Statistique inférentielle au fil de l'ouvrage de Jules Gavarret :
Principes généraux de Statistique Médicale
ou développement des règles qui doivent présider à son emploi (1840).
Histoire et proposition d'activités.

Jean Lejeune
en collaboration avec Denis Lanier et Didier Trotoux
IREM de Basse-Normandie
jeannotlejeune@orange.fr

Dans l'introduction de son ouvrage, *The History of Statistics : The Measurement of Uncertainty before 1900*, Stephen M. Stigler fait remarquer que l'application de nouveaux outils à une discipline ne se transmettait pas immédiatement à d'autres disciplines où elle aurait pourtant pu être très utile. Ainsi écrit-il : « By the 1830s statistical methods were widely used in astronomy,... Yet it is only in the twentieth century that we find these same methods making substantial inroads onto the social sciences. » [S, p. 2].

La question posée par Stigler pourrait aussi s'appliquer à la transmission, dans les années 1830, des outils statistiques à la médecine et à l'épidémiologie et se doubler de la question suivante : pourquoi, dans le cas de l'épidémiologie, fallut-il attendre plus de 80 ans entre l'apparition des nouveaux outils de la statistique inférentielle et leur application effective à l'épidémiologie ?

Aujourd'hui, en médecine et en épidémiologie, pour comparer l'efficacité de divers traitements, il paraît évident de s'appuyer sur les outils de la statistique inférentielle. Concernant l'apparition d'une pathologie et l'efficacité moyenne de divers traitements thérapeutiques sur l'ensemble de la population, ces outils constituent une aide à la décision. Les données d'enquêtes ou d'essais cliniques sont en partie aléatoires parce qu'elles proviennent seulement d'une partie de la population. La décision doit être assortie de la probabilité de se tromper en la prenant. Dans ce cas, il est fait appel à la théorie des tests (Annexe I).

Ces outils permettent aussi de fournir l'encadrement de la valeur numérique d'un paramètre inconnu ou d'une différence de paramètres inconnus concernant toute la population. Cet encadrement est bien sûr aléatoire car calculé à partir d'une partie supposée avoir été tirée de façon aléatoire dans toute la population. Il est assorti du degré de confiance que l'intervalle produit par cet encadrement recouvre bien la quantité inconnue. Dans ce cas, il est fait appel à la théorie des intervalles de confiance (Annexe II).

À la question : « y a t-il eu, au début du XIX^e siècle, transmission à la médecine et à l'épidémiologie des techniques de la statistique inférentielle ? », Edward Huth, dans un article datant de 2006 [H], semble répondre affirmativement puisqu'il fait remonter à la publication en 1840 du livre *Principes généraux de Statistique Médicale ou développement des règles qui doivent présider à son emploi* d'un médecin français, Jules Gavarret (1809-1890), le premier travail sur l'application des statistiques inférentielles aux données thérapeutiques pour aboutir à des jugements critiques sur la valeur des thérapies. Ce jugement est confirmé par Alain-Jacques Valleron qui, dans un rapport de l'Académie des Sciences publié en 2006, dit qu'il fut « sans doute le premier à utiliser la démarche du test statistique afin d'évaluer si des proportions observées étaient significativement différentes. » [V, p. 9].

Dans un premier temps, nous présenterons le contexte dans lequel a été écrit cet ouvrage. Ensuite, à partir d'une lecture linéaire de celui-ci, nous relèverons les problèmes que s'est posé Gavarret concernant les statistiques au sens général et leurs applications dans la discipline médicale et dans l'épidémiologie. Ceux posés dans l'ouvrage seraient sans doute susceptibles d'être utilisés dans le but de faire travailler des élèves et des étudiants sur certains questions relatives aux probabilités et statistiques.

1 Le contexte, les protagonistes.

L'ouvrage de Gavarret comportant un important rappel sur le contexte, nous utiliserons des citations issues de cet ouvrage pour éclairer ce contexte tel qu'il se présentait à l'époque. Les références des citations de Gavarret seront notées [G, p. x], x étant le numéro de page. Les passages en italique dans l'ouvrage le restent dans les citations qui suivent.

1.1 La naissance de la méthode numérique en épidémiologie et les polémiques qui s'ensuivirent.

Au tout début de l'introduction au *Rapport* ... de Valleron [V], il est rappelé que l'épidémiologie « décrit les variations de fréquence des maladies dans les groupes humains, et recherche les déterminants de ces variations. Elle vise à la compréhension des causes des maladies, et à l'amélioration de leurs traitements et moyens de prévention. » Jusqu'à un époque récente, « l'expérience des siècles » et le jugement du praticien suffisait pour déterminer l'influence de certains facteurs sur l'apparition d'une maladie pour un groupe d'individus, avec des affirmations telles qu'en substance : la saison est un facteur important pour l'apparition des maladies pulmonaires.

C'est en 1835 que dans un ouvrage qu'il intitule, *Recherches sur les effets de la saignée dans quelques maladies inflammatoires et sur l'action de l'émétique et des vésicatoires dans la pneumonie* [Lo], Pierre-Charles-Alexandre Louis (1787-1872), éminent clinicien, utilisa des indicateurs statistiques pour « montrer que l'usage de la saignée, pratiquée systématiquement grâce à des sangsues (environ 20 millions de sangsues importées en 1830....) pour traiter les pneumonies ne se justifiait pas. » [V, p. 9].

L'introduction d'indicateurs statistiques, en particulier en médecine, constitue ce qu'on a appelé la "méthode numérique". Elle déclencha à l'Académie des Sciences cette même année puis à l'Académie de Médecine en 1837, une vive polémique entre ceux qui voyaient dans l'utilisation de statistiques, une avancée importante et ceux qui pensaient qu'il fallait surtout se reposer sur l'expérience du praticien en présence d'un malade particulier.

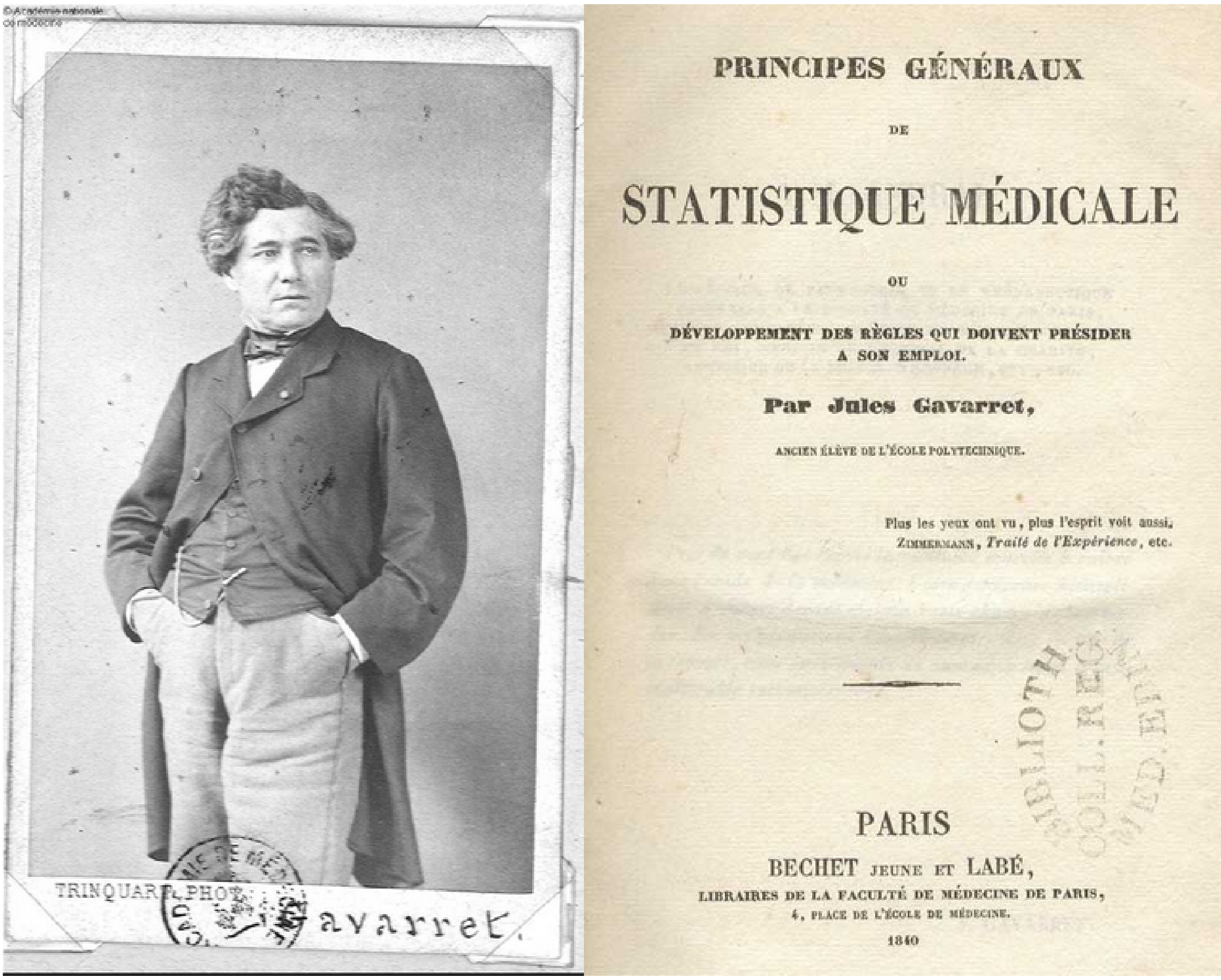
Deux camps étaient en présence. Le premier camp regroupait ceux pour qui « l'esprit de systématisation et de généralisation, loin de servir à l'avancement de la science, ne saurait qu'en enrayer le progrès » [G, p. 5] à cause des erreurs qui ont pu être faites en généralisant hâtivement à partir d'un « groupe trop restreint de phénomènes » [G, p. 4]. Le second camp pensait que « le résultat immédiat d'une telle doctrine serait de condamner le médecin à se renfermer dans la considération exclusive des faits particuliers. » [G, p. 5].

Pour les tenants de la "méthode numérique", il s'agit de bannir la « funeste habitude [...] de se servir de mots vagues et indéterminés, tel que *souvent, rarement* » [G, p. 19] et « de recommander l'emploi des rapports numériques » [G, p. 19].

Cette communication de Louis marque donc la naissance de l'école française dite de la "méthode numérique" et de la statistique en épidémiologie. Il s'agit, à partir d'observations, d'utiliser des indicateurs numériques pour caractériser l'influence d'un facteur ou pour comparer les effets de deux ou plusieurs modalités d'un même facteur sur un groupe d'individus. Ce savant forma beaucoup d'étudiants venus des États-Unis et d'Angleterre et contribua ainsi à développer en dehors de France l'épidémiologie numérique [Li].

1.2 Qui est donc Jules Gavarret ?

Il naquit le 28 janvier 1809 à Astaffort (Lot-et-Garonne) dans une famille bourgeoise et fut admis à l'École polytechnique en 1829 où il eut comme professeur Siméon-Denis Poisson. Nommé sous-lieutenant d'artillerie fin 1831, il démissionna au début de 1833 pour continuer ses études scientifiques. Son compatriote le Dr Gabriel Andral, fervent partisan de la "méthode numérique" l'associa en 1838 à ses recherches sur l'hématologie médicale, sur les altérations du sang dans les cas de maladie et l'incita à faire ses études de médecine. Ils publièrent ensemble de nombreux ouvrages. Il obtint la chaire de physique médicale de la faculté de médecine de Paris le 14 janvier 1844 après avoir soutenu une thèse de concours sur les *Lois générales de l'électricité dynamique*. Il mourut à la retraite le 21 août 1890 après avoir publié des ouvrages sur des sujets aussi divers que *Des images par réflexion et par réfraction*, *De la chaleur produite par les êtres vivants*, *L'Acoustique biologique* ou encore *La télégraphie électrique*.



PRINCIPES GÉNÉRAUX

DE

STATISTIQUE MÉDICALE

OU

DÉVELOPPEMENT DES RÈGLES QUI DOIVENT PRÉSIDER A SON EMPLOI.

Par Jules Gavarret,

ANCIEN ÉLÈVE DE L'ÉCOLE POLYTECHNIQUE.

Plus les yeux ont vu, plus l'esprit voit aussi.
ZIMMERMANN, *Traité de l'Expérience*, etc.

PARIS

BECHET JEUNE ET LABÉ,

LIBRAIRES DE LA FACULTÉ DE MÉDECINE DE PARIS,

4, PLACE DE L'ÉCOLE DE MÉDECINE.

1840

FIG. 1 – J. Gavarret et la page de titre de son ouvrage

2 Une introduction aux préliminaires statistiques et probabilistes.

2.1 Les interrogations de Gavarret.

Devant une série d'observations, issues ou non d'une expérience, doit-on se contenter d'une étude de ces observations (ou d'un rapport numérique calculé à partir de ces observations), ou bien peut-on à partir de ces nombres, découvrir une loi générale sur l'ensemble de la population étudiée dont sont issues ces observations et avec quel degré de certitude ?

– Si on admet qu'il existe une "erreur" entre la valeur de la caractéristique étudiée sur une observation (ou d'une statistique issue d'observations) et la moyenne inconnue de cette caractéristique dans toute la population étudiée, cette "méthode numérique" suffit-elle pour conclure ?

  Gavarret pense que Louis est conscient de cette question des "erreurs" entre les observations et la moyenne inconnue de cette caractéristique dans toute la population étudiée. En effet, il cite une phrase de Louis tirée des *Recherches sur sur les effets de la saignée* : « les erreurs étant les mêmes pour deux groupes de malades traités par des procédés différents, ces erreurs se compensent et peuvent être négligées sans altérer sensiblement l'exactitude des résultats. » [G, p. 20].

  En termes modernes, si μ_1 (resp. μ_2) est la moyenne de la caractéristique étudiée sur l'ensemble de la population ayant subi le traitement M_1 (resp. M_2), si $\overline{x_1}$ (resp. $\overline{x_2}$) est la moyenne des observations numériques sur un échantillon de malades soumis au traitement M_1 (resp. M_2) , l'observation de la statistique $\delta\overline{x} = \overline{x_1} - \overline{x_2}$ peut-elle conduire à une décision quant à la différence $\delta\mu = \mu_1 - \mu_2$?
  À cette question, Louis répond donc oui car il suppose implicitement que, pour chaque observation $x_{1,i}$ issue de la population M_1, $x_{1,i} = \mu_1 + \epsilon_{1,i}$ et de même pour chaque observation $x_{2,j}$ issue de la population M_2, $x_{2,j} = \mu_2 + \epsilon_{2,j}$ ce qui est exact mais son hypothèse se traduit par $\sum_i \epsilon_{1,i} = 0$ et

  $$\sum_j \epsilon_{2,j} = 0.$$

  Dans ce cas, par exemple, l'observation de $\delta\overline{x} > 0$ entrainera la décision $\delta\mu > 0$ car alors $\delta\overline{x} = \delta\mu$. Mais aussitôt après Gavarret écrit : « Mais c'est précisément sur cette compensation des erreurs que les doutes ont été élevés. » [G, p. 20].

Activité 1.

– À cette première critique, Gavarret en ajoute une deuxième. A t-on déterminé « les conditions auxquelles doivent satisfaire les cas particuliers pour qu'on fût autorisé à les disposer les uns à côté des autres dans un même relevé » [G, p. 22], autrement dit, la moyenne observée $\overline{x_1}$ (resp. $\overline{x_2}$) provient-elle d'observations que l'on peut considérer comme ayant la même loi de probabilité ?
– Une troisième critique apparait alors lorsqu'il s'agit de comparer des proportions :

> « [...]croit-on qu'il suffira, pour avoir la véritable mesure de l'influence de la médication expérimentée, d'additionner d'un côté le nombre de morts, de l'autre le nombre total des malades, et de diviser cette première somme par la seconde ? En agissant ainsi, on aura ce que l'on peut appeler la mortalité moyenne. Pour que les indications fournies par cette moyenne aient quelque valeur, le nombre des cas observés est-il indifférent ? une mortalité de un dixième, déduite de la considération de vingt ou trente expériences a-t-elle la même autorité que si elle était déduite de trois cents, quatre cents, mille faits bien constatés ? et si la valeur d'une proposition varie avec les nombres sur lesquels elle s'appuie, à quoi tient un semblable résultat, comment, étant donné la composition d'une statistique, peut-on déterminer le degré de confiance que méritent ses indications ? » [G, p. 22-23].

Gavarret pose ainsi en quelques phrases une grande partie des problèmes de la statistique inférentielle moderne (puissance d'un test, longueur de l'intervalle de confiance, mesure du risque quand une décision est prise....). Il en vient même alors, dès son introduction, à fustiger les conclusions de Louis pour lequel :

> « cent sept cas de pneumonie, quarante quatre d'érysipèle de la face, vingt-trois d'angine gutturale, ont paru suffisants pour établir le peu d'efficacité de la saignée dans le traitement de ces trois phlegmasies aiguës.» [G, p. 24].

Dès le début de son ouvrage, il livre son programme au lecteur :

« **1.** Déterminer ce que l'on doit entendre par *faits semblables* ou *comparables*, propres à rentrer dans la rédaction d'une statistique ;

2. Prouver que toute conclusion déduite *d'un petit nombre de faits* ne mérite en *thérapeutique* aucune confiance, que toute statistique, pour fournir des indications admissibles, doit contenir *plusieurs centaines d'observations* ;

3. Faire voir comment les lois déduites *a posteriori* ne sont jamais vraies que dans certaines *limites d'oscillation*, et donner le moyen de déterminer ces *limites* dans chaque travail particulier. » [G, p. 26].

Il annonce alors que « les principes du *calcul des probabilités*, en nous fournissant le moyen de résoudre d'une manière rigoureuse ces trois grandes difficultés, nous permettront de tracer avec certitude les règles de l'emploi de la statistique en médecine. » [G, p. 27].

Avertissement : à partir de maintenant et jusqu'au chapitre 6, les titres des chapitres et des articles qui constituent ces chapitres sont ceux de l'ouvrage de Gavarret.

3 Nécessité de recourir au calcul des probabilités pour suppléer à l'insuffisance des règles de la logique.

C'est le sous-titre du chapitre I [G, p. 29-56].

3.1 Définition et mesure de la probabilité. - Réponses aux principales objections présentées contre la statistique médicale.

Dans l'article premier [G, p. 29-44], Gavarret répond à ses collègues de l'Académie de Médecine, messieurs Amador et Double.

Première objection : Il y a une grande distinction, selon Amador, entre « l'*acception mathématique* et l'*acception philosophique* du mot *probabilité*. » [G, p. 31]. En clair, Amador semble mettre en cause la possibilité de mesurer le hasard. Gavarret lui répond : « l'*acception mathématique* (pour employer le langage de M. Amador) n'est que la représentation de l'*acception philosophique* du mot *probabilité*, de même qu'une formule algébrique sert tous les jours à représenter les lois de la gravitation. » [G, p. 31].

Deuxième objection : Concernant le calcul des probabilités, Amador « a cru encore devoir s'attaquer à sa certitude en tant que calcul » [G, p. 31] et il s'appuie sur les écrits d'auteurs célèbres pour contester la rigueur de cette théorie. La réponse de Gavarret : il distingue, en citant Poisson, les règles : « qui se démontrent en toute rigueur » [G, p. 33] et la mesure de la probabilité qui « n'est pas une chose absolue, mais relative à nos connaissances acquises sur cet événement. » [G, p. 34]. Suivent plusieurs exemples qui, concernant la réalisation d'un événement, montrent que nos connaissances peuvent modifier « la raison d'y croire » [G, p. 34]. Certains de ces exemples, en fait, illustrent la notion de probabilité bayésienne.

Troisième objection : Amador déclare que « le calcul des probabilités est trop obscur encore pour inspirer aucune confiance » [G, p. 35]. D'autres ont « repoussé son application aux recherches de thérapeutique, sous prétexte qu'avec des probabilités il était impossible de faire une science » [G, p. 35-36]. En guise de réponse Gavarret distingue les « connaissances humaines absolument certaines » [G, p. 36] principalement liées au mathématiques pures et les sciences naturelles qui « ne reposent que sur des séries de probabilités plus ou moins approchées de la certitude » [G, p. 36]. Les tenants de la "méthode numérique", qui n'assortissent pas leurs conclusions de la probabilité quant aux raisons de les croire, voient dans le recours au calcul des probabilités le déni que la médecine soit une science. Gavarret leur répond que « Tant que dans les sciences nous ne pourrons pas *a priori* connaître la nature intime des causes, [...], le calcul des probabilités devra servir de base à nos travaux ; » [G, p. 39].

