



La précognition est-elle démontrée ?

Avner Bar-Hen

► **To cite this version:**

Avner Bar-Hen. La précognition est-elle démontrée ?. Images des Mathématiques, CNRS, 2011, <http://images.math.cnrs.fr/La-precognition-est-elle-demontree.html>. <hal-00592120>

HAL Id: hal-00592120

<https://hal.archives-ouvertes.fr/hal-00592120>

Submitted on 11 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La précognition est-elle démontrée ?

Les pouvoirs paranormaux prouvés par la statistique ?

Le 27 avril 2011, par **Avner Bar-Hen**

Professeur à l'Université Paris Descartes. Président de la Société Française de Statistique (www.sfds.asso.fr). (**page web**)



*En général, quand quelqu'un dit avoir la preuve que les « dons psychiques » existent, telles la télépathie ou la prémonition, la réaction du monde académique se résume à un haussement d'épaule. Quand réaction il y a, évidemment. Le débat actuel a été lancé par Daryl J. Bem, professeur émérite [1] de psychologie à Cornell University dans *The Journal of Personality and Social Psychology* (une revue tout à fait respectable). Il décrit les expériences qu'il a faites afin, dit-il, d'amener ses collègues à au moins considérer la possibilité de l'existence de tels pouvoirs. Du coup, **les uns** crient au scandale dans le *New York Times*. **D'autres** à la mauvaise utilisation généralisée des statistiques en sciences sociales. Et d'autres encore subodorent un canular.*

UNE première version du papier est **disponible**. Notons que le *The Journal of Personality and Social Psychology* publiera en même temps une sévère critique de ses travaux, que l'on peut également **télécharger**, arguant essentiellement que les mauvais tests statistiques ont été utilisés et que des tests idoines invalident les résultats.

Mais que dit le texte de Bem, au juste ? Il relate neuf expériences, toutes inspirées de tests ou d'effets bien établis en psychologie, mais dont l'ordre chronologique des étapes a été renversé afin de mesurer une éventuelle faculté de précognition chez les sujets.

Par exemple : dans une de ces expériences, 100 individus, moitié masculins et moitié féminins, ont été réunis pour jouer à un jeu informatique dans lequel l'écran était séparé en deux, chaque moitié étant cachée par un "rideau". Il s'agissait d'essayer de deviner derrière lequel se cachait une image de nature érotique. Celle-ci pouvait être neutre ou choisie au hasard dans une base de données [2]. Le test est répété 36 fois pour chaque participant, soit 3600 fois en tout.

Les résultats publiés indiquent que, lorsque l'image était érotique (1560 cas), les sujets ont identifié derrière quel rideau elle se cachait dans 53.1 % des cas par comparaison avec les 50 % que l'on pourrait attendre. Le taux de succès n'est que de 49.8% avec les photos non-érotiques.

Si la précognition n'existe pas, ce test revient à tirer à pile ou face 3600 fois. On s'attend grosso modo à 50% de réussite. Notons que même dans un cas parfaitement aléatoire, il est

exceptionnel d'obtenir exactement 50% de « pile » et 50% de « face » : ça n'arrive que moins d'une fois sur 75. Il y aura donc en général un écart à 50% et il faut décider si cet écart « doit » faire penser qu'on n'a pas affaire au « hasard pur ».



Calcul de la probabilité d'exactly 50% de « face »

Si on considère tous les résultats possibles pour 3600 tirages, on a 2^{3600} cas au total. Les cas donnant 1800 « face » et 1800 « pile » exactement, sont au nombre de $\binom{3600}{1800}$ (le nombre de façon de choisir les 1800 tirages « face » parmi les 3600 au total). La probabilité est donc

$$\binom{3600}{1800} / 2^{3600} \approx 0.01329 \dots$$



Pour M. Bem, les résultats montrent que les sujets ont pu identifier derrière quel rideau se cachait la photo érotique dans 53.1 % des cas. Ce pourcentage chute avec les photos neutres, démontrant ainsi que les cobayes répondaient aux stimuli érotiques, avant que ceux-ci n'apparaissent.

