## Chapitre 12. Beaucoup de bruit pour rien ?

Ayant démontré l'inconsistance de l'argument central du rapport d'enquête, reprenons néanmoins l'affaire à la base. Ce qui étonne les enquêteurs, comme ils l'exprimeront à nombreuses reprises, est la faible variabilité selon eux des comptes de basophiles en double ou en triple qui sont transcrits dans le cahier d'expérience d'E. Davenas – ils sont « fabriqués » (*made up*) avait dit W. Stewart. Le même reproche est fait à propos des résultats de l'article de *Nature* de 1988 (correspondant aux expériences réalisées en Israël). Pourquoi – en dehors de tout calcul – les résultats apparaissent-ils souvent « trop bons » ? Ici également, certaines considérations mathématiques vont nous permettre de considérer cette question sous un angle différent.

*Les conséquences trompeuses de l'asymétrie*

Tout d'abord, qu'entend-on ici par des comptes « trop bons » ? Supposons que l'on ait réalisé trois comptes successifs à partir d'un récipient qui contient (de façon certaine) 100 basophiles (pour l'unité de volume considérée). C'est, pour reprendre la terminologie des statisticiens, une urne à partir de laquelle nous procédons à des « tirages » (chaque tirage permet d'obtenir un « échantillon » de l'urne). Nous procédons à trois tirages successifs : nous trouvons 99, 101 et 113. Nous pouvons calculer la moyenne, la variance et l'écart-type (c'est la fameuse erreur d'échantillonnage ou *sampling error* qui est la racine carrée de la variance). Le calcul donne une moyenne de 104,3 avec un écart-type de 7,6 et une variance de 57,3. En général on exprime ce résultat de la façon suivante : moyenne ± écart-type = 104 ± 7,6 (n=3)

Ce que nous cherchons à atteindre est le nombre de basophiles dans le tube initial. Nous en avons ici une approximation. On conçoit que la confiance que nous pouvons avoir dans ce résultat est d'autant plus grande que le calcul a été fait sur un nombre important de tirages. L'écart-type (erreur d'échantillonnage) nous donne une idée de la variabilité entre les différents comptes. S'agissant d'un dénombrement nous savons que moyennant un certain nombre de conditions (voir chapitre précédent), la loi qui s'applique *a priori* est la loi de Poisson. Comme nous l'avons vu, les variables distribuées selon la loi de Poisson ont une variance qui est égale à la moyenne.

Reprenons le calcul ci-dessus obtenu à partir de trois tirages. Sa variance (57,3) est inférieure à sa moyenne (104,3). A-t-il pour autant été « fabriqué » ? Pas forcément, car cette variance fluctue elle-même au hasard. Mais dans quelles limites ? C'est ce que nous allons envisager. Nous supposerons pour cela que nous sommes dans des conditions idéales et que seul le hasard est responsable des résultats obtenus. En d'autres termes, nous supposons qu'il n'y a pas de bruit statistique surajouté à la loi de Poisson.

Nous adoptons la procédure suivante : nous prélevons 3 échantillons et nous calculons ensuite la moyenne et la variance de ces 3 comptes. Nous reproduisons la même opération jusqu'à obtenir 1000 moyennes de 3 valeurs et leurs 1000 variances.

Parmi les 1000 moyennes et leurs 1000 variances correspondantes, quel est le pourcentage des variances ($s^2$) qui seront supérieures à la moyenne (m) et quel sera le pourcentage des variances inférieures à la moyenne, c'est-à-dire :

Pourcentage de comptes triples avec $m > s^2$ (équivalent à $s^2/m < 1$) ?

Pourcentage de comptes triples avec $m < s^2$ (équivalent à $s^2/m > 1$) ?

La première réponse qui vient à l'esprit est : 50 % pour (1) et 50% pour (2). Notre intuition nous suggère qu'étant données les fluctuations, on aura du fait de la loi des grands nombres à peu près autant de valeurs d'un côté que de l'autre.

Sommes-nous si sûrs de ce résultat ? Réalisons une simulation numérique avec des nombres aléatoires générés selon la loi de Poisson. Nous obtenons ainsi 1000 comptes triples (choisis tels que leur moyenne soit 100 et par conséquent leur variance est égale également à 100). Nous calculons ensuite le rapport $s^2/m$.

| | Compte 1 | Compte 2 | Compte 3 | $s^2/m$ |
|---|---|---|---|---|
| Série 1 | 93 | 90 | 117 | 2,19 |
| Série 2 | 97 | 108 | 112 | 0,571 |
| Série 3 | 104 | 107 | 108 | 0,041 |
| Série 4 | 112 | 115 | 110 | 0,056 |
| Série 5 | 99 | 84 | 105 | 1,219 |
| Série 6 | 129 | 95 | 97 | 3,402 |
| Série 7 | 110 | 76 | 97 | 3,12 |
| …………… | | | | |
| Série 1000 | 99 | 99 | 94 | 0,086 |

Nous pouvons maintenant représenter graphiquement sous forme d'un nuage de points ces 1000 valeurs $s^2/m$ puis étudier leur distribution en classes :

