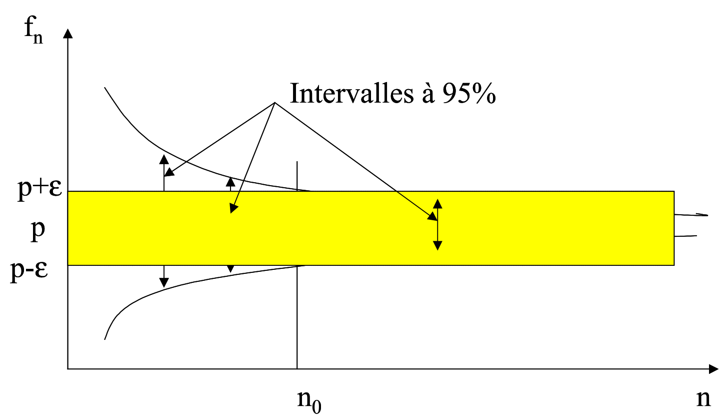
Cette loi joue un rôle très important de lien entre la statistique et le calcul des probabilités, entre les observations et le modèle.

Soit une suite d'épreuves indépendantes à deux issues comme Echouer et Réussir dans un jeu, avoir une Fille ou un Garçon à une naissance, obtenir une pièce Bonne ou Mauvaise dans un contrôle de fabrication. Une telle épreuve binaire est dite épreuve de Bernoulli. L'une des issues (par exemple la seconde) a une probabilité $p$ de se produire et l'autre ($1-p$). La répétition $n$ fois de suite et de façon indépendante de cette épreuve élémentaire est une épreuve composite dite épreuve binomiale (elle fait intervenir le binôme de Newton). On s'intéresse au nombre $x$ de succès (de probabilité $p$) dans la suite des $n$ épreuves indépendantes à deux issues, ou encore à leur fréquence relative $f_n = x/n$. Or $x$ et $f_n$ sont des grandeurs aléatoires : si on recommence l'épreuve on ne trouve pas chaque fois la même valeur. Les valeurs de $x$ peuvent varier par valeurs entières de 0 à n, mais les valeurs extrêmes sont beaucoup moins probables que les valeurs « espérées », proches de $np$. Chaque valeur de x et donc de $f_n$ a une certaine probabilité que l'on peut calculer (en utilisant le développement du binôme de Newton).

Jacques Bernoulli a montré que, pour tout couple de nombres $\varepsilon$ et $\eta$, il existait un nombre de répétitions *n* à partir duquel :

$$\Pr(|f_n - p| > \varepsilon) < \eta$$

Par exemple, il existe un nombre de répétitions $n_0$ à partir duquel la différence $|f_n-p|$ est inférieure à ε avec une probabilité supérieure à 95% (ou supérieure à ε avec une probabilité inférieure à 5%)




En d'autres termes la fréquence de l'événement est aussi proche (en probabilité) que l'on veut de la probabilité de cet événement. C'est la loi faible des grands nombres. Les « fréquentistes » ont cherché à la constituer comme définition même de la probabilité *p* mais se sont heurtés au caractère circulaire de cette définition qui doit utiliser la notion de probabilité (Pr) pour définir la probabilité (p).

La démonstration de Bernoulli repose sur des calculs un peu laborieux faits à partir du binôme de Newton. On peut aussi utiliser des outils plus modernes du calcul des probabilités à base de fonctions caractéristiques (Laplace). On peut aussi approcher cette convergence de la fréquence vers la probabilité par la formule de Bienaymé-Tchebicheff que l'on verra à la leçon suivante. Pour le moment, la meilleure approche est sans doute celle de la simulation (cf.simul1) qui montre comment la fréquence $f_n$ se rapproche de la limite *p* quand n augmente.

## 3.4 Définition axiomatique des probabilités totales

### 3.4.1 Epreuve aléatoire

Une épreuve aléatoire est une action observée ou contrôlée expérimentalement telle que les mêmes causes ne produisent pas les mêmes résultats.

Cette définition semble contredire le déterminisme de bon sens. Plus exactement, la tradition scientifique mécaniste dit à peu près le contraire : les mêmes causes produisent les mêmes effets.