Quatrième objection : D'après Double, « deux observations quelconques, présentent entre elles trop de différences pour pouvoir être considérées comme des unités *comparables*. » [G, p. 40]. Or l'emploi du calcul des probabilités « réclame, comme condition indispensable, de grandes collections de faits ». À cette objection de Double, Gavarret répond que « sans agrégation d'individualités dans des groupes bien délimités, il n'y a pas de science possible ; car il n'y a pas de science sans lois générales » [G, p. 41] pour conclure à la fin de l'Article premier : « On peut compter en médecine ; il ne s'agit plus que de démontrer comment on doit le faire. » [G, p. 44].

Ces quatre objections soulèvent des questions importantes. Entre autres : toute discipline dont l'énoncé d'une conclusion est assortie de la probabilité que cette conclusion soit vraie est-elle une science ? Peut-on considérer des individus comme issus d'une même population ? Peut-on, à partir de l'observation de certains de ces individus, en conclure des lois générales sur toute la population ? Le calcul des probabilités n'est-il utile que parce qu'il est difficile de connaître toutes les causes des événements, et donc utilisé "par défaut" ? Gavarret ne répond pas rigoureusement à ces questions mais il a le grand mérite de les poser.

3.2 Nécessité de recourir au calcul des probabilités pour apprécier l'influence des médicaments.

Dans l'article II [G, p. 44-52] de ce Chapitre premier, pour montrer « *l'insuffisance de la logique pour la solution de question fort importantes.* » [G, p. 44], il prend deux exemples, que nous traduisons en termes modernes :

1. Soit une urne contenant N boules dont K boules blanches et $H = N - K$ boules noires, N et K inconnus. Résultat d'un tirage au hasard de $n_1 = 1000$ boules : $k_1 = 900$ blanches et $h_1 = n_1 - k_1 = 100$ noires.

   Première question : peut-on en conclure que $K > H$?

   Gavarret répond que, suivant les "règles de la logique" : « personne n'hésiterait à affirmer que nécessairement les blanches étaient beaucoup plus nombreuses dans l'urne que les noires ; » [G, p. 46].

   Deuxième question : si on effectue un nouveau tirage au hasard de $n_1^{'} = 1\,000$ boules, et si on note $k_1^{'}$ = nombre de boules blanches et $h_1^{'} = n_1^{'} - k_1^{'}$ le nombre de boules noires, suivant ces même règles, conclut-on que $k_1^{'} > h_1^{'}$?

   À cette deuxième question, Gavarret répond encore oui.

   « De semblables conclusions viendraient spontanément à l'esprit de tout le monde, même de ceux qui nient l'influence des grands nombres de faits sur nos décisions. Mais là s'arrêtent les indications fournies par les règles de la *logique*. » [G, p. 46].

   Troisième question : Soit $p = \dfrac{K}{N}$ la proportion inconnue de boules blanches dans l'urne. Le premier tirage a donné la valeur de la statistique $\widehat{p_1} = \dfrac{k_1}{n_1} = \dfrac{9}{10}$. Gavarret se pose alors la question : « dans *quelles limites d'erreur* le rapport de blanches au nombre total de boules contenues dans l'urne, peut-il différer de l'expression neuf dixièmes fournie par l'expérience ? » [G, p. 47].

   Autrement dit, peut-on calculer ϵ_1 et ϵ_2 tels que $\widehat{p} - \epsilon_1 \leq p \leq \widehat{p} + \epsilon_2$?

   La première question est de fait un problème de test puisqu'à partir d'observations, il s'agit de prendre une décision concernant des paramètres inconnus. Ici il s'agit de décider entre l'hypothèse $K \leq H$ et l'hypothèse $K > H$ à partir de l'observation $k_1 > h_1$.

   La deuxième question à savoir : déduire de l'observation " $k_1 = 900$ blanches et $h_1 = n_1 - k_1 = 100$ noires" que, pour un nouveau tirage de $n_1^{'} = 1\,000$ boules, $k_1^{'} > h_1^{'}$, est une question de calcul des probabilités et de nature différente.

   En effet, ou bien Gavarret considère que la premier tirage a fourni la proportion exacte de boules blanches, et alors il s'agira, en étudiant la distribution d'échantillonnage, de calculer la probabilité que $k_1^{'} > 500$ sachant que $k_1^{'} \sim \mathcal{B}(1\,000\,; 0{,}90)$ où $\mathcal{B}(n, p)$ désigne la loi binomiale de paramètres n et p, ou bien Gavarret considère que le résultat du premier tirage lui permet de décider que $p > 0{,}50$ et il s'agira de calculer la probabilité que $k_1^{'} > 500$ sachant que $k_1^{'} \sim \mathcal{B}(1\,000\,, p)$ avec $p > 0{,}50$, ou bien encore il s'agira de calculer la probabilité que $k_1^{'} > 500$ sachant que $k_1 = 900$. Cette ambiguïté sur la nature de p se rencontre dans d'autres textes que celui de Gavarret.

   La troisième question est encore une question de statistique inférentielle. Sa solution réside en fait dans la construction d'un intervalle de confiance d'une proportion inconnue.

2. Dans la même urne, on effectue un tirage au hasard. Résultat du tirage au hasard de $n_2 = 20$ boules : $k_2 = 18$ blanches et $h_2 = n_2 - k_2 = 2$ noires. Ce tirage a donné pour valeur de la statistique $\widehat{p_2} = \dfrac{k_2}{n_2} = \dfrac{9}{10}$ la même que pour le tirage de $n_1 = 1\,000$ boules. Conclut-on encore $K > H$?

Gavarret affirme que « tout homme à jugement droit se refuserait à énoncer une opinion quelconque, ne se considérant pas comme suffisamment éclairé par un si petit nombre d'épreuves. » [G, p. 47].

Il conjecture donc que, pour une même valeur de la statistique (ici la proportion observée), la taille de l'échantillon a une importance dans la décision qui sera prise et se pose la question de savoir « combien il leur faudrait de tirages pour se croire fondés à regarder le résultat statistique comme représentant, *à très peu près* et dans des *limites données d'erreur possible.* » [G, p. 48].

Cette dernière question contient en fait l'énoncé du problème suivant : quelle doit-être la taille n d'un échantillon pour qu'un intervalle de confiance à x % (x proche de 99 ?) de longueur l donnée, contienne la véritable proportion inconnue p ?

Activités 2.

Il conclut cet article II :

« À propos de l'emploi de la statistique en médecine, il y aura toujours à soulever les deux difficultés suivantes :

1. Une statistique étant donnée, à quelle erreur s'expose-t-on en prenant les rapports qu'elle fournit pour l'expression véritable de l'influence de la médication essayée sur la maladie en observation ?
2. Dans quelles limites d'erreur possible l'influence d'une médication, déduite d'une statistique reposant sur un nombre donné d'observations, indique-t-elle les résultats que l'on aura droit d'espérer plus tard de l'emploi de la même méthode thérapeutique dans la même maladie ? » [G, p. 54].

Les deux derniers énoncés semblent poser le même problème mais en réalité, le premier concerne celui de la précision de l'estimation d'un paramètre inconnu et le second concerne la précision d'une prévision pour un nouvel individu ou un groupe d'individus.

3.3 Coup d'œil général sur le calcul des probabilités. - Faits à chances constantes, faits à chance variable. - Les faits médicaux sont à chance variable.

Dans cet article III [G, p. 52] du chapitre I, pour la recherche de la probabilité des événements, Gavarret distingue le "calcul direct" et le "calcul inverse".

– Le "calcul direct" est celui qui permet d'assigner à un événement le rapport : "nombre de cas favorables" sur "nombre de cas possibles". Mais Gavarret fait remarquer : « Qui oserait, en effet, essayer d'assigner *a priori* toutes les circonstances favorables et contraires à la guérison d'un malade quelconque soumis à une médication connue ? » [G, p. 53]. Ce calcul est donc écarté.
– Le "calcul inverse" est celui qui est « fondé tout entier sur la *loi des grands nombres* » [G, p. 54]. Il « fournit les règles à suivre pour déterminer la probabilité d'un événement quelconque. » [G, p. 54].

En utilisant ce calcul inverse, Gavarret pense qu'il pourra résoudre les deux problèmes suivants :

« *1er Problème.* Déterminer la véritable influence d'une médication donnée dans une maladie également donnée.

2ème Problème. Classer rigoureusement, par ordre d'influence, les diverses médications conseillées dans la même affection. » [G, p. 54].

Gavarret distingue ensuite deux catégories de faits bien distinctes :

« 1 Faits à chance constante ;

2 Faits à chance variable. » [G, p. 54].

Il explique que cette distinction recouvre celle de tirage avec remise (Faits à chance constante) et de tirage sans remise (Faits à chance variable) dans une population finie ou très grande. Puisque « nous n'avons jamais eu la pensée de prétendre que tous les malades auxquels nos moyens bornés d'analyse nous obligent à administrer les mêmes agents curatifs, aient réellement la même chance d'échapper à la mort [...] les faits médicaux doivent être rangés parmi les événements à chance variable. » [G, p. 56].

À cette étape de la réflexion, il est difficile de savoir ce que Gavarret entend exactement par "chance variable". Est-ce seulement une appellation de ce qui est nommé aujourd'hui "variable aléatoire" ? Si, pour expliquer cette notion, il ne se réfère qu'au tirage sans remise, il est évident que l'analogie est insuffisante. Bien que toutes les enquêtes soient en réalité des tirages sans remise (on n'interroge pas deux fois la même personne, ce qui peut arriver lors d'un tirage avec remise), si la probabilité de guérir diffère d'un individu à l'autre, ce n'est pas fondamentalement pour cette raison car la probabilité de tirer une boule blanche est la même au premier ou ou dernier tirage que l'on soit dans le cas d'un tirage avec remise ou sans remise. C'est parce que tous les individus sont différents. Si Gavarret considère qu'une population est un mélange d'individus ayant chacun une probabilité différente de guérir, il fait un premier pas sans le savoir vers ce que l'on nomme aujourd'hui la théorie des modèles aléatoires "mixtes" qui prend en compte l'effet "individu".

Activité 3.

Pour la suite, il attend de la loi des grands nombres la solution à tous ses problèmes.

4 Principes de la loi des grands nombres.

C'est le sous-titre du Chapitre II [G, p. 57-98] qui ne traite pas encore de l'application à la médecine, mais qui annonce les techniques utilisées.

4.1 Conditions auxquelles doivent satisfaire les faits à chances variables pour être comparables.

Dans l'Article premier [G, p. 58-66] de ce chapitre, Gavarret défend l'idée d'utiliser une « moyenne déduite d'un certain nombre d'expériences » [G, p. 60] pour mesurer « l'influence du traitement essayé » [G, p. 60]. Il fait remarquer un peu auparavant que ceux qui critiquent la "méthode numérique", pour certains « citent exactement les nombres de faits observés » [G, p. 60] et pour d'autres « confient ces nombres à leur mémoire » [G, p. 60] ; en clair ils utilisent la "méthode numérique" sans le dire. Mais pour effectuer cette moyenne, contrairement à ceux qui pensent qu'on ne peut utiliser des statistiques que si les conditions sont strictement identiques, Gavarret dit que « l'identité n'est pas nécessaire pour additionner les faits dont se compose une statistique » [G, p. 61].

Selon Gavarret,

> « pour que des observations relatives à des faits de ce genre soient *semblables* ou *comparables*, puissent entrer dans une statistique, il faut et il suffit que l'*ensemble des causes possibles* qui les régissent reste *invariable* pendant toute la durée des épreuves. En d'autres termes, dans chaque cas particulier, le phénomène constaté doit se rattacher à l'intervention d'une ou plusieurs causes faisant partie de cet ensemble *invariable* ». [G, p. 61].

Pour illustrer ce qu'il entend par "un ensemble invariable de causes possibles", il prend deux exemples. **Le premier exemple** est celui des causes d'un naufrage d'un navire.
Parlant des mers sur lesquelles voyagent les navires,

> « si les circonstances favorables ou défavorables dans lesquelles sont placés les vaisseaux qui les parcourent, peuvent se combiner entre elles de mille manières différentes, en sorte que deux vaisseaux ne se trouvent jamais dans une situation identiques ; ces circonstances, cependant dépendent d'une somme de causes possibles invariable dans leur ensemble. » [G, p. 62].

L'événement "naufrage d'un navire" est donc la résultante d'une "combinaison" de causes possibles appartenant à un même ensemble (fini ?) de causes (ce qui laisserait sous-entendre que si on pouvait dénombrer toutes ces causes, il serait possible de calculer la probabilité de réalisation de l'événement).

« D'une part, nous trouvons les phénomènes météorologiques et la disposition du fond des mers ; d'autre part, nous rencontrons la construction des vaisseaux eux-mêmes et l'habilité des marins qui les dirigent » [G, p. 62]. Pour Gavarret, il n'est pas possible de comparer « les naufrages arrivés dans la Méditerranée avec ceux observés dans la mer des Antilles » [G, p. 63] car les modalités des deux premiers facteurs de naufrages sont différentes suivant les lieux.

Le deuxième exemple est celui de « la recherche du rapport entre les filles et les garçons qui naissent dans l'étendue d'une même circonscription territoriale » [G, p. 64]. Il n'y a plus invariabilité des causes possibles si par exemple, « les mœurs, la législation et le bien-être public venaient à être profondément modifiés. » [G, p. 65].

Ces deux exemples sont utilisés à titre pédagogique pour insister sur la nécessité de satisfaire à cette condition pour l'utilisation de la statistique médicale.

4.2 Rapports qui existent entre la fréquence de manifestation des événements et leur chance propre

Dans cet article II [G, p. 66-80], Gavarret s'intéresse à « une série d'observations relativement à deux événements à chance variable, jouissant de la propriété de s'exclure mutuellement, dont un seul arrive nécessairement à chaque épreuve, tels que seraient la *guérison* et la *mort* d'un malade soumis à une médication connue. » [G, p. 66].

Pour lui, le problème consiste à approcher, à partir d'une « série d'observations relativement à deux événements à chance variable » [G, p. 66] « les chances de production des événements étudiés » [G, p. 66] appelées plus loin « lois suivant lesquelles ces phénomènes doivent apparaître. » [G, p. 67]. À la suite du recueil des observations, il reconnaît que « les rapports fournis par la statistique acquièrent une très grande importance. » [G, p. 67], ceux-ci étant « la fréquence de leur manifestation pendant la durée de ce travail » [G, p. 66]. C'est la confirmation de l'apport de la "méthode numérique". Mais il assortit cette reconnaissance de la remarque suivante :

> « Jamais, il est vrai, quelque étendue que soit la série des épreuves tentées, on ne pourra considérer ces rapports comme la traduction rigoureuse et absolue des chances moyennes de ces événements. Mais à mesure que le nombre des observations recueillies deviendra plus considérable, ces rapports et ces chances moyennes tendront à se confondre ; en sorte qu'il arrivera un moment où la différence entre ces quantités sera complètement négligeable. » [G, p. 67-68].

Au début de cet article II, trois notions, "chance de production de l'événement", "loi suivant laquelle le phénomène doit apparaître" et "chance moyenne" se réfèrent certainement au même paramètre dont on cherche la valeur inconnue. Gavarret sous-entend sans doute que l'individu α d'une population de taille N, soumis à la médication a une probabilité p_α de mourir, ce paramètre est certainement $p = \dfrac{1}{N}\sum_{\alpha=1}^{N} p_\alpha$, tandis que le nombre qui représente la "fréquence de leur manifestation" (des événements) est certainement : $f_n = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ où $x_i = 1$ si le i-ème individu d'un tirage de n individus est mort et $x_i = 0$ sinon.

Il est remarquable que Gavarret fasse aussi bien la distinction entre le paramètre certain mais inconnu p et la fréquence observée mais aléatoire f_n. Il est aussi remarquable qu'il annonce la convergence de f_n vers p quand n tend vers l'infini.

Après avoir constaté que :

> « il résulte qu'au moment où un observateur arrête son travail d'expérimentation pour calculer la mortalité moyenne avec les faits qu'il a recueillis, il néglige nécessairement tous les malades qui se présenteront à lui après cette époque. Il s'expose donc à une mortalité plus faible ou plus forte que celle qu'il eût obtenue en ajoutant aux faits qu'il possède déjà, la série des faits qui vont arriver, suivant que cette série présentera des résultats contraires ou favorables à la méthode thérapeutique employée, sans que rien *a priori* puisse lui faire apprécier ni l'étendue ni le sens de l'erreur. Or il est facile de voir que cette erreur, très considérable quand on n'agit que sur de petits nombres, disparaît à peu près complètement quand la statistique est très étendue. » [G, p. 68-69].

Ce problème bien connu aujourd'hui en médecine comme celui de données "censurées à droite", ne sera bien sûr étudié qu'un siècle plus tard. Gavarret se contente de montrer qu'il est important d'avoir un grand nombre d'observations au départ. Il ajoute à cinq séries de résultats donnant toutes une mortalité moyenne de $0{,}400$ d'une part une série où il y a eu 3 morts sur 20 malades (Série 1) et d'autre part une série où il y a eu 8 morts sur 10 malades (Série 2). Il compare ainsi les nouvelles fréquences observées à celle de la série initiale.

Série initiale	En ajoutant la série 1	En ajoutant la série 2
12 morts sur 30 malades	0,200	0,500
120 morts sur 300 malades	0,384	0,413
240 morts sur 600 malades	0,392	0,406
360 morts sur 900 malades	0,394	0,404
480 morts sur 1 200 malades	0,396	0,403

Il conclut avec le dernier exemple : « l'influence des séries particulières sur la valeur de la mortalité moyenne est si peu marquée, qu'on pourrait prendre indistinctement l'une des trois expressions de cette mortalité pour la représenter, car l'erreur ne saurait s'élever à plus de 3 ou 4 sur 1 000 malades. » [G, p. 72].

Activité 4.

> « Il résulte en effet du théorème de Poisson [...], que du moment où l'on possède relativement à un événement une statistique composée d'un très grand nombre de faits comparables, on peut toujours, à l'aide des nombres contenus dans cette statistique, calculer l'*erreur* qu'il y aurait à prendre le rapport de fréquence qu'elle fournit par la véritable chance moyenne de cet événement. » [G, p. 74].

À cet effet, il explicite dans la note A [G, p. 253] la démonstration des principes énoncés dans l'article II du Chapitre II tirée du théorème de Poisson qu'il utilisera. Nous reproduisons ci-dessous l'énoncé du théorème utilisé. Une démonstration actuelle se trouve en Annexe III.

Une expérience amène à la réalisation d'un des deux événements contraires E et F.

> « Cette chance moyenne de l'événement sur laquelle nous n'avons aucune notion, si ce n'est qu'elle est tout aussi invariable que l'ensemble des causes possibles dont elle résulte, nous la désignons, quelle qu'elle soit par la lettre B.
>
> Supposons maintenant qu'une très longue suite d'observations relatives aux événements E et F ait donné les résultats suivants :
>
> $m =$ le nombre de fois que E est arrivé.
>
> $n =$ le nombre de fois que F est arrivé.
>
> $\mu =$ le nombre total des observations faites.
>
> en sorte que : $m + n = \mu$. [...] On peut donc dire en toute certitude, que pour une statistique composée d'un très grand nombre de faits comparables, il y a
>
> la probabilité très grande $P = 1 - \dfrac{2}{\sqrt{\pi}} \displaystyle\int_u^\infty e^{-t^2} dt$
>
> que la chance moyenne (B) de l'événement E est comprise entre les limites
>
> $\begin{cases} \dfrac{m}{\mu} + u\sqrt{\dfrac{2.m.n}{\mu^3}} \\ \dfrac{m}{\mu} - u\sqrt{\dfrac{2.m.n}{\mu^3}}. \end{cases}$ » [G, p. 254-256].

Aujourd'hui, pour un échantillon de taille $n \geq 30$, le calcul d'un intervalle de confiance de niveau $1 - \alpha$ avec $0 < \alpha < 1$ pour une proportion inconnue p s'effectue ainsi :

– Si Z une variable aléatoire qui suit une loi normale centrée réduite, on note z_α le nombre réel tel que : $\mathbb{P}(-z_\alpha < Z < z_\alpha) = 1 - \alpha$ et on note f la fréquence observée de l'événement E pour n expériences.
– un intervalle de confiance de niveau $1 - \alpha$ pour p est donné par $[f - e, \quad f + e]$ avec $e = z_\alpha\sqrt{\frac{f(1-f)}{n}}$.

