Géométrie ou probabilités : une même démarche de modélisation ?(*)

Louis-Marie Bonneval

Objectifs de l'atelier

Un constat : un professeur de mathématiques est souvent plus à l'aise en géométrie qu'en probabilités, surtout si ces probabilités concernent les statistiques inférentielles.
Or une des difficultés des statistiques inférentielles est qu'elles se situent dans la zone floue entre la réalité et le modèle : zone bien connue du physicien, mais moins du mathématicien, qui en général reste en sécurité dans la zone du modèle.
De plus, **souvent le modèle est implicite** : on s'efforcera ici de l'expliciter.
D'où l'idée de l'atelier[1] : pour mieux comprendre la démarche de modélisation en probabilités, la rapprocher de la démarche de modélisation en géométrie.

GÉOMÉTRIE ?	PROBABILITÉS ?
Le sens du mot GÉOMÉTRIE a évolué : initialement « mesure de la terre » et donc branche de la physique, il désigne aujourd'hui une théorie axiomatisée. Cette séparation des rôles, très claire aujourd'hui pour nous, l'était beaucoup moins autrefois, et ne l'est sûrement pas pour nos élèves.	Le mot PROBABILITÉS présente lui aussi une ambivalence : initialement « étude du hasard », il désigne aujourd'hui pour les mathématiciens une branche de la théorie de la mesure.
En géométrie on peut se permettre au lycée de raisonner à l'intérieur du modèle, parce que le passage de la réalité au modèle a en principe été fait auparavant, et parce que le professeur de physique le prend en charge. Mais la question se pose pour l'introduction de toute notion nouvelle : vecteurs, barycentre …	Mais au lycée où se situe le premier contact avec ce domaine, c'est bien le professeur de mathématiques qui a en charge le passage de la réalité au modèle.

Après avoir précisé le sens des mots, on va comparer les deux domaines pour les trois étapes de la modélisation (cf. schéma ci-après) : de la réalité au modèle ; à l'intérieur du modèle ; du modèle à la réalité. Et en fin d'atelier on abordera quelques répercussions possibles sur l'enseignement.

(*) Compte rendu de l'atelier JB9 des Journées Nationales de Pau (24-25 octobre 2003).
(1) Bien entendu, je ne suis pas le premier à comparer géométrie et probabilités : Michel Henry notamment a écrit des choses fort intéressantes là-dessus (séminaire Didatech 1997). Pascal lui-même ne parlait-t-il pas de « géométrie du hasard » ?

I. Modéliser ?

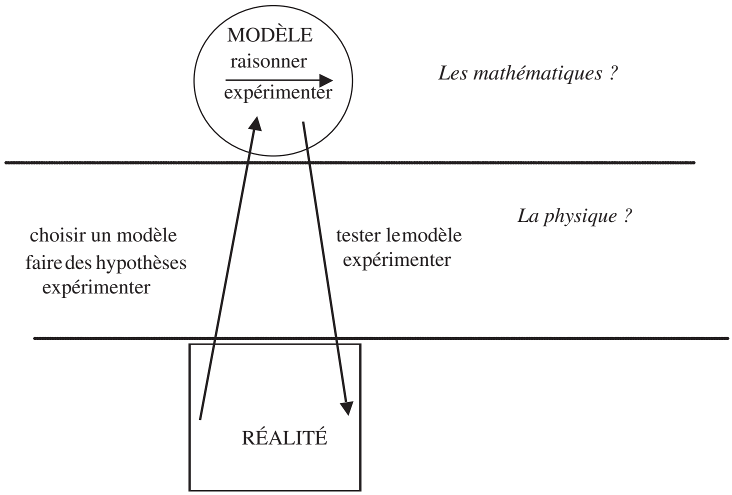
1) Quelques questions (faussement) naïves

Les questions ci-dessous, volontairement ambiguës, voire mal posées, étaient destinées à lancer le débat dans l'atelier.

