Point de vue

LE RISQUE DES STATISTIQUES

Gilberte PASCAL

à

Messieurs les auteurs de l'article
Calcul de risques de première et de seconde espèces à travers un exemple
paru dans le numéro 94

s/c de Monsieur le Rédacteur en Chef de Repères

Chers collègues,

Je souhaiterais réagir à l'article que vous signez dans la dernière livraison de Repères et qui concerne les tests statistiques enseignés actuellement au lycée. Comme j'ai déjà eu naguère l'honneur de collaborer à cette si intéressante revue (à l'occasion des numéros 32 et 46), peut-être le *Comité de Rédaction* ne verra-t-il d'ailleurs pas d'inconvénient à faire part de mes remarques à ses lecteurs, par exemple dans le cadre de la rubrique “Point de vue”.(*)

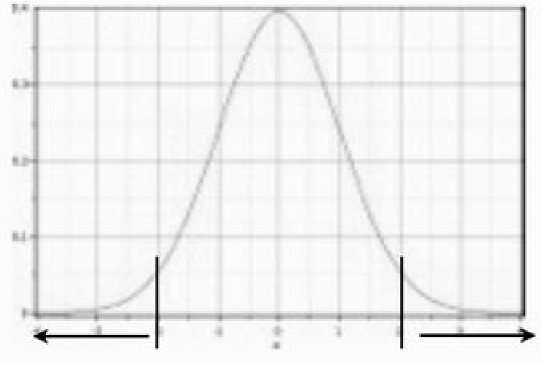
Je dois dire en effet que l'étude que vous présentez à propos des risques liés à la prise de décision a rejoint un sujet qui me préoccupe depuis assez longtemps déjà et dont j'observe l'évolution progressive, sinon dans les pratiques des enseignants, du moins dans les articles qui, comme le vôtre, se proposent de clarifier pour ceux-ci les difficultés de la question.

Il ne s'agit pas, évidemment, de les priver des calculs très intéressants que vous consacrez aux probabilités associées aux diverses définitions utilisées dans les différentes classes. Il en est sans doute qui apprécieront de suivre, sur vingt-cinq pages, quatre fois le même développement, ou qui prendront plaisir à rêver sur des probabilités évaluées jusqu'à la deuxième décimale. Les charmes de la modélisation sont très souvent impénétrables.

En réalité, l'aspect de la question qui m'intéresse plus particulièrement est celui de *l'évaluation du risque*, au sujet de laquelle vous avancez des propositions qui ne peuvent manquer — me semble-t-il — de donner à réfléchir. Mais c'est malheureusement un sujet difficile et je vais tâcher de résumer le problème de la manière la plus claire et la plus simple possible…

(*) *Note de l'éditeur*. La référence à l'article *Chronique d'une correspondance probablement apocryphe*, paru dans le numéro 32, amène à conjecturer que l'auteure du présent texte n'est autre que la personne anonyme écrivant sous un pseudonyme choisi pour laisser penser qu'il s'agirait de la sœur de Blaise Pascal.

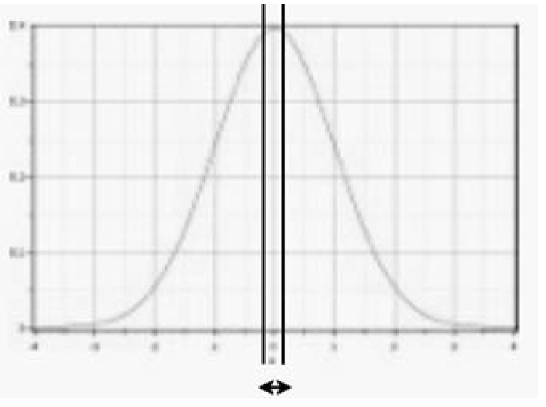
Comme de nombreux professeurs j'ai été amenée (il y a bien longtemps) à enseigner des rudiments de statistiques, alors même qu'en probabilités je n'avais guère appris qu'à reconnaître une loi binomiale et à l'approcher (dans la plupart des cas qui nous occupent) par une *courbe en cloche* de même moyenne et de même écart-type. Autrement dit je savais qu'un tirage d'échantillon ou une série de "pile ou face" se *modélisaient* — comme on dit aujourd'hui — par une loi binomiale et que, dans les calculs, celle-ci pouvait être remplacée avec une bonne approximation, par une loi normale bien choisie.