Des formules simplifiées sont souvent utilisées en lycée et en B.T.S. dans le cas où le niveau de confiance est de 95% (donc $\alpha = 0{,}05$). En effet, z_α est alors égal à $1{,}96$ arrondi à 2, et l'expression $f(1-f)$ est majorée par $\dfrac{1}{4}$, ce qui donne un intervalle de confiance de la forme $[f - \dfrac{1}{\sqrt{n}}, \quad f + \dfrac{1}{\sqrt{n}}]$.

Le théorème de Poisson revient à prendre $e = 2\sqrt{\frac{2f(1-f)}{n}}$. Un calcul à partir des tables de la loi normale centrée réduite montre que,

$\mathbb{P}(-2{,}823 < Z < 2{,}823) = 0{,}9953$. Or $2{,}823 \simeq 2\sqrt{2}$. Il s'ensuit qu'en appliquant la formule utilisée aujourd'hui, on retrouve bien la formule utilisé par Gavarret avec un niveau de confiance de $P = 0{,}9953$, « c'est-à-dire 212 à parier contre 1 » [G, p. 257] et donc un risque $\alpha = 1 - P = 0{,}0047$ $(0{,}47\%)$. Il a en fait repris le niveau utilisé par Poisson pour l'étude de jugements de jury.

Trois exemples suivent pour illustrer l'application du théorème de Poisson.

1. Enfants légitimes nés en France tout entière pendant l'année 1825

$468\,151$	garçons.
$436\,443$	filles.
$904\,594$	naissances.

d'où la fréquence observée $f_n = 0{,}5175$.

> « Les résultats de ce relevé nous conduiraient donc à la proposition suivante :
> Sur $10\,000$ naissances il y a en France $5\,175$ garçons. » [G, p. 76].

Mais l'erreur est égale à $0{,}0015$ d'où :

> « au lieu de la proposition précédente, nous devons réellement admettre celle-ci :
> Sur $10\,000$ naissances, le nombre des garçons varie entre $\left\{ \begin{array}{l} 5\,190 \\ 5\,160 \end{array} \right.$ » [G, p. 76].

2. Naissances légitimes en 1836 dans la ville de Paris

$9\,785$	garçons.
$9\,524$	filles.
$19\,309$	naissances.

d'où la fréquence observée $f_n = 0{,}5068$.

> « Et ce rapport nous conduirait à la proposition suivante :
> Sur $10\,000$ naissances il y a en France $5\,068$ garçons. » [G, p. 78].

Mais l'erreur est égale à $0{,}0102$ d'où :

> « Sur $10\,000$ naissances, le nombre des garçons varie entre $\left\{ \begin{array}{l} 5\,170 \\ 4\,999 \end{array} \right.$ » [G, p. 76].

3. Naissances "hors l'état de mariage" en 1836 dans la ville de Paris

$4\,860$	garçons.
$4\,773$	filles.
$9\,633$	naissances.

d'où la fréquence observée $f_n = 0{,}5045$.

> « Ce qui nous conduirait à la proposition suivante :
> Sur $10\,000$ naissances il y a en France $5\,045$ garçons. » [G, p. 79].

Mais l'erreur est égale à $0{,}0144$ d'où :

> « Au lieu de la proposition précédente, il faudrait donc admettre celle-ci :
> Sur $10\,000$ naissances, le nombre des garçons varie entre $\left\{ \begin{array}{l} 5\,189 \\ 4\,901 \end{array} \right.$ » [G, p. 79].

Activité 5.

4.3 Rapports qui existent entre les résultats fournis par deux longues suites d'observations relatives aux mêmes événements.

Au début cet article III [G, p. 80-98] du chapitre II, Gavarret distingue deux cas :

> « *1er Cas.* Si l'*ensemble des causes possibles*, auxquelles se rapportent ces événements en observation, est resté invariablement *le même* pour les deux statistiques ; [...] cette chance moyenne a dû rester la même pour les deux séries d'observations. Les rapports de fréquence obtenus devront donc se rapprocher beaucoup l'un de l'autre. [...]. Cette différence ne saurait dépasser une certaine *limite* qui dépend des nombres contenus dans les deux statistiques, et qu'on peut toujours calculer quand ces nombres sont connus. » [G, p. 81].
>
> « *2ème Cas.* Supposons que l'*ensemble des causes possibles*, bien que *le même* dans toute l'étendue de chacune des deux statistiques, varie cependant de l'une à l'autre ; [...] la différence entre leurs résultats dépasse cette *limite.* » [G, p. 83-84].

C'est la démarche des tests d'inférence qui est ainsi annoncée. Une traduction en termes modernes se décline ainsi :

Posons p_1 la "chance moyenne" pour la première série d'observations et p_2 pour la deuxième série d'observations.

Nous avons donc affaire à deux hypothèses : l'hypothèse nulle H_0 et l'hypothèse alternative H_1.

H_0 : l'ensemble des causes possibles est le même pour les deux séries statistiques ;

H_1 : l'ensemble des causes possibles n'est pas le même pour les deux séries statistiques.

Dans le premier cas : $p_1 = p_2$ et dans le deuxième cas $p_1 \neq p_2$.

Soit f_{n_1} la fréquence de l'événement pour la série de n_1 observations et f_{n_2} la fréquence de l'événement pour la série de n_2 observations et $d = |f_{n_1} - f_{n_2}|$ la valeur absolue de la différence des fréquences.

Pour décider entre les deux hypothèses, Gavarret propose d'utiliser la règle suivante :

On décide que l'ensemble des causes possibles est le même si $d \leq l$ où l est une limite calculée grâce à la formule de Poisson et on décide que l'ensemble des causes possibles n'est pas le même si $d > l$. La valeur de l est donnée dans une note B dont nous reproduisons une partie ci-dessous [G, p. 264-267]. Une démonstration actuelle se trouve en Annexe IV.

Pour calculer cette limite l qui est « la *limite* des oscillations compatibles avec l'invariabilité de l'ensemble. » [G, p. 84], il s'appuie encore sur la formule de Poisson qu'il explicite dans cette note :

> « Cela posé, une première statistique relative aux deux événement E et F, a fourni les résultats suivants :
>
> $m =$ le nombre de fois que E est arrivé.
>
> $n =$ le nombre de fois que F est arrivé.
>
> $\mu =$ le nombre total des observations recueillies. [...]
>
> Une deuxième statistique relative aux mêmes événements E et F, toujours appuyée sur de très grands nombres, donne les résultats suivants :
>
> $m^{'} =$ le nombre de fois que E est arrivé.
>
> $n^{'} =$ le nombre de fois que F est arrivé.
>
> $\mu^{'} =$ le nombre total des observations recueillies. [...]
>
> On a la probabilité $P = 0{,}9953$ [...], que la différence entre les rapports $\dfrac{m}{\mu}$ et $\dfrac{m^{'}}{\mu^{'}}$, fournis par les deux statistiques, ne dépassera pas la quantité $2\sqrt{\dfrac{2.m.n}{\mu^3} + \dfrac{2.m^{'}.n^{'}}{\mu^{'3}}}$ »

Il observe les conséquences de cette propriété sur trois groupes d'exemples :

1er Groupe d'exemples.

1er Exemple. « Les relevés publiés par l'administration nous enseignent, qu'abstraction faite des procès politiques, les jugements du jury, en France, se sont répartis ainsi qui suit, pendant les années 1832-1833.

1832	1833
4 448 .... condamnés.	4 105 .... condamnés.
3 107 .... acquittés.	2 859 .... acquittés.
7 555 .... accusés.	6 964 .... accusés.

[...] La différence entre les résultats fournis par ces deux relevés ne s'élève qu'à 8 condamnés sur 10 000 accusés.

Mais pendant ces deux années, la législation est restée la même, l'*ensemble des causes possibles* est donc resté *invariable*, si donc nos formules sont vraies, la différence entre les résultats obtenus doit être moindre que la *limite* des variations compatibles avec la permanence des causes qu'elles fournissent ; c'est ce qui a réellement lieu car cette limite est égale à 231 condamnés sur 10 000 accusés. » [G, p. 88-89].

Les fréquences en 1832 et 1833 sont 0,5887 et 0,5895. On a bien $0{,}0008 = d < l = 0{,}0231$.

2ème Exemple. Mouvement annuel de la population de la France. Comparaison 1831-1833.

1831	1833
802 761 nombre total de morts.	812 548 nombre total de morts.
31 758 173 personnes échappées à la mort.	32 094 222 personnes échappées à la mort.
32 560 934 population de la France.	32 906 770 population de la France.

« Mais pendant ces deux années, aucune grande épidémie n'a existé en France, l'*ensemble des causes possibles* de mort est donc resté *invariable*. » [G, p. 90].

Les fréquences en 1831 et 1833 sont 0,024 654 et 0,024 692. Ici $0{,}000\,041 = d < l = 0{,}000\,108$ qui est bien le résultat attendu par Gavarret.

3ème Exemple. Enfants nés dans l'état de mariage en 1824 et 1825.

1824	1825
471 490 nombre de garçons.	468 151 nombre de garçons.
441 488 nombre de filles.	436 443 nombre de filles.
912 978 nombre de naissances.	904 594 nombre de naissances.

« Mais évidemment, pendant ces deux années, l'ensemble des causes possibles n'a pas été perturbé. » [G, p. 90].

Ici encore, la différence est inférieure à la limite calculée.

Activité 6.

Dans les trois exemples, Gavarret constate simplement que sous l'hypothèse nulle, la différence reste bien inférieure à la limite calculée (en réalité, sous cette hypothèse nulle, cet événement n'est pas certain mais il a une probabilité égale à 0,9953). L'utilisation de la règle de décision l'aurait amené à ne pas rejeter l'hypothèse nulle.

2ème Groupe d'exemples.

1er Exemple. « Les relevés publiés par l'administration relativement aux jugements du jury en France, depuis 1825 jusqu'à 1831 inclusivement, abstraction faire des procès politiques, nous fournissent les résultats suivants :

de 1825 à 1830 inclusivement	1831
25 777 nombre de condamnés.	4 098 nombre de condamnés.
16 523 nombre des acquittés.	3 508 nombre des acquittés.
42 300 nombre des accusés.	7 606 nombre des accusés.

[...] Si donc les formules sont vraies, on doit trouver que la legislation de 1831 n'est pas la même que celle des six années précédentes. Or, chacun sait que de 1825 à 1830 les condamnations étaient prononcées à la majorité de sept contre cinq ; tandis que en 1831 elles ne le furent plus qu'à huit contre quatre, ce qui nécessairement dût diminuer le nombre des condamnations. » [G, p. 92].

Les fréquences avant 1831 et en 1831 sont $0{,}6094$ et $0{,}5388$. Ici $0{,}0709 = d > l = 0{,}00175$. Le rejet de l'hypothèse nulle est donc confirmé par les faits.

2ème Exemple. Mouvement annuel de la population de la France. Comparaison 1831-1832.

1831	1832
802 761 nombre total de morts.	933 733 nombre total de morts.
31 758 173 personnes échappées à la mort.	31 815 602 personnes échappées à la mort.
32 560 934 population de la France.	32 749 335 population de la France.

Les fréquences pour 1831 et 1832 sont $0{,}024\,654$ et $0{,}028\,511$. Ici $0{,}003\,857 = d > l = 0{,}000\,112$. Il y a rejet de l'hypothèse nulle.

« La grande épidémie de choléra qui ravagea la France en 1832, vient parfaitement justifier les conclusions auxquelles nous a conduit l'emploi des principes de la loi des grands nombres. » [G, p. 93].

3ème Groupe d'exemples.

Répartition des naissances dans l'état de mariage et hors l'état de mariage pendant les années 1824 et 1825.

Enfants légitimes	Enfants illégitimes
939 641 .... garçons.	71 661 .... garçons.
877 931 .... filles.	68 905 .... filles.
1 817 572.......naissances.	140 566 .... naissances.

Les fréquences pour les enfants légitimes et illégitimes sont $0{,}51\,697$ et $0{,}50\,980$. Ici $0{,}00717 = d > l = 0{,}00391$. Il y a encore rejet de l'hypothèse nulle.

« ...nous devons en conclure que les enfants légitimes ont plus de chance que les enfants illégitimes de naître garçons. Cette proposition, loin de surprendre, pouvait être en quelque sorte prévue *a priori* ; car des documents authentiques prouvent que partout où existe la *monogamie*, les enfants mâles naissent plus nombreux que ceux du sexe féminin, tandis que le contraire a lieu dans les pays où existe la *polygamie*. » [G, p. 94].

Première remarque : le 3^{e} groupe d'exemples (réduit à un exemple) était destiné à illustrer comment la règle énoncée pouvait permettre de « savoir si une cause perturbatrice est ou non survenue dans l'ensemble de toutes celles qui peuvent intervenir pour déterminer la manifestation du phénomène étudié.» [G, p. 88]. Il semble être de même nature que le deuxième groupe où l'existence d'une "perturbation évidente" (changement du mode de vote du jury dans le 1er exemple et existence d'une épidémie de choléra dans le 2e exemple) que l'on connaissait par avance, valide ainsi la décision de rejet de l'hypothèse nulle et donc la méthode utilisée. Cependant la règle de décision appliquée au 3e groupe d'exemples amène à une conclusion plus surprenante, même si Gavarret s'efforce de justifier ici, avec des arguments discutables, l'existence d'une "perturbation évidente". Le sens commun voudrait que l'hypothèse nulle (égalité de la chance moyenne d'avoir un garçon dans le cas d'une naissance légitime ou illégitime) soit vraie. Or la construction du test implique que dans ce cas la probabilité de rejeter l'hypothèse nulle à tort est très petite car égale à $0{,}0047$.

Deuxième remarque : La notion de "chance moyenne" semble précisée dans ce paragraphe. En effet, Gavarret avertit : « il y aurait erreur grave à penser que la notion de la chance moyenne d'un événement déduite d'une longue série d'épreuves, puisse jamais indiquer ce qui arrivera dans tel ou tel cas particulier [...]. Cette chance moyenne ne doit jamais être employée qu'à la prévision des *limites* entre lesquelles peut varier le résultat final d'une grande collection de faits. » [G, p. 82]. Et plus loin : « il tombe sous le sens que jamais une statistique faite ne pourra servir à prévoir quelle est la chance de condamnation de tel ou tel accusé, ni la chance de naissance d'un garçon dans tel ou tel mariage particulier. » [G, p. 83].

Comme nous l'avions pressenti en 4.2. le modèle semble bien celui où chaque individu α dans une population de taille N a une probabilité p_α que l'événement E se réalise. Si on effectue un tirage au hasard d'un individu et que l'on s'interroge sur la probabilité qu'à la suite de ce tirage, l'événement E se réalise, cette probabilité est bien la probabilité moyenne $p = \dfrac{1}{N}\sum_{\alpha=1}^{N} p_\alpha$ et elle pourra être approchée par f_n dans des limites qui sont calculables. Mais si on se fixe un individu particulier α, en aucun cas la statistique ne pourra nous permettre d'approximer p_α. Par cet avertissement, Gavarret répond d'avance aux médecins qui lui objectent que les statistiques ne sont d'aucun intérêt pour la connaissance du malade qu'ils ont en face d'eux. Les statistiques à ce stade ne peuvent donner de conclusions que sur les "chances moyennes" et non sur la "chance" propre d'un individu particulier.

Ce chapitre se termine par quatre conclusions :

1ère Conclusion : Lorsqu'on a fait « une centaine d'observations comparables », on pourra déterminer « quelle différence il peut exister entre le rapport fourni par la statistique et la vraie chance moyenne inconnue. » [G, p. 96].

Il est remarquable qu'à cette époque, Gavarret fasse aussi nettement la distinction entre la statistique (qui est connue mais aléatoire) et le paramètre inconnu mais certain "la vraie chance moyenne inconnue". Cependant la condition imposée "une centaine d'observations comparables" a dû effrayer nombre de médecins et n'est pas en réalité aussi drastique.

2ème Conclusion : Lorque le nombre d'expériences est trop petit « toute loi de manifestation d'un phénomène à chance variable [...] ne mérite aucune confiance. » [G, p. 96].

Actuellement, il existe des moyens de décider à partir de petits échantillons et dire que toute conclusion ne mérite "aucune" confiance est exagéré.

3ème Conclusion : Si les conditions sont les mêmes dans deux séries d'expériences, la différence des fréquences « ne pourra pas dépasser une certaine *limite* assignable » Réciproquement, si la différence des fréquences est inférieure à la limite « on pourra affirmer qu'*aucune perturbation* n'est survenue pendant la durée de l'épreuve » [G, p. 97].

Pour être totalement, rigoureux, Gavarret aurait dû préciser : "avec la probabilité P, la différence des fréquences ne pourra pas dépasser une certaine limite". Aujourd'hui, il est bien connu que si la différence des fréquences est inférieure à la limite, en décidant qu'aucune perturbation n'est survenue pendant la durée de l'épreuve, on s'expose à un "risque de deuxième espèce" (Annexe I) qui peut être important.

4ème Conclusion : Si les conditions ne sont pas les mêmes dans deux séries d'expériences (cause perturbatrice), la différence des fréquences « dépassera une certaine *limite* assignable. ». Réciproquement, si la différence des fréquences est supérieure à la limite « on pourra affirmer qu'il est survenu une perturbation pendant la durée des épreuves » [G, p. 98].

Pour être totalement, rigoureux, Gavarret aurait dû préciser : "avec une probabilité qui dépend à la fois de la taille des échantillons et de la différence inconnue entre les causes perturbatrices, la différence des fréquences dépassera une certaine limite".

De même, si la différence des fréquences est supérieure à la limite, on s'expose au "risque de première espèce" (Annexe I) en affirmant qu'il est survenu une perturbation pendant la durée des épreuves.

Les chapitres suivants traitent de l'application de ces principes.

5 Application des principes de la loi des grands nombres aux recherches de thérapeutique.

C'est dans ce Chapitre III [G, p. 97-162] que Gavarret commence à traiter plus spécifiquement de l'application à la médecine.

5.1 Employer en médecine les principes empruntés au calcul des probabilités, n'est pas consacrer un empiètement des mathématiques sur les sciences médicales. - Causes nécessaires et causes probables. - Exposé des sources d'où dérivent les causes possibles de mort et de guérison qui pèsent sur les malades. - De la bonne foi en médecine.

Dans cet article premier [G, p. 97-115], Gavarret s'emploie à justifier l'usage des probabilités en médecine. Si la logique peut faire défaut, c'est parce que « l'intelligence humaine est trop bornée pour que, livrée à ses propres ressources, elle puisse combiner longtemps ensemble un grand nombre d'idées de manière à en déduire les conséquences qui en découlent nécessairement [...]. Ce n'est donc que comme instrument de raisonnement, comme supplément à la logique, là où elle nous fait défaut, que le calcul des probabilités peut être introduit en médecine. » [G, p. 101]. Ce qui est en réalité sous-tendu par ces phrases est que c'est l'impossibilité matérielle de recenser la combinaison des causes qui amène à utiliser un outil supplémentaire, le calcul des probabilités. Gavarret distingue deux ordres bien distincts pour l'étude des phénomènes dont s'occupent les sciences d'observation. Le premier ordre est celui des faits « qui sont liés à des causes dites *nécessaires*. » [G, p. 102]. Il cite comme exemple « le corps abandonné à lui-même, il tend *nécessairement* à tomber vers le centre de la terre. » [G, p. 102]. « Dans le deuxième ordre viennent se ranger les événements qui ne sont liés qu'à des causes *probables* » [G, p. 104]. Parmi les exemples cités, « une maladie ne doit pas être considérée comme une cause *nécessaire* de mort. » [G, p. 105]. De même, « une médication n'entraine pas *nécessairement* la guérison du malade. » [G, p. 105].

La nature de la variable aléatoire étudiée est clairement annoncée : il s'agit d'une variable de Bernoulli où deux événements peuvent survenir d'une part, la mort, et d'autre part, la guérison du malade. Pour pouvoir faire des statistiques, il faut des observations « relatives à des phénomènes dont la manifestation était due à l'intervention d'une cause quelconque ayant rang dans un ensemble de causes possibles invariable pendant la durée des épreuves. » [G, p. 110]. Ces causes sont réduites par Gavarret à cinq causes principales :

1. les conditions individuelles ;
2. les conditions hygiéniques antécédentes à l'invasion de la maladie ;
3. les conditions hygiéniques pendant le traitement ;
4. la maladie elle-même ;
5. la méthode thérapeutique employée.

Il s'agit de résoudre « deux problèmes fort importants et bien distincts l'un de l'autre :

1. Déterminer [...] la chance moyenne qu'on a de triompher d'une affection connue, en employant telle ou telle médication pour la combattre. [...].

2. [...] classer les différentes médications conseillées par les auteurs contre la même affection, d'après les chances moyennes de guérison qui leur sont propres. » [G, p. 112-113].

