Le QCM et le hasard

Daniel Reisz, Paul-Louis Hennequin,
Philippe Langlois

Les grandes lignes de cet article ont été dessinées par Daniel Reisz quelques semaines avant sa mort. Les deux autres auteurs se sont contentés de les détailler et de les mettre en forme.

Introduction

Il est de tradition en France de regarder le QCM avec une certaine méfiance. Nous n'apprécions guère de voir une note mise sur la seule foi de résultats dont on ne peut pas savoir comment ils ont été obtenus : par un raisonnement approximatif ? par une conjecture réfléchie ? par hasard ? ou pire, par un coup d'œil discret au voisin ?

Aux États-Unis, où l'on est pragmatique, on considère qu'un raisonnement approximatif qui aboutit à un résultat correct mérite d'être pris en compte, et savoir conjecturer est pour eux une qualité appréciée. Ils luttent avec une bonne efficacité contre le coup d'œil au voisin : leurs dispositifs d'organisation et de surveillance des salles d'examen sont très supérieurs aux nôtres.

Reste le délicat problème des réponses faites au hasard ou partiellement au hasard, qui fera l'objet de cet article.

1. Le QCM à réponse Vrai-Faux

Prenons pour commencer une batterie de vingt questions pour chacune desquelles sont proposées deux réponses.

1.1. Sans pénalisation des réponses fausses

On met 1 point pour une réponse exacte, 0 sinon.
Imaginons un candidat ignare, qui répond à toutes les questions, mais au hasard. Sa note est donc une variable aléatoire X, obéissant à une loi binomiale faite d'une succession de 20 épreuves de pile ou face. Son espérance est 10, son écart-type $\sqrt{5}$.

• ***Quelles sont ses chances d'avoir au moins la moyenne ?***

Pour n entier allant de 0 à 20, la probabilité d'avoir la note n est

$$\mathrm{P}(\mathrm{X}=n)=\frac{1}{2^{20}}\binom{20}{n}.$$

Celle d'avoir au moins 10 est

$$\mathrm{P}(\mathrm{X}\geq 10)=\frac{1}{2^{20}}\sum_{n=10}^{20}\binom{20}{n}.$$

Compte tenu de la formule $\binom{20}{n} = \binom{20}{20-n}$, on obtient :

$$\sum_{n=10}^{20} \binom{20}{n} = \frac{1}{2}\sum_{n=0}^{20} \binom{20}{n} + \frac{1}{2}\binom{20}{10} = 2^{19} + 92\,378.$$

La probabilité cherchée[1] est donc :

$$P(X \geq 10) = \frac{1}{2} + \frac{92\,378}{2^{20}} \approx 0{,}5881.$$

Le QCM vrai-faux sans pénalisation des réponses erronées met ainsi presque à égalité le fumiste qui répond au hasard et le candidat moyen, mais scrupuleux, qui ne répond à une question que lorsqu'il est sûr et qui, par suite, n'obtient une note supérieure à 10 que s'il connaît au moins dix réponses.
Cette formule est donc inadaptée à un examen dans lequel il ne s'agit que d'avoir au moins 10/20.

• ***Supposons maintenant que le même QCM serve dans un concours où la barre est à 16/20.***

Comme l'écart de 16 à la moyenne est supérieur à 2 écarts-types, on peut subodorer que notre fumiste a peu de chances. En fait sa probabilité de succès[2] est

$$P(X \geq 16) = \frac{1}{2^{20}} \sum_{n=16}^{20} \binom{20}{n} = \frac{1}{2^{20}} \sum_{n=0}^{4} \binom{20}{n}$$

$$= \frac{1}{2^{20}}\left(1 + 20 + \frac{20 \times 19}{2} + \frac{20 \times 19 \times 18}{2 \times 3} + \frac{20 \times 19 \times 18 \times 17}{2 \times 3 \times 4}\right)$$

ou encore

$$\frac{1}{2^{20}}(1 + 20 + 190 + 1140 + 4845),$$

donc

$$\frac{1}{2^{20}} \times 6196 \approx 0{,}0059.$$

Dans un concours très sélectif, le QCM vrai-faux sans pénalisation, qui a le mérite d'être simplissime à élaborer et à corriger, est finalement une solution valable.
Si la barre est à 15/20, la probabilité de succès du fumiste est

$$P(X \geq 15) = P(X \geq 16) + P(X = 15),$$

soit

(1) Au lieu de faire ce calcul, on obtient tout simplement cette probabilité à l'aide d'une calculatrice ou d'un tableur.
(2) Même remarque que dans la note précédente.