Comment réconcilier ces deux propositions ? En précisant : « une action telle que des conditions *macroscopiquement* (ou *grosso* modo) identiques produisent des résultats différents ». Si je lance un dé plusieurs fois de suite de la même façon, il ne donnera pas en général le même résultat parce que, de fait mon geste a été légèrement différent à chaque fois, ce que l'on pourrait vérifier en plaçant sur mon poignet des capteurs en qualité et nombre suffisants. On en revient à la proposition de Laplace. Mais c'est parce que son démon omniscient n'existe pas et que ces capteurs sont hors de portée de la technologie usuelle que je dois abandonner le modèle mécaniste et déterministe, au profit d'un modèle aléatoire.

Ce n'est donc pas l'épreuve qui est ou n'est pas aléatoire, c'est la manière de la modéliser.

### 3.4.2 Résultat

Définir une épreuve aléatoire suppose que l'on décrive un protocole d'observation ou d'expérimentation, et que l'on précise à quel *résultat* on s'intéresse.

Ex1 : je lance un dé. Je peux m'intéresser a) au chiffre inscrit sur la face supérieure du dé au repos b) ou au fait qu'il tombe ou pas de la table.

Ex2 : je lance deux dés, un vert un rouge. Je peux m'intéresser a) au couple ordonné (vert-rouge) des deux points b) au couple non ordonné des deux points montrés c) à la somme des deux points, d) à leur produit…

Ex3 : Je tire au hasard une personne dans une assemblée. Je peux m'intéresser a) à son nom, b) à son sexe, c) à son âge, d) à sa profession, e) à son revenu…

Ex4 : E = {a,b,c} = ensemble de 3 personnes qui montent dans l'ascenseur au rez de chaussée d'un immeuble de 6 étages. Je peux m'intéresser a) à l'ensemble des personnes qui descendent au $6^{ème}$ b) à leur nombre c) à la répartition nominale de sortie sur les étages 1 à 6.

L'ensemble des résultats possibles de l'épreuve, noté Ω, est la première chose à laquelle il faut s'intéresser. Pour décrire cet ensemble, nous pouvons le faire en compréhension (une phrase non ambiguë) ou en en extension (une liste exhaustive des résultats codés d'une façon explicite). Cette seconde méthode est préférable chaque fois que nous le pouvons.

On appelle Cardinal de Ω, noté CardΩ ou |Ω| le nombre des éléments de l'ensemble Ω.

Attention, il faut toujours décrire avant de compter !

Ex.1a : $\Omega_{1a}$ = ensemble des faces du dé; Ω = {1, 2, 3, 4, 5, 6}; |Ω|=6.

Ex.2a : $\Omega_{2a}$ = ensemble des couples ordonnés = {(1,1), (1,2), (1,3)…(2,1), (2,2)...(6,6)}. Il est plus simple de décrire ces couples comme les cases d'un tableau croisé dans lequel les lignes repèrent le dé vert et les colonnes le dé rouge. On voit mieux qu'un même couple non ordonné correspond à deux couples ordonnés sauf pour les doublons, et que |Ω| = 6x6 = 36.

| | | | | | |
|---|---|---|---|---|---|
| (1,1) | (1,2) | (1,3) | (1,4) | (1,5) | (1,6) |
| (2,1) | (2,2) | (2,3) | (2,4) | (2,5) | (2,6) |
| (3,1) | (3,2) | (3,3) | (3,4) | (3,5) | (3,6) |
| (4,1) | (4,2) | (4,3) | (4,4) | (4,5) | (4,6) |
| (5,1) | (5,2) | (5,3) | (5,4) | (5,5) | (5,6) |
| (6,1) | (6,2) | (6,3) | (6,4) | (6,5) | (6,6) |

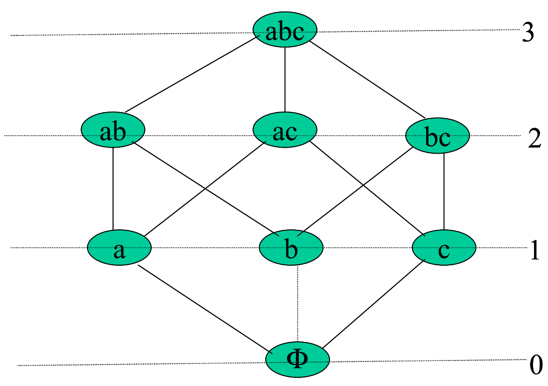
Ex.2b : dans le cas où l'on s'intéresse aux couples non ordonnés, Ω se réduit au sous ensemble des cases jaunes du tableau sous la diagonale y compris celle-ci , et $|\Omega_{2b}|$= 21.