En termes modernes, si nous notons :

- X la variable aléatoire binomiale $= 1$ si mort et $= 0$ sinon ;
- $A, B, C, ...$ les facteurs (non contrôlés) susceptibles d'intervenir dans la guérison ou non et contenus dans les quatre premières causes possibles énumérées plus haut par Gavarret, chaque facteur $A, B, ...$ ayant $n_a, n_b, ...$ modalités (ainsi le facteur sexe mentionné par Gavarret possède 2 modalités, ...mais combien de modalités possède le facteur "influences morales qui ont pu agir sur lui" ?) ;
- M le facteur (contrôlé) : la "médication", méthode thérapeutique appliquée parmi n_m méthodes utilisables ou bien dose de médicament administrée, etc.

Il s'agira donc de construire une fonction f telle que $\mathbb{P}(X = 1) = f(A, B, ..., M)$.

Activité 7.

Enfin, Gavarret aborde ensuite ce qui est nommé aujourd'hui la "reproductibilité" d'une expérience dont tous les facteurs ne sont pas contrôlés : « Il est [...] des sciences dans lesquelles l'homme peut bien étudier et constater les phénomènes qui se passent autour de lui, mais jamais, en général, les produire de toutes pièces. Et même, dans le petit nombre de cas où il est maître de leur donner naissance, les circonstances sont tellement variables suivant les sujets, et les causes de ces variations lui sont si peu connues, que jamais il ne peut acquérir la certitude d'avoir opéré dans des conditions absolument identiques à celles qu'a rencontrées un autre observateur. » [G, p. 114].

5.2 Déterminer par l'expérience la chance de guérison que fournit une modification donnée dans un maladie également donnée.

Dans cet article II [G, p. 115-142], Gavarret insiste longuement sur les conditions auxquelles l'observateur doit se conformer pour l'expérimentation, l'objectif étant de ne faire apparaître que ce qui dépend exclusivement de la médication. Parmi ces conditions au nombre de cinq, nous ne nous attarderons que sur la suivante : « Les malades doivent être pris exclusivement dans la même localité et dans les mêmes classes de la population. » [G, p. 117]. Avec cette précaution, il affirme que cette « première condition suffit pour rendre *invariable* pendant toute la durée des épreuves, l'*ensemble des causes possibles* de mort et de guérison qui se rapportent aux trois premières sources précédemment exposées. » [G, p. 117], c'est-à-dire : les conditions individuelles, les conditions hygiéniques antécédentes à l'invasion de la maladie, les conditions hygiéniques pendant le traitement. Si par exemple, dans un même travail « on allait englober des résultats obtenus dans des hôpitaux destinés à des usages différents ou appartenant à des localités différentes, ou enfin les faits de la pratique extérieure avec ceux de la pratique des hôpitaux » [G, p. 119] cette condition ne serait pas remplie et les « observations, dans ce cas, cesseraient d'être *comparables* entre elles, la statistique contiendrait des quantités hétérogènes, et l'addition des cas ne serait plus légitimes. » [G, p. 119].

En d'autres termes, il s'agit de se mettre dans une situation où on souhaite étudier l'effet d'un facteur "toutes les conditions étant égales par ailleurs". Dans une formulation moderne, le modèle se réduit à $\mathbb{P}(X = 1) = f(M)$ puisque les autres facteurs sont considérés comme invariants dans la population considérée.

En fait, si on ne considère qu'une médication, le problème revient à estimer la mesure exacte de l'influence exercée par la médication essayée, soit $p = \mathbb{P}(X = 1)$. En appliquant les formules de Poisson, Gavarret affirme que :

> « Toute statistique médicale fournit le moyen de déterminer les *limites* entre lesquelles peut osciller au dessus et au dessous de la mortalité moyenne observée, la véritable mortalité moyenne cherchée résultat de la médication essayée. L'énoncé de la mortalité moyenne fournie par l'observation, et des *limites d'erreur possible* qu'on en a déduites constitue une loi thérapeutique. » [G, p. 130].

Ainsi donc, Gavarret ne se contente pas d'une estimation ponctuelle de p inconnu, comme le faisait Louis, mais entend par "loi thérapeutique", la donnée de l'intervalle de confiance pour ce p inconnu. Devant les difficultés observées en 2011 pour obliger les instituts de sondages à publier, en même temps que leurs estimations ponctuelles, les intervalles de confiance qui les accompagnent, on mesure combien Gavarret peut sembler un précurseur.

Il propose entre les pages 142 et 143 une table à deux entrées : une entrée "Mortalité moyenne fournie par la statistique" où sont données 16 mortalités moyennes prises entre $0{,}100\,000$ et $0{,}350\,000$ et une entrée "nombre d'observations", où sont donnés « des relevés composés de $300, 350, 400, 450$, etc...$1\,000$ observations ». Au croisement de ces deux entrées, l' "erreur possible" est calculée en utilisant les formules détaillées plus haut. Pour chacune des 240 cases au croisement des lignes et des colonnes, il a calculé « l'étendue de l'erreur qu'on pourrait commettre en prenant la mortalité moyenne fournie pour la traduction exacte et rigoureuse de l'influence cherchée. » [G, p. 131].

Il explique la façon de se servir de cette table sur un exemple où la mortalité moyenne observée est $0{,}150\,000$ avec 800 observations.

Mortalité moyenne fournie par la statistique	STATISTIQUES de 800 cas	
	Répartition des malades	Erreur possible
0,150 000	120 morts 680 guéris	0,035 707

La conclusion est simple :

« il en résulte que la véritable mortalité cherchée est comprise entre les nombres : $0{,}150\,000$ plus $0{,}035\,707$, c'est-à-dire $0{,}185\,707$ et $0{,}150\,000$ moins $0{,}035\,707$, c'est-à-dire $0{,}114\,293$. » [G, p. 135].

Activité 8.

En d'autres termes, Gavarret fournit, à partir d'un échantillon de taille $n = 800$ pour lequel la fréquence de mortalité observée est de $0{,}15$, un intervalle de confiance de niveau $0{,}9953$ pour la véritable mortalité cherchée p. Il avait fustigé auparavant ceux qui en « se conformant aux principes professés tous les jours par les partisans de la méthode numérique » [G, p. 134] posent la proposition suivant : « Sous l'empire de la médication, il doit mourir $150\,000$ personnes sur $1\,000\,000$ malades ou bien 15 sur 100 malades. » [G, p. 134].

Avec sa méthode, « cesseront de paraître contradictoires des résultats appartenant à divers observateurs et dont les différences rentrent réellement dans les *limites d'oscillation* de la loi thérapeutique établie. » [G, p. 136].

Bien entendu, ce n'est pas la "loi thérapeutique" (ici p) qui oscille mais l'intervalle de confiance. Mais cette erreur existait déjà chez Laplace, sauf à prendre un point de vue bayésien sur cette question.

Il complète cet article par la constatation (p. 137-138) que, pour une même mortalité moyenne observée, les "erreurs possibles" décroissent à mesure que le nombre des observations recueillies augmente. Autre constatation (p. 138-140), les "erreurs possibles" croissent au contraire à mesure que la mortalité moyenne observée se rapproche de 1/2.

Il applique cette formule aux résultats obtenus par Louis dans ses recherches sur la fièvre typhoïde sur 140 cas où 52 sont morts et 88 sont guéris. Il calcule l'erreur possible et conclut :

> « Cet habile observateur, dans ses recherches sur la fièvre typhoïde, a essayé d'éclairer la thérapeutique de cette maladie en se livrant à l'analyse la plus minutieuse de 140 cas de cette maladie. Les sujets observés sont répartis ainsi qu'il suit :
>
> | | |
> |---|---|
> | 52 | morts. |
> | 88 | guéris. |
> | 140 | malades. |
>
> La mortalité moyenne est donc, dans ce cas, égale à $0{,}37143$.
>
> C'est-à-dire qu'en prenant ce rapport par la mesure de l'influence des moyens curatifs employés, on devrait poser en loi qu'il doit mourir $37\,143$ personnes sur $100\,000$ malades ou approximativement 37 sur 100 malades [...]. Si, à l'aide des principes de la loi des [grands] nombres, nous cherchons à déterminer l'étendue de l'*erreur possible* dont est entachée une semblable conclusion, nous la trouvons égale à $0{,}115\,500$.
>
> En sorte que tout ce que nous apprend, en réalité, le travail de Louis, c'est que sous l'empire des moyens curatifs employés dans ses 140 observations, le nombre des morts doit varier [...] approximativement entre 49 et 26 sur 100 malades. » [G, p. 141].

C'est, nous semble-t-il, la première fois que, dans le domaine médical, la "méthode numérique" est enrichie d'un encadrement de l' "erreur" que l'on peut commettre en confondant la "chance moyenne" avec la fréquence d'apparition de l'événement.

5.3 Classer les médications d'après l'influence qu'elles exercent sur une même maladie.

Cet article III [G, p. 143-162] traite de la comparaison des lois.

Soient $M_1, M_2, ..., M_k$, k médications pour la même maladie et $p_1, p_2, ..., p_k$ les probabilités de mourir. Le problème consiste, à partir de l'observation des fréquences de mortalité $f_1, f_2, ..., f_k$ d'aboutir à une décision quant à un ordre possible des p_j. Gavarret pense que pour résoudre ce problème de classification de méthodes thérapeutiques, « on doit les comparer deux à deux, de manière à déterminer quelle est celle qui conserve les avantages sur toutes les autres. » [G, p. 144].

Aujourd'hui, on sait que ce problème n'a pas de solution évidente et que cette comparaison deux à deux, outre qu'elle peut aboutir à des résultats totalement contradictoires au final, entraîne une augmentation du "risque de première espèce", ici le risque de décider à tort que les méthodes n'ont pas la même efficacité.

Pour effectuer cette comparaison deux à deux, il énonce sa méthode :

« On devra donc, une fois les deux relevés médicaux effectués, calculer à l'aide des nombres qu'ils renferment, la *limite* compatible avec l'*invariabilité des causes* de mort et de guérison. On devra ensuite comparer cette *limite* avec la *différence* qui existe entre les mortalités moyennes fournies par les deux statistiques ; il pourra alors arriver deux cas :

1. La *différence* entre les deux mortalités moyennes déduites des deux relevés, est supérieure à la *limite* calculée. On doit nécessairement en conclure que l'une des deux médications vaut mieux que l'autre, et qu'il y a lieu de préférer celle qui a fourni les résultats les plus favorables.
2. La *différence* entre les mortalités moyennes déduites des deux relevés, est inférieure à la *limite* calculée. Alors, bien que l'une des deux médications ait paru avoir une sorte de supériorité, cependant comme les résultats définitifs ne diffèrent entre eux que d'une quantité assez petite pour rentrer dans les *limites des erreurs possibles* dans les conclusions *a posteriori*, tout porte à croire que ces deux médications exercent la même influence et il n'y a pas lieu de préférer l'une à l'autre. » [G, p. 145-146].

En d'autres termes, si p_1 (resp. p_2) est la probabilité de mortalité avec la médication M_1 (resp. M_2), si f_1(resp. f_2) est la fréquence observée des morts à la suite de la médication M_1 (resp. M_2), une limite l est calculée ainsi que la différence $d = |f_1 - f_2|$. Si $d > l$, on rejette $H_0 : p_1 = p_2$ et on décide $p_1 < p_2$ (resp. $p_1 > p_2$) si $f_1 < f_2$ (resp. $f_1 > f_2$). Si $d < l$, on ne rejette pas $H_0 : p_1 = p_2$.

La “méthode numérique”, utilisée par Louis est beaucoup plus simple puisque il décide $p_1 = p_2$ si $f_1 = f_2$ et $p_1 < p_2$ (resp. $p_1 > p_2$) si $f_1 < f_2$ (resp. $f_1 > f_2$).

Pour pouvoir effectuer cette comparaison, « on doit être bien sûr que la thérapeutique seule a pu varier d'une suite d'épreuves à l'autre. » [G, p. 146]. On retrouve la préoccupation que toutes les conditions soient “égales par ailleurs”. Il énonce une condition unique : « les deux statistiques doivent être recueillies dans la même localité et dans des établissements destinés aux mêmes usages. » [G, p. 147].

Suivent alors des exemples où cette condition n'est pas remplie : comparaison de « résultats obtenus dans les hôpitaux ordinaires avec ceux des hôpitaux militaires. Dans ces derniers, en effet, les sujets étant généralement jeunes, vigoureux, et du sexe masculin. » [G, p. 148]. Pour régler le problème de la localisation géographique, « il est bien certain que si tous les malades, passant par le bureau central, étaient distribués au hasard dans les divers services de Paris, il n'y aurait pas de raison pour que les cas graves fussent dirigés vers tel hôpital plutôt que tel autre. » [G, p. 149].

Bien avant les théorisations de Fischer, Gavarret semble avoir l'intuition de ce qui appelé aujourd'hui un plan d'expérience en randomisation totale.

Dans les exemples qui suivent, l'intérêt consiste principalement dans les commentaires faits par Gavarret pour accompagner les décisions prises.

Il reprend la suggestion de Louis consistant à comparer deux groupes de 500 malades.

« *1er Exemple*. Supposons que dans une épidémie, 500 malades, pris au hasard ayant été soumis à une médication, et 500 autres également ayant été pris au hasard, à un traitement différent, on ait obtenu les résultats suivants :

1ère médication	2ème médication
100...morts 400...guéris 500...malades En calculant la mortalité moyenne, nous serions conduits à la proposition suivante : Sous l'influence de cette première médication il doit mourir 20 000 personnes , sur 100 000 malades	150... morts 350 ....guéris 500...malades En calculant la mortalité moyenne, nous serions conduits à la proposition suivante : Sous l'influence de cette seconde médication, il doit mourir 30 000 personnes sur 100 000 malades

[...] En se conformant à la manière de raisonner de Louis, on en conclurait que la première médication est préférable à la seconde [...] » [G, p. 156].

2ème Exemple : « La même expérience ayant été tentée, on est arrivé aux résultats suivants :

1ère médication	2ème médication
100...morts 400...guéris 500...malades	130... morts 370 ....guéris 500...malades
il doit mourir 20 000 personnes sur 100 000 malades	il doit mourir 26 000 personnes sur 100 000 malades

En raisonnant toujours d'après les mêmes principes, on en conclurait nécessairement que la première médication doit être préférée à la seconde. » [G, p. 157-158].

Mais ici $d = 0{,}06$ et $l = 0{,}07694$ et donc $d < l$.

> « Évidemment, puisque la *différence* entre les deux mortalités moyennes est inférieure à cette *limite des erreurs possibles* dans la conclusion *a posteriori*, nous devons reconnaître, que cette variation dans les résultats ne nous enseigne rien, que nous ne sommes pas autorisés à préférer une des deux méthodes à l'autre. Ceci ne veut pas dire que l'une des deux méthodes thérapeutiques employées ne vaille réellement pas mieux que l'autre, mais seulement que les différences constatées entre leurs effets dans 500 expériences tentées de part et d'autre ne sont pas suffisantes pour indiquer laquelle vaut le mieux. » [G, p. 158].

La "méthode numérique", aboutit dans ces deux exemples, à la même conclusion, à savoir la supériorité de la première médication alors que Gavarret remet en cause le seul recours à la comparaison des proportions observées pour conclure à un classement des "médications" et ne donne pas, dans les deux exemples, la même conclusion.

Il est remarquable que, quand l'hypothèse nulle n'est pas rejetée, Gavarret n'affirme pas ici : "les traitements sont identiques" - ce qui est aujourd'hui encore le lot commun de beaucoup d'utilisateurs de statistiques.

Pour appuyer son argumentation, Gavarret modifie ce 2^{e} exemple avec trois versions différentes :

1ère version : les mortalités moyennes sont les mêmes que dans le deuxième exemple mais les effectifs doublent.

1ère médication	2ème médication
200...morts	260... morts
800...guéris	740 ....guéris
1 000...malades	1 000...malades
il doit mourir	il doit mourir
20 000 personnes	26 000 personnes
sur 100 000 malades	sur 100 000 malades

Ici $d = 0{,}06$ et $l = 0{,}05306$. Puisque $d > l$, « ceci nous prouve que la première médication vaut mieux que la seconde, et que $1\,000$ expériences ont mis en évidence une supériorité que 500 n'avaient pu nous faire apprécier. » [G, p. 159].

2ème version : les mortalités moyennes se sont rapprochées l'une de l'autre, avec encore des effectifs doublés.

1ère médication	2ème médication
220...morts	240... morts
780...guéris	760 ....guéris
1 000...malades	1 000...malades
il doit mourir	il doit mourir
22 000 personnes	24 000 personnes
sur 100 000 malades	sur 100 000 malades

Ici $d = 0{,}02$ et $l = 0{,}05321$. Puisque $d < l$, « il n'y a donc pas encore lieu de préférer un des deux traitements à l'autre, la supériorité de l'un sur l'autre, si elle existe, est assez peu marquée que $1\,000$ épreuves n'aient pas pu la mettre en évidence. » [G, p. 160]. Avec la méthode numérique, la première médication aurait été choisie.

3ème version :

1ère médication	2ème médication
250...morts	190... morts
750...guéris	810 ....guéris
1 000...malades	1 000...malades
il doit mourir	il doit mourir
25 000 personnes	19 000 personnes
sur 100 000 malades	sur 100 000 malades

Ici $d = 0{,}06$ et $l = 0{,}05226$. Puisque $d > l$, « il faudrait, en cas pareil, admettre que la seconde médication, qu'on aurait regardée comme inférieure à la première, après 500 épreuves, en se conformant aux principes posés par Louis, lui est cependant supérieure, et que cette supériorité a exigé $1\,000$ faits observés pour être mise en évidence. » [G, p. 161].

Dans cette dernière version, la contradiction apparente entre la conclusion qu'en aurait tiré Louis et celle de Gavarret n'existe pas en réalité car il est évident que si Louis avait dû prendre une décision au regard des 2000 données de cette dernière version, il aurait pris la même que Gavarret.

La différence essentielle entre la "méthode numérique" et celle de Gavarret réside dans le fait que l'observation des fréquences entre deux séries de résultats entraîne immédiatement chez Louis une conclusion certaine quant au classement des méthodes thérapeutiques par l'identification des "chances moyennes" à ces fréquences, alors que chez Gavarret, la distinction entre "chances moyennes" et fréquences est toujours clairement énoncée et la décision concernant le classement des "chances moyennes" prend en compte les fluctuations d'échantillonnage auxquelles sont soumises les fréquences, et s'assortit de beaucoup de précautions dans l'énoncé des conclusions.

Les conclusions de Gavarret appellent cependant quelques remarques :

1. Dans le cas de rejet de l'hypothèse nulle, Gavarret formule la décision : "existence d'une différence entre les deux traitements" comme une proposition certaine alors qu'en réalité la probabilité de se tromper en affirmant ceci est asymptotiquement égale à $0{,}47\%$. Il est prisonnier de la formule de Poisson qui semble ne donner la formule que pour ce risque de première espèce. Il est vrai qu'il explique avant que, pour lui, cette probabilité de se tromper est si petite qu'elle autorise à énoncer cette proposition comme une certitude.

2. Le test utilisé ici, repose en fait sur la construction de l'intervalle de confiance asymptotique de niveau $0{,}9953$ de la différence des probabilités $\delta p = p_1 - p_2$. Il n'y a pas équivalence stricte entre ce test construit à partir de l'intervalle de confiance et le test asymptotique utilisé aujourd'hui pour de grands échantillons qui est légèrement différent. En effet, la loi de la statistique de test étant calculée sous l'hypothèse nulle $p_1 = p_2$, la variance intervenant dans la condition de Liapounov se ramène à :

$$\Big(\frac{1}{n_1} + \frac{1}{n_2}\Big)p^*(1-p^*)$$

où $p^* = p_1 = p_2$ inconnus. p^* sera estimé par $\dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$.

En reprenant les notations de Gavarret et en notant $\mu^* = \mu + \mu'$, $m^* = m + m'$, $n^* = n + n'$, la limite l devient :

$$l = 2\sqrt{\Big(\frac{1}{\mu} + \frac{1}{\mu'}\Big)\Big(\frac{2m^*n^*}{\mu^{*2}}\Big)}$$

En appliquant cette nouvelle formule aux données de la troisième version de l'exemple 2, la limite l ne change guère puisqu'elle est alors égale à : $l = 0{,}05240$ au lieu de $l = 0{,}05226$ avec la formule de Poisson.

Activité 9.