Dès lors — comme vous l'expliquez très bien dans votre article — les calculs et la méthode se ramènent essentiellement à une situation dans laquelle on a choisi un "intervalle de risque" correspondant (par exemple) à une probabilité de 5 % et où l'on pourra dire : « si le calcul a été effectué sous l'hypothèse où la proportion étudiée vaudrait à peu près p_0 et si l'échantillon observé présente une proportion p appartenant à l'intervalle en question, on peut considérer que l'on vient d'assister à un événement qui avait vraiment peu de chances de se produire (moins de 5 %) et il est donc plutôt sage de parier, soit que le test est mal fait, soit que le calcul préliminaire repose sur une mauvaise hypothèse pour la valeur de p_0. »

Cette manière de voir — dont on aura noté qu'elle se garde bien de fixer une quelconque probabilité pour la réussite ou non du "pari" qui constitue la décision finalement prise — n'est au fond rien d'autre que ce que le sens commun pratique en matière de "chance". Plaçons-nous en effet dans le cas d'une loterie qui vendrait un très grand nombre de billets et laisserait donc — sous l'hypothèse raisonnable d'équiprobabilité — une probabilité dérisoire à chaque joueur de gagner. Si mon voisin gagne le gros lot (et que je suis persuadé qu'il n'a pas triché !) n'aurai-je pas tendance à me dire que cela tient à tel point du miracle qu'il devient nécessaire de remettre en cause l'hypothèse d'équiprobabilité... de considérer "qu'il a de la chance"... et qu'une bonne étoile providentielle a obligé le sort à calculer autrement ?...

De manière analogue, mais en quelque sorte inverse, il est tout à fait possible, dans le test résumé précédemment, de se fixer un autre intervalle correspondant à une probabilité inférieure ou égale à 5 % (à partir d'un calcul fait sur l'hypothèse d'une probabilité supposée aux alentours de p_0), mais en prenant par exemple un petit intervalle autour de la valeur maximale... Qu'arrivera-il si l'expérience (tirage de pile ou face ou observation d'un échantillon tiré au hasard) donne un résultat contenu dans cet intervalle ? Il conforterait à merveille, évidemment, l'hypothèse d'une probabilité initiale égale à p_0 ... mais ne penserait-on pas que, si c'est bien le cas, le résultat est "trop beau pour être honnête" ? C'est d'ailleurs une possibilité qui a été évoquée à propos des expériences de Mendel sur les lois de la génétique...



Je ne prends pas ces exemples gratuitement : la question que j'ai laissée en suspens plus haut doit rester présente à notre esprit : « est-il possible de quantifier le *risque* du pari qui consiste à rejeter ou à conserver l'hypothèse testée à propos de la valeur de p_0 ? », ou, si l'on préfère : « peut-on raisonnablement fixer une *probabilité* de *rejeter à tort l'hypothèse* dans le cas où le test a donné un résultat appartenant à l'intervalle choisi pour cela ? ». Les exemples précédents, qui reposent au départ sur des protocoles assez analogues, amènent naturellement à penser que les paramètres rencontrés jusqu'ici ne suffisent pas pour se convaincre que ce soit si facile ! Et on peut même préciser la question : si on ne considère que les deux exemples associés plus haut à la courbe en cloche, si on admet assez aisément, intuitivement, que la première version conduit plus légitiment à rejeter l'hypothèse que la seconde, si on peut trouver (par exemple dans les calculs de votre dernier article…) des arguments convaincants pour se persuader que la deuxième version de l'intervalle de "risque" ne saurait guère fournir un test sérieux, il n'en reste pas moins que ces considérations viennent grandement compliquer le problème de la "probabilisation du risque" !