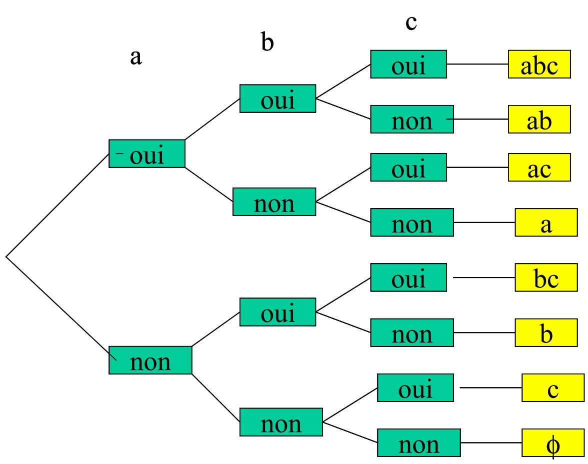
Ex.2c : Ω = {2,3,4,5,6,7,8,9,10,11,12} ; |Ω|=11.

Ex. 3b : Ω = {H,F} ; |Ω| = 2.

Ex. 4a : Ω = ensemble des parties de E = {ϕ,(a),(b),(c),(ab),(ac),(bc), (abc)} ; |Ω| = 8.

On peut générer cet ensemble par un treillis classant les parties par taille et les reliant par leur relation d'inclusion (premier schéma) ou par un arbre qui génère ces parties en répondant chaque fois à la question « est-ce que a est dans cette partie ? », puis « est-ce que b est dans cette partie ? », puis « est-ce que c est dans cette partie (second schéma) : »




Ex.4b : $\Omega = \{0, 1, 2, 3\}$ et $|\Omega| = 4$.

Ex.4c : un résultat est une correspondance qui indique pour chaque personne l'étage où elle descend. Si a descend en 2, et b et c descendent en 4 on notera ce résultat : 244.

$\Omega$ = ensemble des arrangements avec répétitions possibles.

$\Omega = \{111,112,113,114,115,116,121,122,123,124,125,126,\ldots\}$.

Il y a 6 choix pour a, 6 choix pour b, 6 choix pour c; donc $|\Omega| = 6x6x6 = 216$.

On voit qu'il faudra très vite apprendre à *dénombrer* les résultats possibles sans les compter, donc par des formules générales. C'est l'objet de la dernière partie de cette leçon (1.6).

### 3.4.3 Evénement

Un événement (désigné par une grande lettre) est toujours assimilable à une partie de $\Omega$. On peut lui assigner aussi un cardinal qui est le nombre de résultats favorables à cet événement.

Dans 1a, A = obtenir un chiffre pair = $\{2,4,6\} \subset \Omega$ (le signe $\subset$ veut dire inclus dans).

Dans 1a, B = obtenir un chiffre > 2 = $\{3,4,5,6\}$ et $|B| = 4$.

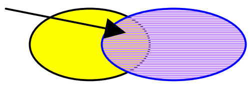
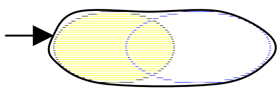
Dans 2a, A = obtenir un doublet = $\{11,22,33,44,55,66\}$ et $|A| = 6$.

Dans 2a, B = obtenir au moins un 6 = ensemble des cases de la dernière ligne et de la dernière colonne du tableau = $\{16,26,36,46,56,66,61,62,63,64,65\}$ et $|B|=11$ (ne pas compter 2 fois 66).

Dans 2b, le même événement B ne comprend plus que les 6 couples non ordonnés de la dernière ligne.

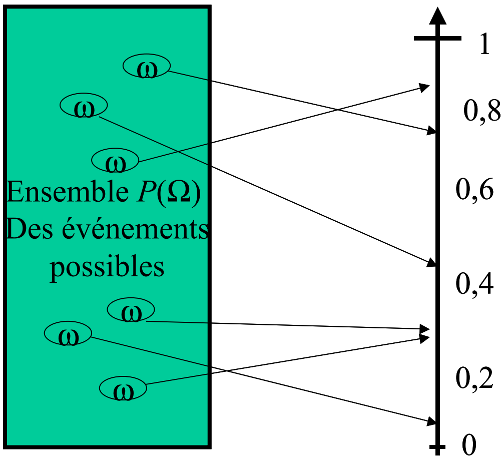
Dans 4c, A = ils descendent tous au même étage = $\{111,222,333,444,555,666\}$.