La conclusion proposée par M. Bem est l'existence d'une influence rétroactive d'un événement futur sur les réponses présentes d'un individu, c'est-à-dire la possibilité d'anticiper un événement futur qui ne pourrait être anticipé par un procédé connu d'inférence. Ce processus (anormal) de transfert d'information ou d'énergie, inexplicable en termes de mécanismes physiques ou biologiques, est appelé effet psi et comporte deux variantes : La précognition (perception d'événements futurs qui prend la forme d'un pressentiment, ie une impression diffuse et lente que quelque chose va se produire) ou la prémonition (expérience spontanée qui implique une action à venir dont le sujet est partie, par exemple certains rêves). Si un don de voyance existe, explique M. Bem, il doit avoir été retenu par la sélection naturelle — et la capacité de « pressentir » un danger ou une occasion de relation sexuelle donnerait un avantage évolutif certain à d'hypothétiques precogs, comme dit John Anderton dans *Minority Report* (ou Bill Murray dans *Ghostbusters*).

Je ne veux surtout pas rentrer dans le débat sur la partie psychologique. Mais essayons de revenir sur la partie statistique et tout particulièrement sur la notion fondamentale de test statistique. La théorie des tests est à la base de nombreuses études médicales, sociales, économiques, etc. Cette controverse est pour nous une occasion de revenir là-dessus.

Un détective probabiliste

Par exemple, imaginons que nous sommes détectives dans un casino et que nous observons depuis son entrée un joueur. Il a parié cent fois de suite sur « rouge » ou « noir » avant le lancer la bille de la roulette. Il a gagné 58 de ces 100 paris. Est-il raisonnable d'attribuer ce surcroît de réussite au hasard ou faut-il expulser ce joueur de votre casino et changer de roulette ?

Pour simplifier on considère qu'il y a autant de numéros rouges que de numéros noirs (ce qui est le cas au casino) et que tous les numéros ont des couleurs (et donc on ne prend pas en compte les subtilités liées à l'absence de couleur du zéro). On suppose aussi que le jeu est honnête, c'est-à-dire que la bille a la même probabilité de s'arrêter sur chacun des numéros. La probabilité que la bille s'arrête sur un numéro rouge est donc de $1/2$. On peut alors calculer la probabilité de réussir 59 fois ou plus sur 100 dans ce cas.

La probabilité d'obtenir un tel résultat sous l'hypothèse « pas de triche » est de 4%.



Calcul de la probabilité

On ajoute les probabilités d'avoir 59, ..., 100 succès. Chacune des probabilités peut être calculée comme la probabilité pour 50 et il faut faire la somme. On peut le faire sans difficulté sur un tableur ou avec un logiciel scientifique (ou à la main en étant un peu plus astucieux).

La théorie des probabilités nous dit que si on expulse tout joueur ayant gagné 59 fois ou plus sur 100, on ne commettra d'injustice que pour moins de 4% des innocents. Si la politique du casino est de se protéger dès que le risque d'une telle injustice à l'encontre d'un innocent est inférieur à 5%, alors il faut expulser ce joueur ! Par contre la probabilité de gagner au moins 58 fois sur 100 ou plus étant sous l'hypothèse « pas de triche » de 6.6% : le risque d'expulser à tort étant supérieur à 5% (risque admissible défini par le casino), on n'expulsera pas.

Plus généralement

Un test statistique considère une hypothèse de référence (appelée habituellement H_0 , lu « H zéro »), et examine si les mesures faites ont pu arriver sous cette hypothèse par le simple jeu du « hasard ». Plus précisément, si la probabilité p d'observer de telles mesures est trop faible, on rejettera l'hypothèse de référence au profit de sa négation (appelée H_1), la tricherie dans l'exemple du casino, la « précognition » dans l'expérience citée (c'est-à-dire l'influence du futur).