$$P(X \geq 16) + \frac{1}{2^{20}}\binom{20}{15};$$

or

$$\frac{1}{2^{20}}\binom{20}{15} = \frac{15\,504}{2^{20}} \approx 0.0148.$$

Au total $P(X \geq 15) \approx 0{,}0207$, ce qui est encore tolérable.

1.2. Avec pénalisation des réponses fausses

On met 1 point pour une réponse exacte, -1 pour une réponse fausse, 0 pour absence de réponse.
L'espérance du fumiste – qui, rappelons-le, répond au hasard à toutes les questions – est maintenant 0, ce qui est assez moral, et il peut avoir aussi bien une note négative qu'une note positive.

• ***Cherchons quelles sont ses chances d'avoir au moins 10/20.***

Notons X le score ; c'est un nombre pair compris entre -20 et 20. Posons $Y = \frac{X+20}{2}$; Y suit une loi binomiale de paramètres 20 et 0,5 et $P(X \geq 10) = P(Y \geq 15)$. Cette probabilité a déjà été calculée à la fin du § 1.1 ; elle est de 0,0207. C'est un risque faible que les concepteurs de l'épreuve peuvent négliger.

• ***Cherchons maintenant quelles sont ses chances d'avoir au moins 14/20.***

La même transformation montre que la probabilité cherchée est celle d'avoir par des réponses au hasard au moins 17/20 dans le système sans pénalisation, soit :

$$\frac{1}{2^{20}}(1+20+190+1140) \approx 0{,}0013.$$

Risque infime, donc.

1.3. Une tactique pour le candidat lambda

Prenons le cas d'un candidat médiocre, mais pas trop éloigné de la barre de 10 : il connaît la réponse à 8 items sur les 20. Sur les 12 qui restent, il répond au hasard, espérant que les dieux seront avec lui. Quelles sont ses chances d'obtenir au moins la note 10 ?

• ***S'il n'y a pas pénalisation des mauvaises réponses :***

Il doit tirer la bonne réponse à au moins 2 items sur les 12. On est encore dans une situation de pile ou face : comment amener au moins 2 fois pile en 12 jets ? La probabilité de l'événement contraire est $\frac{1}{2^{12}}\left(\binom{12}{0}+\binom{12}{1}\right)$, soit encore

$\dfrac{1}{4096}(1+12)=\dfrac{13}{4096}\approx 0{,}0032$. Il a donc une probabilité de 0,9968 de s'en tirer, soit une quasi-certitude.

• ***S'il y a pénalisation des mauvaises réponses :***

Il doit tirer au moins 2 bonnes réponses de plus que de mauvaises, c'est-à-dire tirer la bonne réponse x fois, avec $x \geq 12 - x + 2$, soit $x \geq 7$.
Il doit donc tirer 0,1, 2, 3, 4 ou 5 mauvaises réponses, ce qui donne une probabilité de

$$\frac{1}{2^{12}}\left(\binom{12}{0}+\binom{12}{1}+\binom{12}{2}+\binom{12}{3}+\binom{12}{4}+\binom{12}{5}\right),$$

donc $\dfrac{1}{4096}(1+12+66+220+495+792)\approx 0{,}3872$. Il a donc presque deux chances sur cinq de s'en sortir.

• *Au total, qu'il y ait ou non pénalisation des réponses fausses, le hasard joue dans le QCM Vrai-Faux un rôle trop important pour que ce soit dans un examen un mode d'évaluation fiable.*

2. Les QCM à cinq réponses proposées

Cette situation est beaucoup plus fréquente que le QCM *Vrai-Faux*. On la trouve entre autres dans les tests administrés par les grandes organisations américaines spécialisées ETS et ACT, ou dans le *Kangourou des maths*.
Nous nous placerons dans le cas d'une batterie de 20 items à 5 réponses proposées : une réponse correcte et quatre « distracteurs ».