Puisque un événement est un ensemble (un sous ensemble de $\Omega$) on va pouvoir définir une algèbre des événements grâce aux opérateurs de la théorie des ensembles. Nous donnons dans le tableau qui suit la correspondance entre symbolisme, théorie des ensembles, et langage probabiliste :

| **Symbolisme** | **Langage ensembliste** | **Langage probabiliste** |
|---|---|---|
| $\Omega = \{\omega_1, \ldots \omega_n\}$<br>$\omega \in \Omega$<br>$\omega \notin \Omega$ | $\Omega$ est un ensemble d'éléments.<br>$\omega$ appartient à $\Omega$.<br>$\omega$ n'appartient pas à $\Omega$ | $\Omega$ ensemble des résultats possibles.<br>$\omega$ est un résultat possible.<br>$\omega$ n'est pas un résultat possible. |
| $\|\Omega\| = n$ | N est le cardinal de $\Omega$ | Il y a n résultats possibles. |
| $A \subset \Omega$<br>$(\omega \in A)$♥ $(\omega \in \Omega)$ | A est inclus dans $\Omega$ = A est une partie de $\Omega$ : si $\omega$ est dans A, il est aussi dans $\Omega$. | A est un événement. |
| $\omega \in A$ | L'élément $\omega$ appartient à A. | Le résultat $\omega$ réalise l'événement A |
| $A \in P(\Omega)$<br>$\|P(\Omega)\| = 2^n$ | A appartient à l'ensemble des parties de $\Omega$. | $P(\Omega)$ est l'ensemble des événements possibles. Il y en a $2^n$ |
| $A = \phi$ | A est l'ensemble vide. | Il n'y a pas de résultat qui réalise A. A est l'événement impossible. |
| $A = \Omega$ | A est l'ensemble « plein ». | Tous les résultats possibles réalisent A. A est un événement certain |
| $\|A\| = 1$ | A ne comprend qu'un élément. C'est un singleton. | Un seul résultat réalise A. A est un événement élémentaire. |
| $\overline{A} = \{\omega \in \Omega \ / \ \omega \notin A\}$ | $\overline{A}$ (noté parfois A') est la partie complémentaire de A dans $\Omega$ | $\overline{A}$ est l'événement contraire de A. |
| $C = A \cap B$<br>$(\omega \in C) \Leftrightarrow (\omega \in A)$ et $(\omega \in B)$ | C est l'intersection de A et B : un élément appartient à C si il appartient à A et à B. | C = A et B <br>$\omega$ réalise C.<br>$\Leftrightarrow$ $\omega$ réalise A et B. |
| $A \cap B = \phi$ | A et B sont disjoints : il n'existe pas d'élément commun. | A et B sont incompatibles : il n'y a pas de résultat qui réalise A et B. |
| $D = A \cup B$ | D est la réunion de A et B : un élément appartient à D s'il appartient à A ou (inclusif) à B. | D = A ou B. <br>$\omega$ réalise D.<br>$\Leftrightarrow$ $\omega$ réalise A ou B (ou les deux). |

### 3.4.4 Mesure de probabilité

Une mesure de probabilité est une application qui, à tout événement fait correspondre un nombre réel compris entre 0 et 1 qui est la probabilité de cet événement. Mais pour être une mesure de probabilité cette application doit vérifier un certain nombre de propriétés, dont un petit nombre (2) prises comme axiomes et les autres conséquences de ces axiomes. Ces deux premières propriétés sont évidemment des propriétés souhaitées de la mesure de probabilité. Comparons avec une mesure de température : La mesure de température en degrés Celsius possède aussi son zéro (la glace) et sont 1 (en fait 100° = eau bouillante); mais elle n'est pas additive (un corps à 37° plongé dans un bain à 43° cela ne fait pas 80°!) contrairement aux mesures de longueur ou de probabilité.




**Propriété 1 (Axiome)**

$p(\Omega) = 1$

La probabilité de l'événement certain est égale à 1.

**Propriété 2 (Axiome)**

$A \cap B = \phi$ ♥ $p(A \cup B) = p(A) + p(B)$

Si A et B sont incompatibles, la probabilité de leur réunion est la somme des probabilités de chacun.