En ce sens il y a une analogie entre un test d'hypothèse et un procès : tout suspect est présumé innocent et l'accusation doit apporter la preuve de sa culpabilité avant que la justice décide de le condamner. Cette preuve doit de plus s'appuyer sur des éléments concrets. Dans le cas du test ce sont les données qui permettent ou non de rejeter l'hypothèse. Quand on accepte H_0 , on ne prouve pas qu'elle est vraie, on accepte de conserver H_0 parce qu'on n'a pas pu accumuler suffisamment de preuves contre elle. Accepter H_0 c'est *acquitter faute de preuve*.

Construire un test de l'hypothèse de référence H_0 contre l'hypothèse alternative H_1 , c'est donc établir un critère de décision permettant de choisir entre l'hypothèse H_0 et H_1 : on conserve H_0 sauf si les données conduisent à la rejeter, c'est-à-dire que les données ne soutiennent pas l'hypothèse H_0 .

Être acquitté n'est pas tout à fait la même chose qu'être innocent et il n'y a pas symétrie entre l'hypothèse de référence et l'hypothèse alternative. Un procès dont l'hypothèse de base est la culpabilité du suspect peut être considéré comme douteux. On peut même noter que dans ce cas il vaut mieux ne pas mener d'enquête car il y aurait un risque de trouver des preuves qui pourrait innocenter le suspect. Pour revenir à l'exemple qui nous intéresse, si l'hypothèse de référence est de croire à l'existence de la précognition, il ne faut surtout pas faire d'expérience car cela risquerait de remettre en cause l'hypothèse de référence. Cette manière de faire serait considérée, avec raison comme douteuse, et certainement pas scientifique.

On détermine la loi de la variable aléatoire d'étude (par exemple le nombre de paris réussis) sous H_0 [3]. L'ensemble des valeurs observées pour lesquelles l'hypothèse de référence H_0 est admissible forme par définition la région d'acceptation A (ou plus exactement de non-rejet de H_0) et les autres valeurs constituent la région de rejet R .

La détermination de la région d'acceptation dépend donc d'un risque d'erreur. Dans l'exemple du casino, au seuil 5% on rejettera l'hypothèse H_0 (« pas de triche ») au profit de l'hypothèse H_1 (tricherie) si le joueur gagne 59 fois ou plus sur les 100 lancers. La région d'acceptation est donc $\{0, 1, 2, \dots, 57, 58\}$ et la région de rejet est $\{59, \dots, 99, 100\}$.

Notons aussi que l'on suppose (comme dans le cas de la précognition) qu'un tricheur triche pour gagner et donc la zone de rejet ne comporte que les valeurs « anormalement » élevées de succès.



Test unilatéral ou test bilatéral

Imaginons que l'on veuille tester la qualité d'une balance. Pour cela nous allons peser des masses dont le poids est connu.

Considérons dans un premier temps que la question est posée par un fabricant. La question est de savoir la balance donne le bon poids (hypothèse de référence H_0) et tous les écarts doivent être pris en compte car ils constituent des preuves d'un dérèglement de la balance (hypothèse H_1).

Si nous considérons maintenant que la question est posée pour savoir si un commerçant est honnête (hypothèse H_0) ou cherche à tromper les clients (hypothèse H_1), seuls les poids sur-estimés doivent être pris en compte car une sous-estimation du poids ne peut être considérés comme une preuve de malhonnêteté du commerçant.

Dans le premier cas on parle de test bilatéral : la zone de rejet de l'hypothèse principale est située des deux côtés de la distribution de probabilité de référence.

Dans le deuxième cas on parle de test unilatéral : la zone de rejet de l'hypothèse principale est située d'un seul côté de la distribution de probabilité de référence.

