Analyse de quelques données concernant l'évaluation de travaux d'élèves en probabilités par 5 professeurs de mathématiques francophones enseignant en lycée

Sylvette MAURY et Marie NABBOUT[1]

sylvette.maury@paris5.sorbonne.fr m_nabbout@yahoo.com

Résumé : Est-il aisé pour des professeurs de mathématiques, dans le domaine des probabilités, de s'accorder sur la validité d'une réponse d'élève ? nous abordons ici cette question en demandant à des professeurs de mathématiques de lycée, au cours d'entretiens individuels, d'évaluer des productions d'élèves de terminale ayant résolu des exercices portant sur les notions de probabilité, de probabilité conditionnelle ou d'indépendance. Nous examinons notamment la place que ces professeurs réservent aux raisonnements « intuitifs », en particulier lorsque ceux-ci s'appuient sur des intuitions généralement qualifiées de « correctes ».

1. Introduction

On sait depuis longtemps déjà- et Noizet et Caverni (1978) l'ont mis en évidence de manière magistrale dans un ouvrage déjà ancien, mais qui fait toujours référence- que les divergences entre évaluateurs peuvent être notables. C'est le cas même lorsqu'il s'agit d'évaluer des travaux mathématiques alors qu'en principe, dans ce domaine, on dispose d'une norme à laquelle confronter les productions des élèves. La question se pose d'une manière particulièrement vive dans le champ des probabilités. Plusieurs facteurs convergent en effet pour fonder cette conjecture :

-de nombreux travaux, aux ancrages théoriques pourtant divers, s'accordent pour reconnaître l'importance du recours à « l'intuition » lorsqu'il s'agit de porter des jugements probabilistes ou de résoudre des problèmes de probabilité[2]. Ces intuitions peuvent aller dans le sens de la connaissance scientifique : nous parlerons alors, avec Kanheman et Tversky (1982)

[1] *Faculté des Sciences Humaines et Sociales-Sorbonne,* Université René Descartes-Paris5
Laboratoire EDA, 12 rue Cujas, 75230 Paris Cedex 05

[2] Il existe des théorisations différentes de ce que nous désignons ici par « intuition ». On trouvera une synthèse de travaux sur les intuitions probabilistes dans Maury (1986), dans laquelle sont présentés les recherches

d'intuitions « correctes » ; elles peuvent au contraire s'y opposer : nous parlerons alors d'intuitions « trompeuses»[3]. Qu'en est-il du jugement d'un professeur lorsque la réponse de l'élève est sous-tendue par une (ou des) intuition(s) ? Comment sont prises en compte ces intuitions, en particulier lorsqu'elles sont « correctes », dans l'évaluation des travaux d'élèves par les professeurs de mathématiques ? La question se pose de manière particulièrement cruciale lorsqu'on sait que les intuitions trompeuses peuvent coexister avec des connaissances scientifiques, même chez individus ayant un bon niveau de formation en probabilités (Cohen, 1963 p.20 ; Kahneman et Tversky, 1982 ; Lecoutre, 1985) : il est alors vraisemblable que les professeurs de mathématiques (au moins certains d'entre eux) n'échappent pas à ce phénomène.

-Les raisonnements probabilistes peuvent être qualifiés de « fragiles ». Comme nous l'avons montré dans des travaux antérieurs, ils sont en effet sensibles à des facteurs tels la « formulation »[4] (Maury 1985, Baillé et Maury 1989) ou le « contrat didactique» (Maury, 1987). Cette fragilité, qui concerne a priori les élèves, pourrait trouver un écho chez les professeurs, en particulier lors des activités d'évaluation.

2. Situation du travail présenté

Mentionnons tout d'abord que le corpus sur lequel se fonde cette communication est extrait d'une expérimentation plus large[5], actuellement en cours, dont le but est de cerner certaines pratiques et certaines représentations de professeurs de mathématiques concernant l'enseignement des statistiques et des probabilités en classe terminale

---

d'auteurs majeurs dont nous citons ici quelques travaux : Cohen (1963), Fischbein (1965, 1987), Kahneman et Tversky (1982).

[3] Librement traduit de l'anglais. Les intuitions correctes sont celles que Fischbein (1965) reconnaîtrait comme des « précurseurs de la connaissance scientifique ».

[4] Le terme de « formulation » recouvre ici des éléments assez divers tels le lexique utilisé ou encore certains facteurs plus globaux, de nature rhétorique.