Quel est le « zéro » de la probabilité ? Intuitivement ce devrait être la probabilité de l'événement impossible. Pourquoi cela n'est pas le contenu d'un troisième axiome ? Parce que c'est une propriété que l'on peut déduire des deux axiomes précédents, comme toutes celles qui suivent d'ailleurs :

**Propriété 3** : $p(\phi) = 0$

En effet appliquons l'axiome 2 pour $A=\Omega$ et $B=\phi$. Nous avons bien $\Omega \cap \phi = \phi$ et donc :

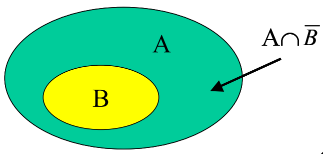
$p(\Omega \cup \phi) = p(\Omega) + p(\phi)$ mais $\Omega \cup \phi = \Omega$ donc cette relation devient :

$p(\Omega) = p(\Omega) + p(\phi)$ ce qui entraîne comme seule solution $p(\phi) = 0$.

**Propriété 4** : si $B \subset A$ alors $p(B) < p(A)$

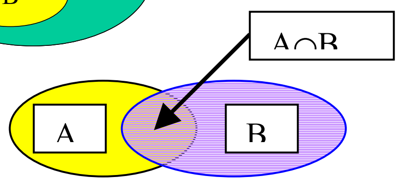
Cette propriété se démontre en appliquant

de nouveau l'axiome 2 à B et $\overline{B} \cap A$.




**Propriété 5** : $p(\overline{A}) = 1 - p(A)$

Se démontre en appliquant l'axiome 2 à A et $\overline{A}$.




**Propriété 6** : $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

Additivité dans le cas général où A et B ne sont pas disjoints : Ajouter les probabilités de A et de B c'est compter deux fois la probabilité de $A \cap B$ ; il faut donc la retirer une fois.

### 3.4.5 Calcul de la probabilité d'un événement

Une conséquence de l'axiome 2 est que la probabilité d'un événement est la somme des probabilités des événements élémentaires qui le composent :

$p(A) = \sum p\{\omega_i\} \ / \ \omega_i \in A$

Dans l'exemple 1a, A = chiffre pair = {2,4,6} $\Rightarrow$ p(A) = 1/6 + 1/6 + 1/6 = 3/6 = ½

On pourrait penser avec juste raison qu'il est plus simple d'obtenir ce même résultat avec la règle classique:

$$p(A) = \frac{|A|}{|\Omega|} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}$$

En effet on voit bien que ces deux règles sont les mêmes ***si les résultats sont équiprobables*****.** Elles ne le sont pas en général comme on peut le voir dans l'exemple suivant :

Ex. 2a : A = obtenir un doublet = {11,22,33,44,55,66}. La probabilité de A serait donc 6 cas favorables / 36 cas possibles = 1/6 si on raisonne sur les 36 couples ordonnés. Mais :

Ex 2b : le même événement A aurait une probabilité 6 cas favorables / 21 cas possibles = 2/7 si l'on raisonne sur les couples non ordonnés. Le résultat dépendrait-il de la façon de décrire Ω ? Y a-t-il contradiction à l'intérieur de cette branche des mathématiques qu'est le calcul des probabilités, ce que d'Alembert lui même avait pensé ? Sinon quel est le raisonnement qui est faux ?

C'est le second car les 21 couples non ordonnés ne sont pas équiprobables. Ce que l'on sait c'est que les faces de chaque dé sont équiprobables et que les deux dés sont indépendants. Cela conduit à dire (cf infra) que les 36 couples ordonnés de $\Omega_{2a}$ sont équiprobables. A chacun des couples non ordonnés correspondent soit deux couples ordonnés, soit un seul dans le cas des doublons. Les 21 couples non ordonnés de $\Omega_{2b}$ ont une probabilité égale à 2/36 pour les uns et 1/36 pour les autres. La formule « cas favorable /cas possibles » ne marche que dans le premier cas. La formule « somme des probabilités des événements élémentaires » marche toujours.

Ex. 4c : peut on calculer la probabilité de A = ils descendent tous au même étage = {111,222,333,444,555,666} ? Oui si nous sommes capable de définir les probabilités des événements élémentaires, c'est à dire des 216 résultats possibles. Il faut rajouter une hypothèse aux données fournies, par exemple que les 3 personnes descendent à un étage « choisi au hasard ». Cette dernière expression voulant dire qu'ils donnent la même chance 1/6 à chacun des 6 étages. Et que l'on ajoute qu'ils prennent des décisions indépendantes. Dans ce cas cela entraîne la même probabilité 1/216 des 216 résultats possibles et la probabilité de A est alors de 6/216 = 1/36.

