

# Les probabilités sans peine ?

Olivier RIOUL

10 mai 2012

Parlons de ce qui fâche : les probabilités forment un domaine mathématique flou et hasardeux par essence, qui fourmille de paradoxes et de calculs contre-intuitifs. Elles nécessitent en préalable une formation solide en théorie de la mesure et de l'intégration de Lebesgue : la simple notion de variable aléatoire fait appel à celle de fonction mesurable sur un espace probabilisé muni d'une tribu — que dire alors des concepts plus avancés de stationnarité et d'ergodicité, si indispensables à l'ingénieur ? Elles demandent, pour être bien faites, une très grande expérience et un langage spécifique, décorrélés des mathématiques ordinaires. Réduites à la modélisation statistique du type vu en Terminale, elles sont très pauvres ; enseignées pour elle-mêmes, elles deviennent trop abstraites. Et l'on peut légitimement craindre leur inflation probable aux concours d'entrée aux Grandes Écoles.

Pour chacune de ces idées reçues, je tenterai quelques possibles démentis, remèdes ou pistes de solutions.

## Table des matières

<b>1</b>	<b>Les probabilités : un domaine hasardeux ?</b>	<b>3</b>
<b>2</b>	<b>Les probabilités : lieu de paradoxes ?</b>	<b>4</b>
2.1	Des résultats corrects mais contre-intuitifs . . . . .	4
2.2	Paradoxes de langage . . . . .	6
<b>3</b>	<b>Les probabilités : une théorie trop difficile ?</b>	<b>8</b>
3.1	Se débarrasser du $\Omega$ . . . . .	9
3.2	Tout axer sur les « v.a. » . . . . .	10
3.3	Isoler les cas discret et continu . . . . .	11
3.4	Tout axer sur les distributions . . . . .	12
3.5	Tout le reste . . . . .	13
<b>4</b>	<b>Les probabilités : un monde à part ?</b>	<b>13</b>
4.1	Deux exemples en analyse . . . . .	14
4.2	Deux exemples en algèbre . . . . .	15
4.3	Deux exemples en arithmétique . . . . .	17
4.4	Deux exemples en géométrie . . . . .	19
<b>5</b>	<b>Les probabilités : ennuyeuses et trop abstraites ?</b>	<b>21</b>
<b>6</b>	<b>Les probabilités : probable inflation aux concours ?</b>	<b>22</b>
<b>7</b>	<b>En guise de conclusion</b>	<b>23</b>
<b>A</b>	<b>Demandez le programme</b>	<b>24</b>
A.1	Première S (septembre 2011) . . . . .	24
A.2	Terminale S (septembre 2012) . . . . .	24
A.3	Maths Sup. (septembre 2013) . . . . .	25
A.4	Maths Spé. (septembre 2014) . . . . .	25
A.5	Programme typique en Grande École . . . . .	25

# 1 Les probabilités : un domaine hasardeux ?

Les probabilités traitant du hasard, on peut craindre un domaine hasardeux par excellence.

Il faut d'abord comprendre pourquoi cela peut faire peur à l'étudiant comme à l'enseignant. Le hasard va à l'encontre de cette idée de belle rigueur qui caractérise *la Mathématique*. Il y a en effet une horreur de l'esprit humain pour le désordre, le flou, l'incertain de notre environnement. Cela s'est reflété dans le passé par le besoin de certitude dans les sciences, allant jusqu'au Déterminisme, conception selon laquelle toute la Nature, notre propre existence comprise, est entièrement déterminée, dans le passé ou dans l'avenir, par des lois immuables. Décrire scientifiquement le *hasard* est donc un gageure : tel le marquis de Laplace, nous pouvons croire qu'il ne reflète que notre incompetence à décrire précisément tous les facteurs qui contribuent à l'état de notre environnement, dans un lieu  $\ell$  et à un instant  $t$  donnés. Le hasard serait donc insaisissable et échapperait à notre connaissance.

Mais voilà que depuis Pascal et Fermat, on découvre – ou on invente... – que le hasard est lui-même soumis aux règles mathématiques, en particulier si l'on considère des configurations *moyennes*. C'est ce qu'on explique généralement par la *loi des grands nombres*. Pour décrire le hasard, il faut alors probablement distinguer deux problèmes... :

1<sup>o</sup>) Le premier problème est d'ordre logique : comment faire pour inventer une définition opérationnelle, la *probabilité*, qui obéira aux axiomes que l'on voudra bien poser et aux théorèmes que l'on en déduira ? La théorie mathématique ne commence vraiment qu'une fois tous les événements considérés affectés de nombres réels du segment  $[0, 1]$  appelés « probabilités ». La rigueur reprend alors tous ses droits.

2<sup>o</sup>) Le deuxième problème est plus physique : comment cette affectation de probabilités aux événements a lieu dans la réalité ? Les événements physiques obéissent-ils aux axiomes mathématiques ? Le désordre et ses fluctuations naturelles sont-elles bien expliquées par la théorie ? Comment alors se définit le vrai hasard, celui dont nous avons la sensation et l'intuition ? Pour toutes ces questions, on est bien obligé de constater une certaine impuissance à trouver des réponses définitives. En tout cas, on sort du domaine mathématique.<sup>1</sup>

Si plonger dans le monde de l'incertain peut inquiéter voire horrifier, c'est une sensation humaine qui ne semble pas provenir des mathématiques elles-

---

1. Je n'ai évidemment pas la prétention de tout traiter et je ne m'attacherai ici qu'à la description mathématique des Probabilités. J'invite l'auditeur de cette conférence désireux d'explorer d'autres axes à se tourner vers les exposés d'Alain MARUANI pour la physique et d'Yves GUIARD pour les sciences humaines.

même. Pour prévenir la peur que l'on pourrait ressentir, démystifions-les : ce seront toujours des mathématiques ordinaires<sup>2</sup> et elles n'ont en elles-mêmes rien d'hasardeux. Après tout, les mathématiques des probabilités sont conçues comme les autres : il y aura toujours des axiomes, des déductions logiques et des théorèmes. L'interprétation physique des modèles mathématiques du hasard est, par contre, sujette à un autre débat.

## 2 Les probabilités : lieu de paradoxes ?

On a eu souvent coutume de mettre en exergue des paradoxes liés au calcul de probabilités, en particulier dans les journaux et ouvrages de vulgarisation. Voici ci-dessous des exemples assez classiques – avec les solutions! – qu'il m'a été donné d'enseigner<sup>3</sup>.

### 2.1 Des résultats corrects mais contre-intuitifs

**Exercice** (von Mises, 1939). *Le professeur parie qu'au moins deux de ses élèves sont nés le même jour (il n'y a pas de jumeaux dans sa classe). À partir de quel effectif a-t-il raison de faire ce pari ?*

*Réponse.* En supposant pour simplifier qu'il y a 365 jours par an et un effectif  $N \leq 365$ , une interprétation linéaire conduirait à dire qu'il faudrait une classe très surchargée de plus de 180 élèves. Mais le nombre d'arrangements de  $n$  dates distinctes parmi 365 est  $= 365 \cdot 364 \cdot (365 - N + 1)$  et la probabilité d'une coïncidence est donc  $P = 1 - \prod_{k=1}^N (1 - k/365)$  soit environ  $1 - \exp(-\frac{N^2}{2 \times 365})$  qui est  $> 1/2$  dès que  $N \geq 23$ , chiffre confirmé par un calcul exact.  $\square$

Cet exercice peut être traité en Première ou en Terminale (c'était déjà le cas lorsque j'étais élève). La difficulté n'est pas particulière aux Probabilités, mais aux mathématiques en général : il s'agit de bien identifier les données du problème (en l'occurrence l'univers des possibles).

—

**Exercice** (Selvin, 1975). *Vous êtes candidat à un jeu télévisé ; la voiture est derrière l'un des trois rideaux. Vous choisissez au hasard un rideau ; le présentateur (qui sait où se trouve la voiture) fait ouvrir un autre rideau derrière lequel se trouve une chèvre, et vous offre la possibilité de changer d'avis : que faites vous ?*

---

2. J'ai déjà rencontré un élève de Terminale qui avait peur de rentrer en classes de mathématiques « spéciales »...

3. Je suppose une connaissance de base de la théorie (disons jusqu'au niveau Licence).

*Réponse.* Beaucoup de gens soutiennent *mordicus* qu'il reste un chance sur deux, donc il n'y a pas d'intérêt particulier à changer. Mais cela suppose en fait que la probabilité initiale (1/3) change par le conditionnement (l'action du présentateur). Or justement, ici ce n'est pas le cas : en effet, si on utilise la méthode d'« inférence bayésienne » : l'hypothèse est  $H$  : « la voiture est derrière le premier rideau »,  $\mathbb{P}(H) = 1/3$ . L'événement supplémentaire est  $E$  : « le présentateur montre que la voiture ne se trouve pas derrière un autre rideau », on a toujours  $\mathbb{P}(E|H) = \mathbb{P}(E|H^c) = \frac{1}{2}$ , et donc les « côtes » sont inchangées :

$$\frac{\mathbb{P}(H|E)}{\mathbb{P}(H^c|E)} = \frac{\mathbb{P}(H)}{\mathbb{P}(H^c)} \cdot \frac{\mathbb{P}(E|H)}{\mathbb{P}(E|H^c)} = \frac{1}{2} \cdot 1 = \frac{1}{2}$$

Par conséquent  $\mathbb{P}(H|E)$  est toujours = 1/3, donc la voiture a deux chances sur trois d'être derrière le troisième rideau, et il est préférable de changer d'avis.  $\square$

La nature contre-intuitive du résultat provient ici de l'utilisation des probabilités conditionnelles (inférence bayésienne), thème souvent délicat pour les étudiants.