Dans le cas d'un test bilatéral, le risque d'erreur (déclarer à tort que la balance n'est pas juste) doit être calculé pour les poids sur-estimés et sous-estimés alors que dans le cas du test unilatéral (déclarer à tort que le commerçant est malhonnête) le risque d'erreur ne doit être calculé que pour les poids sur-estimés. Par symétrie le risque d'erreur dans le test bilatéral est le double du risque d'erreur du test unilatéral.

Pour revenir à l'expérience du casino, l'hypothèse « pas de triche » est unilatéral (on ne peut considérer comme de la triche de perdre très souvent) par contre si la question était « les paris sont-ils non biaisés », nous serions dans le cas d'un test bilatéral.

Si le but était de savoir si le jeu est équilibré entre le joueur et la banque, il faudrait donc construire une zone de rejet qui comprenne les nombres de succès trop grands et les nombres de succès trop petits. On peut calculer que l'on a moins de 5% d'erreur si on rejette H_0 « jeu équilibré » quand le nombre de succès est inférieur à 39 ou supérieur à 61.

On voit que la zone de rejet est différente suivant le caractère bilatéral ou unilatéral du test. Autrement dit le choix précis des deux hypothèses H_0 et H_1 influe sur le choix de cette zone bien que seul H_0 intervienne dans le calcul du niveau de signification d'une zone de rejet donnée.

Le hasard de l'échantillonnage peut fausser les conclusions. Quatre situations doivent être envisagées :

1. l'acceptation de l'hypothèse de référence alors qu'elle est vraie ;
2. le rejet de l'hypothèse de référence alors qu'elle est vraie ;
3. l'acceptation de l'hypothèse de référence alors qu'elle est fautive ;
4. le rejet de l'hypothèse de référence alors qu'elle est fautive.

Dans le premier et le dernier cas, la conclusion obtenue est correcte, mais non dans les deux cas intermédiaires.

Pour reprendre l'analogie du procès, le deuxième cas correspond à l'erreur judiciaire : rejet à tort de l'hypothèse innocence et le troisième cas correspond au cas de l'acquiescement à tort par manque de preuve.

L'erreur qui consiste à rejeter une hypothèse vraie est appelée **erreur de première espèce** (c'est-à-dire, une expulsion injuste par le casino).

La probabilité de rejeter à tort H_1 est appelée **erreur de seconde espèce**.

En résumé, nous avons le tableau suivant : la première colonne considère le cas où H_0 est vraie, la seconde ligne le cas où on affirme que c'est H_1 . Les cases de la diagonale correspondent donc à une bonne réponse et les deux autres à des erreurs.

| | | Réalité | |
|----------|-------|-----------------------|-----------------------|
| | | H_0 | H_1 |
| Décision | H_0 | OK | erreur de 2ème espèce |
| | H_1 | erreur de 1ère espèce | OK |

Pratiquement, on se donne une limite supérieure du risque de première espèce, le plus souvent 5% (appelé *par convention* « significatif »), 1% (« très significatif ») ou 1 pour mille (« hautement significatif »). Cette limite est appelée niveau de signification du test et permet de définir la condition de rejet de l'hypothèse de référence. Le choix d'une région de rejet de H_0 (c'est-à-dire l'ensemble des valeurs qui conduit à rejeter H_0) assez petite permet de réduire le risque de première espèce. La règle étant définie, on peut regarder ce qui se passe au niveau des données : quelle est la probabilité (souvent notée p) d'observer une telle configuration de données sous l'hypothèse H_0 . Dans l'exemple de l'article de Bem, on calcule donc la probabilité d'obtenir au moins 53.1% de succès sous l'hypothèse

d'avoir une chance sur deux de découvrir la bonne image. Cette probabilité p est appelée le niveau de signification *réel* du test et donne la probabilité de rejeter à tort H_0 . On rejette alors l'hypothèse de référence au niveau de signification choisi (par exemple 0.05) si (et seulement si) le niveau de signification réel p est inférieur ou égal au niveau de signification choisi. Dans le cas de l'article de Bem, sous l'hypothèse d'absence de précognition la probabilité de trouver l'image érotique est de 50%, on observe 53.1% de succès parmi les 1560 images. Ce résultat s'obtient en considérant une **loi binomiale** de paramètres 1560 et 0.5. On calcule une probabilité de dépassement d'environ $p = 0.008$. Cette probabilité étant inférieure à 5%, on rejette l'hypothèse de référence et Bem conclut à la réalité de la précognition.