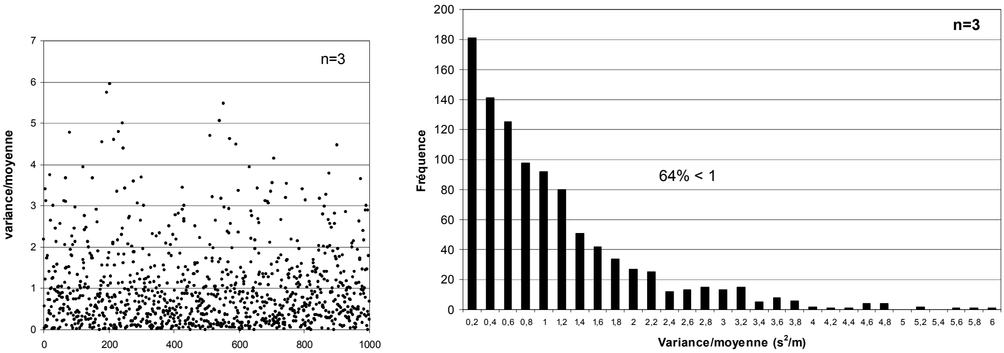


Figure 12.1 Distribution du rapport variance/moyenne de petits échantillons (n=3).
*NB.* Pour cet histogramme et les suivants, chaque valeur en abscisse correspond à la borne supérieure de l'intervalle considéré.

Notre intuition a donc été prise largement en défaut puisque nous trouvons que la loi de répartition de $s^2/m$ est totalement asymétrique. Nous trouvons que la répartition de part et d'autre de la moyenne n'est pas 50/50 mais 64/36. Qui plus est, les valeurs les plus probables correspondent aux valeurs de $s^2/m$ les plus faibles ! *La moyenne des 1000 rapports $s^2/m$ est néanmoins proche de 1* conformément à la loi de Poisson.

Notre intuition (et notre mauvaise connaissance des lois statistiques) nous avait dans un premier temps amené à confondre la *moyenne* d'une variable et son *mode* (c'est-à-dire la valeur qui est la plus fréquente). La moyenne ne correspond au mode que dans le cas des lois de probabilité symétriques. La conclusion paradoxale (et souvent mal comprise car contre-intuitive) est que – du fait de l'asymétrie du rapport variance/moyenne pour les petits échantillons – *la variance est plus fréquemment inférieure à la moyenne.* C'est un résultat fondamental. Nul doute qu'il fera sourire les statisticiens et les mathématiciens pour qui c'est probablement une évidence. Je ne suis pas certain que cette « évidence » était partagée au sein de l'équipe des enquêteurs et – soyons honnêtes – parmi les membres de l'équipe de Clamart.

Poursuivons notre exploration et voyons ce qui se passe dans le cas de comptes en double :

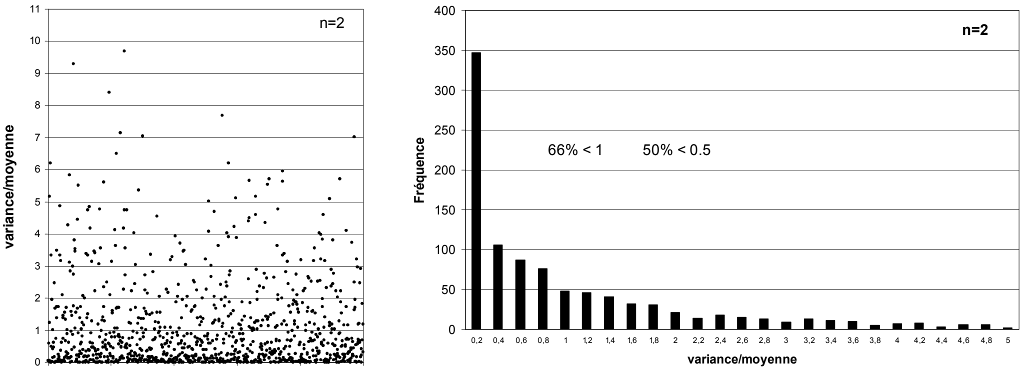


Figure 12.2. Distribution du rapport variance/moyenne de petits échantillons (n=2)

La différence est encore plus accentuée : Dans 66% des cas, les variances sont inférieures à la moyenne. Et dans la moitié des cas le rapport variance/moyenne est inférieur à 0,5. Nous terminerons notre exploration avec n = 10 (Figure 12.3).

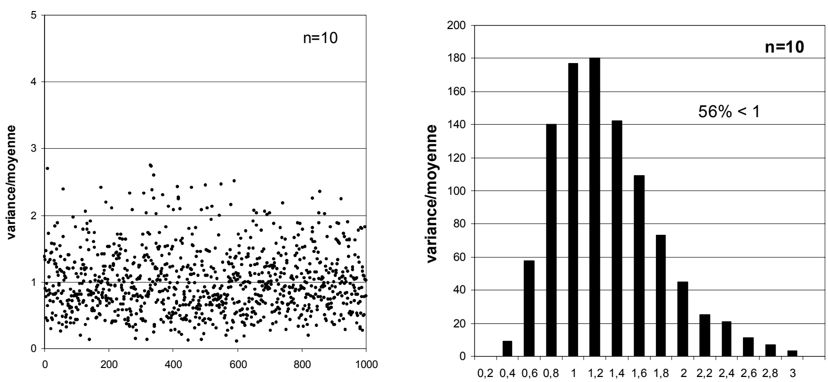


Figure 12.3. Distribution du rapport variance/moyenne de petits échantillons (n=10)

Avec des échantillons de 10 valeurs, on tend vers une loi de distribution plus symétrique (mais nous avons encore 56% des valeurs inférieures à 1). En pratique pour des échantillons supérieurs à 30 la distribution est symétrique (gaussienne).

