6.6. Probabiliser

Si l'élève arrive dans l'enseignement secondaire en ayant déjà un bagage non négligeable en arithmétique et en géométrie, il n'en est certainement pas de même en probabilité. Qui plus est, ce sujet restera généralement absent de l'enseignement des mathématiques jusqu'au début de la cinquième année. Dans cette classe, on devra dégager le concept même de probabilité des activités de traitement numérique de données rencontrées soit dans le milieu extérieur à l'école (journaux, radio et télévision par exemple) soit lors des cours de mathématiques des années précédentes. On arrive ainsi à cette situation tout à fait particulière que la première activité de « probabilisation » proposée aux élèves n'a pas tant pour objectif de construire un modèle d'une situation donnée, en vue de résoudre un problème, que d'introduire le concept (nouveau) de probabilité lui-même.

C'est donc dans le cadre d'une séquence d'enseignement qu'il conviendrait de construire les premières activités de modélisation probabiliste. Une telle séquence devrait *in fine* faire apparaître notamment les faits suivants :

1. Pour que les résultats possibles d'un phénomène aléatoire puissent être étudiés par les méthodes de la théorie des probabilités, il convient que les « chances » d'apparition de ces résultats puissent être quantifiées de façon relativement précise, par une méthode quelconque.
2. Dans certains cas, la quantification dont il vient d'être question résulte de propriétés de symétrie, qui permet de considérer certains résultats comme étant équiprobables.

> Exemple 6.6.1 *On admet qu'au jeu de « pile ou face », les deux possibilités ont même probabilité d'apparition* $\frac{1}{2}$

Cependant, nous savons qu'aucune pièce n'est parfaitement symétrique et qu'en conséquence nous commettons une erreur en acceptant de modéliser de la sorte le jeu de « pile ou face ». Nous le faisons quand même, car nous savons aussi que l'erreur commise est extrèmement faible. Dans certains cas par contre, il peut nous arriver d'utiliser un modèle équiprobable alors que ce modèle serait insuffisant pour une analyse fine.

Exemple 6.6.2 *Les statistiques montrent que la probabilité de naissance d'un garçon est sensiblement supérieure à* $0,5$. *Nous utilisons néanmoins souvent le modèle équiprobable dans un but didactique.*

3. Si aucune propriété de symétrie ne permet d'évaluer *a priori* les probabilités des résultats possibles, la quantification résulte généralement d'une observation d'un grand nombre de réalisations de cette expérience et de la mesure de la fréquence d'apparition des différents résultats. Ceci suppose évidemment que l'expérience soit reproductible à volonté. Dans le cas où l'on estime que des conditions de symétrie existent, la comparaison des probabilités *a priori* avec les fréquences d'apparition permet de confirmer ou d'infirmer cette hypothèse.

   Exemple 6.6.3

   (a) C'est en dénombrant les naissances de garçons et de filles que l'on a constaté que ces deux événements n'étaient pas équiprobables.

   (b) L'observation de longues séries de tirages du « Lotto » permet de se convaincre que la loterie nationale n'a pas trafiqué ses appareils.

   C'est la *loi des grands nombres* qui permet de considérer une probabilité comme étant une fréquence limite. Il serait cependant absurde d'attendre que l'enseignement ait rencontré cette loi — qui se situe à un niveau très conceptuel — avant d'avoir des activités — de type procédural — qui donnent à la notion de probabilité le statut de modèle théorique de celle de fréquence.

4. Enfin, si une expérience n'est pas reproductible et qu'aucune symétrie ne permet d'estimer les chances d'apparition des divers résultats, il vaut mieux ne pas se risquer à des raisonnements probabilistes.

   Exemple 6.6.4 *Il n'y a aucun sens à parler de la probabilité de découvrir un trésor lors du creusement d'un trou au pied de son pommier préféré.*

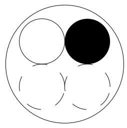
D'autres remarques — mettant en évidence d'autres types de compétences associées aux probabilités — peuvent être formulées concernant ces activités de modélisation qui ont pour objectif de « déterminer » les probabilités des résultats d'une expérience aléatoire.

Par exemple, certaines situations, tout en présentant des symétries qui justifient un *calcul* de probabilités, nécessitent une analyse soigneuse.

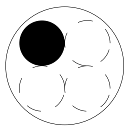
Exemple 6.6.5 *Deux billes noires et deux billes blanches sont contenues dans un tube cylindrique transparent dont le diamètre est tel que les quatre billes se placent nécessairement aux sommets d'un carré lorsqu'on dépose le tube verticalement sur une table. Dans ces conditions, deux configurations sont possibles pour les quatre billes :*

Un des deux raisonnements suivants est-il correct ? Si oui, lequel ?