3. Le test utilisé par Gavarret n'est certes pas le plus puissant qui existe, mais il a le mérite d'avoir un risque de premier espèce qui est celui annoncé (c'est-à-dire $0{,}47\%$). Gavarret, en testant si la différence des proportions est différente ou non de zéro, ne tombe pas dans le piège courant qui consiste à calculer séparément pour p_1 et pour p_2 des intervalles de confiance de niveau $1 - 0{,}0047 = 0{,}9953$ et à décider que $p_1 \neq p_2$ si ces intervalles ne se chevauchent pas.

Activité 10.

4. Lorsque l'hypothèse nulle est rejetée, pour Gavarret la décision $p_1 < p_2$ ou $p_1 > p_2$ est établie sur la seule observation des fréquences f_1 et f_2 alors qu'en toute rigueur, pour avoir le même risque d'erreur de première espèce que celui utilisé par Gavarret, on devrait utiliser la règle suivante :

   Test $H_0 : p_1 = p_2$ contre $H_1 : p_1 < p_2$ avec le risque 0,47%.
   Calcul de l'
   $$l' = 2{,}597\,153\sqrt{\left(\frac{1}{\mu} + \frac{1}{\mu'}\right)\left(\frac{m^*n^*}{\mu^{*2}}\right)}$$
   Rejet de H_0 si $f_2 - f_1 > l'$.
   Test $H_0 : p_1 = p_2$ contre $H_1 : p_1 > p_2$ avec le risque 0,47%.
   Rejet de H_0 si $f_1 - f_2 > l'$.

   En appliquant cette nouvelle formule aux données de la troisième version de l'exemple 2, la limite l' est égale à $0{,}048\,114$. La conclusion en aurait été la même.

5. Alors que Gavarret avait bien posé le problème dans toute sa complexité c'est-à-dire trouver f telle que $\mathbb{P}(X = 1) = f(A, B, ..., M)$ où $A, B, ...$ sont des facteurs non contrôlés et M le facteur "méthode thérapeutique", il s'est bien vite éloigné de ce modèle en se plaçant au mieux dans le cas où l'échantillon sur lequel est expérimentée la méthode M_1 est constitué d'individus qui possèdent pour les facteurs $A, B, ...$, les mêmes modalités que ceux de l'échantillon sur lequel est expérimentée la méthode M_2, et au pire, dans le cas où l'échantillon sur lequel est expérimentée la méthode M_1 est constitué d'individus possédant dans leur ensemble la même combinaison de modalités que ceux des individus constituant l'échantillon sur lequel est expérimentée la méthode M_2. En langage moderne, nous dirions que le plan d'expérience doit être composé de données équilibrées.

6 Quelques considérations sur les constitutions médicales.

Dans ce chapitre IV, [G, p. 162-244], Gavarret utilise les méthodes précédentes sur des exemples réels issus de données épidémiologiques.

6.1 But et importance de l'étude des constitutions médicales.

L'article premier [G, p. 164-183] commence par expliciter cette notion.

« L'étude des variations qui peuvent survenir dans les maladies, et la recherche des causes de ces variations, constituent un point de doctrine encore fort obscur, connu sous le nom de *constitutions médicales*. » [G, p. 164].

Après avoir constaté qu'« il y a eu un instant où les médecins ont espéré rencontrer l'explication de tous les phénomènes observés dans les recherches entreprises sur les cadavres, et trouver au bout de leur scalpel la solution de toutes les obscurités de la science. » [G, p. 166], il exhorte les médecins à imiter l'exemple fourni par les astronomes et les physiciens qui ont dégagé les concepts de force gravitationnelle et de force magnétique pour expliquer des phénomènes. Il les encourage dans un long passage de l'article à définir ce que serait la "force vitale" et à en définir sa loi, en s'inspirant du travail d'autres disciplines scientifiques.

À cette première piste de recherche, il en ajoute une autre. Il rappelle qu'est « sans cesse soulevée la question de savoir si l'art de guérir peut être appuyé sur des principes fixes et invariables, ou si, au contraire, ses préceptes doivent être modifiés même à des intervalles de temps très rapprochés les uns des autres. Pour apprécier à sa juste valeur l'immense influence que les circonstances extérieures au milieu desquelles vivent les populations exercent sur la nature, les études historiques doivent être mises au premier rang. » [G, p. 176-177]. S'ensuit un rappel de faits historiques qui ont modifié l'environnement.

En fait, cet article pose la question de l'influence de l'environnement sur la santé : « rien n'influe plus puissamment sur la santé et même sur la nature des maladies que les conditions sociales et physiques des contrées dans lesquelles on observe. » [G, p. 178], préoccupation que l'on retrouve dans les débats actuels de l'écologie. Ce constat, Gavarret le partage avec Villermé, mais il va plus loin puisqu'il semble penser que cette connaissance de l'environnement peut modifier la pratique médicale.

Activité 11.

Il prolonge cette réflexion en étudiant puis distinguant plusieurs types de maladies.

« L'influence du monde ambiant sur la production des maladies peut être étudiée sous deux points de vue bien distincts :

1. Par rapport aux maladies endémiques et épidémiques, cette étude est connue sous les noms de *constitutions endémiques* et de *constitutions épidémiques* ;
2. Par rapport aux maladies sporadiques : cette dernière étude comprend toutes les questions dont on s'occupe en médecine sous les noms de *constitutions saisonnières* et de *constitutions annuelles*. » [G, p. 183].

6.2 Constitutions endémiques et épidémiques.

Dans cet article II [G, p. 183-187], si Gavarret admet comme démontré les liaisons entre certaines pathologies et leurs causes possibles, par exemple « la liaison entre les fièvres intermittentes comme effet et le voisinage d'un marais comme cause » [G, p. 184], « le typhus des camps et des prisons » et « les temps de disette extrême, [...]. Il est dans l'ordre des névroses, des maladies qui ont aussi régné épidémiquement à certaines époques de l'histoire, et il a été permis de trouver dans les conditions politiques et religieuses de ces peuples les causes véritables de ces effrayants phénomènes. » [G, p. 185], il n'en est pas de même pour d'autres phénomènes comme l'apparition du choléra par exemple. C'est donc « à l'étude comparative et approfondie des phénomènes et des circonstances desquelles on les observe, qu'il faut demander la solution de ces hautes questions. » [G, p. 187].

6.3 Conditions saisonnières et annuelles.

Dans cet article III [G, p. 187-244], Gavarret constate que c'est « une opinion presque généralement admise » d'admettre une influence saisonnière sur l'apparition des maladies ainsi « l'érysipèle se montrerait de préférence pendant l'automne et le printemps ; la bronchite pendant l'hiver, le printemps et l'automne. » [G, p. 188].

Il regrette que les praticiens n'aient « jamais songé à étayer leurs doctrines sur de grandes collections d'observations. » [G, p. 190] et répète ce qu'il n'a cessé d'énoncer depuis le début de son ouvrage :

> « En effet, les résultats de deux longues séries d'expériences relatives au même événement, pouvaient différer entre eux dans certaines limites assignables, sans que, pour cela, l'ensemble des causes possibles présidant à la manifestation de cet événement, eût varié d'une statistique à l'autre. Ce n'est, avons-nous dit alors, qu'autant qu'elle dépasse certaines *limites* que la différence constatée dans les résultats de l'observation, annonce une perturbation dans cet ensemble de causes possibles. » [G, p. 193].

.

Ainsi, la prise de décision à la suite d'observations ne peut survenir que si la variable de test appartient à une "zone critique", annonçant la théorie de Neyman-Pearson.

Il va dans la suite appliquer ce principe à plusieurs exemples.

6.3.1 Constitutions saisonnières.

Il commence d'abord par présenter la méthode générale qu'il utilisera sur des exemples.

Soit un événement qui se produit de façon aléatoire, ou bien dans l'intervalle de temps I_1, ou bien dans l'intervalle de temps I_2, ces deux intervalles étant bien sûr disjoints. Il s'agit de décider si l'événement : " il a lieu dans l'intervalle I_1" et l'événement : " il a lieu dans l'intervalle I_2" ont la même chance de se produire.

On observe que l'événement se produit n_1 fois dans l'intervalle de temps I_1 et n_2 fois dans l'intervalle de temps I_2. Le "rapport de fréquence" de l'événement dans l'intervalle I_1 relativement à l'intervalle I_2 est pour Gavarret égal à $\frac{n_1}{n_1 + n_2}$.

Mais,

> « la loi des grands nombres indique les *limites* de variation entre lesquelles ce rapport de fréquence peut osciller au-dessus et au-dessous de la fraction $0{,}50$, sans que les chances des deux événements cessent d'être égales. Ces *limites* dépendent uniquement du nombre des épreuves tentées, et sont d'autant plus rapprochées que ce nombre est plus grand. [...] Il peut arriver trois cas importants à considérer :
>
> **1er Cas.** Le rapport de fréquence que l'on considère est compris entre les *limites* déterminés. On doit en conclure que ces deux événements ont la même chance de production l'un que l'autre, qu'il n'existe pas de cause permanente que favorise particulièrement la manifestation de l'un d'eux.
>
> **2ème Cas.** Le rapport de fréquence que l'on considère est plus petit que la *limite inférieure.* Cela indique que les deux événements n'ont pas la même probabilité ; que celui qui est arrivé le moins souvent a une plus faible chance que l'autre ; qu'il existe une *cause* permanente favorable à la manifestation de ce dernier.
>
> **3ème Cas.** Le rapport de fréquence que l'on considère est plus grand que la *limite supérieure*. On doit en conclure que l'événement auquel il se rapporte a une plus forte chance de production que l'autre, que sa manifestation est favorisée par une *cause* permanente. » [G, p. 196-199].

Une reproduction de la note E page 289 montre que c'est exactement cette région critique qui a été utilisée par Gavarret.

> « Si, d'une part, on appelle :
>
> p la chance propre de l'événement E
>
> q la chance propre de l'événement contraire F
>
> Si, d'autre part, une statistique, composée d'un très grand nombre d'observations relatives à ces événements, a donné les résultats suivants :
>
> m = le nombre de fois que E est arrivé.
>
> n= le nombre de fois que F est arrivé
>
> μ = le nombre total d'observations recueillies.

Il existera entre ces quantités les relations suivantes :

il y a la probabilité P $= 0{,}9953$ c'est-à-dire 212 à parier contre 1 que le rapport $\frac{m}{\mu}$ est compris entre $p - 2\sqrt{\frac{2pq}{\mu}}$ et $p + 2\sqrt{\frac{2pq}{\mu}}$ » [G, p. 289-290].

En prenant $p = q = 0{,}50$, on obtient que le rapport $\frac{m}{\mu}$ est compris entre $0{,}50 - \sqrt{\frac{2}{\mu}}$ et $0{,}50 + \sqrt{\frac{2}{\mu}}$.

En termes modernes, on pose p probabilité que l'événement se produise dans l'intervalle I_1 et donc $q = 1 - p$ la probabilité qu'il se produise dans l'intervalle I_2. Soit p_0 un probabilité de référence. Il s'agit de tester :

$H_0 : p = p_0$ contre $H_1 : p \neq p_0$ avec le risque α de se tromper en décidant H_1.

Suite à n observations, f_n est la fréquence d'apparition de l'événement dans l'intervalle I_1.

On rejette H_0 si et seulement si $f_n \notin \left[p_0 - z_\alpha \frac{p_0(1-p_0)}{n},\ \ p_0 + z_\alpha \frac{p_0(1-p_0)}{n}\right]$.

En effet, comme plus haut, $2\sqrt{2} \simeq z_\alpha$ avec $\alpha = 0{,}0047 = 1 - 0{,}9953$ et on retrouve la formule de Poisson.

Il assortit son test de la règle de décision suivante : si la statistique de test $\frac{m}{\mu}$ est inférieure à $0{,}50 - \sqrt{\frac{2}{\mu}}$ (resp. supérieure à $0{,}50 + \sqrt{\frac{2}{\mu}}$) alors $p < 0{,}50$ (resp. $p > 0{,}50$).

Mais pour pouvoir utiliser cette méthode, Gavarret demande que les deux intervalles aient même étendue. Si par exemple l'étendue de I_2 est plus petite que l'étendue de I_1, il remplace par exemple n_1 par $n_1^{'} = n_1 \times \frac{I_2}{I_1}$. Il applique alors sa méthode à plusieurs exemples.

1er Exemple :

« Sur 13 900 individus atteints de dysenterie au Bengale, de 1820 à 1825, le docteur Annesley a trouvé qu'il y en avait eu :

2 400	pendant la saison froide
4 500	pendant la saison chaude et sèche
7 000	pendant la saison chaude et humide
13 900	pendant les trois saisons réunies » [G, p. 201].

Pour traiter cet exemple, il supposera que la condition de l'égale étendue des trois saisons est vérifiée et il effectue des groupements pour obtenir des comparaisons deux à deux.

Comparaison de la saison chaude et humide avec la saison chaude et sèche :

4 500	pendant la saison chaude et sèche.
7 000	pendant la saison chaude et humide.
11 500	pendant les deux saisons réunies.

Il calcule « les limites entre lesquelles le rapport de fréquence aurait pu osciller au dessus et au dessous de 0,50, sans que les causes eussent changé d'une saison à l'autre, on les trouve égales à $\{0{,}48\,681 ; 0{,}51\,319\}$ [...]. Or, le rapport $0{,}60\,870$ fourni par la statistique, surpasse notablement la *limite supérieure* ; le calcul s'accorde donc avec la simple inspection des résultats obtenus, pour nous indiquer une plus grande intensité dans les causes de dysenterie au Bengale pendant la saison chaude et humide que pendant la saison chaude et sèche ; voilà déjà une première vérification de l'exactitude des principes de la loi des grands nombres. » [G, p. 203-204].

Comparaison de la saison chaude et humide avec la saison froide :

7 000	pendant la saison chaude et humide.
2 400	pendant la saison froide
9 400	pendant les deux saisons réunies.

Pour Gavarret « il est évident, pour tout le monde, que les causes de dysenterie, quelles qu'elles soient, ont dû agir, avec plus d'intensité pendant la saison chaude et humide que pendant la saison froide. » [G, 204-205].

Les limites sont égales à $\{0{,}48\,542\,; 0{,}51\,458\}$. Or, le rapport de fréquence $0{,}74\,468$ fourni par la statistique, « dépasse très notablement la *limite supérieure*. Dans ce cas-ci, comme dans le précédent, les principes de la loi des grands nombres et la simple inspection des résultats obtenus s'accordent parfaitement. » [G, p. 205].

Activité 12.

Comparaison de la saison chaude et sèche avec la saison froide :

4 500	pendant la saison chaude et sèche.
2 400	pendant la saison froide.
6 900	pendant les deux saisons réunies.

Le rapport de fréquence est égal à $0{,}34\,783$. Pour Gavarret « Ce rapport est trop notablement inférieur à la moitié du nombre total des cas observés, pour qu'il soit possible de méconnaître une moindre intensité dans les causes de dysenterie pendant la saison froide que pendant la saison chaude et sèche. » [G, p. 206].

Les limites sont égales à $\{0{,}48\,298\,; 0{,}51\,702\}$. Donc le rapport de fréquence $0{,}34\,783$ fourni par la statistique « est très notablement moindre que la *limite inférieure*. C'est donc encore un cas où les résultats de calcul s'accordent parfaitement avec les indications incontestables de simple bon sens. » [G, p. 206-207].

Les conclusions de Gavarret concernant les deux premiers regroupements vont dans le même sens : les résultats du test statistique confirment ce que le "bon sens" médical. La conclusion qu'il tire du troisième regroupement est plus ambiguë car ici le "bon sens" auquel on compare la décision issue du test est fondé sur le fait que le rapport des fréquences soit « très notablement inférieur à la moitié du nombre des cas observés », c'est-à-dire sur la même statistique que celle qui a servi au test.

2ème Exemple :

Travail publié en 1834 sur la marche et les effets du choléra dans le département de la Seine, relativement aux entrées des malades dans les hôpitaux de Paris.

Jours de la semaine	Nombre de jours	Nombre de malades
Dimanches	27	1833
Lundis	27	2075
Mardis	27	1947
Mercredis	27	1978
Jeudis	27	2004
Vendredis	27	1971
Samedis	27	1969
	189	13 777

« De ce tableau, la commission a cru devoir tirer la conclusion suivante (Page 143 du rapport) :

« L'influence des excès des dimanches et des premiers jours de la semaine sur la partie de la classe ouvrière admise aux hôpitaux, se trouve indiquée par l'augmentation des entrées les lundis, mercredis et jeudis, la diminution des mardis paraissant être une conséquence de la forte augmentation des lundis. »

La commission admet donc comme démontré que l'intempérance a dû favoriser la production du choléra. » [G, p. 209].

Gavarret établit d'abord « qu'il n'y a que deux jours de la semaine pendant lesquels une partie très considérable de la classe ouvrière se livre réellement à des excès, comparativement à son genre de vie ordinaire. Chacun sait que ces deux jours sont le dimanche et le lundi. » [G, p. 210].

Il fait ensuite le raisonnement suivant : ou bien l'effet est immédiat et il aurait dû se faire sentir aussi bien le dimanche que le lundi, ou bien l'effet est différé et le pic devrait se situer symétriquement par rapport à la fin de la semaine c'est-à-dire le mercredi et le jeudi. Or il n'en est rien et il opère en plusieurs temps pour rechercher si l'on peut mettre sur le seul compte des fluctuations d'échantillonnage les différences observées.

Premier temps :

Il a rangé ensemble les jours « qui ont immédiatement suivi les débauches commises » [G, p. 212] et il compare les 6 000 entrées pour le groupe **Lundis, Mardis, Mercredis** avec les 5 773 pour le second groupe **Vendredis, Samedis, Dimanches**. En appliquant la même méthode que dans l'exemple précédent, il constate que la valeur de la statistique de test à savoir $0{,}50\,964$ est à l'intérieur de l'intervalle $[0{,}48\,697\ ;0{,}51\,303]$ et par conséquent « nous devons conclure que les documents recueillis n'indiquent nullement l'intervention de la cause perturbatrice admise par la commission. » [G, p. 213].

Deuxième temps :

Il forme un groupe composé des jours de débauches, les **dimanches** et les **lundis** et puis des **mercredis** et des **jeudis** qui sont les plus chargés parmi ceux qui restent. Le second groupe est constitué des **mardis, vendredis, samedis**.

Mais ces deux périodes n'étant pas de même étendue, il multiplie par 3/4 le nombre d'entrées 7 890 relatif au premier groupe, pour obtenir un résultat réduit à 81 jours d'observation, et par suite comparable à celui du second groupe. La valeur de la statistique de test à savoir $0{,}50\,131$ est à l'intérieur de l'intervalle $[0{,}48\,698\ ;0{,}51\,302]$ et par conséquent « rien ne saurait autoriser à considérer la débauche comme une cause déterminante du choléra. » [G, p. 216].

Au lieu de cet artifice consistant à multiplier les cas de choléra du premier groupe par 3/4, Gavarret avait pourtant en main la formule énoncée plus haut lui permettant de tester $H_0 : p = 4/7$ contre $H_1 : p \neq 4/7$.

Les bornes de l'intervalle auraient alors été : $4/7 \pm 2\sqrt{\dfrac{2 \times \frac{4}{7} \times \frac{3}{7}}{13777}}$ d'où l'intervalle $[0{,}55\,951\ ;0{,}58\,335]$. Là encore, la valeur de la statistique de test à savoir $0{,}57\,269$ est encore à l'intérieur de l'intervalle et conduit donc à la même conclusion.

Activité 13.

Troisième temps :

Il reprend les groupes formés par la commission c'est-à-dire un groupe **lundis, mardis, mercredis et jeudis** et un groupe **vendredis, samedis, dimanches**. En appliquant le même artifice, il trouve que la valeur de la statistique de test se trouve encore dans l'intervalle et il en conclut que la commission « s'est laissé induire en erreur par l'existence d'une différence qui réellement ne prouve rien par elle-même, puisqu'elle ne s'élève pas au dessus de la limite des erreurs possibles en pareil cas. » [G, p. 218].

En n'utilisant pas l'artifice, la valeur de la statistique de test se trouve encore dans l'intervalle, d'où la même conclusion.

Activité 14.

Quatrième temps :

La commission a pensé que le maximum d'entrées (hors dimanches et lundis) devait être attribué aux "excès" commis les dimanches et les lundis. Gavarret a donc comparé les **jeudis** aux **samedis** qui eux correspondent au minimum d'entrées (hors dimanches et lundis). Là encore la valeur de la statistique de test à savoir $0{,}50\,440$ se situe à l'intérieur de l'intervalle $[0{,}47\,756\ ; 0{,}52\,244]$ et donc « l'accroissement correspondant aux jeudis est un simple effet de l'irrégularité avec laquelle se déroulent les événements à chance variable [....] et par suite ne saurait en aucune façon justifier les conclusions de la commission. » [G, p. 219].