En résumé nous constatons que la distribution des *variances des petits échantillons* (n=2, 3,…) est fortement asymétrique (ici nous sommes partis d'une loi de Poisson, mais une loi gaussienne donnerait des résultats similaires). La conséquence est que si l'on essaye de vérifier la conformité de comptages à partir des variances de petits échantillons (comme c'est souvent le cas), on risque de conclure que les résultats sont « trop bons ». Voici par exemple une simulation informatique de 10 comptes de basophiles conformes à la loi de Poisson :

Il s'agit de puits qui sont censés contenir le même nombre de basophiles (100). Chaque résultat est donné par la moyenne ± écart-type.

| | | | |
|---|---|---|---|
| Puits 1 : | 117 ± 6 | Puits 6 : | 99 ± 3 |
| Puits 2 : | 92 ± 2 | Puits 7 : | 110 ± 8 |
| Puits 3 : | 101 ± 13 | Puits 8 : | 106 ± 16 |
| Puits 4 : | 95 ± 6 | Puits 9 : | 96 ± 8 |
| Puits 5 : | 94 ± 3 | Puits 10 : | 93 ± 5 |

Mentalement, nous calculons les variances en prenant le carré de l'écart-type. Nous constatons qu'à part pour les puits 3 et 8, la variance est très souvent inférieure à la moyenne et fréquemment très faible. La suspicion nous gagne. Nous avons appris en effet que dans ce genre de comptages la variance *doit* être égale à la moyenne et que c'était d'ailleurs un moyen de vérifier que des comptages étaient sans biais. Les valeurs auraient-elles été « arrangées » ? Pour les variances supérieures à la moyenne, on pourrait imaginer que les volumes prélevés n'étaient pas exacts ou toute autre explication (bruit statistique). Mais pour les variances inférieures à la moyenne, la seule explication est que « de l'ordre a été introduit ». Rappelons que ces considérations sont élaborées à partir de valeurs générées consécutivement par un programme informatique (elles n'ont pas été choisies).

*Application aux résultats d'Israël de février-mars 1987 (cas ou n=3)*

Forts de nos nouvelles connaissances, reprenons maintenant les comptes de basophiles si contestés et calculons les rapports $s^2/m$ (variance / moyenne) du tableau 1 de l'article de *Nature* (ce sont les 4 expériences faites en Israël par E. Davenas et dont les résultats sont en annexe 2) puis déterminons la distribution des valeurs :

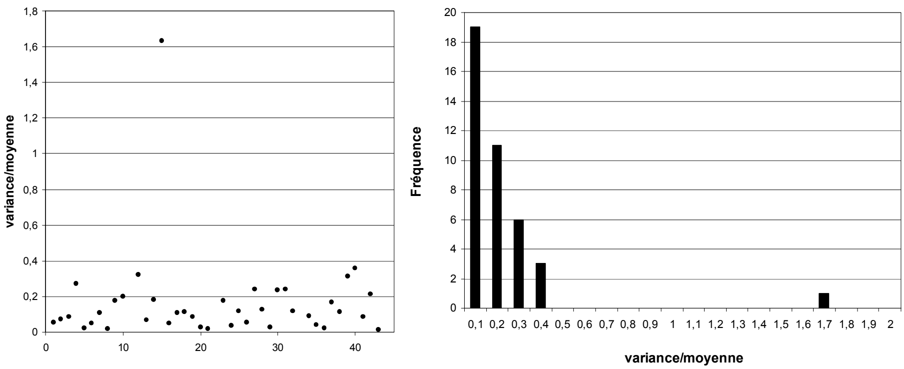


Figure 12.4. Distribution du rapport variance/moyenne des expériences d'Israël de février-mars 1987 (petits échantillons avec n=3)

Nous retrouvons donc la même allure fortement asymétrique avec les probabilités les plus élevées pour les classes de valeur les plus faibles. Il est difficile de « construire » ces résultats. Le lecteur peut essayer de simuler des résultats en inventant des comptes triples, il constatera qu'il est très difficile d'obtenir une telle distribution, surtout si on n'a pas à l'esprit la nécessité d'obtenir une distribution asymétrique. C'est – pour ceux qui en auraient douté – un argument en faveur de la « sincérité » des comptages réalisés en Israël.

Nous faisant l'avocat du diable, nous constatons que la moyenne des rapports variance/moyenne n'est pas 1, mais seulement 0,16. Toutefois, en prenant en compte les résultats de l'article de H. Gérard *et al*, le rapport moyenne/variance pour environ 80 basophiles devrait être nettement inférieur à 1, de l'ordre de 0,34 en moyenne, ce qui nous rapproche de 0,16 sans toutefois l'atteindre.

Il n'est pas impossible que certaines valeurs considérées comme aberrantes aient été recomptées, « par précaution », car s'écartant trop des deux autres comptes. En d'autres termes, on ne peut objectivement écarter un « effet expérimentateur » sur les comptes triples. En moyenne, cette procédure *ne change rien* au résultat de l'expérience car les valeurs qui s'écartent « trop » se produisent avec une probabilité égale dans un sens ou dans l'autre. A partir du moment où on ne connaissait pas « l'étiquette » des dilutions testées (expériences *à l'aveugle*) ces résultats ne pouvaient être biaisés dans le sens d'une plus grande production de résultats en faveur d'un effet des hautes dilutions. Car, rappelons enfin que le but de ces expériences *n'était pas de vérifier la validité de la loi de Poisson sur les comptes répétés* mais d'évaluer une éventuelle différence entre les tubes « actifs » et les tubes « inactifs » et ceci avec une méthode la plus précise possible.