J'avoue que pour ma part je me suis posé la question. J'avoue que j'ai fini par conclure que la réponse à cette question était négative. J'avoue enfin que je me suis donc systématiquement efforcée d'éviter de soulever le problème... et que je me suis soigneusement cantonnée à un discours du type « on prend la décision de rejeter l'hypothèse a*u seuil de risque* 5 % », en précisant au besoin que cela ne devait pas être traduit par une expression du type : « en prenant cette décision je n'ai pas plus de 5 % de chances de me tromper ». Les années ont passé. Les programmes du secondaire ont trouvé important d'adopter en leur sein les statistiques inférentielles, les manuels et les irem ont proposé des exemples, des activités et des problèmes, et on a pu constater peu à peu, sinon une disparition de l'expression "au seuil de risque 5 %", du moins un commentaire systématique du type, précisément de celui que j'ai déjà cité : « en prenant cette décision je n'ai pas plus de 5 % de chances de me tromper » !

Pour votre part vous placez, dès la troisième page, votre article sous le signe d'un tel choix, puisque vous écrivez :

> « Remarquons qu'en décision statistique, avec le raisonnement ci-dessus, on a dans notre exemple 5% de chances de rejeter à tort l'hypothèse H_0. Par la nature de ce raisonnement, on peut donc être amené à se tromper en prenant la décision de rejeter H_0, mais cette erreur est très peu probable, ici égale à 5%, ce qui prouve statistiquement l'intérêt pratique de la démarche mathématique de la prise de décision. »

Quel est donc l'ensemble des événements sur lequel est définie la probabilité qui vaut ici exactement 5 % ? Quel est donc l'événement contraire de celui qui revient à "rejeter à tort" ? Et dont la probabilité est ainsi évaluée à 95 % ? Est-ce autre chose que l'événement qui consiste à "accepter avec raison l'hypothèse H_0" ? Mais comment diable peut-on affecter une telle probabilité de 95 % à une *valeur isolée* de p_0, alors même que chacun sait qu'il ne saurait, en vérité, y avoir la moindre chance que (par exemple) la pièce soit *exactement équilibrée* ?… Et qu'on ne peut donc, en toute logique, que *rejeter avec raison* l'hypothèse selon laquelle la probabilité testée puisse valoir p_0 !

Mais la question n'est pas de savoir comment des auteurs qui connaissent manifestement des probabilités en arrivent à écrire de telles affirmations. Elle n'intéresserait guère, s'il en existait encore, que quelqu'anthropologue féru de transposition didactique (*). Le vrai problème est celui de comprendre pourquoi on peut être amené à penser, en toute bonne foi, qu'il est possible de quantifier le "risque d'erreur" et de lui affecter la seule valeur qui semble naturelle, compte tenu de la démarche suivie, et qui est évidemment celle qui sert à fixer le "seuil de risque" caractérisant le choix constitutif du test ...

La réponse à la question revient essentiellement à clarifier un malentendu fondamental entre statistiques et probabilités, ma-lentendu que les articles offerts aux enseignants et censés les aider à dominer cette partie des programmes se gardent bien de signaler, lorsqu'ils ne s'ingénient pas, tout bonnement, à l'obscurcir un peu plus…

(*) *Note de l'éditeur*. En faisant allusion à la théorie dite de la transposition didactique, qui se place sous l'égide de la "Théorie anthropologique du didactique", l'auteure de ces lignes semble sous-entendre qu'il y aurait donc «trahison du savoir savant» dans le passage incriminé au «savoir enseigné»... En tout état de cause, l'évocation de cette théorie de la fin du 20ème siècle, comme les références qui suivent à Poincaré ou à Bayes, confirment, s'il en était besoin, qu'il ne s'agit pas véritablement de la sœur de Blaise Pascal.