Cinquième temps :

Il compare les $2\,075$ entrées des **lundis** avec les $1\,833$ des **dimanches** et là, la valeur de la statistique de test $0{,}53\,096$ se trouve en dehors de l'intervalle $[0{,}47\,738\ ; 0{,}52\,262]$.

Gavarret remarque que cette fréquence « ne surpasse que de très peu la limite supérieure » et par conséquent « la cause permanente [...] ne doit voir qu'une faible intensité » et constate que « cette cause n'a réellement existé que pour les lundis comparativement aux dimanches. » [G, p. 224].

Une enquête plus approfondie lui a permis de constater que

> « les ouvriers de Paris ont l'habitude d'entrer le moins possible à l'hôpital le dimanche et de renvoyer au lendemain leur présentation au bureau central. Si l'on considère, en outre que le dimanche, le bureau central n'étant ouvert que deux heures au lieu de cinq, et les consultations des hôpitaux étant fermées, il y a réellement moins de facilité à y entrer ce jour-là ; on verra qu'il n'est pas besoin de recourir à l'influence des débauches pour se rendre compte de l'augmentation correspondante des lundis. » [G, p. 221-222].

Pour éviter des calculs aux médecins, Gavarret construit une table pour « vérifier si le rapport de fréquence relatif à l'un des deux événements observés s'écarte assez notablement de la fraction $0{,}50$, pour qu'il y ait lieu d'admettre l'existence d'une cause permanente favorable à la production de de celui des deux phénomènes qui s'est manifesté le plus souvent dans une longue suite d'épreuves. » [G, p. 225]. Cette table donne pour les échantillons de taille allant de 300 à 4000 par pas de 50, les deux "limites de l'erreur possible".

Il applique cette table à deux exemples fictifs concernant des cas de bronchites :

1er exemple :

1 400 cas du 21 mars au 21 juin.
960 cas du 21 juin au 21 septembre.
2 000 cas du 21 mars au 21 septembre.

La table donne comme "limites d'oscillations" $[0{,}468\,377\ ;0{,}531\,623]$. La fréquence de production de la maladie au printemps, égale à $0{,}520$ « est comprise entre ces deux limites, on en conclut que la différence constatée est négligeable. » [G, p. 228].

2ème exemple :

1 650 cas du 21 mars au 21 juin.
1 350 cas du 21 juin au 21 septembre.
3 000 cas du 21 mars au 21 septembre.

La table donne comme "limites d'oscillations" $[0{,}474\,180\ ;0{,}525\,820]$. Le rapport de fréquence de la maladie, égal à $0{,}550$, « étant plus grand que pendant le printemps ces causes sont plus intenses que pendant l'été. » lG, p. 229].

Quelques remarques :

– l'exemple des entrées dans les hôpitaux illustre assez bien la qualité des conclusions tirées de l'exécution des tests quand l'hypothèse nulle n'est pas rejetée, mais dans le cas où elle est rejetée de peu, le doute s'instaure et Gavarret doit compléter son enquête par une observation sur le terrain des conditions de l'expérience ;
– un traitement rigoureux et possible du problème est bien sûr celui utilisant les processus ponctuels mais il est encore trop tôt pour utiliser cet outil. Cependant, au travers des tâtonnements observés (quels groupes constituer ? doivent-ils être de même taille ?), la recherche est bien celle de l'équirépartition des événements sur les jours de la semaine. Actuellement, un test du chi-deux donnerait une réponse à ce problème avec une valeur de la statistique de test égale à $16{,}01$ pour 6 degrés de libertés et une p-value égale à $0{,}01367$. Avec le risque pris par Gavarret à savoir $\alpha = 0{,}0047$, l'hypothèse d'une équirépartition sur les 7 jours de la semaine ne serait pas rejetée, alors qu'avec le même risque, on rejette celle de l'équirépartition entre les dimanches et les lundis. Mais on sait qu'il n'est pas équivalent de tester tous les groupements de jours deux à deux et de tester globalement l'équirépartition ; il n'est pas non plus équivalent de tester globalement l'équirépartition et d'effectuer des tests sur des groupes choisis a priori par le statisticien, même si ceux-ci semblent refléter des situations extrêmes ;

Activité 15.

– encore une fois, l'hypothèse de l'indépendance des événements pour pouvoir rigoureusement appliquer la loi des grands nombres n'est même pas évoquée alors que dans le cas de maladies, il est difficile de ne pas envisager l'effet contagion.

6.3.2 Constitutions annuelles.

Pour Gavarret, les statistiques devraient être utiles pour « savoir si, à deux époques différentes, le lieu de l'observation et le traitement restant les mêmes, la mortalité de la pneumonie présente des variations assez considérables pour admettre que les causes de mort qui pèsent sur la maladie aient changé d'une époque à l'autre. » [G, p. 232].

> « Malheureusement les recueils d'observations publiés jusqu'ici sont tout à fait insuffisants pour fournir la moindre notion sur ce sujet important. Il nous a été impossible de trouver dans les ouvrages de médecine des relevés assez étendus pour nous servir d'exemple d'application. Cependant, comme la même manière de procéder doit être suivie pour apprécier l'influence des

âges, des *sexes* ou des *localités*, sur la mortalité d'une maladie, nous trouverons dans le rapport de la commission du choléra des matériaux authentiques et très précieux. » [G, p. 235].

Ce sera l'objet de la fin de l'ouvrage.

Il s'agit d'appliquer le test $H_0 : p_1 = p_2$ contre $H_1 : p_1 \neq p_2$ construit en comparant la différence observée d des fréquences avec la limite l calculée, sur les données de mortalité du choléra à Paris en 1832.

1. **Influence des sexes sur la mortalité du choléra à Paris.**

Mortalité pour les hommes	Mortalité pour les femmes
7 975 morts du choléra	8 597 morts du choléra
360 965 personnes échappées à la mort	381 598 personnes échappées à la mort
368 940 habitants mâles	390 195 habitants du sexe féminin

Ici $d = 0{,}000\,417$ inférieure à $l = 0{,}000\,949$. Gavarret en conclut : « C'est donc à tort que la commission s'est appuyée sur cette variation pour poser qu'en règle générale les femmes étaient plus que les hommes exposées à mourir du choléra. » [G, p. 237].

Sur un autre exemple, celui des 696 érysipèles observés dans les hôpitaux de Paris, où 364 existaient chez les femmes et 332 existaient chez les hommes, Gavarret ne peut conclure car il pense qu'il manque des données, « le nombre d'hommes et le nombre de femmes qui ont servi à établir cette proportion. » [G, p. 238].

Activité 16.

2. **Influence de la densité de la population sur la mortalité de choléra à Paris.**

La densité moyenne étant de 43 mètres carrés par habitant, le rapport de la commission contient un tableau où se trouvent d'un côté les quartiers qui sont au dessus de cette moyenne et de l'autre les quartiers qui sont au dessous.

Quartiers au dessus de 43 mètres carrés	Quartiers au dessous de 43 mètres carrés
6 422 morts du choléra	10 150 morts du choléra
283 348 personnes échappées à la mort	459 215 personnes échappées à la mort
289 770 habitants	469 365 habitants

La fréquence de mort dans les quartiers "entassés" est égale à $0{,}021\,625$ alors que la fréquence de mort dans les quartiers "non entassés" est égale à $0{,}022\,162$.

> « Il semblerait donc que l'épidémie a été moins meurtrière dans les quartiers où les habitants sont les plus entassés, ce qui contrarie toutes les idées reçues. Mais avant de porter une jugement aussi grave, calculons la *limite* des différences, entre les deux mortalités moyennes, compatibles avec l'invariabilité des mêmes causes. » [G, p. 240].

Ici $d = 0{,}000\,537$ inférieure à $l = 0{,}000\,979$. Gavarret en conclut que la statistique « n'a pas mis en évidence l'influence du tassement des individus. » [G, p. 240].

Il fait remarquer que le classement met dans le même groupe le quartier des Tuileries « si opulent » et celui de la Porte Saint-Martin, et, à côté des Champs-Élysées « nous rencontrons le faubourg Saint-Marcel, etc., etc. » [G, p. 240].

Cet exemple illustre avec éclat combien une conclusion établie uniquement à partir de l'observation des fréquences peut être sujette à caution.

3. **Influence de l'orientation des arrondissements de Paris sur la mortalité du choléra.**
   Il subdivise les douze arrondissements en un premier groupe de six arrondissements de la rive droite et un second des six arrondissements de la rive gauche.

Parmi les six premiers arrondissements	Parmi les six derniers arrondissements
5 196 morts du choléra	11 376 morts du choléra
378 194 personnes échappées à la mort	364 369 personnes échappées à la mort
383 390 habitants	375 745 habitants

   Ici $d = 0{,}016\,723$ supérieure à $l = 0{,}000\,951$. Gavarret en conclut : « Ce résultat doit engager à chercher très sérieusement quelles sont les influences d'exposition propres aux six derniers arrondissement de Paris qui ont pu entraîner une si grande différence dans l'intensité de l'épidémie de choléra. » [G, p. 242].

Le travail se termine par l'énoncé de dix propositions [G, p. 245-247] qui ne sont que la reprise de celles déjà émises et par des notes techniques qui illustrent les calculs effectués dans l'ouvrage.

7 Les transmissions ou les non-transmissions.

L'apport de Gavarret à l'utilisation des statistiques en médecine semble considérable puisque pour la première fois, comme le fait remarquer Huth, il a mis en œuvre la suggestion de Pierre Simon Laplace qui en 1825, dans son *Essai Philosophique sur les probabilités*, émettait l'idée d'appliquer le calcul des probabilités pour découvrir le meilleur traitement à une maladie.

7.1 Quelles transmissions ?

Nul doute que Gavarret a bénéficié d'une transmission "institutionnelle" dans la première moitié du XIXᵉ siècle, puisque Poisson a été son professeur. Il a remarquablement adapté à l'épidémiologie naissante les techniques statistiques utilisées par Poisson pour les jugements de jurys et surtout en a compris le sens. Une étude plus complète serait nécessaire pour savoir si dans la première moitié du XIXᵉ siècle, d'autres disciplines que la médecine ont pu approcher les travaux de Poisson et en extraire des enseignements quant à l'application des statistiques inférentielles à leur discipline, réalisant ainsi une transmission "transversale".

La biographie de Gavarret atteste que ses sujets d'intérêt étaient très variés, comme beaucoup de savants de cette époque qui ne se cantonnaient pas à leur domaine de prédilection. Logiquement, on pourrait s'attendre à une réponse positive à cette question.

7.2 Quelles non-transmissions ?

7.2.1 Non-transmission à l'intérieur de la médecine

Même si David E. Lilienfeld & Abraham M. Lilienfeld mentionnent dans *The French Influence on the Development of Epidemiology* que "Dedicating his *Essay on the Philosophy of Medical Science* to Louis, Bartlett gave examples of both the calculation an use of confidence limits for mortality rates, based on a publication by Jules Gavarett [*sic*] in 1840", il faut bien constater que Gavarret a disparu très longtemps de la mémoire des épidémiologistes.

Plusieurs explications sont possibles.

a) Au regard de ses objectifs, Gavarret a échoué sur au moins deux problèmes : celui du classement des thérapeutiques et celui de la prévision de l'apparition d'une maladie. Mais il a introduit avec justesse la notion d'intervalle de confiance et celle de test pour toute décision quant à un traitement thérapeutique, notions qui sont aujourd'hui universellement appliquées.

b) L' imperméabilité du corps médical à des apports extérieurs pendant une certaine époque. Une sociologie du corps médical pourrait peut être fournir une explication.

À l'intérieur du milieu médical, Gavarret a été aussitôt contesté dans un rapport de la Faculté de Médecine, publié en 1840, principalement sur trois points :

– les constitutions saisonnières : ce rapport explique qu'il n'y avait nul besoin de statistiques pour affirmer que certaines maladies soient plus fréquentes en hiver et au printemps qu'en été
– l'utilisation des probabilités : ce rapport explique que ce qui est applicable au tirage de boules ne peut s'appliquer en médecine car il y a des causes qui produisent toujours les mêmes effets
– l'incertitude que calcule Gavarret sur les résultats de Louis à partir de ses formules : ce rapport argumente que Gavarret s'appuie sur des nombres qui ne reflètent pas toute la réalité et qu'il oublie beaucoup d'autres facteurs.

Une étude plus attentive de l'ouvrage de Gavarret pourrait faire découvrir qu'il a été conscient, en particulier sur le troisième point, que la présence de plusieurs facteurs pour prédire un phénomène était un obstacle que tentera de franchir Ronald A. Fisher avec la technique des plans d'expériences mais cela beaucoup plus tard (1935).

c) Le rôle de l'idéologie scientifique dominante. Valleron fait remarquer que « La fin du XIX^{e} siècle et le début du XX^{e} siècle furent en revanche relativement silencieux parce que la démarche épidémiologique est fondée essentiellement sur l'analyse de la variabilité, alors que l'idéologie scientifique dominante de cette époque était le déterminisme. » [V, p. 7].

d) Le rôle de la collecte des données : les conditions mises par Gavarret pour pouvoir utiliser les techniques de statistiques inférentielles, à savoir entre autres - plus d'une centaine d'observations, observations faites dans les mêmes conditions - ont dû faire reculer ceux qui auraient pu tenter de se servir de ces techniques.

e) La relative contradiction groupe-individu. Celle-ci était un argument majeur des opposants à la méthode numérique arguant que chaque individu était particulier et ne pouvait pas être assimilé à un élément d'une population. Gavarret avait déjà prévu cette objection en précisant bien (voir notre 2^{e} remarque dans le paragraphe 4.3. et [G, p. 82-83]) que l'étude ne portait que sur les chances moyennes et non sur les chances propres à chaque individu.

Mais comme le fait remarquer Valleron, « l'épidémiologie fut quasiment absente de France jusqu'à la fin des années 1950. » [V, p.10].

7.2.2 Non-transmission à l'intérieur des mathématiques.

Ce constat est plus étonnant. Il a fallu attendre 1934 pour que Jerzy Neymann définisse le concept d'intervalle de confiance (sans mentionner ni le rôle de Gavarret, ni celui de Poisson) et pour voir apparaître les premiers grands travaux des fondateurs de la statistique inférentielle.

Doit-on y voir un désintérêt des probabilistes "purs" pour ce qui pouvait leur sembler n'être que des applications de formules à des données numériques ? Et pourtant, c'est William S. Gosset (1876-1937) et Ronald A. Fisher (1890-1962), deux chercheurs d'abord préoccupés d'applications, qui donnèrent un élan décisif aux statistiques inférentielles. Des recherches ultérieures permettront sans doute de répondre à la question.

8 Bibliographie

[G] Jules GAVARRET, *Principes généraux de Statistique médicale ou développement des règles qui doivent présider à son emploi*, Paris : Béchet jeune et Labé, 1840.

[H] Edward J. HUTH, « Jules Gavarret's *Principes Généraux de Statistique Médicale* : a pioneering Text on the statistical Analysis of the Results of Treatments », *in : The James Lind Library Bulletin : Commentaries on the History of Treatement Evaluation*, Oxford : The James Lind Library, 2006.

[La] Jean-Marc LANGE, « L'émergence des bio-statistiques au XIX^{e} siècle », *in : Epidémiologie pour une éducation raisonnée à l'incertitude*, Paris : Vuibert-ADAPT-SNES, 2006, p. 39-49.

[L] Pierre-Simon (de) LAPLACE, *Théorie analytique des probabilités*, 3^{e} éd. Paris, Vve Courcier, 1820.

[L] Pierre-Simon (de) LAPLACE, *Essai Philosophique sur les probabilités*, 5^{e} éd., Paris : Bachelier, 1825.

[Li] David E. LILIENFELD & Abraham M. LILIENFELD, « The French Influence on the Development of Epidemiology », *in* Abraham M. LILIENFELD, *Times, Places and Persons : Aspects of the History of Epidemiology*, Baltimore : Johns Hopkins University Press, 1980.

[Lo] Pierre-Charles-Alexandre LOUIS, *Recherche sur les effets de la saignée dans quelques maladies inflammatoires*, Paris : Baillière, 1835. Cet ouvrage est une version augmentée d'un mémoire publié dans les *Archives générales de médecine* de 1828, n^{0}18, p. 321-336.

[P] Siméon-Denis POISSON, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités*, Paris : Bachelier, 1837.

[S] Stephen M. STIGLER, *The History of Statistics : The Measurement of Uncertainty before 1900*, Cambridge (Mass.) ; London : The Belknap Press of Harvard University Press, 1986.

[V] Alain-Jacques VALLERON (éd.), *L'épidémiologie humaine : condition de son développement en France, et rôle des mathématiques. Rapport sur la science et la technologie n^{0}23*, Les Ulis : EDP Sciences, 2006.

Table des matières

Annexe I

Théorie des tests :

Rappel très simplifié sur un exemple .

Test de l'efficacité d'un remède sur des malades atteint d'un rhume.

p_0 : probabilité de guérir dans les huit jours avec un placebo
p_1 : probabilité de guérir dans les huit jours avec le remède.

Deux hypothèses possibles :
- H_0 hypothèse nulle : $p_0 = p_1$ (pas d'effet du remède).
- H_1 hypothèse alternative : $p_0 \neq p_1$ (effet du remède)

Suite à une expérimentation où n_0 seront traités avec le placebo et n_1 avec le remède, observer le nombre de malades guéris : k_0 avec le placebo et k_1 avec le remède. A partir de cette observation, deux décisions seront possibles :
- D_0 : "H_0 est vraie" ou encore "pas d'effet du remède",
- D_1 : "H_1 est vraie" ou encore "effet du remède".

Démarche à utiliser :
- Construire une fonction qui dépend des observations et pour laquelle, quand l'hypothèse nulle est vraie, il est possible de calculer la loi de probabilité et ceci sans connaître les paramètres inconnus. Cette fonction appelée "statistique de test" sera notée φ ;
- Fixer $\alpha =$ risque de se tromper en décidant D_1 (c.a.d. "alors que H_0 est vraie"). C'est le risque dit "de première espèce" ;
- Déterminer une "zone critique" notée C_α telle que, lorsque l'hypothèse nulle est vraie, alors : $\mathbb{P}(\varphi \in C_\alpha) \leq \alpha$;
- Etablir la règle de décision suivante : rejet de H_0 (décision D_1) si et seulement si $\varphi \in C_\alpha$;
- Effectuer l'expérience et, suivant la valeur prise par φ, énoncer la décision. La construction du test garantit que la probabilité de se tromper en rejetant l'hypothèse nulle est inférieure à α.
- Evaluer le risque dit "de deuxième espèce". C'est le risque de se tromper en ne rejetant pas l'hypothèse nulle (décision D_0), c'est-à-dire $\mathbb{P}(\varphi \notin C_\alpha)$. Il est noté β.

Avant d'utiliser l'exemple, rappelons les notations suivantes :

Z désigne une variable aléatoire suivant une loi normale centrée réduite.
On note $Z \sim \mathcal{N}(0,1)$.

F_Z désigne la fonction de répartition de la loi normale centrée réduite (souvent notée Π ou Φ), c'est-à-dire : $F_Z(z) = \mathbb{P}(Z \leq z)$ avec $Z \sim \mathcal{N}(0,1)$.

z_α est le nombre réel tel que : $\mathbb{P}(-z_\alpha \leq Z \leq z_\alpha) = 1 - \alpha$ ou encore $\mathbb{P}(|Z| \geq z_\alpha) = \alpha$.
Remarque : $F_Z(z_{2\alpha}) = 1 - \alpha$.

Utilisation de la démarche :
- On pose :
  $f_0 = \frac{k_0}{n_0}$, $f_1 = \frac{k_1}{n_1}$ et $d = f_0 - f_1$;
  $f^* = \frac{k_0+k_1}{n_0+n_1}$, $s_d = \sqrt{\frac{1}{n_0} + \frac{1}{n_1}}\sqrt{f^*(1-f^*)}$ et $\varphi = \dfrac{d}{s_d}$
  On peut montrer que si $p_0 = p_1$, alors φ suit approximativement une loi normale centrée réduite $Z \sim \mathcal{N}(0,1)$. Les conditions $n_0 \geq 30$, $n_1 \geq 30$, $n_0 p_0 \geq 5$, $n_0(1-p_0) \geq 5$, $n_1 p_1 \geq 5$ et $n_1(1-p_1) \geq 5$ permettent une bonne approximation.

– On prend $\alpha = 0{,}05$ c'est-à-dire 5% de chance de se tromper en disant que le remède fait de l'effet.