—

**Exercice.** *On attend le bus dans un trafic perturbé où les arrivées des bus suivent un processus de Poisson, avec en moyenne un bus toutes les  $T$  minutes. Trouver le temps moyen d'attente du bus.*

*Réponse.* La réponse naïve est  $T/2$  car il y a un bus toutes les  $T$  minutes. Ce serait vrai si les arrivées des bus étaient uniformément réparties, car l'attente suivrait un loi uniforme entre 0 et  $T$ . Mais le processus étant poissonnien, les intervalles entre les arrivées consécutives de bus suivent des v.a. i.i.d. exponentielles de paramètre  $\lambda = T$  (moyenne de chaque intervalle). La loi exponentielle étant sans mémoire, sachant qu'on arrive à un certain instant  $t$ , l'attente moyenne restante suit encore une loi exponentielle de même paramètre  $\lambda = T$ . Le temps moyen d'attente du bus est donc  $T$ , le double de ce qu'on obtient par un raisonnement naïf.  $\square$

Ce très joli résultat peut être vérifié expérimentalement (et pour les plus malchanceux d'entre nous, quotidiennement). Il n'est contre-intuitif qu'au premier abord.

—

**Exercice.** *Vous êtes (encore) candidat à un jeu télévisé; on vous présente deux enveloppes contenant des sommes d'argent, vous en choisissez une et prenez connaissance de son contenu. Vous avez la possibilité de garder celui-ci ou de choisir l'autre enveloppe. Que faites vous ?*

*Réponse.* <sup>4</sup> Là encore, on a tendance à répondre qu'il n'y a rien de particulier à

4. Merci à Aslan TCHAMKERTEN de m'avoir signalé cet exercice.

faire. Mais contrairement aux apparences, il existe une méthode (probabiliste !) pour faire son choix de sorte d'avoir raison avec une probabilité  $> 1/2$ . Notons  $x_1 < x_2$  les deux montants inconnus du candidat,  $X$  le montant observé (v.a. binaire avec  $\mathbb{P}(X = x_1) = \mathbb{P}(X = x_2) = 1/2$ ). Sans même aucune connaissance de la façon dont ont été choisies les deux montants, on tire au hasard un nombre réel  $Y = y$  selon une loi telle que la probabilité de se trouver dans un intervalle de longueur  $> 0$  est toujours  $> 0$  (par exemple, une v.a.  $Y$  gaussienne, indépendante de  $X$ ). Si le montant observé est  $\geq y$ , on le garde ; s'il est  $< y$ , on prend l'autre enveloppe (de montant  $X'$ ). La probabilité d'avoir raison est

$$\mathbb{P}(X1_{X \geq Y} + X'1_{X < Y} = x_2) = \mathbb{P}(X = x_1)\mathbb{P}(x_1 < Y) + \mathbb{P}(X = x_2)\mathbb{P}(x_2 \geq Y) = \frac{1+p}{2} > \frac{1}{2}$$

où on a noté  $p = \mathbb{P}(x_1 < Y \leq x_2) > 0$ . L'espérance de gain qui en résulte est

$$\mathbb{E}(X1_{X \geq Y} + X'1_{X < Y} = x_2) = x_1 \frac{1-p}{2} + x_2 \frac{1+p}{2} = \frac{x_1 + x_2}{2} + (x_2 - x_1) \frac{p}{2} > \frac{x_1 + x_2}{2} \quad \square$$

**Exercice.** *Même jeu télévisé, mais on sait à l'avance qu'un des montants de l'enveloppe est égal au double de l'autre. Quitte ou moitié/double ?*

*Réponse.* Disons  $x_1 = s$  et  $x_2 = 2s$ . Un raisonnement fallacieux consiste à dire que sachant  $X = x$ , l'espérance de l'autre montant  $\frac{2x+x/2}{2} > x$  et qu'il faut donc toujours choisir l'autre enveloppe. . . . C'est faux car sachant  $X = x$ ,  $X'$  devient une v.a. déterministe  $= 2x$  si  $x = s$  et  $= x/2$  si  $x = 2s$ , et  $\mathbb{E}(X'|X = x)$  a la même valeur ( $= 2x$  si  $x = s$  et  $= x/2$  si  $x = 2s$ ). On ne peut pas moyenniser ces deux valeurs. Une réponse possible est d'utiliser la méthode ci-dessus, qui assure une espérance de gain  $> 3s/2^5$ . □

Il est clair que dans ces deux exercices, la difficulté est également renforcée par le langage utilisé, en particulier la question vague "qui tue" : « que faites-vous ? ». On ne pourrait pas imaginer ce genre de question à l'écrit d'un concours, difficilement à l'oral.

## 2.2 Paradoxes de langage

L'interprétation du langage (utilisé dans l'énoncé) est souvent déterminant pour la résolution. Quelques exemples :

---

5. Pour les deux variantes du jeu des enveloppes, on peut faire l'hypothèse que les deux montants (ou la somme  $s$ ) ont été choisis selon une loi de probabilité connue (ou avec une contrainte, par exemple de montant minimum). Dans ce cas on peut affiner la méthode exposée pour maximiser l'espérance du gain, en moyenne sur cette loi.

**Exercice** (Bertrand, 1888). *Dans un cercle donné, quelle est la probabilité  $P$  qu'une corde du cercle choisie au hasard soit de longueur supérieure au côté d'un triangle équilatéral inscrit ?*

*Réponse.* Tout dépend de ce qu'on entend par « choisi au hasard ». Si on choisit une direction au hasard à l'aide d'un rayon du cercle et un point au hasard sur ce rayon pour définir la corde perpendiculaire en ce point à ce rayon, un côté du triangle équilatéral inscrit sera obtenu lorsque le point choisi est au milieu du rayon, donc  $P = \frac{1}{2}$ . Si ce sont les deux extrémités de la corde qui sont choisies au hasard, il s'agit de la probabilité pour qu'une extrémité ayant été choisie, considérant le triangle équilatéral inscrit de sommet cette extrémité, l'autre extrémité de la corde se trouve entre les deux autres sommets du triangle, d'où  $P = \frac{1}{3}$ . Si on choisit un point au hasard dans le cercle définissant une corde dont ce point est le milieu, la corde sera plus petite que le côté d'un triangle équilatéral inscrit si le point est choisi à l'extérieur du cercle inscrit à ce triangle, qui est de rayon moitié du rayon du cercle initial ; donc  $P = \frac{1}{4}$ . Si on choisit au hasard la longueur de cette corde entre 0 et deux fois le rayon, sachant que le côté d'un triangle équilatéral inscrit est  $\sqrt{3}$  fois le rayon, on trouve  $P = 1 - \frac{\sqrt{3}}{2}$ .  $\square$

Cet exemple montre qu'il y a parfois aucune hypothèse a priori préférable à une autre et qu'il *faut* préciser la façon dont un choix au hasard est fait.

**Exercice** (Gardner, 1954). *Votre voisin vous a dit qu'il avait deux enfants. Lorsque vous sonnez à sa porte, une fille ouvre. Quelle est la probabilité que l'autre enfant soit un garçon ?*

*Réponse.* Il manque ici beaucoup de données : Tout d'abord, il est entendu par défaut que les naissances sont indépendantes et qu'il y a une chance sur deux d'avoir une fille (ou un garçon). Bref, sans connaissance supplémentaire, on fait l'hypothèse d'équiprobabilité et d'indépendance (même si ce n'est pas forcément tout à fait vrai en pratique). Ce peut être une règle par défaut universelle pour les exercices.

Maintenant, si (disons) l'aînée est une fille, il y a une chance sur deux pour que le deuxième soit un garçon (les naissances sont supposées indépendantes). Mais si un des deux enfants est une fille, il y a deux chances sur trois pour que l'autre soit un garçon (les possibilités sont (G,F) (F,G), (F,F)). Disons que par défaut si on ne dit rien, c'est qu'on est dans le deuxième cas (on n'a pas spécifié l'ordre de naissance).