## 3.5 Probabilités conditionnelles et indépendance

### 3.5.1 Probabilités conditionnelle

Toutes les probabilités d'événements évoquées jusqu'ici sont toujours relatives au même ensemble des possibilités : Ω. Supposons que pour une raison ou une autre on ait l'information qu'un certain événement B s'est réalisé. Alors l'ensemble des possibles s'est réduit à A. Il faudra donc réévaluer les probabilités des autres événements, de A par exemple, *sachant que B s'est réalisé*. Cette probabilité *conditionnelle* de A sachant B se note $p_B(A)$.

Voici par exemple les données de contrôle de fabrication de 3 ateliers A, B, C d'une même usine pour lesquelles les 5000 pièces (Ω) sont classées à la sortie de production en conformes à la norme

(N) ou Mauvaises (M):

| | A | B | C | Total |
|---|---|---|---|---|
| Normales | 850 | 2250 | 1400 | 4500 |
| Mauvaises | 150 | 250 | 100 | 500 |
| Total | 1000 | 2500 | 1500 | 5000 |

Si je tire une pièce au hasard dans le lot complet des 5000 pièces produites (Ω), la probabilité qu'elle soit mauvaise est p(M) = 500/5000 = 10% ; qu'elle soit bonne p(N) = 90%. La probabilité qu'elle provienne de A est p(A) = 1000/5000 = 1/5 = 20% ; qu'elle provienne de B : 50% ; de C : 30%.

Si je sais que la pièce provient de A, cette information *modifie* la probabilité qu'elle soit mauvaise qui devient : $p_A(M)$ = 150/1000 = 15% . Il reviendrait au même d'écrire :

$$p_A(M) = \frac{150/5000}{1000/5000} = \frac{p_\Omega(A \cap M)}{p_\Omega(A)} = \frac{p(A \cap M)}{p(A)}$$

qui exprime la probabilité conditionnelle directement en fonction des probabilités non conditionnelles.

Pour C la probabilité est également modifiée, mais à la baisse : $p_C(M)$ = 1/15 = 6,66%.

### 3.5.2 Indépendance

Remarquons que la probabilité d'avoir une pièce mauvaise n'est pas modifiée par le fait qu'elle provienne de B, puisque $p_B(M)$ = 250/2500 = 10% = p(M). On dira que M et B sont deux événements indépendants, alors que M et A comme M et C étaient dépendants en probabilité. Cette indépendance est une notion symétrique car on peut vérifier que $p_M(B) = p(B)$. Le fait de savoir que l'un des événements s'est produit ne modifie pas la probabilité de l'autre.

En résumé : Pour deux événements A et B quelconques on a toujours :

$p(A\cap B) = p_B(A).p(B) = p_A(B).p(A)$

Mais s'ils sont indépendants la condition « sachant A » ou « sachant B » ne change rien

Et l'on a : $p(A\cap B) = p(A).p(B)$

Cette règle de la *multiplication* pour A ***et*** B dans le cas de l'indépendance est à comparer avec la règle d'*addition* pour A ***ou*** B dans le cas de l'incompatibilité.

Cette dernière propriété - l'incompatibilité - n'est pas à confondre avec l'indépendance. La première ne fait pas intervenir de calcul des probabilités et dit simplement qu'il n'y a pas de résultat qui puisse réaliser les deux événements (l'intersection est vide et donc de probabilité nulle). La seconde propriété d'indépendance dit que cette probabilité de l'intersection est juste égale au produit des probabilités de A et de B.

### 3.5.3 Théorème de Bayes

La formule de Bayes est une application particulière aux causes d'un événement d'une conséquence immédiate des deux écritures de $p(A \cap B) = p_B(A).p(B) = p_A(B).p(A)$. On en déduit immédiatement qu'une des probabilités conditionnelles peut s'exprimer en fonction de l'autre :

$p_B(A) = p_A(B).p(A)/p(B)$

Mais comme $B = (B \cap A) \cup (B \cap \overline{A})$

$p(B) = p(B \cap A) + p(B \cap \overline{A}) = p_A(B).p(A) + p_{\overline{A}}(B).p(\overline{A})$

Donc $p_B(A) = \dfrac{p_A(B).p(A)}{p_A(B).p(A) + p_{\overline{A}}(B).p(\overline{A})}$

Bayes en a tiré une généralisation à plus de deux événements, et Laplace une interprétation en terme de probabilité qu'un événement E ait pour cause une des causes possibles:

Si l'événement E est l'effet d'une série de n causes disjointes $A_j$ , alors la probabilité dite a posteriori (après observation de E) de la cause $A_i$ est :

$$p_E(A_i) = \frac{p_{A_i}(E).p(A_i)}{\sum_j p_{A_j}(E).p(A_j)}$$

ce que Laplace réduit à $\dfrac{p_E(A_i)}{\sum_j p_{A_j}(E)}$

si les probabilités a priori des causes sont égales à savoir si $\forall i, p(A_i) = 1/n$ .