– On sait que $\mathbb{P}(|Z| > z_\alpha) = \alpha$ donc en prenant comme "zone critique", $C_\alpha =]-\infty, -z_\alpha[\cup]z_\alpha, +\infty[$, on a bien $\mathbb{P}(\varphi \in C_\alpha) \leq \alpha$. D'après les tables, $z_{0{,}05} = 1{,}96$, donc dire que la statistique appartient à la "zone critique" signifie que $\dfrac{|d|}{s_d} > 1{,}96$ ou encore que $|d| > l$ avec $l = 1{,}96 s_d$. On rejette l'hypothèse nulle si et seulement si $|d| > l$.

– Exécution du test.
$n_0 = 200,\ \ k_0 = 140,\ \ f_0 = 0{,}70\ ,\ n_1 = 210,\ \ k_1 = 164,\ \ f_1 = 0{,}780\,952.$
$d = -0{,}08095,\ \ s_d = 0{,}04325869,\ \ l = 1{,}96 \times 0{,}04325869 = 0{,}08478$
$|d| = 0{,}08095 < 0{,}08478 = l$

Décision : D_0 c'est-à-dire non-rejet de H_0.
On ne conclut pas à un effet du remède.

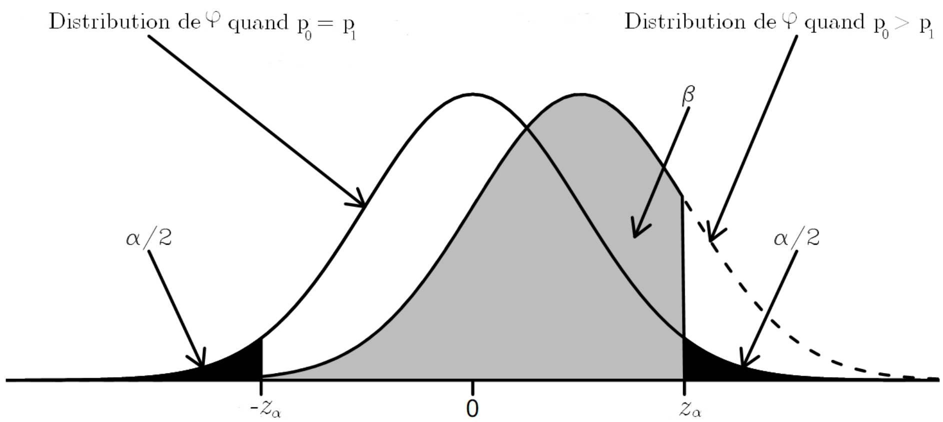
– Le risque de se tromper vaut $\beta = \mathbb{P}(\varphi \notin C_\alpha) = \mathbb{P}(-z_\alpha \leq \varphi \leq z_\alpha)$.
Ici $\beta = \mathbb{P}(-1{,}96 \leq \varphi \leq 1{,}96)$.

Dans le cas où on supposait $p_0 = p_1$ (hypothèse H_0), φ suivait loi normale centrée réduite, on avait alors $\mathbb{P}(-1{,}96 \leq \varphi \leq 1{,}96) = 1 - 0{,}05 = 0{,}95$.
Pour calculer le "risque de deuxième espèce", il faut se placer dans le cas où on se trompe en décidant $p_0 = p_1$, donc (puisqu'on se trompe) dans le cas où $p_0 \neq p_1$ (hypothèse H_1). Mais alors dans ce cas, φ suit une loi de probabilité qui **dépend de** n_0, n_1, p_0, p_1 et cette loi **n'est pas la loi normale centrée réduite**.

Pour un risque de première espèce donné α, le "risque de deuxième espèce" β dépend de n_0, n_1, p_0, p_1 et de α. Il n'est pas égal à α. Il peut être grand et même proche de $1 - \alpha$ (et ici proche de $0{,}95$!).

Pour cette raison, lorsque la décision D_0 est prise, il faut, quand c'est possible, assortir cette décision de l'évaluation du "risque de deuxième espèce", et sinon, il est alors préférable de dire qu'on ne rejette pas l'hypothèse nulle. Dans le cas de ce test de comparaison entre deux probabilités, l'expression exacte de ce "risque de deuxième espèce" n'est pas donnée dans cette annexe car trop complexe mais il existe des tests (voir Activités 2) où elle peut être exprimée aisément.

Les deux risques



Retour à l'article p.1 Retour à l'article p.18

Annexe II

Intervalle de confiance.

Rappel très simplifié sur un exemple.

p proportion inconnue de patients ayant eu un effet secondaire suite à l'absorption d'un remède. Un échantillon de n patients sera observé. Sur ces n patients, k ont eu un effet secondaire.

Démarche à effectuer :

– Déterminer le niveau de confiance $1-\alpha$ avec $0<\alpha<1$.
– Trouver deux fonctions φ_1 et φ_2 de k (donc pour lesquelles le résultat sera aléatoire) telles que $\mathbb{P}(\varphi_1 \leq p \leq \varphi_2) \geq 1-\alpha$.
– A la suite de l'expérience, appliquer les deux fonctions à k. L'intervalle de confiance de niveau $1-\alpha$ pour p sera $[\varphi_1(k)\,,\,\varphi_2(k)]$.

On pose :

$$\varphi(k)=\frac{\frac{k}{n}-p}{\sqrt{\frac{\frac{k}{n}(1-\frac{k}{n})}{n}}},$$

On peut démontrer que, si les conditions suivantes sont réalisées : $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, alors φ suit approximativement une loi normale centrée réduite notée $\mathcal{N}(0,1)$.

Si Z suit une loi normale centrée réduite, les tables permettent de calculer z_α vérifiant :
$\mathbb{P}(-z_\alpha \leq Z \leq z_\alpha)=1-\alpha$.

Donc $\mathbb{P}(-z_\alpha \leq \varphi \leq z_\alpha)=1-\alpha$. En remplaçant φ par son expression, on trouve :

$\mathbb{P}(f-e \leq p \leq f+e)=1-\alpha$ avec $f=\dfrac{k}{n}$ et $e=z_\alpha\sqrt{\dfrac{\frac{k}{n}(1-\frac{k}{n})}{n}}=z_\alpha\sqrt{\dfrac{f(1-f)}{n}}$.

On aura ici $\varphi_1(k)=f-e$ et $\varphi_2(k)=f+e$.

L'intervalle de confiance de niveau $1-\alpha$ pour p sera $[f-e\ ,\ f+e]$.

Dans un échantillon de $n=1100$ patients ayant absorbé le remède, 583 ont eu un effet secondaire. On cherche un intervalle de confiance de niveau $0{,}95$ pour p.

Ici $z_{0{,}95}=1{,}96$

$f=0{,}53$ et $e=1{,}96\sqrt{\frac{f(1-f)}{n}}=0{,}0294$

Ici l'intervalle de confiance est : $[0{,}5006\ ;0{,}5594]$

Programme B.T.S. : $[f-\frac{1}{\sqrt{n}}\ ,\ \ f+\frac{1}{\sqrt{n}}]$ intervalle de confiance de niveau $0{,}95$ pour p.

Ici $[0{,}49985\ ;0{,}56015]$ intervalle de confiance de niveau $0{,}95$ pour p.

Retour à l'article

Annexe III

Une démonstration actuelle est la suivante :

Soit p la probabilité de réalisation de E et f_n la fréquence observée de l'événement E à la suite de n expériences indépendantes . On pose :

$Y_n = \dfrac{f_n - p}{\sqrt{\dfrac{p(1-p)}{n}}}$, $T_n = \sqrt{\dfrac{p(1-p)}{f_n(1-f_n)}}$, $U = 1$ variable aléatoire constante.

$Y_n T_n = \dfrac{f_n - p}{\sqrt{\dfrac{f_n(1-f_n)}{n}}}$

Z une variable aléatoire de loi normale $\mathcal{N}(0,1)$

D'après le théorème de Moivre-Laplace :

$Y_n \xrightarrow{\mathcal{L}} Z$ (convergence en loi vers la loi normale).

Mais :

$f_n \xrightarrow{p.s.} p$ (convergence presque sûre d'après la loi forte des grands nombres)

Donc $T_n \xrightarrow{p.s.} U$

Puisque U est une variable aléatoire constante :

$Y_n T_n \xrightarrow{\mathcal{L}} ZU$ donc $Z_n \xrightarrow{\mathcal{L}} Z$.

Ceci entraîne :

$$\mathbb{P}\left(-z_\alpha \leq \frac{f_n - p}{\sqrt{\dfrac{f_n(1-f_n)}{n}}} \leq z_\alpha\right) \longrightarrow \mathbb{P}(-z_\alpha \leq Z \leq z_\alpha) = 1 - \alpha$$

Ou encore :

$$\mathbb{P}\left(f_n - z_\alpha\sqrt{\frac{f_n(1-f_n)}{n}} \leq p \leq f_n + z_\alpha\sqrt{\frac{f_n(1-f_n)}{n}}\right) \simeq 1 - \alpha$$

Il est alors facile de voir que ce résultat est aussi celui de Gavarret.

Posons : $\alpha = \dfrac{2}{\sqrt{\pi}}\displaystyle\int_u^{\infty} e^{-t^2}dt$

Le changement de variable $t = \dfrac{x}{\sqrt{2}}$ donne :

$\dfrac{\alpha}{2} = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{u\sqrt{2}}^{+\infty} e^{-\frac{x^2}{2}}dx = 1 - F_Z(u\sqrt{2})$ où F_Z est la fonction de répartition de la loi normale centrée réduite.

Donc $u\sqrt{2} = F_Z^{-1}(1 - \dfrac{\alpha}{2}) = z_\alpha$

Gavarret pose : $s = u\sqrt{\frac{2.m.n}{\mu^3}}$, ce qui lui donne l'intervalle de confiance de niveau P pour B de la forme : $[\frac{m}{\mu} - s,\ \frac{m}{\mu} + s]$.

Puisque $\frac{m}{\mu} = f_n$, $\frac{n}{\mu} = 1 - f_n$ et on a bien $s = z_\alpha\sqrt{\dfrac{f_n(1-f_n)}{n}}$. En prenant $\alpha = 0{,}0047$ (donc $P = 0{,}9953$), $z_\alpha \simeq 2\sqrt{2}$, l'intervalle de confiance de niveau $1-\alpha$ pour p (ici égal à B) est bien le même que celui annoncé par Gavarret.

Retour à l'article

Annexe IV

Une démonstration actuelle possible utilise une généralisation du théorème central limite.

On pose $X_i^{(1)} \sim \mathcal{B}(p_1)$ pour $1 \leq i \leq n_1$ (resp. $X_k^{(2)} \sim \mathcal{B}(p_2)$ pour $1 \leq k \leq n_2$) la variable qui vaut 1 si E est arrivé dans la première série (resp. dans la deuxième série)

Soit $f_{n_1} = \dfrac{1}{n_1}\sum_{i=1}^{n_1} X_i^{(1)}$ (resp. $f_{n_2} = \dfrac{1}{n_2}\sum_{k=1}^{n_2} X_k^{(2)}$) la fréquence d'apparition de E dans la première série (resp. la deuxième série).

Soit $Y_j = \dfrac{1}{n_1}X_j^{(1)}$ si $1 \leq j \leq n_1$ et $Y_j = -\dfrac{1}{n_2}X_{j-n_1}^{(2)}$ si $n_1 + 1 \leq j \leq n_1 + n_2$.

On pose encore :

$$W_{(n_1,n_2)} = \frac{Y_1 + \cdots + Y_{n_1} + Y_{n_1+1} + \cdots + Y_{n_1+n_2} - \mathbb{E}(Y_1 + \cdots + Y_{n_1} + Y_{n_1+1} + \cdots + Y_{n_1+n_2})}{\sqrt{var(Y_1 + \cdots + Y_{n_1} + Y_{n_1+1} + \cdots + Y_{n_1+n_2})}}$$

Les variables Y_j n'ont pas toutes la même loi mais sous des conditions dites de Liapounov qui sont ici vérifiées, on a encore :

$$W_{(n_1,n_2)} \xrightarrow{\mathcal{L}} Z$$

où Z est une variable aléatoire de loi $\mathcal{N}(0,1)$.

Mais :

$$W_{(n_1,n_2)} = \frac{(f_{n_1} - f_{n_2}) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$$

Si on pose :

$$T_{(n_1,n_2)} = \frac{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}{\sqrt{\dfrac{f_{n_1}(1-f_{n_1})}{n_1} + \dfrac{f_{n_2}(1-f_{n_2})}{n_2}}}$$

D'après la loi forte des grands nombres :

$$T_{(n_1,n_2)} \xrightarrow{p.s.} U$$

où U est la variable aléatoire constante égale à 1.

Alors :

$$W_{(n_1,n_2)} T_{(n_1,n_2)} \xrightarrow{\mathcal{L}} Z.U = Z$$

et donc :

$$Z_{(n_1,n_2)} = \frac{(f_{n_1} - f_{n_2}) - (p_1 - p_2)}{\sqrt{\dfrac{f_{n_1}(1-f_{n_1})}{n_1} + \dfrac{f_{n_2}(1-f_{n_2})}{n_2}}} \xrightarrow{\mathcal{L}} Z$$

Sous l'hypothèse nulle, $p_1 = p_2$ et par conséquent :

$$\mathbb{P}\left(|f_{n_1} - f_{n_1}| \leq z_\alpha \sqrt{\frac{f_{n_1}(1-f_{n_1})}{n_1} + \frac{f_{n_2}(1-f_{n_2})}{n_2}}\right) \simeq 1 - \alpha$$

En reprenant $\alpha = 0{,}0047$, on retrouve la formule de Gavarret.

Retour à l'article

Activité 1.

20 personnes (identifiées par leur numéro $\alpha = 1, ..., 20$) notent leur tension supérieure $\xi_{1,\alpha}$ en l'absence de traitement avec un tensiomètre individuel. Les nombres ne sont connus que d'eux-mêmes. Les résultats sont les suivants (données fictives) :

Tableau 1

α	1	2	3	4	5	6	7	8	9	10
$\xi_{1,\alpha}$	12	13	16	11	14	13	15	16	15	12
α	11	12	13	14	15	16	17	18	19	20
$\xi_{1,\alpha}$	15	15	12	16	15	14	14	12	17	13

Un médicament administré à tous est censé faire baisser la moyenne du groupe de ces 20 personnes. Cette moyenne **avant** traitement, notée μ_1 n'est pas connue du médecin.

Les 20 personnes notent leur tension $\xi_{2,\alpha}$ **après** le traitement avec un tensiomètre individuel. Les nombres ne sont connus que d'eux-mêmes. La moyenne **après** traitement notée μ_2 n'est pas connue du médecin.

Les résultats sont les suivants (données fictives) :

Tableau 2

α	1	2	3	4	5	6	7	8	9	10
$\xi_{2,\alpha}$	14	14	18	12	12	12	15	15	13	12
α	11	12	13	14	15	16	17	18	19	20
$\xi_{2,\alpha}$	16	15	11	18	13	16	14	11	15	14

Le médecin se contente d'interroger $n = 5$ personnes au hasard **avant** le traitement.

Il note $x_{1,i}$ ($i = 1, ..., n$) la tension correspondant à la i-ème personne tirée et $\overline{x_1}$ la moyenne de la tension de ces $n = 5$ personnes.

Il interroge ensuite $n = 5$ personnes au hasard **après** le traitement. Il note $x_{2,i}$ ($i = 1, ..., n$) la tension correspondant à la i-ème personne tirée et $\overline{x_2}$ la moyenne de la tension de ces $n = 5$ personnes.

1.1. Tirer au hasard 5 nombres compris entre 1 et 20. Attribuer à chacun de ces nombres la tension correspondant au tableau 1 et calculer la moyenne observée $\overline{x_1}$. Tirer de nouveau au hasard 5 nombres compris entre 1 et 20. Attribuer à chacun de ces nombres la tension correspondant au tableau 2 et calculer la moyenne observée $\overline{x_2}$. Au vu de ces deux séries de 5 résultats, qu'auraient décidé les tenants de la "méthode numérique" ?

1.2. Calculer maintenant μ_1 et μ_2. Construire un tableau de 20 nombres $\epsilon_{1,\alpha}$ ($\alpha = 1, ..., 20$) qui sont les écarts des tensions **avant** traitement à la moyenne μ_1. Construire de même un tableau de 20 nombres $\epsilon_{2,\alpha}$ $\alpha = 1, ..., 20$) qui sont les écarts des tensions **après** traitement à la moyenne μ_2. Les "erreurs" se compensent-elles ? Pouvait-on prévoir le résultat ? Que penser alors de la décision des tenants de la "méthode numérique" ?

1.3. Pour les 5 personnes tirées au hasard **avant** traitement,

$$\text{calculer } e_{1,i} = x_{1,i} - \mu_1 \ (i = 1, ..., 5) \text{ et } \sum_{i=1}^{5} e_{1,i}.$$

Pour les 5 personnes tirées au hasard **après** traitement,

$$\text{calculer } e_{2,i} = x_{2,i} - \mu_2 \ (i = 1, ..., 5) \text{ et } \sum_{i=1}^{5} e_{2,i}.$$

Les “erreurs” se compensent-elles ? Si ce n’est pas le cas, pouvez-vous trouver cependant des tirages de 5 individus avant le traitement et des tirages de 5 individus après le traitement pour lesquels les “erreurs” se compensent.

Gavarret avait-il raison en émettant des doutes sur le fait que les “erreurs” puissent se compenser ?

Dans le cas où “les erreurs” se compensent, pourquoi peut-on dire que $\overline{x_1} > \overline{x_2}$ (réciproquement $\overline{x_1} < \overline{x_2}$) entraîne $\mu_1 > \mu_2$ (réciproquement $\mu_1 < \mu_2$) ? Pour quelle raison cette implication ne peut elle pas se produire dans tous les cas ?

1.4. Dans le cas où vous auriez été dans la situation du médecin qui ne peut obtenir que 5 données avant traitement et 5 données après traitement, auriez-vous procédé comme lui ? Pourquoi ?

Retour à l’article Indications pour cette activité

Activités 2. - Sommaire

On trouve dans les pages suivantes les énoncés de six activités qui reprennent des questions soulevées par Gavarret dans les deux premiers articles du chapitre premier de son ouvrage.
L'objet de ce chapitre est de mettre en évidence la « Nécessité de recourir au calcul des probabilités pour suppléer l'insuffisance des règles de la logique » comme l'indique le sous-titre.
Un exemple de questions soulevées par Gavarret est illustré par l'extrait de texte suivant [G., p. 46-47].
On peut accéder aux énoncés de ces activités ou revenir à l'article grâce aux liens placés sous l'extrait, dans le bas de cette page.

1° Une urne contient un nombre inconnu de boules blanches et de boules noires ; on fait mille tirages successifs d'une boule chacun. Sur ces mille tirages, faits complètement au hasard, il sort neuf cents boules blanches et cent boules noires. A la suite d'un semblable résultat, personne n'hésiterait à affirmer que nécessairement les blanches étaient beaucoup plus nombreuses dans l'urne que les noires ; personne n'hésiterait à admettre que si toutes les boules sorties étaient remises dans l'urne et mêlées à celles qui y sont restées, dans mille nouveaux tirages, faits au hasard, on obtiendrait beaucoup plus de blanches que de noires. De semblables conclusions viendraient spontanément à l'esprit de tout le monde, même de ceux qui nient l'influence des grands nombres de faits sur nos décisions. Mais là s'arrêtent les indications fournies par les règles de la *logique*. Supposons, en effet, qu'on veuille aller plus loin, et qu'on se dise : Le résultat statistique montre que sur dix boules sorties, il y a neuf blanches et une noire, dans *quelles limites d'erreur* le rapport de blanches au nombre total des boules contenues dans l'urne, peut-il différer de l'expression neuf dixièmes fournie par l'expérience ? En n'appelant à son secours que les règles de la logique, qui se croirait assez éclairé pour répondre à une semblable question ? Voilà une difficulté bien faite pour arrêter l'esprit le plus juste et le plus clairvoyant.

2° Une urne contient un nombre inconnu de boules blanches et de boules noires ; on fait complètement au hasard vingt tirages successifs d'une boule chacun ; sur les vingt boules sorties, dix-huit sont blanches et deux seulement noires. Ici, comme dans l'exemple précédent, sur dix boules sorties on a amené neuf blanches et une noire. La seule différence, et elle est grave, n'existe que dans le nombre des tirages effectués. Conclura-t-on de ce résultat que certainement l'urne contenait avant les tirages beaucoup plus de blanches que de noires ?