Il manque néanmoins une information sur la probabilité de qui répond à la porte. Disons par exemple (par défaut) que l'un ou l'autre des enfants, avec la même probabilité  $1/2$ , répond à la porte. On applique donc la méthode d'« inférence bayésienne » : l'hypothèse est  $H$  : « l'autre enfant est un garçon » (sachant

qu'il y a une fille) et donc  $\mathbb{P}(H) = 2/3$ . Mais il y a un événement supplémentaire  $E$  : « une fille ouvre » tel que  $\mathbb{P}(E|H) = 1/2$ ,  $\mathbb{P}(E|H^c) = 1$  et on cherche à déterminer  $\mathbb{P}(H|E)$  : alors

$$\frac{\mathbb{P}(H|E)}{\mathbb{P}(H^c|E)} = \frac{\mathbb{P}(H)}{\mathbb{P}(H^c)} \cdot \frac{\mathbb{P}(E|H)}{\mathbb{P}(E|H^c)} = 2 \cdot \frac{1}{2} = 1$$

d'où  $\mathbb{P}(H|E) = \frac{1}{2}$ . □

Ainsi, dans le cas d'hypothèses reflétant la plus grande méconnaissance des données, il se trouve que c'est la réponse naturelle qui est la bonne : « une chance sur deux, voyons ! ». On voit que le problème est une question d'interprétation mathématique du langage commun, il faut parfois préciser les hypothèses faites ou si c'est possible, adopter un principe de méconnaissance maximale par défaut.

D'autres énoncés de « paradoxe » restent encore sujets à polémiques et controverses aujourd'hui (paradoxe de la Belle au bois dormant, paradoxe de l'Apocalypse, etc.), faute de s'être mis d'accord pour clarifier leur interprétation mathématique.

Ces controverses n'ont pas pour origine les mathématiques elles-mêmes, seulement de l'*interprétation* qui est faite de l'énoncé. Comme ailleurs en sciences, il s'agit de faire preuve de la plus grande vigilance en essayant de poser des énoncés univoques. D'ailleurs, ne pourrait-on pas dire logiquement que s'il y avait réellement paradoxe de contenu en théorie des Probabilités, celle-ci n'étant que purement mathématique, il impliquerait paradoxe de *toutes* les mathématiques ?

### 3 Les probabilités : une théorie trop difficile ?

Depuis Kolmogorov et son approche axiomatique de la Théorie des Probabilités, il est bien reconnu que les outils mathématiques ainsi donnés aux concepts fondamentaux des probabilités sont puissants et efficaces. C'est pourquoi la plupart des livres et manuels sur ce sujet, dès le niveau de la Licence, obligent à se familiariser d'abord avec la théorie de la *mesure* et de l'*intégration* avant d'aborder le calcul de probabilités proprement dit<sup>6</sup>.

Le passage obligé par la théorie de la mesure et de l'intégration constitue évidemment une difficulté initiale importante, qui s'avère rebutante pour de

---

6. La théorie de la mesure la plus adaptée aux probabilités n'est d'ailleurs pas nécessairement celle préconisée par les mathématiciens « purs » (restreinte aux mesures de Radon sur les espaces topologiques localement compacts) mais plutôt celle de la mesure « abstraite » (avec le théorème d'unicité de Carathéodory) qui permet ensuite d'appliquer les résultats à des processus aléatoires généraux.



nombreux étudiants. L'étudiant (comme l'enseignant, d'ailleurs) doit alors se farcir ces « tribus » barbares, ces « clans » redoutés, ces «  $\sigma$ -algèbres » hermétiques, ces espaces boréliens, mesurables, probabilisables, probabilisés... On est assez loin des préoccupations quotidiennes de l'ingénieur, et la théorie de la mesure est souvent perçue – par les étudiants comme les enseignants! – comme ennuyeuse et trop longue.

À tel point qu'on se pose sérieusement la question si, pour être explicite sans perdre du temps en classes préparatoires, on ne va pas simplement abandonner les probabilités continues et se restreindre aux probabilités discrètes (tout ensemble discret est mesurable...). Ce serait dommage, car il est déjà prévu d'enseigner les densités de probabilité (gaussiennes, exponentielles, uniformes) en Terminale, en vue notamment d'aborder l'approximation gaussienne (pour ne pas dire le *théorème central limite*). Ces considérations constituent tout de même, comme son nom l'indique, un théorème central<sup>7</sup> pour tout ingénieur qui se respecte.

Car vue d'une École d'ingénieurs, quel est le but de l'enseignement des Probabilités? Il s'agit de rapidement bénéficier des outils de calcul de probabilités pour des besoins pratiques (sans pour cela nécessairement abandonner la saveur et l'intérêt d'une étude suffisamment rigoureuse des concepts). Dans les besoins pratiques, il y a naturellement l'hypothèse normale et les lois continues. Par ailleurs, l'exposé systématique de la théorie de la mesure et de l'intégration n'est peut-être pas un préliminaire indispensable pour commencer à « faire des probabilités » : les résultats utiles liés à la théorie de la mesure peuvent-ils être démontrés au fur et à mesure des besoins sans technicité excessive?

Voici résumées un certain nombre de pistes de solutions explorées pendant plus de dix ans dans mes cours, et exposées dans mon livre [4]. Si elles ne conduisent pas toujours à une rigueur parfaite (ce point devrait être améliorable), elles tendent en pratique à atteindre l'objectif premier : être rapidement capable de « faire des probabilités ».

### 3.1 Se débarrasser du $\Omega$

Dans de nombreux manuels, on se familiarise avec les probabilités en se référant à un univers de réalisations possibles souvent noté  $\Omega$ , et en manipulant à l'aide de diagrammes ensemblistes des probabilités des ensembles d'une tribu  $\mathcal{A}$  de parties de  $\Omega$ . C'est ainsi qu'on définit rapidement les probabilités conditionnelles par la formule  $\mathbb{P}(A|B) = \mathbb{P}(A \cap B) / \mathbb{P}(B)$ , ainsi que la notion d'événements indépendants : on dit que «  $A \perp B$  » si  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . Cela donne

7. Le terme *central limit theorem* (dû à G. Pólya) désigne à l'origine un théorème central de la limite, et non comme on le voit trop souvent un « théorème de la limite centrale » (d'ailleurs, la limite en question n'est pas spécialement centrale, mais normale).

lieu à de multiples exercices plus ou moins subtiles de probabilités « élémentaires » dont la résolution passe par des choix plus ou moins judicieux de l'espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ .

Mieux (ou pire) : en admettant ou en construisant un espace probabilisé produit (infini) permettant de modéliser la même expérience répétée une infinité dénombrable de fois, on peut très rapidement énoncer l'indépendance d'une infinité d'événements et démontrer les lemmes de Borel-Cantelli. C'est ainsi qu'on prouve, par exemple, que tout événement probable, même très peu probable, se réalisera presque sûrement une infinité de fois.

Bien que ce pan du cours de probabilités a le mérite de mettre en perspective des éléments de théorie des ensembles, il est critiquable : Borel-Cantelli est déjà d'une compréhension délicate<sup>8</sup> mais surtout, certaines notions – même les plus simples – apparaissent un peu artificielles : pourquoi tel choix de  $\Omega$  serait-il plus naturel que tel autre pour résoudre tel ou tel exercice ? que déduire du fait que tirer un as ou un pique d'un jeu de cartes sont deux événements indépendants ?

Je soutiens qu'il est possible, et même salutaire à un niveau relativement simple, de se passer du fameux  $\Omega$ . En effet, il est souvent plus intuitif et plus commode d'analyser un problème de calcul de probabilités en identifiant d'abord les *quantités* variables aléatoires, puis en raisonnant sur leurs distributions de probabilité de ces variables aléatoires<sup>9</sup>. Comme on va le voir maintenant, nul besoin de  $\Omega$  pour cela.

### 3.2 Tout axer sur les « v.a. »

On définit classiquement, selon l'approche de la théorie de la mesure, une *variable aléatoire* (v.a.)  $X$  comme une application mesurable de  $\Omega$  vers un espace probabilisable (typiquement  $\mathbb{R}$  dans le cas d'une v.a. réelle). On écrit donc  $X(\omega)$  où  $\omega \in \Omega$ . Cependant, au regard des applications pratiques, le rôle du  $\omega$  dans l'espace probabilisé  $\Omega$  ne correspond à rien de bien palpable et reste donc très mystérieux. Par exemple, supposons que l'on modélise des échantillons d'un bruit thermique par une suite de v.a. normales corrélées : à quoi correspond  $\omega$  ?

Ce qui est important en pratique, c'est que l'utilisation d'une variable aléatoire  $X$  est entièrement déterminée par la donnée des probabilités  $\mathbb{P}\{X \in A\}$ , et l'utilisation de la lettre  $X$  ne constituera finalement qu'une commodité de notation. Cela suffit largement pour « faire des probabilités », tout du moins au niveau souhaité (jusqu'à la loi des grands nombres et même les processus

---

8. La notion de réalisation une infinité de fois presque sûrement fait appel à celle de limite supérieure d'ensembles.

9. On peut toujours, en exercice, retrouver les définitions élémentaires (ensemblistes) à l'aide de variables aléatoires définies par des fonctions indicatrices.

aléatoires). Cette approche, qui est celle suivie par Cramér [2] dans son ouvrage de référence, est plus simple et plus directe que celle utilisant le mystérieux  $\omega$ .

Une variable (ou vecteur) aléatoire  $X$  sera, dans cette approche, *défini(e)* par la donnée des *probabilités*  $\mathbb{P}\{X \in A\}$  de sous-ensembles  $A$  de  $\mathcal{X}$  appelés *événements*. Le fait qu'un résultat satisfait à une propriété donnée correspond à la réalisation d'un certain événement.