2.1. Le problème des réponses faites au hasard

Comment peut-on lutter contre la tentation classique du candidat qui ne connaît pas la bonne réponse : cocher au hasard et faire confiance au destin ?

Statistiquement, la solution est simple : s'arranger pour que l'espérance de gain d'une réponse au hasard soit nulle. Si on attribue 1 à chaque bonne réponse et $-r$ à chaque réponse fausse, l'espérance de gain d'une réponse au hasard est $\dfrac{1-4r}{5}$. Elle est nulle si et seulement si $r = \dfrac{1}{4}$.

Dans la pratique, les seules situations rencontrées sont l'absence de pénalisation ou cette pénalisation de $\dfrac{1}{4}$ par réponse fausse.

2.2. Premier cas : pas de pénalisation des réponses fausses

Ce cas est notamment celui de l'examen américain ACT, qu'ont passé environ 1 850 000 candidats en 2014 (Cf. encadré).

How ACT calculates the multiple-choice test scores and the Composite score

1. First we count the number of questions on each test that you answered correctly. We do not deduct any points for incorrect answers. (There is no penalty for guessing.)

• ***Prenons d'abord le fumiste absolu, qui répond au hasard à toutes les questions.***

Sa probabilité de réponse juste à un item est $\frac{1}{5}$. Son score est encore une variable aléatoire Y obéissant à une loi binomiale, mais celle-ci n'est plus une succession de « pile ou face », la probabilité de succès de chaque épreuve étant $\frac{1}{5}$. L'espérance est cette fois de 4, l'écart-type étant $\sqrt{20 \times \frac{1}{5} \times \frac{4}{5}}$, soit $\frac{4}{\sqrt{5}} \approx 1{,}789$. La note 10/20 est donc à $\frac{10-4}{1{,}789}$ écarts types de l'espérance (soit environ 3,3) ; on peut donc s'attendre à une probabilité très faible d'avoir au moins 10/20.

La probabilité d'avoir n réponses justes est $\mathrm{P}(\mathrm{Y}=n)=\binom{20}{n}\left(\frac{1}{5}\right)^{n}\left(\frac{4}{5}\right)^{20-n}$. Donc la probabilité d'avoir au moins la moyenne est

$$\mathrm{P}(\mathrm{Y} \geq 10)=\sum_{n=10}^{20}\binom{20}{n}\left(\frac{1}{5}\right)^{n}\left(\frac{4}{5}\right)^{20-n}.$$

La machine fournit $\mathrm{P}(\mathrm{Y} \leq 9) \approx 0{,}9974$, d'où $\mathrm{P}(\mathrm{Y} \geq 10)=1-\mathrm{P}(\mathrm{Y} \leq 9) \approx 1-0{,}9974$, soit $\mathrm{P}(\mathrm{Y} \geq 10) \approx 0{,}0026$.

On peut donc considérer que, même sans pénaliser les réponses fausses, le QCM à cinq réponses possibles par item n'a guère de pitié pour les fumistes.

En fait le vrai risque pour la fiabilité du test n'est pas là. Il est dans le candidat qui sait des choses et complète au hasard quand il ne sait pas.

• ***Reprenons le cas d'un candidat qui, sur les 20 items, connaît la réponse à 8 d'entre eux et qui, sur les 12 items restants, répond au hasard.***

Quelles sont ses chances d'obtenir au moins la note 10 ? Nous avons toujours un schéma de Bernoulli, avec 12 expériences indépendantes de probabilité $\frac{1}{5}$, dont on veut qu'au moins deux réussissent. Le mieux est de calculer la probabilité de l'événement contraire : 0 ou 1 réussite ; sa valeur est

$$\left(\frac{4}{5}\right)^{12} + 12 \times \frac{1}{5} \times \left(\frac{4}{5}\right)^{11} = \left(\frac{4}{5}\right)^{12} \times 4 \approx 0{,}2749.$$

Sa probabilité de succès est donc 0,7251, soit près de trois chances sur quatre.