GÉOMÉTRIE	PROBABILITÉS
1) Quelle est la distance Paris-Pau au mètre près ? 2) Quand on mesure une longueur, obtient-on toujours un nombre rationnel ? 3) Paris et Grenoble sont-elles équidistantes de Bordeaux ? 4) On plante deux piquets verticaux. Sont-ils parallèles ? 5) Les rayons du Soleil sont-ils parallèles ?	1) Quelle est la probabilité à 10^{-5} près qu'une pièce d'un euro tombe sur pile ? 2) Quand on estime une probabilité, obtient-on toujours un nombre rationnel ? 3) La naissance d'un garçon et celle d'une fille sont-elles équiprobables ? 4) On jette deux dés : les deux résultats sont-ils indépendants ? 5) Lors des accouchements successifs d'une mère, les sexes des enfants sont-ils indépendants ?

2) De la Terre (des réalités concrètes) au Ciel (des idées abstraites)

Le schéma ci-dessous est indispensable pour clarifier les questions précédentes. Il servira de guide pour l'atelier (on suivra les flèches).



3) Qu'est-ce qu'un modèle ?

Un **modèle** est une représentation **simplifiée** et **idéalisée** de la réalité.
Il a un rôle **d'explication**. Mais cela ne suffit pas à caractériser un modèle scientifique (les mythes, notamment religieux, ont un pouvoir explicatif très fort. À l'inverse, certains modèles scientifiques heurtent le bon sens).

Un **modèle scientifique** permet, par le **raisonnement** ou l'**expérimentation**, de **prévoir** et d'**agir**. Il se caractérise par :

- des **hypothèses** simplificatrices,
- des **paramètres** issus de l'observation.

Un travail essentiel du scientifique consiste à :

- Formuler les **hypothèses** issues de l'observation et de l'expérimentation.
- **Mesurer ou estimer** les valeurs des paramètres : ce qui pose la question de la **précision**.
- **Tester la validité** des hypothèses : ce qui pose aussi la question de la **précision**.

GÉOMÉTRIE	PROBABILITÉS
La Terre est-elle ronde ? Oui, en première approximation. Mais d'autres modèles existent (ellipsoïde, géoïde, …) **On plante deux piquets verticaux : sont-ils parallèles ?** Oui, s'il ne sont pas trop éloignés… **Les rayons du soleil sont-ils parallèles ?** Oui, si on accepte les hypothèses (chacune discutable) que : • Le Soleil est un point ; • La lumière du jour suit une trajectoire rectiligne issue de ce point ; • Ce point est à l'infini. Mais comment pourrait-on **tester** ce parallélisme ?	**Lors d'une naissance, a-t-on autant de chances d'avoir un garçon qu'une fille ?** Oui en première approximation. Mais d'autres modèles existent : 105 garçons pour 100 filles, 107 garçons pour 100 filles, … **Lors des accouchements successifs d'une mère, les sexes des enfants sont-ils indépendants ?** Oui en première approximation. Mais comment le **tester** ?

II. De la réalité au modèle

1) Choisir un modèle selon le problème posé

GÉOMÉTRIE	PROBABILITÉS
1) Quelle est la superficie de ma table ? 2) Quel est le plus court chemin entre Moscou et New York ? 3) Quelle est la longueur des côtes de Bretagne ?	1) Quelle est la probabilité que ma pièce tombe sur pile ? 2) Quelle est la probabilité qu'une famille de 5 enfants ait au moins trois garçons ? 3) Quelle est la probabilité qu'une aiguille de longueur connue rencontre les rainures d'un parquet d'espacement connu ?