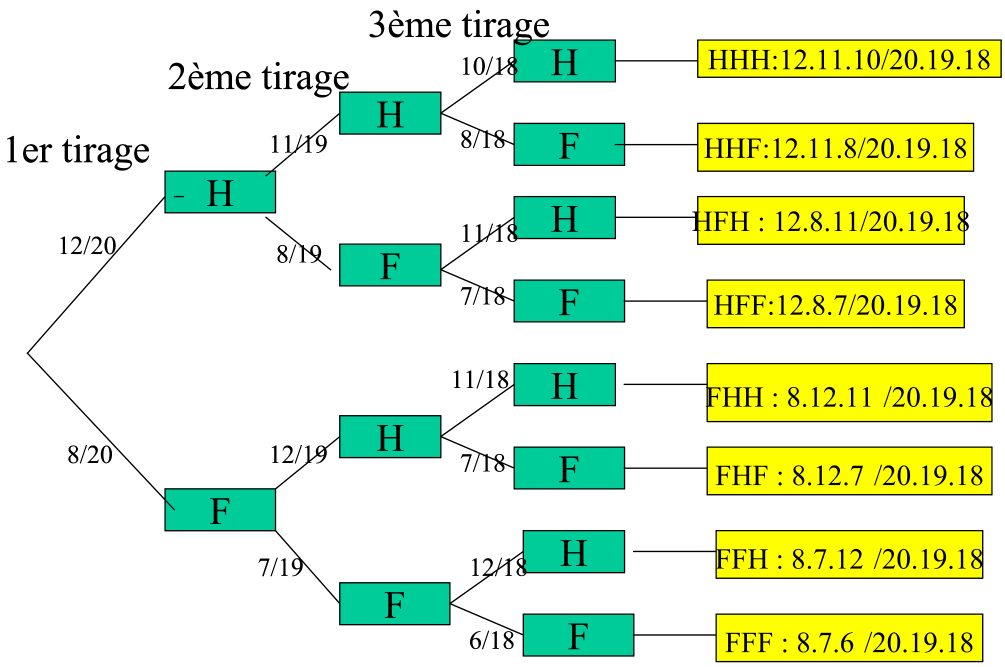
LAPLACE : « *PRINCIPE. Si un événement peut être produit par un nombre n de causes différentes, les probabilités de l'existence des causes prises de l'événement sont entre elles comme les probabilités de l'événement prises de ces causes, et la probabilité de l'existence de chacune d'elles est égale à la probabilité de l'événement prise de cette cause, divisée par la somme de toutes les probabilités de l'événement prises de chacune de ces causes* ».

Cette formule de la probabilité des causes dépend fortement d'un principe très discuté et contesté de « raison insuffisante », à savoir que ne sachant rien des causes on les suppose également probables a priori. On peut aussi, comme la plupart des bayesiens modernes, utiliser cette formule sans ce principe, pour réviser des probabilités subjectives estimées a priori.

### 3.5.4 Epreuves répétées. Tirages.

La plupart des problèmes de calcul des probabilités font appel à des épreuves composées, obtenues par une répétition d'épreuves élémentaires. La règle du produit des probabilités élémentaires est alors celle qui s'applique systématiquement.

Il est pratique de figurer cette succession d'épreuves élémentaires par un arbre, binaire par exemple si les épreuves ont 2 issues. Les arcs (ou branches) représentent les épreuves élémentaires. Ils sont évalués par des probabilités conditionnelles. Les nœuds représentent les résultats des épreuves. Exemple d'un tirage sans remise de trois personnes dans un groupe comprenant 12 hommes et 8 femmes :




## 3.6 Dénombrement

Voici quelques éléments de dénombrement utiles pour dénombrer les cas possibles ou favorables à un événement dans les exercices de calcul des probabilités.