1. *Fatalement, au moins une bille doit noire doit être adjacente à une bille blanche. Nous pouvons donc toujours placer ces deux billes comme suit :*

   Il y a alors deux façons de placer les deux dernières billes dans les cases vides. Chacune des configurations A et B est donc de probabilité $\frac{1}{2}$.
2. *Commençons par placer une bille noire.*

   La seconde bille noire peut être placée en trois endroits différents dont une seule correspond à la configuration B. La probabilité de celle-ci est donc $\frac{1}{3}$.

Lorsque nous cherchons à déterminer une probabilité à partir de l'observation d'un grand nombre de réalisations d'une expérience, deux points de vue peuvent être adoptés. Dans certains cas, il est raisonnable d'adopter comme probabilité la fréquence observée elle-même.

Exemple 6.6.6 *Lorsqu'on jette une punaise sur une table, la punaise peut atterrir « pointe en haut » ou « pointe en bas ». Si h_N est le nombre d'événements « pointe en haut » en N jets, il est raisonnable de choisir $\frac{h_N}{N}$ — ou une valeur arrondie de cette fraction — comme probabilité de cet événement. Il convient évidemment de rester conscient de ce que en considérant N' jets, avec $N' \neq N$ le résultat aurait été différent.*

Des constatations de ce genre et le besoin d'éclaircir la situation, amèneront — lorsque la théorie aura été suffisamment développée —, à déterminer des *encadrements probabilistes* (encore appelés *intervalles de confiance*). On pourra alors choisir une valeur de N assurant (avec une probabilité fixée) la précision souhaitée.

Dans d'autres cas, nous pouvons chercher à aller plus loin en essayant de déterminer une véritable *loi de probabilité* jouant le même rôle que les lois des sciences naturelles. En clair, il s'agit de trouver *une expression analytique* permettant de prédire de façon *raisonnablement précise* les fréquences observées.

EXEMPLE 6.6.7 *Lors d'une des journées du championnat de football belge de mars 1993, 31 matches ont été joués dans les promotions A à D. Le relevé des nombres de buts inscrits par les équipes visitées nous fournit le tableau suivant :*

Nombre de buts marqués	*0*	*1*	*2*	*3*	*4*	*5*
Nombre de matches	*8*	*9*	*7*	*4*	*1*	*2*

Nous ne nous intéressons qu'aux buts marqués par les équipes visitées afin de pouvoir considérer ces nombres de buts comme des événements indépendants. (On pourrait s'intéresser séparément aux nombres de buts marqués par les équipes visiteuses.)

Au total, quarante-neuf buts ont été marqués, de sorte que le nombre moyen de buts par rencontre est $\frac{49}{31} = 1,58\ldots$*. La question posée est alors de trouver une formule donnant en fonction de* n *la probabilité pour qu'au cours d'un match l'équipe visitée marque* n *buts.*

Une réponse acceptable en première approximation est connue depuis longtemps : le nombre de buts marqués peut-être modélisé par une loi de Poisson dont le paramètre vaut le nombre moyen de buts. Autrement dit, la probabilité de marquer n *buts serait calculée (dans le cas de nos données) par*

$$p_n = e^{-1,58}\frac{(1,58)^n}{n!}$$

On aurait ainsi par exemple $p_0 = 0.2058\ldots$*,* $p_1 = 0,3253\ldots$*,* $p_2 = 0,2571\ldots$*, ... Ces valeurs permettent de compléter le tableau précédent à l'aide d'une dernière ligne qui indique en fonction de* n *la valeur moyenne du nombre de matches au cours desquels* n *buts sont marqués. Afin de tenir compte des petits effectifs, nous regroupons en une seule les colonnes correspondant à* $n > 2$*.*

Nombre de buts marqués	*0*	*1*	*2*	*>2*
Nombre observés de matches	*8*	*9*	*7*	*7*
Nombre moyen de matches	*6,4*	*10,1*	*8*	*6,6*

> *La comparaison des deux dernières lignes du tableau ne permet pas de considérer comme absurde l'idée que le nombre de buts inscrits au cours d'un match obéit effectivement à une loi de Poisson. Si l'hypothèse est admise (des techniques plus fines permettent de la tester de façon plus précise), nous pouvons en relevant les scores d'un plus grand nombre de matches contrôler la stabilité du paramètre de la loi de Poisson, examiner s'il est différent pour les divisions supérieures et les divisions inférieures, comparer avec les résultats des équipes visiteuses, ...*

L'utilisation de lois de probabilités données sous forme analytique est fréquente dans les sciences naturelles, l'exemple le plus connu étant celui de la *loi normale*. Celle-ci intervient aussi de façon interne aux probabilités en fournissant par exemple une approximation de la loi binômiale dans le cas de tirages suffisamment nombreux. Nous sommes alors en présence d'une *seconde modélisation* : la loi binômiale modélise de façon discrète une situation expérimentale et la loi normale modélise de façon continue la loi binômiale.

Références

[43], [28], [53], [54], [55], [56], [52], [60], [63], [104], [106], [109], [113], [126], [124], [5].