*Application aux résultats de l'enquête de* Nature *de juillet 1988 (cas où n=2)*

Etudions comme ci-dessus la distribution du rapport $s^2/m$ pour l'expérience qui avait été comptée en double et à l'aveugle (cas où n=2). Il s'agit de l'expérience F comptée le 7 juillet (voir annexe).

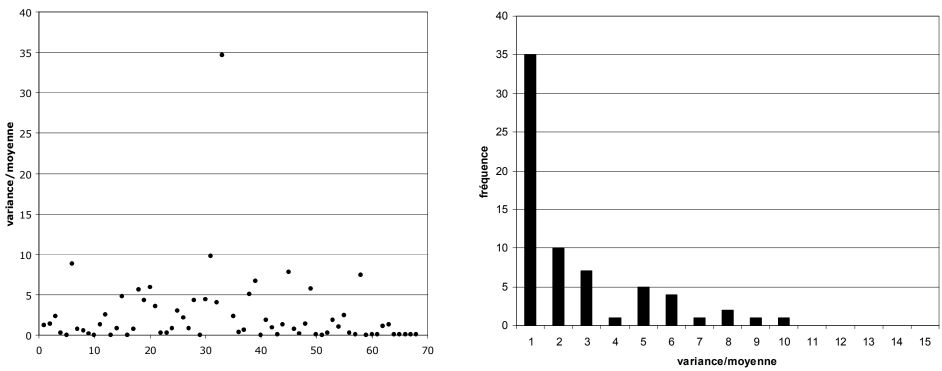


Figure 12.4. Distribution du rapport variance/moyenne de l'expérience F réalisée pendant l'enquête de Nature de juillet 1988 (petits échantillons avec n=2)

On remarque immédiatement la différence d'échelle du rapport variance/moyenne en comparaison avec les expériences d'Israël ou avec les expériences simulées pour n=2. Comme on peut le voir, le problème n'est plus ici d'expliquer des variances trop faibles par rapport à la moyenne mais d'expliquer des variances *beaucoup trop élevées* par rapport à la moyenne ! Le bruit statistique surajouté est manifestement très grand (la moyenne du rapport variance/moyenne est 2,4). Ce dernier peut provenir d'erreurs dans le pipetage des volumes ou d'une remise en suspension de l'échantillon qui était défectueuse.

Redisons, au risque d'ennuyer le lecteur, 1) que ces étapes techniques avaient été prises en charge par W. Stewart et 2) que c'est cette expérience qui a servi à construire la fameuse courbe censée décrire la distribution gaussienne attendue lorsque les comptes étaient faits à l'aveugle.

*Existe-t-il un phénomène physique qui pourrait expliquer la variance trop faible des comptes de basophiles ?*

Nous avons vu dans le chapitre précédent que si les particules (cellule, bactéries…) que l'on compte ont tendance à se repousser alors elles sont sous-dispersées et la variance des comptes observée est plus faible que la variance attendue. Par quel mécanisme, les basophiles pourraient-ils avoir tendance à se maintenir à distance les uns des autres ?

Pour expliquer cette « anomalie », nous devons évoquer un mécanisme qui concerne les basophiles mais pas les autres cellules (rappelons qu'ils ne représentent que 1% des globules blancs dans le test de dégranulation) et si possible seulement les basophiles qui sont comptés, c'est-à-dire les basophiles colorés (non activés).

La solution pourrait résider précisément dans la coloration des basophiles. En effet, le bleu de toluidine colore les basophiles de façon particulière : le colorant est bleu mais les basophiles sont colorés en rouge. On parle alors de métachromasie. D'une façon plus générale, la métachromasie est la propriété de certains colorants qui colorent des structures tissulaires avec une teinte différente de celle de la solution colorante initiale. Cette propriété ne s'observe que pour certains colorants électropositifs tels que le bleu de toluidine. Quant à la réaction métachromatique, elle est la marque de structures polycationiques sur lesquelles se fixe le colorant. En effet, la presque totalité des granules des basophiles est constituée d'une matrice de mucopolysaccharides acides qui sont très électronégatifs. C'est la *densité importante des charges négatives* qui est responsable d'une modification de la longueur d'onde d'émission du colorant (du bleu vers le rouge) du fait d'une « agrégation » des molécules de colorant.

Le bleu de toluidine est en quelque sorte un marqueur de structures ayant une densité importante de charges négatives ; les granules des basophiles non colorés ne portent plus de charges électronégatives (voir annexe 1). Comme chacun sait, les charges de même signe se repoussent. Par conséquent, pendant les quelques minutes où les cellules sont laissées à sédimenter dans l'hématimètre, les forces de répulsion électrostatiques qui s'exercent entre les basophiles auraient tendance à faire se repousser légèrement les basophiles entre eux. Du fait de ces *forces répulsives* s'exerçant entre basophiles, les distances d'un basophile avec ses plus proches voisins seraient par conséquent plus régulières que ne le voudrait le hasard. Et ceci serait d'autant plus vrai que la concentration des basophiles serait élevée comme l'indique l'article de H. Gérard *et al* puisque nous avons affaire à une force de répulsion dont l'intensité diminue avec la distance.