### 3.3 Isoler les cas discret et continu

Si une variable aléatoire ne prend pas, à proprement parler, une ou plusieurs valeurs numériques (par exemple un résultat qualitatif comme pile ou face, ou la couleur du ciel), on peut généralement adopter un codage numérique (comme 0 pour pile et 1 pour face, ou le code RGB d'une couleur). On se ramène ainsi au cas de variables aléatoires *réelles* (en abrégé v.a.r., dans  $\mathbb{R}$ ) ou de vecteurs aléatoires réels<sup>10</sup> (v.a.r. dans  $\mathbb{R}^n$ ). Bien que l'on élimine ainsi le cas d'un espace d'arrivée fonctionnel, cela couvre presque tous les cas donnant lieu à des modélisations pratiques, ce qui est bien suffisant pour démarrer. De la sorte, les mesures considérées sont concrètes (non abstraites) selon la terminologie traditionnelle, on peut même se restreindre à considérer des événements boréliens (la seule tribu à considérer est celle engendrée par les intervalles/pavés)<sup>11</sup>.

Une variable (ou vecteur) aléatoire peut être *discrète* (à valeurs dans un ensemble fini ou infini dénombrable) ou *continue* (à valeurs dans un continuum de valeurs comme un intervalle ou un domaine de l'espace). Il y a aussi d'autres possibilités : variables « dégénérées » ou « mixtes » (discrètes/continues). Dans un premier temps, il est facile de dégager la partie *discrète* d'une v.a. quelconque (ensemble au plus dénombrable de masses ponctuelles dites de Dirac). Il est plus difficile, mais possible (en admettant le théorème de Radon-Nikodym) d'isoler sa partie (absolument) continue (c'est-à-dire définie par une densité). Il reste alors une partie « dégénérée », c'est-à-dire entièrement concentré sur un volume de mesure nulle.

Il est difficile d'imaginer un seul exemple d'une v.a.r. dégénérée utile dans la pratique des mathématiques appliquées. Sa fonction de répartition  $P(X \leq x)$  serait :

---

10. Le cas d'une v.a.  $X$  *complexe* est un cas particulier d'un couple aléatoire  $(\operatorname{Re} X, \operatorname{Im} X) \in \mathbb{R}^2$  où on identifie  $\mathbb{C}$  à  $\mathbb{R}^2$ .

11. Notons au passage que R. Solovay a montré qu'on ne peut obtenir aucun exemple explicite d'ensemble non mesurable sans faire intervenir l'axiome du choix général, et qu'on peut donc parfaitement imposer que tout ensemble et toute fonction est mesurable (au sens de Lebesgue) en restant compatible avec l'axiome du choix dépendant et donc toutes les mathématiques utiles pour les applications. La notion même de mesurabilité n'a alors plus d'intérêt et peut disparaître de l'enseignement.

- 1° croissante de 0 à 1, comme toute fonction de répartition ;
- 2° continue partout, car il n'y a pas de masse ponctuelle ;
- 3° presque partout dérivable de dérivée nulle, puisque concentrée sur un ensemble de mesure nulle (et donc en ce sens presque partout constante) !

On peut donc bien parler de cas pathologique<sup>12</sup>. En résumé, on peut fort bien, sans trop de perte de généralité, se limiter exclusivement aux v.a. discrètes ou continues.

### 3.4 Tout axer sur les distributions

On peut d'ailleurs adopter un formalisme unifié pour étudier en même temps les propriétés des cas discret et continu, ce qui permet de rendre les calculs généraux plus agréables. Ce formalisme unifié utilise la notion centrale de *distribution de probabilité*  $p(x)$  : par exemple, la probabilité que  $X \in A$  s'écrit sous la forme :

$$\mathbb{P}\{X \in A\} = \int_{x \in A} p(x)$$

où la sommation peut être discrète (pour une variable  $X$  discrète où  $p(x) = \mathbb{P}(X = x)$ ) ou continue (sommation intégrale pour une variable  $X$  continue) ; dans ce dernier cas,  $p(x)$  désigne une densité de probabilité. Cette formule est d'un grand intérêt pratique lorsqu'on veut traiter les cas discret et continu ensemble sans devoir tout réécrire deux fois.

Chaque fois que c'est possible, les notions importantes (comme les changements de variable, l'indépendance et le conditionnement) peuvent être exprimées directement sur les distributions de probabilité, plutôt que de faire le détour – comme c'est souvent l'usage – par les fonctions de répartition ou les probabilités d'ensembles. Le formalisme obtenu est plus direct et plus simple.

Par exemple, si l'on tenait à présenter les v.a. comme des applications mesurables définies sur le mystérieux  $\Omega$ , l'espérance est une intégrale (de Lebesgue-Stieltjes, par exemple) par rapport à la mesure de la variable  $\omega$  :

$$\mathbb{E}(X) = \int X(\omega) dP(\omega).$$

Ici, avec le formalisme basé sur la distribution  $p(x)$ , l'espérance  $\mathbb{E}(X)$  est définie par la formule utile pour les calculs pratiques :

$$\mathbb{E}(X) = \int_{x \in \mathcal{X}} x \cdot p(x)$$

---

12. Un exemple classique est appelé *escalier du diable*. Les v.a. singulières en dimension  $n$  sont en revanche importantes à considérer ; cela correspond en pratique à des v.a.r. concentrés sur une variété de dimension  $< n$ .

(sommation discrète ou intégrale, suivant le cas). Les deux formules sont équivalentes, mais la dernière exprime clairement  $\mathbb{E}(X)$  comme une valeur moyenne de  $X$ , c'est-à-dire une sommation des valeurs de  $x$  pondérées par la distribution de probabilité  $p(x)$ .

### 3.5 Tout le reste

On dispose ainsi de tous les outils indispensables pour aborder les grands théorèmes du calcul des probabilités : la *loi (faible) des grands nombres* et la convergence vers la loi gaussienne ou *théorème central limite*.

Avec la même approche, une introduction aux *processus aléatoires* est possible, où l'on expose les notions de *stationnarité* et d'*ergodicité* en liaison avec la *loi forte* des grands nombres<sup>13</sup>. Les notions plus avancées – et plus difficiles, comme celles résultant des différentes définitions de convergence (étroite ou presque sûre), de la construction de Kolmogorov ou des lois du tout ou rien (comme Borel-Cantelli) – sont ainsi exposées à la fin et non au début.

Par ailleurs, l'espace des v.a.r. de carré intégrable ( $L^2$ ) est naturellement muni d'un produit scalaire, pour lequel c'est un espace de Hilbert ; mais on peut se rassurer : la notion d'espace de Hilbert (essentiellement son caractère complet) n'est utile que pour résoudre des problèmes d'approximation aux moindres carrés et peut être passé sous silence dans une première approche.

Enfin, ne fantasmons pas : les mouvements browniens, les martingales, le calcul différentiel stochastique, l'intégrale de Itô ... ne seront accessibles qu'à un niveau bien supérieur. Les probabilités enseignées en CPGE resteront vraisemblablement des apprentissages fondamentaux faisant suite au lycée (voir annexe A de ce document).

## 4 Les probabilités : un monde à part ?

L'étude des probabilités passe par l'introduction d'un nouveau langage propre qu'il s'agit d'assimiler. Comme tous les langages d'une spécialité mathématique, celui des probabilités utilise un vocabulaire spécifique souvent constitué de mots courants auxquels on donne un sens précis : *expérience, événement, probabilité, espérance, moments, lois...*

Il n'y a là, me semble-t-il, rien de très différent de ce qui se passe dans d'autres domaines mathématiques, si ce n'est que ce vocabulaire est peut être plus proche du langage courant et donne lieu à des raisonnements plus intuitifs. Qu'on en juge en comparant aux notions classiques de *corps, anneau, idéal*,

---

13. Une première introduction aux chaînes de Markov est également possible (voir l'exposé de Roger MANSUY).

*trace, base, noyau, relèvement, polarisation...* qui sont clairement plus éloignées du sens courant ! Ce rapport des probabilités avec le langage humain constitue même une force de la théorie qui est rendue ainsi plus accessible à l'ingénieur pour d'innombrables applications importantes.

C'est aussi ce langage et ces applications qui font la richesse des probabilités et qui la rendent incontournable dans les enseignements. Il faut souligner d'ailleurs qu'actuellement les Grandes Écoles sont un peu pris à la gorge pour « faire passer » en quelques semaines trois années de programme en probabilités<sup>14</sup>.

L'absence des probabilités en CPGE suscite en effet depuis très longtemps une incompréhension unanime et notoire des chercheurs, y compris en sciences humaines. Le réel apprentissage des mathématiques de nos étudiants commençant en Sup et s'arrêtant au mieux à la fin de la première année de Grande École, les CPGE sont responsables d'une partie très importante de la formation d'un ingénieur ; raison de plus pour ne pas y négliger l'enseignement des probabilités... Il n'y a en effet aucune raison objective de penser que les commerciaux – qui ont des cours de probabilités en classes préparatoires – soient amenés à utiliser plus souvent la loi des grands nombres que les ingénieurs, bien au contraire !