### 3.6.1 Binôme de Newton

$(a+b)^2 = a^2 + 2ab + b^2$

$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$

$(a+b)^n = a^n + nab^{n-1} + \{n(n-1)/1.2.\}a^2b^{n-2} + \ldots$

$+ \{n(n-1)\ldots(n-k)/1.2.3\ldots k\}a^kb^{n-k}$

$+\ldots+na^{n-1}b + b^n$

Pour a = 1 et b = 1 on obtient une décomposition du nombre total ($2^n$) des résultats possibles d'une épreuve de Bernoulli (deux issues A et B) répétée n fois, en des termes successifs (termes entres {}) appelés « coefficients binomiaux » et aujourd'hui « combinaisons ».

$C_n^k = \dfrac{n!}{k!(n-k)!}$ = nombre de résultats comprenant k fois A et n-k fois B

= nombre de parties à k éléments prises dans un ensemble à n éléments

Pascal a conçu vers 1665 une méthode de génération de ces coefficients binomiaux à l'aide d'un tableau triangulaire à partir de 2 propriétés :

1. On a toujours une seule partie vide dans un ensemble : $C_n^0 = 1$
2. On peut trouver les parties à 2 éléments dans un ensemble à 4 éléments (a,b,c,d) en prenant d'abord les parties à 2 éléments d'un ensemble à 3 éléments (a,b,c), à savoir ab, ac, et bc, et en rajoutant les parties à 2 éléments que l'on peut faire en rajoutant d aux partie à 1 élément de (a,b,c), ce qui donne ad, bd, cd. Soit plus généralement la formule de récurrence : $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$

Il suffit d'appliquer cette règle pour trouver les coefficients d'une ligne en fonction de ceux de la ligne précédente :

| n k= | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | | | | | | | | 2 |
| 2 | 1 | 2 | 1 | | | | | | | 4 |
| 3 | 1 | 3 | 3 | 1 | | | | | | 8 |
| 4 | 1 | 4 | 6 | 4 | 1 | | | | | 16 |
| 5 | 1 | 5 | 10 | 10 | 5 | 1 | | | | 32 |
| 6 | 1 | 6 | 15 | 20 | 15 | 6 | 1 | | | 64 |
| 7 | | | | | | | | | | |

### 3.6.2 Principes généraux de dénombrement

On peut résoudre la plupart des questions de dénombrement avec la méthode suivante.

Le problème consiste très souvent à former des ***« mots » de longueur p*** à l'aide d'un ***« alphabet » de n signes***.

Si l'ordre de ces signes dans les mots n'importe pas pour distinguer deux mots on parle de combinaisons. Si l'ordre des signes importe, on parle d'arrangements. Les arrangements peuvent être représentés par un arbre; les combinaisons par un treillis.

Selon que la répétition du même signe dans un mot est possible ou non, on a des combinaisons ou des arrangements avec ou sans répétitions.

Le tableau suivant résume donc les 4 cas principaux :

| | L'ordre importe<br>Arrangements | L'ordre n'importe pas<br>Combinaisons |
|---|---|---|
| La répétition est possible. | $E_n^p = n^p$ | $K_n^p = C_{n+p-1}^p$ |
| La répétition n'est pas possible. | $A_n^p = \dfrac{n!}{(n-p)!}$ | $C_n^p = \dfrac{n!}{p!(n-p)!}$ |

Rappelons que n! = n.(n-1)(n-2)(n-3)….3.2.1 se dit « factorielle n » et 0! = 1.

Le nombre de permutations de n signes $P_n$ est un cas particulier d'arrangements $A_n^p$ pour lequel p = n. Donc $P_n$ = n!.

Quelques exemples d'application :

- Combien de parties peut on faire avec l'ensemble : E = {a,b,c,d,e} ?

Une partie est constituée de la présence ou de l'absence de chaque élément de E. Construire une partie, c'est faire un mot de 5 signes pris dans l'alphabet {oui, non} dans lequel l'ordre n'importe pas et la répétition est possible : $E_2^5 = 2^5 = 32$ donc |P(E)| = 32.

- Combien de tiercés gagnants possibles dans l'ordre avec 15 chevaux au départ ?

Un tel tiercé est un mot de 3 signes pris dans l'alphabet {1,2,3,….,15} dans lequel la répétition est impossible : donc il y a $A_{15}^3$ = 15!/12! = 15.14.13 = 2730 solutions.

- Combien de tiercés possibles dans le désordre : 6 fois moins puisque l'on obtient 3! tiercés dans l'ordre à partir d'un tiercé dans le désordre. Autre raisonnement, aller directement aux combinaisons : la solution est $C_{15}^3$ = 15!/3!12! = 455.