• ***Plus généralement, prenons le candidat qui connaît la réponse à n items et répond au hasard aux 20 − n autres.***

Il fait donc $20 - n$ expériences indépendantes de probabilité $\frac{1}{5}$; le bénéfice qu'il peut escompter a pour espérance $\frac{20-n}{5}$, ce qui est loin d'être négligeable, d'autant plus que le risque est nul.

2.3. Second cas : pénalisation des réponses fausses

On met 1 par bonne réponse, $-\frac{1}{4}$ par réponse fausse, 0 par absence de réponse. C'est notamment le cas du célèbre SAT américain (Cf. encadré).

SAT Scoring

Correct Answers	Incorrect Answers	Omitted
+1 point for questions you get correct	-1/4 point subtracted for incorrect multiple-choice	0 points subtracted for questions you don't answer

Nous avons vu que, dans cette situation, l'espérance de gain d'une réponse au hasard est nulle. Mais, dans certains cas, il pourrait *a priori* être intéressant de « tenter le coup ».

• ***Prenons encore une fois le cas d'un candidat qui, sur les 20 items, connaît la réponse à 8 d'entre eux et qui, sur les 12 items restants, répond au hasard. Quelles sont ses chances d'obtenir au moins la note 10 ?***

S'il tombe juste x fois, sa note sera $8 + \left(x - \frac{1}{4}(12 - x)\right) = \frac{5}{4}x + 5$. L'inégalité $\frac{5}{4}x + 5 \geq 10$ équivaut à $x \geq 4$. Il doit donc tirer au moins 4 réponses correctes sur les 12. La probabilité de l'événement contraire est d'avoir 0, 1, 2 ou 3 bonnes réponses, soit :

$$\left(\frac{4}{5}\right)^{12} + 12 \times \frac{1}{5}\left(\frac{4}{5}\right)^{11} + \frac{12 \times 11}{2} \times \left(\frac{1}{5}\right)^{2}\left(\frac{4}{5}\right)^{10} + \frac{12 \times 11 \times 10}{2 \times 3} \times \left(\frac{1}{5}\right)^{3}\left(\frac{4}{5}\right)^{9} \approx 0{,}7946.$$

Sa probabilité de succès est donc 0,2054, soit une chance sur cinq, situation non désespérée, mais nettement moins favorable que dans le système sans pénalisation.

• ***Prenons maintenant un bon candidat qui connaît la réponse à 17 des items, mais qui veut accéder à une prestigieuse formation pour laquelle il lui faudrait 18.***

Il répond donc au hasard aux 3 questions litigieuses. S'il tombe juste x fois, sa note pour l'ensemble des 3 questions sera $x-\frac{1}{4}(3-x)=\frac{5}{4}x-\frac{3}{4}$. Il lui faut donc 2 bonnes réponses au moins sur les 3, ce qui a une probabilité de $\frac{1}{5^3}+3\times\frac{1}{5^2}\times\frac{4}{5}=\frac{13}{5^3}\approx 0{,}104$, ce qui est bien mince.

• On peut donc conclure que le système utilisé pour le SAT, à la différence de celui d'ACT, protège relativement bien contre les tentatives d'exploitation du hasard.

Remarque 1

On rencontre assez souvent des QCM offrant le choix entre quatre réponses au lieu de cinq. Pour qu'une réponse faite au hasard ait une espérance nulle, il faut alors, en attribuant toujours 1 à une réponse correcte, attribuer $-\frac{1}{3}$ à une réponse fausse. Faire n réponses au hasard revient donc à faire n expériences indépendantes, la probabilité de réussite de chacune étant 0,25.

Cela se résout de façon analogue aux cas précédents, aux précédents, dans une situation un peu plus simple.

Remarque 2

Nous avons souvent utilisé des probabilités du type $\mathrm{P}(\mathrm{X}\geq m)$, donc $1-\mathrm{P}(\mathrm{X}\leq m-1)$, ce qui revient à travailler sans le dire sur la fonction de répartition de la loi binomiale, ce qui se fait couramment dans les classes bien que l'expression ne soit pas au programme.