2) Mesurer les paramètres du modèle

GÉOMÉTRIE	PROBABILITÉS
Quelle est la distance Paris-Pau au mètre près ? La précision demandée est abusive, parce que à la précision du mètre, le modèle du point n'est pas valide pour représenter une ville. **Bordeaux est-elle équidistante de Paris et Grenoble ?** Poser cette question c'est d'emblée se placer dans un modèle : les villes sont assimilables à des points, le sol est assimilable à un plan. Choisir ce modèle implique une certaine **précision** dans la mesure des longueurs, de l'ordre de 10 km : au-delà le modèle ne serait plus acceptable (taille des villes, rotondité de la Terre). Pour répondre on dispose d'une **représentation** de la réalité (une carte géographique). Une mesure sur la carte, autrement dit une **expérimentation à l'intérieur du modèle**, permet de répondre à la question. La réponse sera fonction de la **précision choisie**.	**Quelle est la probabilité à 10^{-5} près qu'une pièce d'un euro tombe sur pile ?** La précision demandée est abusive, parce que estimer à 10^{-5} près la probabilité en question suppose de la lancer 10 milliards de fois. Sans parler de la difficulté pratique de la chose (plus de 3 siècles, à raison d'un jet par seconde), on ne peut plus garantir que l'épreuve se passe dans les mêmes conditions (usure de la pièce). **La naissance d'un garçon et celle d'une fille sont-elles équiprobables ?** On ne peut pas calculer le rapport du nombre de cas favorables au nombre de cas possibles. Mais depuis Jakob Bernoulli (*l'Art de conjecturer*, 1713), il y a une autre façon de mesurer une probabilité : c'est de répéter l'épreuve, et d'observer la fréquence d'apparition de l'issue étudiée. La **précision** de la mesure dépend alors du nombre n d'observations : on sait qu'elle varie comme $\frac{1}{\sqrt{n}}$. La réponse **sera fonction de la précision choisie** : oui, à 10^{-1} près, non à 10^{-3} près.

3) Évaluer la précision d'une mesure

Évaluer la précision d'une mesure suppose d'admettre que plusieurs mesures d'une même grandeur x fourniront des résultats différents.
Dès lors mesurer x apparaît comme une épreuve aléatoire e. Le résultat de la mesure est une variable aléatoire X. La loi de probabilité de X, son espérance μ, son écart-type σ sont *a priori* inconnus.
Effectuer n fois la mesure est une épreuve répétée e^n, si on fait l'hypothèse que les n mesures sont indépendantes. Un résultat de e^n est un **échantillon** de X, de taille n.

La moyenne des mesures $\overline{X} = \frac{1}{n}\sum X_i$ dépend de l'échantillon : c'est une variable aléatoire relative à e^n.

Si n est assez grand ($n > 30$), on sait (théorème central limite) que $\overline{X}$ suit sensiblement une loi normale, d'espérance μ et d'écart-type $\frac{\sigma}{\sqrt{n}}$. Il en résulte qu'il y a plus de 95 % de chances que $\left|\overline{X}-\mu\right|$ soit inférieur à $2\frac{\sigma}{\sqrt{n}}$.

Donc la moyenne a de fortes chances d'être une valeur approchée de μ avec une incertitude $2\frac{\sigma}{\sqrt{n}}$.

Mais σ est inconnu. On l'estime par σ_{n-1}, valeur observée de $\sqrt{\frac{\sum\left(X_i-\overline{X}\right)^2}{n-1}}$.

Remarque : le **niveau de confiance** 95 % est usuel, mais on peut en choisir un autre, par exemple 90% ou 99%. On constate alors que ce qu'on perd en confiance on le gagne en précision, et inversement.

On voit que les **statistiques** sont un outil incontournable dans la démarche de modélisation.