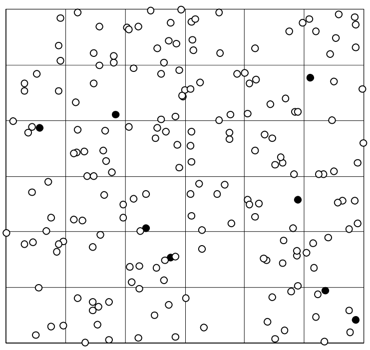

Figure 12.5. Si la position de chaque basophile (boules noires) est indépendante de celles des autres basophiles et des globules blancs (boules blanches), alors la loi de Poisson doit s'appliquer lorsqu'on dénombre une série d'échantillons prélevés dans une même « urne ». Toutefois, si une force de répulsion (ou un mécanisme conduisant à un effet identique) s'exerce, alors les distances entre les basophiles seront plus régulières que ne le prévoit la loi de Poisons. La conséquence en sera que la variance des comptes diminuera car « de l'ordre aura été introduit » dans le système. C'est ce que suggèrent les données empiriques obtenues sur des comptes de basophiles. La nature de cette force répulsive n'est pas connue. Il pourrait s'agir d'une force électrostatique (force de Coulomb) prenant sa source dans la matrice glycoprotéique sulfatée des granules des basophiles qui possède de nombreuses charges électronégatives. C'est précisément parce que la densité des charges électronégatives de cette matrice est élevée que le bleu de toluidine subit le phénomène de métachomasie lorsqu'il colore les basophiles. Cette force répulsive diminuant avec la distance, ceci expliquerait pourquoi ce phénomène serait surtout visible aux fortes concentrations cellulaires, comme l'ont constaté les auteurs de l'article de H. Gérard *et al.*

D'autres mécanismes peuvent être suggérés pour expliquer les observations de H. Gérard *et al.* [1] Bien qu'il n'existe actuellement aucune certitude sur les raisons de cet écart à la loi de Poisson, ceci permet néanmoins d'illustrer l'idée qu'il est parfois simpliste de plaquer sans précaution une loi mathématique sur un phénomène physique ou biologique complexe.

*Pourquoi un effet à hautes dilutions n'a-t-il pas été mis en évidence dans les trois expériences à l'aveugle des deux derniers jours de l'enquête ?*

Tout d'abord, soyons clair, il est possible que, même réalisées dans de meilleures conditions, les expériences au centre de la polémique auraient été négatives. Il faut être totalement étranger au domaine de la biologie et de la médecine expérimentale pour s'étonner du fait. Bien évidemment, il arrive parfois que l'on teste des hypothèses « massives » pour lesquelles l'usage de méthodes statistiques raffinées n'est pas nécessaire. Mais, même dans ce cas, on n'est pas à l'abri de circonstances imprévues. L'expérimentation à la paillasse n'est jamais un long fleuve tranquille. Et comme l'expliquera J. Benveniste :

> « […] Tout ce qui semble avoir intéressé les gens de Nature, c'est que l'expérience pouvait, une fois, ne pas réussir. Mais cela, nous le savions ! Nous n'avions pas besoin d'eux pour le savoir ! Et j'ai l'impression que leur but a été de pousser le système à sa limite, de créer des conditions de travail et d'effectuation invraisemblables pour parvenir, enfin, à une expérience ratée ! »[2]

Pour se prémunir des différents biais d'interprétation, il est important de décider *a priori* (c'est à dire avant de connaître le résultat) ce que l'on considérera comme conditions expérimentales acceptables. Par exemple, dans le cas des hautes dilutions d'anti-IgE, l'expérience accumulée en près de trois ans de travail, avait permis de définir – entre autres conditions – qu'il était nécessaire d'avoir des contrôles et un premier pic valides (c'est à dire inclus dans des limites prédéfinies) avant même de considérer les résultats à hautes dilutions. Il est étonnant de constater que les experts sont tombés des nues lorsqu'ils ont appris (ou feint d'apprendre) que dans certaines expériences, on n'observait pas de dégranulation avec l'antiserum anti-IgE :

> « Nous avons été surpris d'apprendre que les expériences ne "marchaient" pas toujours. […] Il semble également que certains sangs qui "ne dégranulent pas" sont souvent rencontrés ; nous avons appris que dans ce cas les résultats sont consignés mais ne sont pas inclus dans les analyses destinées à être publiées. »[3]

Supposons que l'on teste l'effet d'un médicament sur une population de patients. Il est tout a fait possible – c'est même habituel – que le médicament soit sans effet chez certains. On ne s'étonne pas de cela. C'est pour ces raisons que les statistiques sont utilisées afin d'analyser non pas un résultat individuel mais les résultats obtenus sur des populations de patients. Nous sommes ici dans le même cas de figure. Ce qui est important de savoir est si, *globalement*, sur l'ensemble des expériences réalisées, un effet significatif est obtenu en présence de hautes dilutions. Et pour J. Benveniste, l'expérience cumulée par son équipe pendant plusieurs années incluant de nombreuses expériences à l'aveugle avec une analyse statistique appropriée avait un poids sans commune mesure avec ces trois expériences négatives réalisées dans de mauvaises conditions expérimentales et, qui plus est, avec une unique série de dilutions d'anti-IgE. Il ne s'agissait pas ici de démontrer un théorème de mathématique qu'un seul contre-exemple suffit à invalider.