Remarque 3

On rencontre parfois dans les QCM des items à cinq réponses proposées où la bonne réponse n'est pas forcément unique. Cette pratique est de moins en moins courante, à cause des nombreuses difficultés que présente son traitement, difficultés dont nous allons donner une idée.

Il faut distinguer deux cas selon que l'énoncé précise ou non le nombre de réponses valables. En voici un exemple :
Cocher dans cette liste les nombres premiers : 17☐ 35☐ 57☐ 67☐ 87☐
Cocher dans cette liste les deux nombres premiers : 17☐ 35☐ 57☐ 67☐ 87☐

Dans le premier cas, derrière l'item à cinq réponses se dissimule un ensemble de cinq questions à réponses Vrai-Faux, à traiter indépendamment les unes des autres. Mais quel poids accorder à un tel item par rapport à un item standard à une seule bonne réponse parmi cinq ? En outre, dans une question de ce genre, le candidat peut estimer plausible qu'il doit cocher au moins une case et au plus quatre ; donc ces cinq items ne sont pas vraiment indépendants.

Dans le second cas, les réponses aux cinq sous-questions ne sont pas indépendantes : il faut donc traiter l'item en bloc. Mais que faire du candidat qui coche correctement une des cases, mais pas l'autre (le lecteur pourra vérifier qu'un choix fait au hasard donne six chances sur dix d'aboutir à cette situation) ?

Remarque 4

Dans le système sans pénalisation type ACT, le candidat lambda qui est sûr de 8 réponses et veut atteindre la note 10 a tout intérêt à répondre au hasard aux 12 autres : chaque réponse supplémentaire apporte un plus ou n'apporte rien.
Dans le système SAT, compte tenu des pénalités, on peut se demander si le candidat n'a pas intérêt à ne faire de réponse au hasard qu'à m questions, m étant compris entre 2 et 12. Une généralisation de ce problème est étudiée dans l'annexe. La réponse est plus complexe qu'on ne pourrait *a priori* le penser.

3. Que se passe-t-il dans la réalité ?

Le destin d'un lycéen américain de classe terminale se joue en grande partie sur les tests SAT et/ou ACT. Les responsables d'ACT, qui visent surtout les candidats moyens, glissent discrètement sous le tapis le rôle des hasards heureux. Les responsables du SAT, visant le haut de gamme, croient par leur système de pénalisation s'être mis à l'abri des aléas. Mais qu'en est-il réellement ?

La grosse difficulté vient du choix des distracteurs, ces réponses fausses proposées à côté de la bonne. Les distracteurs d'un item n'ont en effet pas tous le même niveau de vraisemblance, si bien qu'un candidat astucieux, après avoir coché toutes les réponses dont il est sûr, peut souvent éliminer pour chacun des items restants une ou deux, voire trois réponses manifestement invraisemblables et cocher au hasard une de celles qui reste. Il fait ainsi passer pour les items douteux sa probabilité de réussite de $\frac{1}{5}$ à $\frac{1}{4}$ ou $\frac{1}{3}$, voire $\frac{1}{2}$, ce qui améliore sérieusement ses chances. Sur un item du SAT où il peut écarter à coup sûr deux réponses, par exemple, son espérance de gain, au lieu d'être nulle, est $\frac{1}{3}+2\times\frac{1}{3}\times\left(-\frac{1}{4}\right)=\frac{1}{6}$.

En outre, quantité d'experts autoproclamés proposent des recettes plus ou moins miraculeuses. En voici quelques-unes, pour cinq réponses étiquetées A, B, C, D, E :

- ne pas choisir deux fois de suite le même numéro ; si on sait que la réponse à la question k est D, ne pas choisir D pour les questions $k+1$ et $k-1$;
- si les réponses proposées sont numériques, éliminer la plus grande et la plus petite ;
- éliminer les corps étrangers : parmi $x+2y$, $2xy$, $2x-y$, $x+\frac{y}{2}$, écarter $2xy$;
- éliminer les réponses de type inhabituel ;
- en cas de doute choisir la réponse la plus longue ;
- en cas de doute[(3)] choisir B (sic).