Soyons donc clairs et donnons donc une réponse claire. Celle-ci tient en une seule sentence : l'étude du *risque de se tromper* dans un test EST et N'EST PAS un problème de *probabilité des causes*.

Ecoutons pour commencer Henri Poincaré qui s'est donné la peine, un jour, d'expliquer à des béotiens en quoi consiste les notions de *probabilité des effets* et de *probabilité des causes* (c'était dans le contexte de l'affaire Dreyfus, pour laquelle il s'agissait de démonter les raisonnements quelque peu farfelus de l'accusation) :

> « Comme exemple de probabilité des effets, on choisit d'ordinaire une urne contenant 90 boules blanches et 10 boules noires. Si l'on tire au hasard une boule de cette urne, quelle est la probabilité pour que cette boule soit noire ? C'est évidemment 1/10.
>
> Les problèmes de probabilité des causes sont beaucoup plus compliqués, mais beaucoup plus intéressants.
>
> Supposons par exemple deux urnes d'aspect extérieur identique ; nous savons que l'une contient 90 boules blanches et 10 boules noires, et l'autre au contraire 90 boules noires et 10 boules blanches. Nous tirons au hasard une boule de l'une des urnes, sans savoir de laquelle, et nous constatons qu'elle est blanche. Quelle est la probabilité pour que ce soit dans la première urne que nous ayons puisé ?
>
> Dans ce nouveau problème, l'*effet* est connu, on a constaté que la boule tirée était blanche ; mais la *cause* est inconnue, on ne sait pas dans quelle urne on a fait le tirage. […]
>
> Ce sont donc les formules dites de *probabilité des causes* qu'il convient d'appliquer. La probabilité est de 9/10, mais c'est parce que nous supposons qu'il n'y a *a priori* aucune rai-

son pour qu'on soit tombé sur l'une des urnes plutôt que sur l'autre. Mais les choses auraient été bien différentes si nous avions eu 11 urnes, dont 10 composées comme la première et une seulement comme la seconde. *A priori*, la probabilité pour que l'on tombe sur une urne où les blanches dominent aurait été déjà grande, et les résultats auraient dû être notablement modifiés.

Pour pouvoir calculer, d'après un événement constaté, la probabilité d'une cause, il nous faut donc plusieurs données :

1° Il faut savoir quelle était *a priori*, avant l'événement, la probabilité de cette cause,

2° Il faut savoir ensuite quelle serait, pour chacune des causes possibles, la probabilité de l'événement constaté. (C'est ainsi que dans l'exemple cité il faut connaître la composition des urnes.) »

Chacun aura reconnu là une explication et une application classiques de la "formule de Bayes". Je ne rappelle ce point de vue que pour ré-expliquer le regard que porte naturellement un probabiliste sur un test : on ne peut *évaluer numériquement* une probabilité des causes que si on se donne l'ensemble des probabilités susceptibles de gouverner l'ensemble des causes possibles.

Effectuer un test revient à effectuer l'équivalent d'un tirage dans une urne, et prendre une décision revient à choisir — en fonction de la probabilité des causes — dans quel type d'urne on a tiré. Et dans ce contexte, toujours d'un point de vue probabiliste, la *mesure* du "risque de se tromper" n'est qu'une autre manière de désigner cette *probabilité a posteriori* de l'hypothèse que l'on a retenue ou non.