*Une correction de dernière minute*

On se souvient qu'une phrase rapportant les résultats de la 4ème expérience réalisée pendant la semaine d'enquête avait été supprimée du rapport d'enquête (elle était présente dans les épreuves d'imprimerie). Cette phrase disait en substance que les effets constatés aux hautes dilutions n'étaient rien d'autre que des fluctuations statistiques mais que cette explication ne s'appliquait pas à tous les résultats et en particulier à la fameuse 4ème expérience.

En fait, cette phrase supprimée au dernier moment ne représente qu'une partie d'un paragraphe beaucoup plus large qu'il est particulièrement intéressant de reproduire en totalité car il concerne – encore – la question de l'échantillonnage :

> « Les valeurs des contrôles sont utilisées pour normaliser les comptes obtenus avec les produits à hautes dilutions. En dépit de la règle adoptée par le laboratoire de présenter les résultats sous la forme de pourcentages de dégranulation par les substances diluées par rapport aux contrôles, il semble ne pas avoir été pris en compte que l'erreur de comptage est la somme statistique (la racine carrée de la somme des carrés) de l'erreur d'échantillonnage d'un puits donné et l'erreur estimée de la moyenne des échantillons contrôles. Dans le cas particulier de la première expérience, par exemple, nous estimons que l'erreur d'échantillonnage attendue est de 14%. Il semble évident que de nombreux pics considérés comme significatifs à Clamart 200 (*sic*) sont bien compris dans l'intervalle de deux déviations standards autour de la ligne 0% d'achromasie, même si on ne prend pas compte des autres sources d'erreur possibles (par exemple, un comptage défectueux des basophiles). »[4]

Ce passage – même s'il n'a pas été retenu dans la version publiée – confirme la véritable idée fixe des enquêteurs vis-à-vis de l'erreur d'échantillonnage. Ici, J. Maddox tente de semer la suspicion (par des arguments très techniques). En substance, il suggère que les chercheurs de Clamart considèrent comme positives des dégranulations qui ne sortiraient pas en fait du bruit de fond statistique.

Toutefois, la démarche du calcul proposée est dans ce cas tout à fait exacte et pertinente et elle contraste avec le calcul erroné ci-dessus (nous invitons le lecteur intéressé à se reporter à la note [5] de fin de chapitre pour une explicitation des calculs de J. Maddox). Cette incohérence pourrait peut-être s'expliquer par le fait que les calculs avec la formule inexacte ont été réalisés sur place par W. Stewart qui était venu à Clamart avec un micro-ordinateur – un Macintosh – et qui dès le premier jour s'est employé à saisir des données provenant des cahiers d'expériences d'E. Davenas.

J. Maddox ayant finalement décidé de ne pas publier ce passage, on ne peut l'accuser de mauvaise foi, mais il a manifestement été tenté une fois de plus de confirmer ses *a priori* sur une expérience qui – avant même de parler d'effets de hautes dilutions – était de mauvaise qualité (expérience A ; voir chapitre 9). Si J. Maddox avait suivi le même raisonnement sur les expériences B et C (voir chapitre 9) – qu'il a préféré ne pas montrer dans le rapport d'enquête – il aurait dû reconnaître qu'il ne s'agissait pas « d'artefacts du bruit statistique ». C'est peut-être la raison pour laquelle J. Maddox a préféré supprimer ce passage.

*Du bon usage de l'ironie*

Les enquêteurs auront donc fait preuve d'une rare insistance sur la question de l'erreur d'échantillonnage. N'ayant pas découvert le tricheur supposé au cours de leur enquête, c'était la seule donnée objective qu'ils rapportaient de leur expédition. De plus, n'explicitant pas leurs calculs et les données à partir desquels ces derniers avaient été menés, utilisant de surcroît l'autorité que leur conférait *Nature*, il était difficile pour le lecteur de leur rapport de mettre en doute un argument présenté comme un théorème.

« L'évidence » de leur calcul ne leur suffisant pas, les enquêteurs manieront la dérision vis-à-vis de J. Benveniste. Ils noteront ainsi dans le rapport à propos de J. Benveniste et de l'erreur d'échantillonnage :

> « Ironiquement, il est lui-même l'un des auteurs d'un article publié en 1981, dans lequel cette question a été abordée de façon à peu près similaire (Petiot, J.F., Sainte-Laudy, J. et Benveniste, J. Ann. Biol. Clin. 39, 355 ; 1981) […]. Ce bref article considère exclusivement les conséquences des erreurs d'échantillonnage (pas les autres types d'erreurs) sur l'interprétation des mesures de basophiles non dégranulés après contact de suspensions cellulaires de globules blancs avec les allergènes par l'intermédiaire des IgE membranaires. »[6]

L'ironie est parfois une arme à double tranchant. En effet, il est dommage que les enquêteurs n'aient pas lu plus attentivement ce « bref article » de Petiot *et al* qu'ils citent avec un plaisir non dissimulé. Ils auraient alors pris connaissance de l'information suivante :

> « L'expérience de Gérard et coll. ainsi que la nôtre nous ont montré que cet estimateur de la variance [i.e., la moyenne des basophiles] est biaisé supérieurement. Ceci réduit donc le risque de première espèce, certainement inférieur à celui énoncé, c'est à dire le risque de faux positifs. »[7]

Les auteurs formulent ici en termes clairs une notion répandue dans le « milieu » des utilisateurs du test de dégranulation qui aurait dû pour le moins mettre les enquêteurs en garde : la variance constatée des comptes de basophiles est en pratique  plus faible que ne le voudrait la loi de Poisson. Cette phrase ne semble pas avoir éveillé l'intérêt des investigateurs. Ils ne se priveront pas en effet d'expliquer que « Les résultats n'ont pas les marges d'erreurs de l'amplitude attendue. » et que « les comptes répétés sont plus proches les uns des autres que ce qu'autorise la distribution sous-jacente. »[8].