(3) L'auteur de ce conseil prétend avoir vérifié que la fréquence de la réponse B est significativement plus élevée que les autrs

On serait tenté d'en ajouter une : *ne jamais oublier que les conseilleurs ne sont pas les payeurs*.

Conclusion

Nous pensons avoir convaincu le lecteur que ce thème est une mine d'exercices sur la loi binomiale. Pour une telle loi, dès que le nombre des tirages est un peu élevé les calculs à la main deviendraient très lourds, mais heureusement il y a les machines !

Annexe : Tableau de la probabilité d'obtenir la moyenne en fonction du nombre d'items choisis

Nous avons évoqué dans la remarque 4 du § 2 le problème du candidat au SAT qui est sûr de 8 réponses sur les 20 et veut maximiser ses chances d'avoir au moins 10 en répondant au hasard à m items sur les 12 qui restent. Quel est pour m le meilleur choix possible ? On est tenté de dire que c'est $m = 12$, mais est-ce si sûr ? Est-on d'ailleurs sûr que la probabilité de succès augmente avec m ?

Posons le problème sous une forme plus générale :

On fait une suite de m tirages indépendants, répondant aux règles suivantes :
probabilité de succès 1/5, gain 1 ; probabilité d'échec 4/5, pénalité 1/4.
Quelle est la probabilité d'obtenir au moins le total entier k (k > 0)?

Le nombre X de tirages favorables est une variable aléatoire binomiale de paramètres m et $\frac{1}{5}$. On gagne X points et on en perd $\frac{m-\mathrm{X}}{4}$, ce qui donne un solde de $\mathrm{X}-\frac{m-\mathrm{X}}{4}=\frac{5\mathrm{X}-m}{4}$. Ce qui nous intéresse est donc $\mathrm{P}\left(\frac{5\mathrm{X}-m}{4}\geq k\right)$, autrement dit $\mathrm{P}\left(\mathrm{X}\geq\frac{4k+m}{5}\right)$ ou mieux, en désignant par $\lceil u\rceil$ le plus petit entier au moins égal à u, $\mathrm{P}\left(\mathrm{X}\geq\left\lceil\frac{4k+m}{5}\right\rceil\right)$.

Voyons ce qui se passe pour $k = 1$, $k = 2$ et $k = 3$.

	$k = 1$		$k = 2$		$k = 3$	
m	$\left\lceil\frac{4+m}{5}\right\rceil$	$P\left(X \geq \left\lceil\frac{4+m}{5}\right\rceil\right)$	$\left\lceil\frac{8+m}{5}\right\rceil$	$P\left(X \geq \left\lceil\frac{8+m}{5}\right\rceil\right)$	$\left\lceil\frac{12+m}{5}\right\rceil$	$P\left(X \geq \left\lceil\frac{12+m}{5}\right\rceil\right)$
1	1	0,2000				
2	2	0,0400	2	0.0400		
3	2	0,1040	3	0.0080	3	0.0080
4	2	0,1808	3	0.0272	4	0,0016
5	2	0,2627	3	0.0579	4	0,0067
6	2	0,3446	3	0.0989	4	0,0170
7	3	0,1480	3	0.1480	4	0,0333
8	3	0,2031	4	0.0563	4	0,0563
9	3	0,2618	4	0.0856	5	0,0196
10	3	0,3222	4	0.1209	5	0,0328
11	3	0,3826	4	0.1613	5	0,0504
12	4	0,2054	4	0.2054	5	0,0726

On constate que, pour k donné, la probabilité d'obtenir au moins k points en m coups croît tant que $4k + m$ ne franchit aucun multiple de 5, baisse lorsqu'un tel multiple est franchi, puis recommence à croître, etc. Il n'y a donc pas croissance constante de la probabilité de réussite en fonction du nombre d'essais. En particulier, quand on ne veut qu'un point supplémentaire, il y a une chute spectaculaire quand on passe de 11 essais à 12, le maximum étant d'ailleurs atteint pour $k = 1$.