Seulement voilà, si on reprend sous cet angle l'exemple du test de la pièce de Buffon, on s'aperçoit très vite que :

1° dans la mesure où l'équilibre ou non de la pièce est représenté *a priori* par un nombre réel p compris entre 0 et 1, il devient nécessaire d'envisager ici une série de tirages dans une urne choisie au hasard *parmi une infinité d'urnes* qui contiendraient, chacune, une proportion de piles égale à l'une des valeurs possibles de p,

2° le résultat du test constitue un événement dont on cherche à connaître la cause sous une forme probabiliste, c'est-à-dire en affectant une probabilité (*a posteriori*) non pas à telle ou telle valeur de p, mais sous la forme d'une distribution correspondant à une loi continue : il n'est pas question de parier, par exemple, que $p = 0{,}50$ car cette possibilité a une probabilité nulle, on ne peut que déterminer des probabilités affectées à des intervalles,

3° une telle modélisation n'est possible que si on se donne à l'avance – *a priori*, si l'on préfère – une loi de probabilité sur les valeurs possibles du nombre p, destinée à traduire l'équilibre (ou plutôt les déséquilibres) de la pièce...

On voit notamment que ce point de vue inclut une assez grande part d'arbitraire : si on postule au départ que toutes les valeurs de p comprises entre 0 et 1 sont envisageables avec la même probabilité, on partira par exemple d'une *distribution a priori* qui sera uniforme... mais cela n'aura pratiquement aucun sens d'un point de vue purement mécanique ! il sera donc préférable de trouver une distribution plus satisfaisante... ce qui n'évite évidemment pas non plus d'introduire des hypothèses plus ou moins justifiées... En tout état de cause, on aboutit ainsi à une théorie relativement praticable, à condition de trouver ces valeurs *a priori* et de se lancer dans des calculs un peu plus compliqués... Le lecteur intéressé pourra se reporter, pour en savoir un peu plus, à la documentation consacrée à ce que l'on désigne généralement sous le nom de "statistiques bayesiennes".

Cela étant, il s'agit là de la seule piste légitime pour parvenir à une *quantification acceptable du risque de se tromper* en tirant une conclusion à partir d'un test. C'est-à-dire pour tenter d'exprimer par une *probabilité* le risque de conclure à tort (par exemple) que la pièce de Buffon est déséquilibrée ! Et, comme on le voit, il y a peu d'espoir que la probabilité de ce risque soit effectivement égale à 5 % et, pour tout dire, il est même peu probable que la probabilité annoncée soit véritablement crédible, étant donné la part d'arbitraire qui présiderait à son calcul "bayesien"...

« Certes, me direz-vous, vous nous exposez là les rudiments bien connus par tous ceux qui ont un peu compris la formule de Bayes des probabilités *a posteriori*, mais vous n'irez jamais jusqu'à pouvoir mettre en faux la phrase : « il n'y a pas plus de 5 % de chances de rejeter à tort l'hypothèse », puisque justement, *si l'hypothèse est vraie*, le calcul montre qu'il n'y a pas plus de 5 % de chances de tomber dans l'intervalle qui amène la décision de rejet ! Il y a là, bien sûr une double négation, mais chacun aura compris que cette affirmation est indéniable, et que l'on a d'ailleurs tout fait pour cela. »

Je vous répondrai alors ce que j'ai déjà dit au début de ces quelques pages. D'abord vous êtes bien trop modeste, car il y a en réalité 0 % de chances que vous rejetiez à tort l'hypothèse, car il n'y a *aucune chance* pour que la probabilité d'avoir pile ou face soit précisément égale à 0,5. Pas plus qu'un lanceur de fléchettes n'a de chances d'atteindre exactement le point géométrique situé au centre de la cible... Dès lors, vous auriez 100 % de chances d'accepter à tort le fait que la pièce soit équilibrée ! Les considérations les plus élémentaires en probabilité vous obligent à ne conclure que sous une forme du type « la probabilité de la pièce est *à peu près* 0,5 » et il vous faut quantifier la manière dont vous gérer cet *à peu près*, avant même de quantifier quoi que ce soit pour parler du *risque lié au fait de décider par rapport à cet à peu près*.