On note donc que p est le risque réel de première espèce et qu'il peut être sensiblement inférieur au risque de première espèce choisi. Cette attitude est donc conservatrice. Dans l'exemple du casino le risque choisi était de 5% mais le risque réel p est d'environ 4%.

Un bon test possède des risques de première et de deuxième espèce faibles. Attention : ces risques ne sont pas indépendants, mais on doit choisir un compromis. Si on rejette toujours H_0 le risque de deuxième espèce est nul (un procès qui condamne toujours ne laisse pas échapper de coupable) mais l'erreur de première espèce sera importante (beaucoup d'innocents sont condamnés). Réciproquement un procès qui innocenté toujours l'accusé ne condamne pas d'innocents (risque de première espèce égal à zéro) à un risque de deuxième espèce important (les coupables innocenté-e-s).

Limite du principe du test

La théorie des tests statistiques est extrêmement puissante et utilisée. Elle permet dans une large mesure (et au prix d'efforts empiriques certains) de tester des hypothèses en présence de beaucoup d'inconnues (par exemple l'effet de tel médicament). On ne sera pas étonné s'il s'agit probablement des théorèmes mathématiques les plus souvent utilisés... voire abusés. C'est pourquoi il est important de comprendre les limites intrinsèques des tests statistiques. Voici un rapide florilège.

Une différence significative n'équivaut pas à une différence ayant du sens.

Considérons le cas de la discrimination contre les femmes à l'embauche et supposons que l'on désire un niveau de signification de 5%. Dans une entreprise de 10 employés, l'hypothèse H_0 d'avoir la même probabilité d'embaucher un homme ou une femme est à rejeter s'il y a deux femmes ou moins parmi les 10 employés. Dans une entreprise de 1000 employés, on rejette cette hypothèse à partir de 527 hommes (pour un niveau de signification inchangé à 5%).

Notons que le test est unilatéral car une entreprise n'employant que des femmes ne pourra être considérée comme pratiquant la discrimination contre les femmes. Si la question était de savoir si il y a une discrimination contre les hommes ou les femmes, le test serait bilatéral car tous les écarts à la parité seraient considérés comme des « preuves ».

Détail

Soit Z le nombre de femmes parmi les n employés. Sous H_0 , Z est distribué comme une loi binomiale de paramètres n et $p = 0.5$, Sous H_1 , Z est distribué comme une loi binomiale de paramètres n et $p < 0.5$. Ce sont les petites valeurs de Z qui conduisent à rejeter H_0 de discrimination et le test est donc unilatéral. Dans le premier cas $n = 10$, dans le deuxième $n = 1000$.

Dans l'entreprise de 10 salariés l'hypothèse H_0 (de référence) de non-discrimination est rejetée s'il y a moins de 20% de femmes, alors que dans l'entreprise de 1000 salariés, l'hypothèse H_0 est rejetée s'il y a moins de 47.4%. En effet, les fluctuations dues au hasard [4] diminuant avec la taille de l'échantillon, de plus petits écarts entre les chances d'embaucher un homme ou une femme sont détectés. On peut donc avoir des écarts statistiquement significatif (à un seuil donné) mais qui n'ont que peu de sens en pratique (la législateur peut décider de ne pas punir un biais de 2.6% seulement). On observe plus souvent des règles considérant qu'il y a discrimination s'il y a moins de 40% de femmes indépendamment de la taille de l'entreprise.