La différence d'appréciation de l'importance de l'erreur d'échantillonnage selon les enquêteurs ou selon J. Benveniste est assez savoureuse. J. Benveniste qui balaie d'un revers de main toute considération « théorique » à partir du moment où les faits la contredisent, est dans la lignée de Claude Bernard pour qui : « La méthode expérimentale [...] n'est rien d'autre qu'un raisonnement à l'aide duquel nous soumettons méthodiquement nos idées à l'expérience des faits »[9] ou encore pour qui, selon une formule proche, « Quand le fait qu'on rencontre est en opposition avec une théorie régnante, il faut accepter le fait et abandonner la théorie, lors même que celle-ci, soutenue par de grands noms, est généralement adoptée ». Nous sommes donc en présence d'une démarche que l'on pourrait qualifier de pragmatique. Pragmatisme que l'on a plutôt coutume d'attribuer aux chercheurs anglo-saxons. Selon ce cliché, en effet, ces derniers ne s'embarrassent pas de théories du moment que « ça marche ».

En revanche, l'esprit cartésien qui veut que toute observation s'inscrive dans un cadre théorique serait plutôt l'apanage d'une tradition française. Et lorsqu'il y a contradiction entre faits et théorie – lorsque « ça marche » alors que la théorie s'y oppose – que font les cartésiens ? Et bien selon Descartes : « Et les démonstrations de tout ceci sont si certaines qu'encore que l'expérience nous semblerait faire voir le contraire, nous serions néanmoins obligés d'ajouter plus de foi à notre raison qu'à nos sens ».[10] Et ironiquement, c'est une attitude ultra-cartésienne que les enquêteurs d'outre-Manche et d'outre-Atlantique vont adopter : comme les faits ne s'accordent pas à leurs attentes alors, en bons cartésiens, ils rejetteront les faits.

*En guise de résumé des chapitres 11 et 12*

Pour le lecteur qui a survolé les deux chapitres précédents, voici un résumé de ce qui a été abordé à propos des reproches d'ordre scientifique que l'on pouvait adresser aux enquêteurs de *Nature* :

1) Mauvaises pratiques méthodologiques et scientifiques :

- Les enquêteurs ont commis une *erreur de calcul* qui minimise la variance de la différence des comptes en double et « dramatise » cette variance, plus faible que ne le voudrait la théorie (loi de Poisson) ;

- Concernant ces mesures répétées, les enquêteurs n'ont pas tenu compte du fait que la variance des *petits échantillons statistiques* (comptes en triple ou en double) a plus fréquemment des valeurs faibles du fait de *l'asymétrie de la loi de distribution des variances* (tout en respectant bien entendu la loi de Poisson en moyenne). La conséquence de cette asymétrie est que les résultats des comptes de petites séries peuvent sembler « trop bons » ;

- Les enquêteurs ont mis systématiquement en exergue dans leur rapport des expériences qui pourtant ne remplissaient pas les *critères de qualité* et ne devaient donc pas être prises en considération.

2) Absence de prise en compte des connaissances du domaine de recherche :

- C'était une *notion connue* (et publiée) des chercheurs qui utilisaient le test de dégranulation que l'erreur d'échantillonnage est en moyenne plus faible que ce que prévoit la loi de Poisson, en particulier lorsque la densité cellulaire est élevée ;

- Il est possible que cette sous-dispersion des comptes de basophiles *s'explique par un phénomène physique* qui tendrait à faire se repousser les basophiles entre eux par exemple pour des raisons de charges électrostatiques.

En conclusion, n'étant pas parvenus à débusquer celui qui « jouait un tour dans le dos de Benveniste », les enquêteurs de *Nature* se sont rabattus sur des arguments techniques d'ordre statistique car selon eux les résultats étaient « trop bons ». Pourtant cet argument central du rapport d'enquête n'était lui non plus pas pertinent.

*Notes de fin de chapitre*

---

[1] M. Schiff (Un cas de censure dans la science, p. 237) a suggéré que des forces répulsives à longue distance appelées forces de Frölich pourraient jouer un rôle dans la sous-dispersion des comptes de basophiles. Ce type de force à longue distance parait en effet jouer un rôle dans l'interaction entre les globules rouges (Rowland et al. A Frölich interaction of human erythrocytes. *Physics Letters* 1981 ; 82A : 436). Il ne semble pas toutefois que ce type de forces puisse jouer un rôle dans le cas qui nous intéresse car en ce qui concerne les globules rouges, cette force n'est observée que si les cellules sont vivantes. Rappelons en effet que la solution colorante pour les basophiles fixe dans le même temps les cellules par l'éthanol.

[2] P. Alfonsi, Au nom de la science, p. 33.

[3] J. Maddox, W. Stewart, J. Randi. « High-dilution » experiments a delusion. *Nature*, 28 juillet 1988, p. 287.

[4] Epreuves d'imprimerie (« *composed proofs* ») du 25 juillet 1988.