On peut malgré tout craindre une sorte d'isolement. Les probabilités ne sont-elles pas ce cheveu sur la soupe, ce pan des mathématiques appliquées (pour ne pas dire impures) qui étaient si souvent absentes des traités de mathématiques ? Ne pourrait-on craindre une scission entre probabilités et le reste des mathématiques ?

Tout dépend peut-être de la façon d'enseigner. L'expérience du mathématicien, en tout cas, démontre clairement que les probabilités sont entièrement imbriquées dans le reste des mathématiques : analyse, algèbre, arithmétique, géométrie... Je présente ci-dessus deux exemples simples dans chacun de ces domaines. Comme l'écrivait Cédric VILLANI dans une conférence LIESSE il y a un an, « les théorèmes ne se mettent pas dans des cases ».

## 4.1 Deux exemples en analyse

**Théorème** (Weierstraß). *Toute fonction continue sur un segment est limite uniforme d'une suite de polynômes.*

*Démonstration probabiliste (Bernstein, 1912).* On se ramène à une fonction  $f$  continue sur  $[0, 1]$  ; soit  $X$  une v.a. binomiale de paramètre  $p$  et de longueur  $n$  et

---

14. On peut d'ailleurs en dire autant de l'intégration et de l'analyse de Fourier, ainsi que de l'analyse complexe qui est carrément abandonnée à Télécom ParisTech.

posons

$$B_n(p) = \mathbb{E}\left(f\left(\frac{X}{n}\right)\right).$$

Clairement  $B_n(p) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f\left(\frac{k}{n}\right)$  est un polynôme en  $p$ . Pour  $\varepsilon > 0$  donné, majorons la différence  $|B_n(p) - f(p)|$  :

$$\begin{aligned} |B_n(p) - f(p)| &= \left| \mathbb{E}\left(f\left(\frac{X}{n}\right) - f(p)\right) \right| \leq \mathbb{E}\left(\left|f\left(\frac{X}{n}\right) - f(p)\right|\right) \\ &= \mathbb{E}\left(\left|f\left(\frac{X}{n}\right) - f(p)\right| \mathbf{1}_{\left|\frac{X}{n} - p\right| < \varepsilon}\right) + \mathbb{E}\left(\left|f\left(\frac{X}{n}\right) - f(p)\right| \mathbf{1}_{\left|\frac{X}{n} - p\right| \geq \varepsilon}\right) \\ &\leq \max_{\left|\frac{k}{n} - p\right| < \varepsilon} \left|f\left(\frac{k}{n}\right) - f(p)\right| + 2\|f\|_\infty \mathbb{P}\left(\left|\frac{X}{n} - p\right| \geq \varepsilon\right) \end{aligned}$$

où on a utilisé le fait que  $f$  est bornée. La fonction  $f$  étant continue sur le compact  $[0, 1]$ , elle y est uniformément continue, et le premier terme peut être rendu arbitrairement petit, indépendamment de  $p$ . Par l'inégalité de Bienaymé-Chebyshev, le deuxième terme est majoré par  $2\|f\|_\infty \frac{p(1-p)}{n\varepsilon^2} \leq \frac{\|f\|_\infty}{2n\varepsilon^2}$  qui tend uniformément vers zéro. Ainsi  $B_n \rightarrow f$  uniformément sur  $[0, 1]$ .  $\square$

**Théorème** (formule de Stirling).

$$\frac{n^{n+1/2}}{n!} e^{-n} \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}}.$$

*Démonstration probabiliste.* Soit  $(X_n)_n$  une suite i.i.d. de v.a. poissoniennes de paramètre 1 (égal à la moyenne et la variance). La v.a.  $S_n = X_1 + X_2 + \dots + X_n$  suit une loi de Poisson de paramètre  $n$ , et par le théorème central limite,  $\frac{S_n - n}{\sqrt{n}}$  converge en loi vers  $Y \sim \mathcal{N}(0, 1)$ . Comme la fonction  $x \mapsto x^-$  est continue,

$$\mathbb{E}\left\{\left(\frac{S_n - n}{\sqrt{n}}\right)^-\right\} \xrightarrow{n \rightarrow \infty} \mathbb{E}(Y^-) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 y e^{-y^2/2} dx = \frac{1}{\sqrt{2\pi}}$$

Mais

$$\mathbb{E}\left\{\left(\frac{S_n - n}{\sqrt{n}}\right)^-\right\} = \sum_{k=0}^n \binom{n-k}{\sqrt{n}} \frac{n^k}{k!} e^{-n} = \sum_{k=0}^n \frac{n^{k+1/2}}{k!} e^{-n} - \sum_{k=1}^n \frac{n^{k-1/2}}{(k-1)!} e^{-n}$$

Tous les termes s'éliminent en cascade sauf  $\frac{n^{n+1/2}}{n!} e^{-n}$ .  $\square$

## 4.2 Deux exemples en algèbre

Le premier en combinatoire, l'autre en algèbre linéaire.

**Théorème** (Sperner, 1928). *On peut trouver au plus*

$$\max_k \binom{n}{k} = \binom{n}{\lfloor n/2 \rfloor}$$

*parties de  $\{1, 2, \dots, n\}$  dont aucune n'est contenue dans aucune autre.*

*Démonstration probabiliste.* Soit  $\mathcal{A}$  une telle famille de parties et  $\mathcal{C}_\sigma$  la chaîne :

$$\mathcal{C}_\sigma = \{\emptyset, \{\sigma_1\}, \{\sigma_1, \sigma_2\}, \dots, \{1, 2, \dots, n\}\}$$

où la permutation  $\sigma$  de  $\{1, 2, \dots, n\}$  est choisie au hasard suivant une distribution uniforme. Soit  $X$  le nombre de parties de  $\mathcal{A}$  dans la chaîne  $\mathcal{C}_\sigma$  : clairement  $X \in \{0, 1\}$  sinon il y aurait au moins deux parties dans  $\mathcal{A}$  qui seraient incluses l'une dans l'autre. Or

$$X = |\mathcal{A} \cap \mathcal{C}_\sigma| = \sum_{A \in \mathcal{A}} \mathbf{1}_{A \in \mathcal{C}_\sigma}$$

d'où

$$\mathbb{E}(X) = \sum_{A \in \mathcal{A}} \mathbb{E}(\mathbf{1}_{A \in \mathcal{C}_\sigma}) = \sum_{A \in \mathcal{A}} \mathbb{P}(A \in \mathcal{C}_\sigma)$$

et puisque  $\mathcal{C}_\sigma$  ne contient qu'un seul ensemble de cardinal donné  $|A|$ , qui est choisi au hasard parmi toutes les  $\binom{n}{|A|}$  parties de même cardinal  $= |A|$ ,

$$\mathbb{P}(A \in \mathcal{C}_\sigma) = \frac{1}{\binom{n}{|A|}}.$$

Finalement, puisque  $X \leq 1$ ,

$$\frac{|\mathcal{A}|}{\max_k \binom{n}{k}} \leq \sum_{A \in \mathcal{A}} \frac{1}{\binom{n}{|A|}} = \mathbb{E}(X) \leq 1. \quad \square$$

**Théorème** (Ky Fan, 1950). *Pour toutes matrices  $A, B \in S_n^{++}(\mathbb{R})$  (symétriques définies positives) et tous réels positifs  $\lambda, \mu$  tels que  $\lambda + \mu = 1$ ,*

$$\det(\lambda A + \mu B) \geq (\det A)^\lambda (\det B)^\mu.$$

*Démonstration probabiliste (Cover, Thomas, 1988).* Soit  $X_0$  et  $X_1$  deux vecteurs gaussiens centrés de matrices de covariance respectives  $A$  et  $B$ ,  $\Theta$  une v.a. binaire à valeurs dans  $\{0, 1\}$ , indépendante de  $(X_0, X_1)$ , de loi de Bernoulli  $(\lambda, \mu)$ . La matrice de covariance du vecteur aléatoire  $Y = X_\Theta$  est

$$\mathbb{E}({}^t Y Y) = \mathbb{E}\{\mathbb{E}({}^t X_\Theta X_\Theta | \Theta)\} = \lambda \mathbb{E}({}^t X_0 X_0) + \mu \mathbb{E}({}^t X_1 X_1) = \lambda A + \mu B.$$



On utilise des résultats sur l'entropie de vecteurs aléatoires définis par des densités [3]. À matrice de covariance fixée, l'entropie est maximale pour une loi normale :

$$h(Y) \leq h(Z)$$

où  $Z$  suit une loi normale de matrice de covariance  $= \lambda A + \mu B$ . Par ailleurs, en appliquant l'inégalité de l'information :

$$h(Y) \geq h(Y|\Theta) = \lambda h(X_0) + \mu h(X_1).$$

L'entropie d'un vecteur gaussien  $X$  de densité  $f_X$  et de matrice de covariance  $C$  se calcule explicitement :

$$h(X) = - \int f_X \ln f_X = \frac{1}{2} \mathbb{E}(\mathop{\text{tr}}\limits^t X C^{-1} X) + \ln \sqrt{(2\pi)^n \det C} = \frac{n}{2} \ln(2\pi e) + \frac{1}{2} \ln \det C.$$

On remplace alors les entropies par leurs valeurs dans l'inégalité  $\lambda h(X_0) + \mu h(X_1) \leq h(Z)$ . □

### 4.3 Deux exemples en arithmétique

Commençons par un résultat célèbre de Hardy et Ramanujan, qui établit que l'ordre de grandeur de  $\omega(n)$ , le nombre de diviseurs premiers distincts de  $n$ , est  $\ln \ln n$ . Ainsi, un nombre choisi au hasard entre 1000 et 5000000000 n'aura usuellement que 2 ou 3 facteurs premiers distincts <sup>15</sup>.