Ensuite rien ne vous permet, avec cette seule stratégie, de donner un sens crédible à la légitimité de votre test. Ce que vous affirmez c'est seulement la chose suivante :

> « en supposant la pièce équilibrée, j'ai moins de 5 % de chances de prononcer la phrase "je rejette l'hypothèse", et ce sera évidemment à tort, puisque je suppose la pièce équilibrée ! »

Mais, comme je l'ai laissé entendre plus haut, la seule chose que vous avez faite est de décider à l'avance d'un événement qui a une probabilité inférieure à 5 % de se produire, et qui vous amènera à prononcer la décision de rejet. Cette simple démarche n'explique en rien – et encore moins ne permet de mesurer – la légitimité de ce choix. Considérons, pour changer, un test exactement semblable à votre test initial, mais décidons que nous ne gardons comme "intervalle de risque" que la moitié droite de celui qui est schématisé sur la figure 1. Nous pourrons dire :

> « sachant que le calcul est fait sous l'hypothèse d'équilibre, sachant qu'il n'y a alors pas plus de 2,5 % de chances de tomber dans cet intervalle et de décider (à tort, évidemment) le rejet, je n'ai que 2,5 % de chances de décider (à tort, évidemment) de ce rejet ».

Bien sûr que cette phrase ne sera pas vraiment incorrecte, mais quel poids lui accorderez-vous pour la crédibilité de votre test, puisque chacun croira que celui-ci est rudement efficace, alors même que *plus la pièce sera déséquilibrée* d'un côté, moins vous aurez tendance à la rejeter ! La vérité c'est que votre affirmation ne fait que souligner le fait suivant : *sachant que la pièce est équilibrée* la probabilité de prononcer la phrase de rejet est de 5 %. C'est une phrase qui n'exprime rien d'autre qu'une *probabilité des effets*, et vous la faites passer pour une phrase exprimant une *probabilité des causes*. Elle est lue par tout un chacun sous la forme

> « je prends un risque inférieur ou égal à 5 % si je parie que la pièce n'est pas équilibrée ».

Lorsqu'elle n'est pas directement résumée sous la forme :

> « puisque la proportion observée est dans l'intervalle de risque, la pièce n'a pas plus de 5 % de chances de ne pas être équilibrée ».

Elle fonctionne donc sous la forme : « la probabilité de prononcer la phrase de rejet ET que ce soit à tort, est inférieure ou égale à 5 % ».

En définitive, la "base logique du raisonnement en statistiques" est effectivement non déterministe, mais elle n'est pas non plus exactement probabiliste. Alors même que pratiquement tous les textes qui développent les calculs liés à l'approximation normale des lois binomiales s'ingénient à laisser croire qu'il est possible d'associer une valeur numérique au r*isque de se tromper*. Les statisticiens ont accepté le fait que la vision purement bayesienne des probabilités des causes, comme dit Poincaré, reposait en fait sur une grande part d'arbitraire. Ils ont au contraire choisi une approche qui rend leurs tests crédibles en combinant les contraintes résumées sous le nom de "risques de première et seconde espèce".

Paradoxalement, votre article apporte tous les éléments permettant d'en mesurer l'importance... à condition de poser soi-même la problématique, d'en tirer soi-même les éléments pertinents... et de ne pas se laisser piéger dans la façon dont vous présentez d'entrée de jeu la question du risque, qui ne parvient qu'à brouiller les pistes!

C'est sans doute dans l'air du temps de vouloir donner le sentiment que les mathématiques sont capables de *mesurer les risques*, mais il est dommage — me semble-t-il — qu'un article destiné aux professeurs — et pour leur expliquer les secrets des statistiques inférentielles — ne prenne pas la peine d'en expliquer clairement les difficultés, en quelque sorte philosophiques, et prenne autant le risque d'encourager dans l'usage de formulations pour le moins maladroites...

Votre dévouée,
Gilberte Pascal