4) La crise des irrationnels

GÉOMÉTRIE	PROBABILITÉS
Quand on mesure une longueur, peut-on trouver un nombre irrationnel ? Dans la conception élémentaire d'une longueur, mesurer une longueur consiste à faire le rapport entre la longueur choisie et la longueur choisie comme unité. Comment une mesure de longueur pourrait-elle alors être autre chose qu'un rapport d'entiers ? **La diagonale du carré** montre que deux longueurs ne sont pas toujours commensurables. Cet exemple est interne à un modèle mathématique : un carré, sa diagonale, sont des concepts mathématiques.	**Quand on estime une probabilité, peut-on trouver un nombre irrationnel ?** Dans la conception élémentaire d'une probabilité, mesurer une probabilité consiste à faire le **rapport entre le nombre de cas favorables et le nombre de cas possibles** (supposés également possibles). Comment une probabilité pourrait-elle alors être autre chose qu'un rapport d'entiers ? **L'aiguille de Buffon** Si L est la distance entre les rainures et a la longueur de l'aiguille, des hypothèses naturelles (à expliciter) conduisent à une probabilité $\frac{2a}{\pi L}$. En particulier si $a=\frac{L}{2}$, la probabilité est $\frac{1}{\pi}$. Cela a beaucoup surpris les contemporains de voir apparaître π dans une probabilité. Remarque : on peut **SIMULER** l'aiguille de Buffon. Quel est l'intérêt de la simulation ? Certainement pas de confirmer le modèle : en effet elle a été construite sur ce modèle. Elle confirme le calcul,

Quand il s'agit d'objets concrets, la longueur est définie et mesurée avec une certaine **précision**, et la mesure est un nombre décimal.	ou plutôt elle infirmerait le calcul si les résultats étaient très différents (à préciser) de la valeur théorique attendue. Elle pourrait remplacer le calcul si on ne savait pas le faire : il s'agirait alors d'une **estimation** à l'intérieur du modèle. Quand il s'agit d'épreuves concrètes, la probabilité est définie et estimée avec une certaine **précision**, et la mesure est un nombre rationnel, souvent même décimal. C'est d'ailleurs ce qui permet de représenter une épreuve par une **urne de Bernoulli**, et aussi de **simuler**.

5) Des modèles concurrents pour une même réalité

GÉOMÉTRIE	PROBABILITÉS
Ératosthène contre Anaxagore Ératosthène, deux siècles avant J.-C., à partir d'observations sur l'inclinaison des rayons du Soleil en deux endroits, calcule le rayon de la Terre. Deux siècles avant lui, Anaxagore, partant des mêmes observations, mesure la distance Terre-Soleil. Qu'est-ce qui fait dire qu'Ératosthène a raison et qu'Anaxagore a tort ?	**d'Alembert contre Fermat** Dans l'article « Croix ou Pile » de l'Encyclopédie, d'Alembert cherche la probabilité d'obtenir face en jetant une pièce au plus deux fois. Il prétend que, contrairement aux « principes ordinaires », les trois issues FF, PF, PP sont équiprobables. Qu'est ce qui fait dire qu'il se trompe ?

III. Dans le modèle

1) Expérimenter ou raisonner ?

GÉOMÉTRIE	PROBABILITÉS
Exemple : distance Terre-Lune ? Une mesure directe est impossible. On va donc **choisir un modèle** : la Lune est une sphère. Démarche **expérimentale** à l'intérieur du modèle : on fait un dessin à l'échelle, et on **mesure sur le dessin** (*niveau Sixième*). Démarche de **raisonnement** (c'est plus difficile) : on utilise le théorème de Thalès (niveau Quatrième) ou on mesure l'angle et on utilise la **trigonométrie** (*niveau Première S*).	Exemple : probabilité qu'une famille de 5 enfants ait au moins 3 enfants *consécutifs* du même sexe ? Une statistique est difficile. On va donc **choisir un modèle** : le plus simple est celui de l'équiprobabilité des deux sexes assorti de l'indépendance des naissances successives. Démarche **expérimentale** à l'intérieur du modèle : on **simule** (*niveau Seconde*). Démarche de **raisonnement** (c'est plus difficile) : on fait appel aux **graphes probabilistes** (*niveau Terminale ES spécialité*).

2) Une même théorie pour différentes réalités ?