**Théorème** (Hardy & Ramanujan, 1920). *Pour tout  $\varepsilon > 0$ ,*

$$\frac{1}{n} \left| \left\{ N \leq n ; (1 - \varepsilon) \ln \ln N < \omega(N) < (1 + \varepsilon) \ln \ln N \right\} \right| \xrightarrow{n \rightarrow \infty} 1.$$

Hardy avait apparemment une sainte horreur des probabilités (jugées trop vagues) et pour les mathématiques appliquées en general. Mais Turan a grandement simplifié la preuve de Hardy-Ramanujan en interprétant leur résultat comme l'énoncé d'une loi des grands nombres (il existe d'ailleurs un théorème central limite qui affine le résultat, étudié par Erdős et Kac en 1939). Hardy et Wright, dans leur célèbre traité de théorie des nombres, ont repris la preuve de Turan ci-dessous... en prenant soin d'effacer toute trace des probabilités!

---

15. Hardy aimait à citer l'anecdote suivante sur Ramanujan : « *I remember once going to see [Ramanujan] when he was lying ill at Putney. I had ridden in taxi cab number 1729 and remarked that the number seemed to me rather a dull one, and that I hoped it was not an unfavorable omen. "No, Hardy, no Hardy," he replied, "it is a very interesting number; it is the smallest number expressible as the sum of two cubes in two different ways.* » Selon Hardy, un nombre tel que  $1729 = 7 \cdot 13 \cdot 19$ ,  $\omega(1729) = 3$ , est assez typique et donc « terne ». La réponse de Ramanujan est d'un tout autre registre :  $1729 = 1^3 + 12^3 = 9^3 + 10^3 \dots$

*Démonstration probabiliste (Turan, 1934).* Soit  $N$  une variable aléatoire entière uniforme sur  $\{1, 2, \dots, n\}$  où  $n \geq 2$ . Sa factorisation en nombre premiers s'écrit :  $N = \prod_p p^{X_p}$ , et on a  $\omega(N) = \sum_{p \leq n} X_p^*$  où  $X_p^* = \min(X_p, 1)$ . Puisque qu'il y a  $\lfloor \frac{n}{p} \rfloor$  multiples de  $p$  inférieurs ou égaux à  $n$ ,  $X_p^*$  suit une loi de Bernoulli de paramètre  $P(X_p^* = 1) = P(X_p \geq 1) = \frac{1}{n} \lfloor \frac{n}{p} \rfloor$ . Sachant que <sup>16</sup>  $\sum_{p \leq n} \frac{1}{p} = \ln \ln n + O(1)$ , il vient  $E(\omega(N)) = \sum_p E(X_p^*) = \sum_p \frac{1}{n} \lfloor \frac{n}{p} \rfloor = \ln \ln n + O(1)$ , car enlever les crochets revient à commettre une erreur  $\leq \frac{1}{n} \sum_{p \leq n} 1 \leq 1$ . De plus, grâce à l'encadrement  $\frac{1}{k} - \frac{1}{n} \leq \frac{1}{n} \lfloor \frac{n}{k} \rfloor \leq \frac{1}{k}$  :

$$\begin{aligned} \text{Var}(\omega(N)) &= \sum_{p \leq n} E(X_p^*) - E^2(X_p^*) + \sum_{\substack{p, p' \leq n \\ p \neq p'}} E(X_p^* X_{p'}^*) - E(X_p^*) E(X_{p'}^*) \\ &= \sum_{p \leq n} \frac{1}{n} \lfloor \frac{n}{p} \rfloor - \left( \frac{1}{n} \lfloor \frac{n}{p} \rfloor \right)^2 + \sum_{\substack{p, p' \leq n \\ p \neq p'}} \frac{1}{n} \lfloor \frac{n}{pp'} \rfloor - \frac{1}{n} \lfloor \frac{n}{p} \rfloor \frac{1}{n} \lfloor \frac{n}{p'} \rfloor \\ &\leq \sum_{p \leq n} \frac{1}{p} + \sum_{p, p' \leq n} \frac{1}{pp'} - \left( \frac{1}{p} - \frac{1}{n} \right) \left( \frac{1}{p'} - \frac{1}{n} \right) = O(\ln \ln n). \end{aligned}$$

Par l'inégalité de Bienaymé-Chebyshev

$$\mathbb{P}\left(\left| \frac{\omega(N) - E(\omega(N))}{\ln \ln n} \right| \geq \varepsilon\right) \leq \frac{\text{Var}(\omega(N))}{(\ln \ln n)^2 \varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

Il est facile de voir qu'on peut remplacer  $E(\omega(N))$  par  $\ln \ln n$ , et même par  $\ln \ln N$ , car  $\ln \ln n - 1 < \ln \ln N \leq \ln \ln n$  pour tout  $N > n^{1/e}$ , le reste donnant une contribution  $\leq \mathbb{P}(N \leq n^{1/e}) \rightarrow 0$ .  $\square$

En 1948, Shannon utilise un argument de moyenne d'ensemble pour prouver l'existence (sans preuve constructive) d'au moins un « bon » code correcteur d'erreurs pour les communications numériques arbitrairement fiables [3]. Le même argument peut être utilisé pour prouver des résultats plus élémentaires. Par exemple :

**Théorème (Erdős, 1965).** *Tout ensemble  $A = \{a_1, a_2, \dots, a_n\}$  de  $n$  entiers non nuls contient  $> n/3$  entiers  $\{a_{i_k}\}_k$  tels que  $a_{i_k} + a_{i_l} \neq a_{i_m}$  pour tous  $k, l, m$ .*

*Démonstration probabiliste.* Soit  $p$  un nombre premier de la forme  $p = 3k + 2$ , plus grand <sup>17</sup> que tous les  $2|a_i|$  et soit  $X$  un v.a. uniforme à valeurs dans  $\mathbb{Z}_p^*$

16. Voir [1] pour une preuve « probabiliste ».

17. Il y en a bien une infinité de nombres premiers de la forme  $3k + 2$  : car s'il n'y en avait qu'un nombre fini  $p_1, p_2, \dots, p_N$ , le nombre impair  $M = 6p_1 p_2 \dots p_N - 1$  n'aurait que des facteurs premiers impairs distincts des  $p_i$ , donc de la forme  $3k + 1$ , d'où  $M = -1 \pmod 3 = 1 \pmod 3$ , ce qui est impossible.

(entiers non nuls modulo  $p$ ). Puisque  $a_i \neq 0 \pmod p$ , la v.a.  $X_i = a_i \cdot X \pmod p$  à valeurs dans  $\mathbb{Z}_p^*$  suit aussi une loi uniforme. Soit  $\mathbf{B}$  le sous-ensemble aléatoire de  $A$  constitué des entiers  $a_i$  tels que  $X_i \in \{k+1, k+2, \dots, 2k+1\}$ . Sa taille moyenne est

$$\mathbb{E}(|\mathbf{B}|) = \mathbb{E}\left(\sum_{i=1}^n 1_{k < X_i \leq 2k+1}\right) = \sum_{i=1}^n \mathbb{P}(k < X_i \leq 2k+1) = \sum_{i=1}^n \frac{k+1}{p-1} = \frac{n(k+1)}{3k+1} > \frac{n}{3}$$

donc il existe au moins une valeur  $X = x$  conduisant à un ensemble  $B$  de taille  $> n/3$ . Les éléments  $a_i \in B$  sont ceux pour lesquels  $a_i x \pmod p \in \{k+1, k+2, \dots, 2k+1\}$ . Si on avait  $a_{i_k} + a_{i_l} = a_{i_m}$  pour trois entiers de  $B$ , en multipliant par  $x$  modulo  $p$  on trouverait deux entiers  $\in \{k+1, k+2, \dots, 2k+1\}$  de somme (modulo  $3k+2$ ) égale à un autre entier  $\in \{k+1, k+2, \dots, 2k+1\}$ , ce qui est impossible.  $\square$

#### 4.4 Deux exemples en géométrie

**Théorème** (Gram, 1874). Soit  $\alpha_i$  ( $0 \leq i \leq n$ ) la somme des angles solides intérieurs des  $i$ -faces<sup>18</sup> d'un polyèdre convexe en  $n$  dimensions. Alors

$$\sum_{i=0}^n (-1)^i \alpha_i = 0.$$

En particulier ( $n = 2$ ), on retrouve que la somme des angles d'un polygone à  $N$  côtés dans le plan est  $\alpha_0 = \alpha_1 - \alpha_2 = \frac{N}{2} - 1$ , soit  $(N-2)\pi$  radians.