[5] Il s'agit de la thèse en cours de Marie Nabbout, sous la direction de Sylvette Maury. L'expérimentation complète concerne un effectif beaucoup plus important de professeurs libanais enseignant dans des écoles

Nous examinons ici les données recueillies auprès de 5 professeurs de mathématiques lors d'un entretien individuel (enregistré) au cours duquel il leur est demandé d'évaluer les réponses d'élèves à 8 exercices (4 réponses d'élèves pour chacun des exercice[6]). Nous recueillons et transcrivons leur commentaires ainsi que la note qu'ils attribuent à chacune des 4x8 = 32 réponses. Nous nous intéressons ici essentiellement à la distribution des évaluations quantitatives (notes) en essayant de rapporter l'analyse aux caractéristiques des 8 situations (S1 à S8) respectivement engagées dans les 8 exercices et à celle des réponses. La classification des situations, qui fait l'objet du paragraphe 3, repose sur une analyse conceptuelle. Celle des réponses d'élèves, classées en 4 catégories, est présentée au paragraphe 4.

3. Classification des situations

Elle est fondée, comme nous l'avons annoncé, sur l'identification du concept probabiliste en jeu dans chacune des situations (indépendance, probabilité conditionnelle, quantification de la probabilité). S'agissant du concept d'indépendance, il est connu (Maury, 1985 ; Steinbrig, 1986), tant du point de vue d'une analyse didactique que d'une analyse épistémologique, qu'il convient d'en distinguer deux « aspects », respectivement désignés par « indépendance chronologique » et « indépendance formelle »[7]. Les 8 situations se répartissent ainsi en 4 classes, respectivement référées à l'indépendance chronologique, l'indépendance formelle, la quantification, la probabilité conditionnelle. La distribution des situations selon les 4 classes est alors la suivante :

francophones, et la partie expérimentale impliquant directement les enseignants comprend 4 entretiens individuels.

[6] Les mêmes réponses sont évidemment communiquées à tous les enseignants.

[7] Selon les auteurs, on pourra également trouver les termes de « indépendance a priori » (qui peut avoir cependant un sens plus large qu'indépendance chronologique) et « indépendance stochastique ».

-*Indépendance chronologique* : S1 (jets successifs d'une pièce)[8], S5 (rotations successives d'une roulette autour de son axe), S7 (tirages successifs dans une urne).

-*Indépendance formelle*: S4 (lancer d'un dé), S6 (tirage d'une carte)[9].

-*Quantification de la probabilité* : S2 (situation de « prévision »[10] avec deux urnes), S8 (situation de « prévision » avec deux roulettes).

-*Probabilité conditionnelle* : S3 (tirages successifs dans une urne, sans remise).

4. Classification des réponses d'élèves

Nous retenons 4 catégories, respectivement désignées par « faux », « juste institutionnel », « juste intuitif », « divers », que nous avons caractérisées de la manière suivante :

-Dans les réponses classées « faux », on peut généralement repérer le mécanisme (la conception) qui est source de l'erreur. Il peut s'agir, par exemple, de la confusion entre « probabilité » et « représentativité »[11] ou de celle entre « indépendance » et « incompatibilité ». La réponse est donc « fausse » au sens où elle est sous-tendue par une conception identifiable qui ne correspond pas à la (ou à l'une des) conception(s) institutionnelle(s) et où elle ne relève pas d'une intuition correcte.

-Dans les réponses « juste institutionnelles », c'est la conception et/ou l'un des formalismes correspondant au programme que l'on retrouve. A titre d'exemple nous pouvons citer l'utilisation correcte d'un arbre, *telle qu'elle est préconisée dans le programme*, pour modéliser directement une situation de tirage sans remise.

---

[8] Le type d'expérience aléatoire retenu dans chaque situation est indiqué entre parenthèse.

[9] La situation S6, qui ne relève pas de l'indépendance chronologique, peut cependant bénéficier d'un traitement « intuitif » correct. Il s'agit en effet d'une situation dans laquelle l'espace fondamental est obtenu à partir du croisement de deux critères « indépendants », la « hauteur » et la « couleur » d'une carte.

[10] Nous reprenons ici la terminologie utilisée dans Maury (1986), elle-même empruntée à Piaget et Inhelder.

[11] Pour des exemples et une bibliographie concernant cette confusion, on pourra consulter Maury (1985, 1986).

-Les réponses « justes intuitives » sont argumentées par ce que nous reconnaissons comme des « intuitions correctes » (par exemple l'intuition de l'indépendance dans les situations relevant de l'indépendance chronologique).

-La catégorie « divers », qui comprend des réponses ne trouvant pas leur place dans les catégories précédentes, réunit en fait 4 cas. On y trouve des réponses où l'élève recourt à un « formalisme » en usage dans l'institution mais où sa réponse est non pertinente ; On y trouve également des réponses qui contiennent une justification pertinente accompagnée d'un élément « perturbateur ». Nous avons recensé 3 sortes d'éléments perturbateur : il peut s'agir d'un argument pertinent mais non canonique, de l'ajout d'une justification erronée ou encore de la présence d'un argument « hors sujet ». Des exemples sont donnés dans l'annexe.