On observe une analogie de **démarche** entre géométrie et probabilités.
Or il est fascinant de constater que les **modèles** eux-mêmes présentent une parenté, et peuvent relever d'une **même théorie**.
En effet l'ensemble des variables aléatoires réelles sur un même univers peut être muni d'une structure d'espace vectoriel euclidien.
Dès lors les outils conceptuels développés pour faire de la géométrie permettent de faire des probabilités : la covariance apparaît comme un produit scalaire, la corrélation comme un cosinus, la formule développée de la variance est une conséquence du théorème de Pythagore, …
C'est d'ailleurs avec les outils de la géométrie euclidienne que Karl Pearson en 1900 a établi le théorème du khi-deux (qu'on va utiliser plus loin).

IV. Juger la pertinence du modèle

GÉOMÉTRIE	PROBABILITÉS
Comment tester l'**équidistance** ? • Pour deux distances On mesure les deux distances (à la précision dont on dispose), et on calcule l'écart. Sous l'hypothèse d'équidistance, il est nul. D'où la règle de décision : s'il est suffisamment petit, on acceptera l'hypothèse d'équidistance. Sinon on la rejettera. Mais où placer la barre (la valeur critique) ? Cela dépend de la **confiance** qu'on a dans le modèle. • Pour plusieurs distances : Par exemple on veut **savoir si un hexagone est régulier** (une tomette provençale …) On mesure les longueurs des 6 côtés. Comment évaluer si ces longueurs sont assez proches les unes des autres ? On peut par exemple calculer leur moyenne m, puis calculer $\sum(d_i - m)^2$: c'est la variance de (d_i), multipliée par 6. Sous l'hypothèse de régularité, cette quantité serait nulle. D'où la règle de déci-	Comment tester l'**équiprobabilité** ? • Pour deux issues (alternative) (*cf. programme de Seconde*). Par exemple on veut savoir **si une (véritable) pièce est équilibrée**. Ou bien **si la naissance d'un garçon et celle d'une fille sont équiprobables**… On mesure l'une des probabilités : autrement dit on répète n fois l'épreuve, on observe la fréquence, et on calcule son écart à 0,5. On démontre que, sous l'hypothèse d'équiprobabilité, cet écart a 95% de chances d'être inférieur à $\frac{1}{\sqrt{n}}$. Si on a confiance à 95% dans le modèle, on choisira donc cette valeur comme seuil de décision pour accepter ou rejeter l'hypothèse. • Pour k issues (*cf. programme de TS et TES*) Par exemple on **veut savoir si un (véritable) dé est régulier**. $k = 6$. On mesure les probabilités (à la précision dont on dispose) : autrement dit on répète n fois l'épreuve, et on observe les fréquences f_i. On calcule $$d^2 = \sum(f_i - p_i)^2 :$$

sion : si elle est suffisamment petite, on acceptera l'hypothèse d'équidistance. Sinon on la rejettera. Mais où placer la barre (la valeur critique) ? Cela dépend de la **confiance** qu'on a dans le modèle.

Comment tester le parallélisme ?
On n'a pas de niveau, de rapporteur, d'équerre, …, mais seulement un mètre.
On peut porter sur chaque segment deux longueurs égales (à la précision dont on dispose), appeler I et J les milieux des deux diagonales, et mesurer la distance IJ. Sous l'hypothèse de parallélisme, elle est nulle. D'où la règle de décision : si elle est suffisamment petite, on acceptera l'hypothèse de parallélisme. Sinon on la rejettera. Mais où placer la barre (la valeur critique) ? Cela dépend de la **confiance** qu'on a dans le modèle.

Comment tester l'orthogonalité ?
On peut placer A à l'intersection, B et C sur chaque droite, et calculer

$$BC^2 - AB^2 - AC^2.$$

Sous l'hypothèse d'orthogonalité, cette quantité est nulle. D'où la règle de décision : si elle est suffisamment petite, on acceptera l'hypothèse d'orthogonalité. Sinon on la rejettera. Mais où placer la barre (la valeur critique) ? Cela dépend de la **confiance** qu'on a dans le modèle.

sous l'hypothèse d'équiprobabilité (c'est-à-dire $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$), on démontre qu'il y a au moins 95% de chances que d^2 soit inférieur à $\frac{2}{n}$ (*voir bulletin APMEP n° 441, p. 516*).
Si on a confiance à 95% dans le modèle, on choisira donc cette valeur comme seuil critique.
Si on a confiance à 99% dans le modèle, on placera la barre plus haut. Si on a 90% de confiance dans le modèle, on la placera moins haut.