*Démonstration probabiliste* (Welzl, 1994). Soit  $P$  le polyèdre en question,  $f_i$  son nombre de  $i$ -faces pour  $i = 0$  à  $n$  (ainsi  $f_n = 1$  et  $\alpha_{n-1} = f_{n-1}/2$ ).

Soit  $H$  un hyperplan dont la direction est choisie au hasard (le vecteur normal à  $H$  est choisi comme un point de la sphère  $S^n$  selon une distribution uniforme) et considérons le projeté orthogonal du polyèdre  $P$  sur  $H$ ; c'est un autre polyèdre  $P'$  à  $n-1$  dimensions. Soit  $F'_i$  son nombre de  $i$ -faces ( $i = 0$  à  $n-1$ ). La loi de la direction de projection étant uniforme, la probabilité qu'une  $i$ -face de  $P$  ( $i < n-1$ ) ne soit pas projetée sur une  $i$ -face de  $P'$  est égale au double de l'angle solide intérieur. En sommant les probabilités complémentaires pour chacune des  $f_i$   $i$ -faces de  $P$ , on trouve le nombre moyen de  $i$ -faces de  $P'$  :

$$\mathbb{E}(F'_i) = f_i - 2\alpha_i \quad (i \leq n-2)$$

18. Les 0-faces sont les sommets, les 1-faces sont les arêtes, etc. ; la  $n$ -face est l'hyper-volume du polyèdre. L'angle solide est supposé normalisé à 1 (mesure de l'angle total). Ainsi  $\alpha_n = 1$ , et  $\alpha_{n-1}$  est égal à la moitié du nombre d'hyperplans qui délimitent le polyèdre.

La relation d'Euler-Poincaré pour  $P'$  s'écrit  $\sum_{i=0}^{n-1} (-1)^i F'_i = 1$  et celle pour  $P$  s'écrit  $\sum_{i=0}^n (-1)^i f_i = 1$ . Puisque  $F'_{n-1} = 1$ ,  $f_n = 1$  et  $\alpha_{n-1} = f_{n-1}/2$  :

$$\begin{aligned} 0 &= \sum_{i=0}^{n-1} (-1)^i \mathbb{E}(F'_i) - 1 = \sum_{i=0}^{n-2} (-1)^i (f_i - 2\alpha_i) + (-1)^{n-1} - 1 \\ &= \sum_{i=0}^{n-1} (-1)^i (f_i - 2\alpha_i) + (-1)^{n-1} - \sum_{i=0}^n (-1)^i f_i = -2 \sum_{i=0}^{n-1} (-1)^i \alpha_i. \quad \square \end{aligned}$$

**Théorème** (Zubkov, 1979). *Dans l'espace  $\mathbb{R}^n$  de base canonique  $(e_1, e_2, \dots, e_n)$ , soit  $C = [0, 1]^n$  l'hypercube unité de  $\mathbb{R}^n$ ,  $H_t$  l'hyperplan d'équation  $x_1 + x_2 + \dots + x_n = t$ , et  $S$  le simplexe défini par l'enveloppe convexe des  $n + 1$  points  $0$  et  $e_1, 2e_2, \dots, ne_n$ . Alors les volumes  $(n - 1)$ -dimensionnels suivants sont égaux :*

$$\text{vol}(C \cap H_t) = \text{vol}(S \cap H_t).$$

*Démonstration probabiliste.* Soit  $X_1, X_2, \dots, X_n$  des v.a. i.i.d. exponentielles de paramètre  $\lambda > 0$ , réordonnons les  $X_i$  par ordre croissant :  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  et posons  $(Y_1, Y_2, \dots, Y_n) = (X_{(1)}, X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(n-1)})$ . La loi exponentielle étant sans mémoire :  $\mathbb{P}(X_i \geq t + x | X_i \geq t) = \mathbb{P}(X_i \geq x)$ , le vecteur  $(X_{(2)} - X_{(1)}, X_{(3)} - X_{(1)}, \dots, X_{(n)} - X_{(1)})$  suit la même loi que le vecteur ordonné de  $n - 1$  v.a. exponentielles i.i.d. de paramètre  $\lambda$ , indépendamment de  $X_{(1)}$ . De proche en proche, on voit que  $Y_i$  est indépendant de  $Y_1, Y_2, \dots, Y_{i-1}$  et que  $\mathbb{P}(Y_i \geq y) = \mathbb{P}(X_{(i)} \geq X_{(i-1)} + y) = \mathbb{P}(\min X_j \geq y)$  où le minimum porte sur les  $n - i + 1$  v.a. exponentielles i.i.d. restantes de paramètre  $\lambda$ , d'où  $\mathbb{P}(Y_i \geq y) = e^{-(n-i+1)\lambda y}$  :  $(Y_1, Y_2, \dots, Y_n)$  suit donc la même loi que  $(\frac{X_n}{n}, \dots, \frac{X_2}{2}, X_1)$  et

$$\mathbb{P}(\forall i, X_i \leq 1) = \mathbb{P}(X_{(n)} \leq 1) = \mathbb{P}(Y_1 + Y_2 + \dots + Y_n \leq 1) = \mathbb{P}\left(\frac{X_n}{n} + \frac{X_{n-1}}{n-1} + \dots + X_1 \leq 1\right)$$

Cela se réécrit (en divisant par  $\lambda^n$ ) :

$$\int_C e^{-\lambda(x_1 + \dots + x_n)} dx_1 \dots dx_n = \int_S e^{-\lambda(x_1 + \dots + x_n)} dx_1 \dots dx_n$$

ou encore

$$\int_0^{+\infty} \text{vol}(C \cap H_t) e^{-\lambda t} dt = \int_0^{+\infty} \text{vol}(S \cap H_t) e^{-\lambda t} dt$$

pour tout  $\lambda > 0$ , ce qui montre que la transformée de Laplace des deux volumes (qui sont clairement des fonctions de  $t$  continues à support compact  $\subset \mathbb{R}_+$ ) sont égales. On conclut par injectivité de la transformée de Laplace<sup>19</sup>.  $\square$

19. La transformée de Laplace de  $f$  est  $F(\lambda) = \int_0^\infty f(t) e^{-\lambda t} dt$  ( $\lambda > 0$ ). Il existe des preuves élémentaires que cette transformée est injective pour des fonctions  $f$  continues à support compact : Par exemple, si  $F = 0$ , on obtient par dérivation que les moments de  $f$  sont tous nuls, donc  $\int P f = 0$  et  $\int f^2 = \int f(f - P)$  pour tout polynôme  $P$  ; le théorème d'approximation de Weierstrass permet alors de montrer que  $f = 0$ .

## 5 Les probabilités : ennuyeuses et trop abstraites ?

Les souvenirs de lycée en probabilités laissent parfois un sentiment d'ennui. On se souvient de raisonnements flous sur les cartes à jouer, les tirages de boules dans les urnes et autres types de dénombrements pas très joyeux : le titre « *Les probabilités sans les boules* » d'un recueil d'exercices de Terminale par Gérard FRUGIER est symptomatique à cet égard. Peut-être que les arbres pondérés préconisés aujourd'hui par les réformateurs des programmes pourraient avoir plus d'intérêt (on pense par exemple à leur utilisation pour des procédés de codage comme l'algorithme de Huffman).

Il y a également un sentiment d'inachevé lorsqu'on se souvient des raisonnements de pure statistique descriptive (comme les écarts-type, inter-quartiles et autres rectangles à moustache) pour introduire l'« intuition » des probabilités dans le Secondaire. Il ressort une grande impression de pauvreté au regard de la vraie théorie. Cela pose aussi le problème de l'enseignement des *statistiques* : doivent-elles être enseignées avant, après (comme une continuation naturelle) ou en même temps que la théorie des probabilités ? Quelle place accorder aux problèmes pratiques de modélisation chez les étudiants ? Réciproquement, les probabilités enseignées servent-elles vraiment à faire comprendre des notions pratiques de modélisation mathématique ?

Tout semble être une question d'équilibre entre la théorie et les applications. Même à un niveau élémentaire, les probabilités peuvent être bien appliquées : des simulations, par exemple à l'aide du logiciel `scilab` accompagnent régulièrement les cours de première année d'Écoles d'ingénieurs<sup>20</sup>. Par la suite, certains problèmes délicats de simulation (à commencer par les méthodes de type « Monte Carlo ») nécessitent déjà un bon bagage théorique en Probabilités<sup>21</sup>. Par ailleurs, il est toujours possible de rendre la théorie des probabilités très abstraite sans référence apparente avec la réalité ou les simulations : mais cela se fait généralement à un niveau bien supérieur.

Il est vrai que le calcul numérique a peu ou prou disparu des épreuves écrites de concours (à cause des progrès technologiques et de la peur des tricheries) et on peut le regretter. Il reste quand même les TIPE dont le rôle formateur est important et n'est pas remis en question. Les probabilités pourront donc aider le futur ingénieur à comprendre les principes de la modélisation mathématique du monde qui nous entoure – ce serait un comble si elles ne le permettaient pas !