[5] Dans ce texte J. Maddox estime que l'équipe de J. Benveniste considère que des points expérimentaux sont jugés significatifs (c'est-à-dire qu'il y a une dégranulation) là où pour lui ce n'est que bruit statistique. Il attribue cette différence au fait que l'équipe de J. Benveniste n'a pas conscience de l'importance de considérer non seulement la variabilité des contrôles mais également la variabilité liée au puits test.

Voyons ce qu'il en est en prenant cette expérience n°1 comme exemple. Deux puits témoins T1 et T2 avaient été réalisés pour cette première expérience et avaient été comptés plusieurs fois : T1 (78, 82 et 83 basophiles) et T2 (89 et 93). Le but est donc de comparer un compte x de basophiles du puits test aux comptes des puits témoins. Selon J. Maddox – et cette fois nous ne pouvons qu'être d'accord avec cette approche – il faut tenir compte de la variabilité de x. Par conséquent la valeur de la variable normalisé est :

$$\frac{x-(\frac{T1+T2}{2})}{\sqrt{x+1/4(T1+T2)}}$$

Etant donné que sous l'hypothèse nulle, x, T1 et T2 ne sont pas différents, T qui est la moyenne de T1 et T2 constitue donc le meilleur estimateur de x. L'écart-type de x – T devient alors :

$$\sigma=\sqrt{T+1/4(T1+T2)}=\sqrt{T+T/2}=\sqrt{3T/2}$$

On peut généraliser facilement à n valeurs témoins T1, T2, …, Tn :

$$\sigma=\sqrt{\frac{n+1}{n}\times T}$$

Si on rapporte à la moyenne T (pour avoir un %), l'écart-type relatif (de la différence entre un compte x et la moyenne des comptes dans les tubes témoins) est :

$$\sigma=\sqrt{\frac{(n+1)}{n}\times\frac{1}{T}}$$

Pour l'expérience 1, la moyenne des 5 comptes des contrôles est 85, d'où un écart-type de 14 % (c'est bien la valeur estimée par J. Maddox). J. Maddox calcule ensuite un seuil minimal à 28 % pour que le compte d'un puit test soit considéré significativement différent (c'est-à-dire 2 fois l'écart-type). J. Maddox s'étonne que ce seuil soit plus grand que le seuil de dégranulation placé sur la figure 1 par E. Davenas (vers 20 %). On peut faire plusieurs remarques : tout d'abord le facteur $\sqrt{\frac{n+1}{n}}$ tend rapidement vers 1 lorsque n augmente. Ainsi pour l'expérience 1, si on ne tient pas compte de ce facteur on trouve un écart-type relatif égal à $\sqrt{\frac{n+1}{n}\times\frac{1}{T}}$ = 12 %. La différence par rapport à 14 % n'est donc pas énorme. A l'U200 de l'Inserm, pour placer la limite à partir de laquelle la dégranulation était considérée comme significative, un test unilatéral au seuil alpha 5 % avait été choisi (comme dans l'article de Gérard *et al*). En effet, on ne s'intéresse qu'aux basophiles en moins, c'est-à-dire qui ont dégranulé (test unilatéral). Ceci conduit, dans le cas de l'expérience 1, à 1,67 x 12 = 20 % d'erreur d'échantillonnage. C'est ce seuil qui était porté sur le graphique de l'expérience 1 de l'article de *Nature* de juin 1988. J. Maddox, quant à lui, avec un test bilatéral et un écart-type de 14%, parvient à un seuil de dégranulation de 2 x 14% = 28 %. Avec un test unilatéral, le seuil serait dans ce dernier cas de 1,67 x 14= 23 %.

On peut faire plusieurs remarques sur ces calculs : 1) tout d'abord les différences sont peu importantes ; 2) ces calculs sont en fait plutôt vains puisque nous avons vu qu'en pratique on constate que la variance des comptes est plus faible que ne prévoit la loi de Poisson ; 3) la présente discussion n'avait plus guère d'objet en 1988 car de façon pragmatique, seules les dégranulations supérieures à 30% étaient considérés à l'U200 comme « biologiquement significatives » ; 4) les tests statistiques étaient réalisés non pas au sein de chaque expérience mais sur une série d'expériences réalisées dans les mêmes conditions expérimentales.

Le calcul de cette limite de significativité tire sa justification de l'utilisation clinique du test de dégranulation des basophiles. Pour un patient donné, il est important de fixer une limite et de définir des critères de décision permettant de conclure si le test est positif ou non. C'est le propos des deux articles de 1981, celui de Petiot *et al* et celui de Gérard *et al*. Le seuil de dégranulation matérialisé par la ligne pointillé visible sur les graphiques de l'article de *Nature* de juin 88 sur laquelle se focalise J. Maddox n'était plus guère représentée au laboratoire qu'à titre « historique ».

[6] J. Maddox, J. Randi, W. Stewart. "High dilution" experiments a delusion. *Nature*, 28 juillet 1988, p. 288.

[7] J.F. Petiot, J. Sainte-Laudy, J. Benveniste. Interprétation du résultat d'un test de dégranulation des basophiles humains. *Ann Biol Clin* 1981 : 39 : 355–359.

[8] J. Maddox, J. Randi, W. Stewart. "High dilution" experiments a delusion. *Nature*, 28 juillet 1988, p. 290.

[9] C. Bernard. Introduction à l'étude de la médecine expérimentale (1865).

[10] R. Descartes. Principes de la philosophie (1644).