Comment tester l'indépendance ?
Par exemple on jette deux dés, dont on admet qu'ils sont réguliers, et on se demande si les deux résultats sont indépendants. Cela revient ici à tester l'équiprobabilité des 36 couples.
On répète n fois l'épreuve (n aussi grand que possible, compte tenu des contraintes de temps, de coût, …), et on relève les fréquences d'apparition f_{ij}, ce qui fournit un tableau de contingence.

On calcule $d^2 = \sum \left(f_{ij} - \frac{1}{36} \right)^2$: sous l'hypothèse d'indépendance, on démontre que cette quantité a 95% de chances d'être inférieure à $\frac{1,4}{n}$. Si on a confiance à 95% dans le modèle, on choisira donc cette valeur comme seuil critique.

V. Aspects didactiques

1) Des objets de référence

GÉOMÉTRIE	PROBABILITÉS
Les figures de référence : • triangle, parallélogramme, … • polygones (polyèdres) réguliers.	Les épreuves de référence : • Urne de Bernoulli. • Pièce équilibrée.

• configurations de Thalès, de Pythagore.	• Dé régulier(2). • Jeu de cartes bien battu. Ces épreuves sont des épreuves idéales, leur nom est une façon imagée de désigner une épreuve aux issues équiprobables.

2) Donner du sens, motiver

GÉOMÉTRIE	PROBABILITÉS
Exercice sec Un triangle isocèle a pour côtés 50,5 ; 50,5 ; 100. Calculer la longueur de la hauteur principale.	**Exercice sec** On répète cinq fois une alternative à issues équiprobables. La probabilité d'obtenir au moins trois succès dépasse-t-elle 50 % ?
Problème contextualisé On dispose d'une corde de 100 m tendue au ras du sol entre deux piquets. On allonge la corde d'un mètre, et on la soulève en son milieu. Une personne debout peut-elle passer dessous ?	**Problème contextualisé** Parmi les familles de cinq enfants, celles qui ont au moins trois garçons sont-elles majoritaires ?
Remarque Comment pourrait-on résoudre ce problème sans utiliser le théorème de Pythagore ? On fait une figure à l'échelle $\frac{1}{500}$, et on mesure.	**Remarque** Comment pourrait-on résoudre ce problème sans utiliser la loi binomiale ? On simule au tableur ou à la calculatrice.

3) Des questions à plusieurs niveaux

GÉOMÉTRIE	PROBABILITÉS
1) Une équerre rectangle a pour petits côtés 5,2 cm et 8,3 cm. Quelle est la longueur du grand côté ? (réponse 9,8 cm)	1) On lance 100 fois une pièce de monnaie équilibrée. Quelle est la probabilité d'obtenir Pile plus de 60 fois ? (réponse 0,018)
2) Une équerre a pour côtés 5,2 cm, 8,3 cm et 9,8 cm. Est-elle rectangle ?	2) On a lancé 100 fois une pièce de monnaie, et on a obtenu Pile 70 fois. La pièce est-elle équilibrée ?
3) Une équerre a pour côtés 5,2 cm, 8,3 cm et 9,8 cm. Quel est l'angle opposé au grand côté ?	3) On lance 100 fois une pièce de monnaie, et on obtient Pile 70 fois. Quelle est la probabilité d'obtenir Pile si on la lance une nouvelle fois ?

(2) On devrait dire « Lancer régulier de dé », car le modèle intègre non seulement les caractéristiques du dé, mais aussi la façon dont il est lancé, l'atmosphère dans laquelle il baigne, la surface sur laquelle il roule, … Il en est de même des autres épreuves citées.