5. Analyse des résultats concernant l'évaluation quantitative

Rappelons que chaque professeur, au cours d'un entretien individuel, était invité à évaluer les 32 réponses d'élèves. On lui demandait en particulier d'attribuer à chaque réponse une note comprise entre 0 et 5. Nous proposons ici une analyse de la distribution des notes basée sur la considération de l'étendue[12] de la « sous-distribution » relative à chacun des exercices. Pour chaque exercice nous parlerons ainsi de :

-*convergence des évaluations*, si l'étendue de la sous-distribution des notes est inférieure ou égale à 1.

-*divergence des évaluations*, si l'étendue est supérieure ou égale à 2,5 (remarque : la divergence implique que certains professeurs accordent la moyenne alors que d'autres ne l'accordent pas).

-*« neutralité » des évaluations*, si l'étendue est strictement comprise entre 1 et 2,5 (c'est à dire lorsque l'on ne peut pas parler de convergence ni de divergence).

[12] Le choix d'une autre mesure de la dispersion n'aurait guère modifié les résultats.

Les résultats obtenus sont assez tranchés :

-Les évaluations sont convergentes pour l'intégralité des réponses classées fausses (13 réponses) et la moyenne des notes, pour l'ensemble des exercices concernés et l'ensemble des professeurs, est très faible (égale à 0,1).

-Les évaluations ne sont jamais divergentes pour les réponses classées « juste institutionnel » (5 réponses) ; Elles sont convergentes dans 3 cas et neutres dans 2 cas, la moyenne variant entre 4,2 et 4,9 selon les exercices (moyenne globale : 4,64). On peut remarquer dans les 2 cas de neutralité, que la non-convergence provient d'une seule note et n'est pas due au même individu.

-En revanche, les résultats ne sont presque jamais convergents (sauf une fois) pour les réponses classées « intuitif juste » (8 réponses). Ils sont divergents 4 fois, neutres 3 fois, avec des moyennes oscillant entre 0,5 et 4,3 pour les 7 situations de non-convergences.

-Enfin les évaluations sont toujours divergentes dans les 6 réponses classées « divers », dont la moyenne varie, selon l'exercice de 1,1 à 3. Il semble donc que les professeurs apprécient diversement la présence d'un argument non pertinent à côté d'un argument pertinent, accordent un statut différent à la présence d'un commentaire qui n'est pas faux mais qui est hors sujet ou encore évaluent différemment des réponses pertinentes mais non canoniques, ou des réponses non pertinentes mais référées au formalisme en cours dans l'institution.

Au total, les résultats sont quand même rassurants dans le sens où l'on n'observe pas de divergence dans les évaluations à l'occasion des réponses classées « faux » ou « juste institutionnel ». En revanche le recours au formalisme institutionnel ou à des raisonnements intuitifs est diversement apprécié, de même que les réponses qui contiennent des éléments « perturbateurs ».

6.Conclusion

L'étude, qui fait l'objet de cet article où sont examinées les évaluations produites par plusieurs professeurs, est conduite en référence à l'analyse conceptuelle d'exercices de probabilités soumis à des élèves (fictifs) et à celle de leurs réponses (supposées). Les commentaires des professeurs (recueillis et transcrits), qui éclairent leur comportement d'évaluation, ne sont pas exploités ici car nous nous limitons dans l'article à la présentation des résultats les plus saillants[13].

L'analyse des évaluations quantitatives (notes) fournies par les divers évaluateurs met clairement en lumière des divergences dans leurs appréciations de certaines productions d'élèves. Comme nous l'argumentions dans l'introduction, certaines de ces divergences semblent pouvoir être attribuées à des spécificités des probabilités, notamment au rôle important que les « intuitions » jouent dans ce domaine. Il est vraisemblable que le phénomène de divergence traduise, au moins en partie, des différences entre les conceptualisations du champ par les divers professeurs. On notera cependant qu'une bonne convergence des évaluations est observée lorsqu'il s'agit de noter des réponses « franchement fausses » ou « franchement justes » (au sens de l'institution).

Pour nous, bien entendu, le constat de divergences ou de convergences dans les évaluations ne constitue pas une fin en soi. L'évaluation doit être rapportée aux représentations de l'enseignant concernant les probabilités, les élèves, l'enseignement des probabilités. L'analyse des commentaires que nous avons recueillis, lors de la résolution de la tâche d'évaluation par les enseignants, nous permettra d'avancer dans cette voie. Elle constitue d'ailleurs l'un des objectifs de la recherche évoquée dans la note 4, actuellement en cours.