Multiplicité des tests : quand il y en a un (test) ça va. C'est quand il y en a beaucoup qu'il y a des problèmes.

Un autre problème dans cet article provient de la répétition des tests. Cette question est bien connue dans les problèmes en grande dimension (génomique ou traitement d'image par exemple). Imaginons que l'on fasse 100 000 tests (pour détecter des gènes jouant un certain rôle ou des pixels erronés pour garder les deux exemples). Considérons que pour chaque test on rejette l'hypothèse de référence si le risque d'erreur est inférieur à 5%. Le nombre d'erreurs peut être vu comme un exemple de loi binomiale de paramètres 100 000 et 0.05 et donc le nombre « moyen » d'erreurs (appelé également **espérance mathématique** du nombre d'erreurs) est de 5000 erreurs. Pour un phénomène rare, le taux d'erreur final est beaucoup trop élevé pour être acceptable : on peut avoir beaucoup plus de fausses alertes que de vraies détections. Les tests multiples sont un sujet actuel de recherches actives et un lecteur intéressé peut par exemple consulter un **article de base**.

Pourquoi 5% ?

On peut discuter sans fin sur le seuil de 5% mais il semble que les raisons soient surtout historiques : jadis il fallait préparer des tables donnant les valeurs de statistique (par exemple le nombre de succès) maximales admissibles en fonction de la taille de l'échantillon (le nombre de paris) pour chaque niveau de test choisi. Ce problème calculatoire n'est plus d'actualité et il est donc préférable d'indiquer le niveau du test que de fixer une limite (par construction arbitraire) qui va décider si le test est ou non significatif (significatif dans le sens que le risque d'erreur de première espèce est assez faible si on rejette H_0).

Dans la critique de l'article on note que les auteurs proposent une autre approche dite **bayésienne** (cette approche nécessite également des choix tout aussi contestables mais c'est une autre histoire... peut-être l'objet d'un prochain article dans cette rubrique !)

Quelle conclusion ?

Comme indiqué au début, je ne souhaite pas rentrer dans le débat sur la **précognition** mais de toutes les façons cette expérience a besoin d'être à nouveau effectuée par d'autres chercheurs, nombreux à s'intéresser à ces travaux et à vouloir les reproduire. On verra bien alors leurs résultats. Au passage on peut noter que la Cour Suprême américaine a récemment jugé que la preuve ne pouvait uniquement s'appuyer sur un niveau de signification statistique (voir par exemple **ici** ou **là**)

La deuxième conclusion qu'on entend classiquement est qu'on peut faire dire ce qu'on veut à la statistique. On peut cependant noter que si Mr Bem avait construit une voiture beaucoup de gens auraient eu des doutes sur ses capacités et peu seraient montés dedans car on ne s'improvise pas ingénieur. Peut-être peut-on étendre cette idée et se dire que statisticien est un métier et que M. Bem aurait pu se demander s'il n'aurait pas été utile de collaborer avec un statisticien. C'est la collaboration entre les précogs et John Anderton qui fait la force de la « Précrime ».

P.S. :

La rédaction d'Images des maths, ainsi que l'auteur, remercient pour leur relecture attentive, les relecteurs dont le pseudonyme est le suivant : Thierry Monteil, François Sauvageot et Claire Lacour.

Notes

[▲1] c'est-à-dire retraité mais actif.

[▲2] Les psychologues disposent pour ce faire de bases de données étalonnées !

[▲3] C'est souvent difficile d'avoir la loi sous H_1 il faudrait comprendre le mécanisme de la tricherie ou de la précognition.

[▲4] Les fluctuations sont mesurés techniquement par la notion de variance.

► Crédits images

Pour citer cet article : **Avner Bar-Hen**, **La précognition est-elle démontrée ?**. *Images des Mathématiques*, CNRS, 2011. En ligne, URL : <http://images.math.cnrs.fr/La-precognition-est-elle-demontree.html>