Il faudra bien entendu conserver à l'esprit le fait que l'enseignement des probabilités doit être utile aux sciences physiques : les probabilités ne se retrouvent évidemment pas qu'en algèbre linéaire ou en calcul différentiel et intégral, mais

---

20. Voir par exemple le témoignage exposé de Sylvie MÉLÉARD à l'École Polytechnique.

21. Voir l'exposé d'Eric MOULINES.

aussi dans de nombreuses théories physiques<sup>22</sup>. Tout comme on peut accepter que, comme pour le reste du programme en mathématiques, les probabilités sont également riches de beaux raisonnements, et contribuent à leur manière à l'apprentissage de la logique, du raisonnement, de la démonstration, et de la rigueur.

## 6 Les probabilités : probable inflation aux concours ?

Certains professeurs de classes préparatoires ont, semble-t-il, une sainte frayeur des probabilités aux concours. On craint une dérive inflationniste qui rendrait omniprésente les probabilités abstraites dans toutes les épreuves de mathématiques<sup>23</sup>.

Cependant, si l'on regarde de près les épreuves actuelles des concours, on peut s'apercevoir que les probabilités y sont déjà présentes, bien que cachées. En ce qui concerne le concours que je connais le mieux (le concours commun Mines-Ponts), on y trouve dans les années récentes :

- une allusion à un théorème ergodique (épreuve filière PC, 2011)
- une preuve du théorème central limite (épreuve PC/PSI, 2010)
- le problème des moments et la loi log-normale (épreuve MP, 2009)
- des matrices aléatoires (épreuve PSI, 2009)
- des matrices stochastiques (épreuve PC/PSI 2007, MP 2006)
- des séries génératrices de v.a. entières (loi de Poisson) (épreuve PC/PSI 2004)
- ...

sans compter le nombre important de sujets de TIPE qui se basent d'ores et déjà sur des notions de probabilités.

Une des raisons probables de cette présence « cachée » est le besoin important des probabilités dans la plupart des Écoles d'Ingénieurs. De ce point de vue, introduire explicitement les probabilités serait un moyen d'arrêter l'hypocrisie. Par exemple, si un programme raisonnable en probabilités existait en classes préparatoires, il est certain que le problème des moments et la loi log-normale de l'épreuve MP de 2009 serait rendu caduc ou traité en moins d'un quart d'heure<sup>24</sup>.

De plus, les probabilités sont parfois, comme on l'a vu ci-dessus, des outils très utiles à d'autres domaines mathématiques (analyse, algèbre, géométrie,

---

22. Voir l'exposé d'Alain MARUANI.

23. Ce serait également la raison pour laquelle il serait envisagé d'éviter d'enseigner un minimum de théorie de la mesure en CPGE (voir l'exposé de Laurent DECREUSEFOND).

24. C'est à peine le temps qu'il faut pour traiter du même problème en exercice de première année à l'École Polytechnique.

arithmétique...). Il est donc parfois possible de gagner du temps grâce à elles!

## 7 En guise de conclusion

L'introduction des probabilités dans les programmes des classes préparatoires aux grandes écoles semble acquise. J'espère que la réflexion exposée ici montre que cela n'apporte pas que des inconvénients, surtout au regard des besoins du futur ingénieur : il faudra donc s'adapter bon gré mal gré. On peut même souhaiter des influences favorables pour les autres domaines enseignés, en mathématiques ou dans les sciences physiques. Par ailleurs, le nouveau mode de raisonnement introduit enrichit nécessairement la palette du candidat aux concours.

Le principal problème pratique des enseignants sera leur formation, dès 2012-2013. Les Grandes Écoles seront naturellement mises à contribution par le biais de stages LIESSE. Les manuels actuels posent également problème : ceux qui ne sont pas simplement d'un niveau très basique (BAC ou prépas commerciales) apparaissent parfois trop abstraites au niveau master/Grandes Écoles. Des références intermédiaires sont souhaitables (c'est une des raisons qui m'ont poussé à rédiger le livre [4] à partir de notes de cours).

## Références

- [1] P. Billingsley, *Probability and measure*, J. Wiley & Sons, 1995.
- [2] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946 (réimpression 1999).
- [3] O. Rioul, *Théorie de l'information et du codage*, Hermes-Science Lavoisier, 2007.
- [4] O. Rioul, *Théorie des probabilités*, Hermes-Science Lavoisier, 2008.

# Annexe

## A Demandez le programme

### A.1 Première S (septembre 2011)

- Intuition basée sur la « statistique descriptive » et des simulations sur logiciel exploitant la moyenne et écart-type, médiane et écart inter-quartile d'une série statistique.
- Approche heuristique de la loi des grands nombres pour interpréter la moyenne et la variance d'une loi de probabilité discrète dans le cas d'un grand nombre  $N$  de répétitions d'expériences « identiques et indépendantes ».
- Lois discrètes à nombre fini de possibilités, surtout la loi binaire (schéma Bernoulli) menant par répétition à des lois binomiales *via* des raisonnements sur des « arbres pondérés » ; de manière annexe, on voit la loi géométrique (*tronquée!*).
- Intervalle de « fluctuation » pour réaliser un test d'hypothèse (sans le dire).

### A.2 Terminale S (septembre 2012)

Le programme est beaucoup plus ambitieux :

- Le raisonnement sur les arbres pondérés est là encore privilégié et étendu aux probabilités conditionnelles  $\mathbb{P}(A|B)$  noté  $P_B(A)$ , événements indépendants, formule des probabilités totales, simulation de marche aléatoire.
- Et surtout, les lois à densité sont introduites (à partir de la notion intuitive d'aire pour définir les intégrales), pour des densités continues sur un intervalle borné  $[a, b]$ . On définit à cette occasion une v.a.r.  $X$  comme une fonction de l'univers  $\Omega$  dans  $\mathbb{R}$ , définissant une probabilité par la formule  $\mathbb{P}(X(\omega) \in J) = \int_J p(x) dx$ , d'espérance  $\mathbb{E}(X) = \int_a^b xp(x) dx$ .
- Exemples : loi uniforme sur  $[a, b]$ , loi exponentielle avec propriété de loi sans mémoire, loi normale, ces deux exemples étant pourtant sur des intervalle non bornés.
- Méthode de Monte-Carlo, calculs admis d'intégrales donnant les moyenne et variance.
- Convergence de la loi binomiale vers la loi de Gauss (Théorème de Moivre) et intervalles de confiance, notamment à 5% et 1%, règles des  $\sigma$ ,  $2\sigma$  et  $3\sigma$ . On insiste bien sur les distinctions entre intervalles de fluctuation et de confiance (l'intervalle de confiance est aléatoire).



### A.3 Maths Sup. (septembre 2013)

... ?...

### A.4 Maths Spé. (septembre 2014)

... ?...

### A.5 Programme typique en Grande École

Ce qui est prévisible en CPGE constitue tout ou partie du programme suivant (en **gras** les choses plus difficiles qui resteront probablement aux Grandes Ecoles) :

- Probabilité sur un espace dénombrable, loi d'une v.a. (probabilité image)
- Conditionnement et indépendance ensembliste, formule de Poincaré
- Lois de Bernoulli, binomiale, géométrique, de Poisson, etc. Lois conditionnelles, v.a. indépendantes et leur somme.
- Moments et série génératrice. **Entropie.**
- **Existence et unicité d'une mesure de probabilité sur une tribu borélienne (voire jusqu'à la théorie de la mesure de Lebesgue)**
- Fonctions de répartition, lois à densité, simulation par inversion de la fonction de répartition (voire méthode du rejet).
- Espérance (**en lien avec la notion d'intégrale à rapport à une mesure**)
- Variables aléatoires intégrables et de carré intégrable, variance et covariance, corrélation. Régression linéaire.
- Lois uniforme, exponentielle, normale, gamma, de Cauchy, Bêta, de Pareto, etc.
- Inégalités de Bienaymé-Chebyshev, Cauchy-Schwarz, **Jensen, Hölder, Minkowski...**
- **Espérance conditionnelle : cas discret ou cas à densité, voire jusqu'au conditionnement par rapport à une tribu. Méthode des moindres carrés.**
- Vecteurs aléatoires (lien avec Fubini-Tonelli), matrice de covariance, vecteurs gaussiens
- **Densités conditionnelles, produits de convolution.**
- Calcul de loi et recherche de densité par la méthode de la fonction de répartition ou par le théorème de la loi image.
- **Espace probabilisé produit infini pour modéliser des expériences indépendantes en nombre infini, les deux lemmes de Borel-Cantelli.**
- Convergence en loi, en probabilité, en moyenne (en relation avec le théorème de convergence dominée), presque sûre
- Loi faible **et forte** des grands nombres, méthode de Monte-Carlo
- **Fonctions caractéristiques ou transformée de Laplace, théorème de Paul**

**Lévy.**

- Théorème central limite, intervalles de confiance
- **Introduction aux marches aléatoires, processus de branchement, files d'attente, chaînes (discrètes) de Markov.**

Pour les Grandes Écoles elles-même, il y aura nécessité de réorganisation complète de leur enseignement dans le domaine en **septembre 2015**. Les Grandes Écoles (conformément à leur souhait dans la plupart des cas) conserveront leurs enseignements actuels de première année mais bénéficient de bases avancées leur permettant d'aller plus vite et plus loin (lignes en gras ci-dessus).