[13] A ce stade du travail, certains phénomènes dont nous pensons qu'ils interviennent dans l'évaluation produite n'ont pas été explicitement pris en considération. Ainsi par exemple, pour un exercice donné, le fait que parmi les 4 réponses prévues dans le protocole expérimental les 4 catégories de réponses soient ou non représentées, peut jouer sur l'évaluation. On peut attendre un effet comparable à celui des « distracteurs » dans un QCM.

Références bibliographiques

BAILLE J., MAURY S., (1989) : Etude du discours argumentatif dans la résolution de problèmes probabilistes, in MONTEIL J-M., FAYOL M., *La psychologie expérimentale et ses applications*, pp221-228, Grenoble : PUG.

COHEN J., (1963), *Hasard, adresse et chance. La psychologie du pari et du jeu*, Paris : PUF.

FISCHBEIN E., (1965), *The Intuitive Source of Probabilistic Thinking in Children*, Dordrecht: D. Reidel.

FISCHBEIN E., (1987), *Intuition in Science and Mathematics. An Educational Approach*, deuxième édition 1994, Dordrecht, Boston, Lancaster, Tokyo : D. Reidel.

KAHNEMAN D., TVERSKY A., (1982), On the Study of Statistical Intuition, *Cognition, 11*, pp.143-157.

LECOUTRE M-P., (1985), Jugements probabilistes chez des adultes : pratique des jeux de hasard et formation en théorie des probabilités, in *Bulletin de psychologie, 38*, pp. 891-899.

LECOUTRE M-P., FISCHBEIN E., (1998), Evolution avec l'âge de misconceptions dans les intuitions probabilistes en France et en Israël, *Recherches en didactiques des mathématiques*, 18-3, pp.

MAURY S.,(1985), Influence de la question dans une épreuve relative à la notion d'indépendance, *Educational Studies in Mathematics, 16*, pp283-301.

MAURY S., (1986), *Contribution à l'étude didactique de quelques notions de probabilités et de combinatoire, à travers la résolution de problèmes*, Thèse d'état, Montpellier : Université des Sciences et Techniques du Languedoc.

MAURY S., (1987), Procédures dans la résolution de problèmes probabilistes, in VERGNAUD et Coll., *Didactique et acquisition des connaissances scientifiques*, pp.293-304, Grenoble : La Pensée Sauvage.

NOIZET G., CAVERNI J-P., (1978), *Psychologie de l'évaluation scolaire*, Paris :PUF.

STEINBRIG H., (1986), L'indépendance stochastique : Un exemple de renversement du contenu intuitif d'un concept et de sa définition mathématique formelle, *Recherche en didactique des mathématiques, 7,2-3*, pp.133-169.

Annexe : exemple de réponses de la catégorie « divers »

Argumentation non pertinente / « formalisme » institutionnel

Résumé de la situation : une roulette découpée en 6 secteurs égaux : 4 bleus et 2 rouges. On fait tourner 2 fois de suite la roulette.

Question adressée à l'élève : l'événement « elle s'arrête d'abord sur rouge puis sur bleu » est-il plus probable que l'événement « elle s'arrête d'abord sur bleu et puis sur rouge » ?.

Résumé de la réponse :

1er événement : rouge, P=2/6 ; bleu, P=4/5 ; Probabilité totale=1/3x4/5=4/15

2ème événement : bleu, P=4/6=2/3 ; rouge, P=2/5 ; Probabilité totale=2/3x2/5=4/15

La probabilité est la même.

Ajout d'un élément « hors sujet »

Même situation que la précédente.

Résumé de la réponse :

Les 2 événements sont aussi probable l'un que l'autre, les chances des secteurs rouges ou bleus n'influant pas sur l'ordre d'arrivée. Mais il serait plus probable qu'il y ait 2 fois la couleur bleue (hors sujet).

Argument pertinent mais non canonique :

Résumé de la situation : 2 sacs contenant respectivement 1 boule bleue et 4 rouges, 3 bleues et 6 rouges.

Question adressée à l'élève : quel est le sac préférable à la réalisation de l'événement « tirer une boule bleue » ?

Résumé de la réponse : l'élève préfère le deuxième sac car 3/6 est supérieur à ¼

Ajout d'une justification erronée

Résumé de l'énoncé de l'exercice : 2 roulettes comportant respectivement 6 secteurs bleus et 4 rouges, 3 bleus et 2 rouges.

Question adressée à l'élève : quelle roulette choisir pour que l'événement « l'aiguille indique la couleur bleue » soit le plus probable ?

Résumé de la réponse : la surface de bleu est la même dans les deux roulettes (6/10=3/5) de même pour le rouge. Mais je crois que l'événement sera plus probable avec la deuxième roulette parce que ses secteurs sont plus larges.