Activités 2.1

Activités 2.2

Activités 2.3

Activités 2.4

Activités 2.5

Activités 2.6

Retour à l'article

Activité 2.1

Cette activité reprend les questions posées par Gavarret pages 46 et 47.

1° Une urne contient un nombre inconnu de boules blanches et de boules noires ; on fait mille tirages successifs d'une boule chacun. Sur ces mille tirages, faits complètement au hasard, il sort neuf cents boules blanches et cent boules noires. A la suite d'un semblable résultat, personne n'hésiterait à affirmer que nécessairement les blanches étaient beaucoup plus nombreuses dans l'urne que les noires ; personne n'hésiterait à admettre que si toutes les boules sorties étaient remises dans l'urne et mêlées à celles qui y sont restées, dans mille nouveaux tirages, faits au hasard, on obtiendrait beaucoup plus de blanches que de noires. De semblables conclusions viendraient spontanément à l'esprit de tout le monde, même de ceux qui nient l'influence des grands nombres de faits sur nos décisions. Mais là s'arrêtent les indications fournies par les règles de la *logique*. Supposons, en effet, qu'on veuille aller plus loin, et qu'on se dise : Le résultat statistique montre que sur dix boules sorties, il y a neuf blanches et une noire, dans *quelles limites d'erreur* le rapport de blanches au nombre total des boules contenues dans l'urne, peut-il différer de l'expression neuf dixièmes fournie par l'expérience ? En n'appelant à son secours que les règles de la logique, qui se croirait assez éclairé pour répondre à une semblable question ? Voilà une difficulté bien faite pour arrêter l'esprit le plus juste et le plus clairvoyant.

2° Une urne contient un nombre inconnu de boules blanches et de boules noires ; on fait complètement au hasard vingt tirages successifs d'une boule chacun ; sur les vingt boules sorties, dix-huit sont blanches et deux seulement noires. Ici, comme dans l'exemple précédent, sur dix boules sorties on a amené neuf blanches et une noire. La seule différence, et elle est grave, n'existe que dans le nombre des tirages effectués. Conclura-t-on de ce résultat que certainement l'urne contenait avant les tirages beaucoup plus de blanches que de noires ?

2.1.1. Donner un exemple de composition d'urne comprenant N boules dont K blanches et H noires telle que le résultat d'un tirage de $n_1 = 1000$ boules au hasard puisse donner $k_1 = 900$ blanches et $h_1 = 100$ noires, **mais** avec cependant $K < H$ (donner des valeurs numériques pour K, N et H).

2.1.2. Même question avec $K = H$.

Retour au sommaire des activités 2. Indications pour cette activité

Activité 2.2

2.2.1. Gavarret, à partir de l'observation de $k_1 = 900$ blanches et $h_1 = 100$ noires, en a déduit que $K > H$. Aurait-il déduit le même résultat s'il avait observé $k_1 = 901$ blanches et $h_1 = 99$ noires ? Même question avec $k_1 = 902$ blanches et $h_1 = 98$ noires ? Ne peut-on pas en déduire que c'est l'observation d' "**au moins** 900 **blanches**" qui amène à la décision : $K > H$?

2.2.2. Dans ce cas, à partir des N, K et H proposés en **2.1.**, calculer la probabilité que le tirage de $n_1 = 1000$ boules au hasard donne au moins 900 blanches.

Retour au sommaire des activités 2. Indications pour cette activité

Activité 2.3

Le résultat d'un premier tirage de n_1 boules a fourni Gavarret une information sur la composition de l'urne. Une des utilisations possibles de cette information est la suivante : la véritable proportion de boules blanches est celle fournie par le premier tirage c'est-à-dire 90%.

a) En se servant de cette interprétation, calculer la probabilité que suite à un nouveau tirage de $n_2 = 1000$ boules dans cette urne, il y ait plus de blanches que de noires.

Que pensez-vous de la phrase de Gavarret : « A la suite d'un semblable résultat, personne n'hésiterait à affirmer que nécessairement les blanches étaient beaucoup plus nombreuses dans l'urne que les noires ; personne n'hésiterait à admettre que si toutes les boules sorties étaient remises dans l'urne et mêlées à celles qui y sont restées on obtiendrait beaucoup plus de blanches que de noires » ?

b) Quelle serait la probabilité de tirer au moins 950 blanches sur les 1000 boules tirées ? Comment expliquer ce paradoxe ?

Retour au sommaire des activités 2.   Indications pour cette activité

Activité 2.4

Dans la suite de son ouvrage, on verra que Gavarret considère de fait un événement comme « absolument certain » [G, p. 257] si sa probabilité est supérieure ou égale à $0{,}9953$.

Dans la suite, on dira qu'un événement est "impossible au sens de Gavarret" si sa probabilité est inférieure ou égale à $0{,}0047$.

2.4.1. a) Si dans l'urne, il y a autant de blanches que de noires, c'est-à-dire $K = H$, montrer que dans un tirage de 1000 boules, le fait de tirer au moins 900 blanches est "impossible au sens de Gavarret".

Si dans l'urne il y a autant de blanches que de noires, calculer :

- la probabilité qu'il y ait au moins 546 blanches.
- la probabilité qu'il y ait au moins 543 blanches.
- la probabilité qu'il y ait au moins 542 blanches.
- la probabilité qu'il y ait au moins 541 blanches.

b) Montrer alors que $A = 543$ est le plus petit des entiers x tels que, si dans l'urne il y a autant de blanches que de noires, l'événement "on a tiré au moins x boules blanches à la suite de 1000 tirages" est "impossible au sens de Gavarret".

2.4.2. a) Si la proportion de blanches dans l'urne est 49%, calculer la probabilité de tirer au moins 543 blanches.

Si la proportion de blanches dans l'urne est 48%, calculer la probabilité de tirer au moins 543 blanches.

On pose $\mathbb{P}_p(X \geq 543)$ la probabilité d'avoir au moins 543 boules blanches parmi 1000 tirées dans une urne contenant une proportion $p = \dfrac{K}{N}$ de blanches.

A la suite des résultats précédents, on conjecture que la probabilité de tirer au moins 543 blanches parmi 1000 est une fonction croissante de la proportion p de blanches dans l'urne.

b) Montrer rigoureusement cette conjecture.

c) En déduire que si c'est l'hypothèse $K < H$ qui est vraie, on a $\mathbb{P}_p(X \geq 543) < 0{,}0047$.

Traduction : si $X \geq 543$ est réalisé, en décidant que $K > P$ alors que c'est l'hypothèse $K \leq H$ qui est vraie, on prend donc un risque de se tromper inférieur à $\alpha = 0{,}0047$.

2.4.3. Soit maintenant le problème suivant :

On doit décider entre l'hypothèse $H_0 : K \leq H$ et l'hypothèse $H_1 : K > H$.

On prend un risque au plus égal à $\alpha = 0{,}0047$ de se tromper en décidant $K > H$ alors que c'est l'hypothèse $K \leq H$ qui est vraie. $\alpha = 0{,}0047$ est dénommé "risque de première espèce".

On effectue $n = 1000$ tirages dans l'urne et on observe x blanches.

D'après les résultats précédents, on obtient la règle de décision suivante :

Règle de décision 1 : on rejette l'hypothèse nulle (et donc on accepte l'hypothèse alternative en décidant "$K > H$") si et seulement si $x \geq 543$.

Soit la règle de décision suivante :

Règle de décision 2 : on rejette l'hypothèse nulle (et donc on accepte l'hypothèse alternative en décidant "$K > H$") si et seulement $\mathbb{P}_{0,50}(X \geq x) \leq 0{,}0047$.

Montrer que les deux règles de décision sont équivalentes.

2.4.4. On note toujours $p = \dfrac{K}{N}$ et $\mathbb{P}_p((X < 543)$ la probabilité de ne pas rejeter l'hypothèse nulle $K \leq H$ alors qu'en réalité on a $K > H$. C'est le risque de se tromper en décidant $K \leq H$ appelé "risque de deuxième espèce". Ce risque dépend de p $(0{,}50 < p < 1)$ et est noté $\beta(p)$.

a) Calculer $\beta(p)$ pour $p = 0{,}51; 0{,}52; 0{,}53; 0{,}54; 0{,}55; 0{,}60$.

Montrer rigoureusement que, plus il y a de boules blanches dans l'urne, plus le risque de se tromper en disant "il y a moins de boules blanches que de boules noires" diminue.

Ici, le “risque de première espèce” est égal à $\alpha = 0{,}0047$. Actuellement, le “risque de première espèce” est souvent pris égal à $\alpha = 0{,}05$ (5%).

Soit maintenant le nouveau problème de test.

On doit décider entre l’hypothèse $H_0 : K \leq H$ et l’hypothèse $H_1 : K > H$.

On prend un risque au plus égal à $\alpha = 0{,}05$ de se tromper en décidant $H_1 : K > H$ alors que c’est l’hypothèse $H_0 : K \leq H$ qui est vraie.

On effectue $n = 1000$ tirages dans l’urne et on observe x blanches.

Soit A^* le plus petit entier $x \leq n$ tel que $\mathbb{P}(X \geq x) \leq 0{,}05$.

On utilise une des deux règles de décision :

Règle de décision 1 : on rejette l’hypothèse nulle $H_0 : K \leq H$ (et donc on accepte l’hypothèse alternative H_1 en décidant “$K > H$”) si et seulement si $x \geq A$.

Règle de décision 2 : on rejette l’hypothèse nulle $H_0 : K \leq H$ (et donc on accepte l’hypothèse alternative H_1 en décidant “$K > H$”) si et seulement $\mathbb{P}_{0,50}(X \geq x) \leq \alpha$.

b) Calculer A^* dans le cas où on prend $\alpha = 0{,}05$.

Avec cette règle de décision, lorsque $x < A^*$, on ne rejette pas $K \leq H$ puisqu’on ne prend la décision “ $K > H$ est vrai” que lorsque $x \geq A^*$. On note $\beta^*(p) = \mathbb{P}_p((X < A^*)$ la probabilité de ne pas rejeter l’hypothèse nulle avec un “risque de première espèce” $\alpha = 0{,}05$.

c) Calculer $\beta^*(p)$ pour $p = 0{,}51; 0{,}52; 0{,}53; 0{,}54; 0{,}55; 0{,}60$. Pour ces valeurs de p, comparer les résultats de fonctions β et β^*. Peut-on généraliser ce résultat à tout p tel que $0{,}50 \leq p \leq 1$?

Qu’en concluez-vous concernant la comparaison des risques dans les deux cas suivants : celui où le risque de première espèce est $\alpha = 0{,}0047$ et celui où le risque de première espèce est $\alpha = 0{,}05$?

2.4.5. Dans son texte, Gavarret dit qu’à la suite du tirage de 1000 boules et après l’observation de 900 blanches “personne n’hésiterait à affirmer que nécessairement les blanches étaient **beaucoup** plus nombreuses dans l’urne que les noires” (c’est nous qui soulignons). Or les tests proposés précédemment concernaient la question de confirmer ou infirmer la proposition : “nécessairement les blanches étaient plus nombreuses dans l’urne que les noires”.

En l’absence de précision dans le texte sur l’interprétation de “beaucoup”, on peut, parmi une infinité de possibles, proposer par exemple trois “interprétations” pour le mot “beaucoup” :

- Première interprétation : “beaucoup” signifie “il y a au moins trois fois plus de blanches que de noires dans l’urne” ;
- Deuxième interprétation : “beaucoup” signifie “il y a au moins quatre fois plus de blanches que de noires dans l’urne” ;
- Troisième interprétation : “beaucoup” signifie “il y a au moins neuf fois plus de blanches que de noires dans l’urne”.

Pour chacune de ces “interprétations”, en prenant toujours le même risque de première espèce que Gavarret, c’est-à-dire $\alpha = 0{,}0047$, construire la règle de décision, énoncer la décision qui aurait été prise à la suite du tirage de 1000 boules et après l’observation de 900 blanches et évaluer le “risque de deuxième espèce” à la suite du résultat.

Retour au sommaire des activités 2. Indications pour cette activité

Activité 2.5

À la fin de l'alinéa 1, Gavarret cherche en fait, à la suite du tirage de 1000 boules et après l'observation de 900 blanches, la valeur de ϵ tel que $|p - \widehat{p}| < \epsilon$. $\widehat{p}$, estimation de p, est fournie par la phrase : « sur dix boules sorties, on a amené neuf blanches et une noire » donc $\widehat{p} = 0{,}90$. N'ayant pas précisé avec quelle probabilité il cherche cet encadrement, on peut supposer que cette probabilité est celle qu'il attribue à des événements "certains" c'est-à-dire qu'elle est égale à $1 - \alpha$ où $\alpha = 0{,}0047$.

2.5.1. Contourner la « difficulté bien faite pour arrêter l'esprit le plus juste et le plus clairvoyant », en calculant ϵ et en déduire un intervalle de confiance pour p de niveau $0{,}9953$.

2.5.2. Gavarret ajoute encore une interrogation :

> **Essayons, en effet, de pénétrer plus avant, et demandons aux spectateurs combien il leur faudrait de tirages pour se croire fondés à regarder le résultat statistique comme représentant, *à très peu près* et dans des *limites données d'erreur possible*, la véritable composition de l'urne.**

Pour un ϵ donné, trouver la formule qui donne approximativement la taille de l'échantillon qui permettra d'obtenir $|p - \widehat{p}| \leq \epsilon$ avec une probabilité supérieure à $1 - 0{,}0047 = 0{,}9953$.

Retour au sommaire des activités 2.

Indications pour cette activité

Activité 2.6

Dans l'alinéa 2, il considère que la proportion observée reste la même mais ceci à la suite d'un tirage de 20 boules.

On examine quatre cas correspondant à des interprétations de " beaucoup plus de blanches que de noires" (dans l'urne) :

- 1er cas : tester $p \leq 0{,}50$ contre $p > 0{,}50$;
- 2ème cas : tester $p \leq 0{,}75$ contre $p > 0{,}75$;
- 3ème cas : tester $p \leq 0{,}80$ contre $p > 0{,}80$;
- 4ème cas : tester $p \leq 0{,}90$ contre $p > 0{,}90$.

2.6.1. Calculer les coefficients binomiaux :

$$\binom{20}{20}, \quad \binom{20}{19}, \quad \binom{20}{18}, \quad \binom{20}{17}, \quad \binom{20}{16}.$$

2.6.2. On note Y la variable aléatoire qui, à un tirage de 20 boules dans une urne où la proportion de blanches est égale à $p = \dfrac{K}{N}$, associe le nombre y de blanches.

Pour $p = 0{,}50$; $0{,}75$; $0{,}80$; $0{,}90$, calculer $\mathbb{P}_p((Y \geq y)$ avec $y = 16, 17, 18, 19$ et 20. Ici, dans la mesure où la taille de l'échantillon est inférieure à 30, on n'utilisera pas l'approximation normale mais la formule de la loi binomiale :

$$\mathbb{P}_p(Y \geq y) = \sum_{i=y}^{20} \binom{20}{i} p^i (1-p)^{20-i}.$$

2.6.3. On note A_p le plus petit entier $y \leq n$ tel que $\mathbb{P}_p(Y \geq y) \leq 0{,}0047$.

ATTENTION. Si $\mathbb{P}_p(Y = n) > \alpha$, pour tout $y \leq n$ on a $\mathbb{P}_p(Y \geq y) > \alpha$ et il ne sera pas possible de trouver A_p. La probabilité de prendre la décision "$K > H$" sera toujours supérieure au "risque de première espèce" accepté et il ne sera pas possible de construire un test ayant α pour "risque de première espèce".

a) Pour chaque cas, en prenant toujours le même "risque de première espèce" que Gavarret, c'est-à-dire $\alpha = 0{,}0047$, construire la règle de décision, énoncer la décision qui aurait été prise à la suite du tirage de 20 boules et après l'observation de 18 blanches et quand il est possible de construire un test ayant un "risque de première espèce" $\alpha = 0{,}0047$, évaluer le "risque de deuxième espèce" à la suite du résultat.

b) Comparer alors l'évaluation de ce risque avec celui obtenu en tirant 1000 boules. Peut-on conjecturer que le "risque de deuxième espèce" diminue quand la taille de l'échantillon s'accroît ?

c) Prouver alors cette conjecture. Gavarret a-t-il raison en ne s'estimant pas suffisamment « éclairé par un si petit nombre d'épreuves » ?

2.6.4. Aujourd'hui, il est d'usage d'utiliser le "risque de première espèce" $\alpha = 0{,}05$ alors que Gavarret utilisait implicitement le risque $\alpha = 0{,}0047$.

1er problème : On souhaite toujours tester $H_0 : p \leq 0{,}50$ contre $H_1 : p > 0{,}50$. Supposons qu'à la suite du tirage de 1000 boules dans une urne, on ait tiré 532 blanches c'est-à-dire une proportion observée de $53{,}2\%$ de blanches.

Qu'en conclurait-on aujourd'hui ?

Qu'en aurait-on conclu avec le risque pris par Gavarret ?

2ème problème : On souhaite encore tester $H_0 : p \leq 0{,}50$ contre $H_1 : p > 0{,}50$. Supposons qu'à la suite du tirage de 2000 boules dans une urne, on ait tiré 1064 blanches c'est-à-dire encore une proportion observée de $53{,}2\%$ de blanches.

Qu'en conclurait-on aujourd'hui ?

Qu'en aurait-on conclu avec le risque pris par Gavarret ?

2.6.5. En gardant le même nombre $n = 1000$ de tirage dans l'urne et le même risque $\alpha = 0{,}0047$, montrer qu'il n'est pas équivalent d'effectuer le test $H_0 : K \leq H$ contre $H_1 : K > H$ et le test $H_0 : K \geq H$ contre $H_1 : K < H$.

Donner les décisions obtenues à la suite de chacun des deux tests lorsque le nombre de boules blanches tirées est respectivement égal à 550, 502, 480 et 420 et comparer ces décisions.

Retour au sommaire des activités 2. Indications pour cette activité

Activité 3

Cette activité représente une tentative d'illustration de la notion de chance variable.

Une population est composée de N individus numérotés $1, 2, ..., \alpha, ..., N-1, N$ possédant la pathologie donnée. L'événement "mort à la suite d'une pathologie pour l'individu numéro α" est un événement aléatoire dont la réalisation est assimilée au tirage d'une boule blanche dans une urne contenant K_α boules blanches et H_α noires.

Calculer la probabilité de l'événement : "l'individu numéro α meurt de la pathologie".

Calculer la probabilité de l'événement : "un individu tiré au hasard dans la population meurt de la pathologie".

Retour à l'article Indications pour cette activité

Activité 4

Compléter le tableau en ajoutant deux lignes :

- 6ème ligne avec la série initiale 600 morts sur 1500 malades.
- 7ème ligne avec la série initiale 840 morts sur 2100 malades.

Les mortalités moyennes en ajoutant la série 1 et en ajoutant la série 2 se rapprochent. Pouvait-on prévoir le résultat ?

Retour à l'article Indications pour cette activité

Activité 5

Dans les trois exemples précédents, comment se transformeraient les énoncés des résultats si, pour construire des intervalles de confiance, Gavarret avait utilisé ceux obtenus par les formules utilisées en classe de lycée et de B.T.S.

Quelle observation peut-on faire à partir de ces résultats concernant le rapport entre le degré de confiance à accorder à l'intervalle et la précision ?

Une application courante de la théorie des intervalles de confiance de niveau donné $1 - \alpha$ consiste dans son utilisation pour tester l'hypothèse "lorsqu'un enfant naît, il y a autant de chance que ce soit un garçon qu'une fille" contre l'hypothèse "lorsqu'un enfant naît, les chances d'être un garçon ou une fille sont différentes" avec un "risque de première espèce" égal à α. On rejette l'hypothèse nulle si et seulement si $\frac{1}{2}$ n'appartient pas à l'intervalle construit.

A l'aide de cette règle, pour chacun des trois exemples, effectuer ce test, à la fois dans le cadre du "risque de premier espèce" pris par Gavarret et de celui utilisé dans les classes de lycée et de B.T.S.

Retour à l'article Indications pour cette activité

Activité 6

Pour le 3ème exemple du 1er groupe d'exemples (paragraphe 4.3), calculer la différence observée d et la limite l en utilisant la formule de Poisson. Obtient-on la même conclusion que Gavarret ?

Retour à l'article

Indications pour cette activité

Activité 7

On note p_x la probabilité qu' un malade atteint d'une pathologie donnée soit guéri avec la dose x d'un médicament. On cherche à modéliser la relation entre p_x et x par $p_x = f(x)$ où f est une fonction donnée.

7.1. Dans un premier temps, on essaie la fonction f définie ainsi :
$f(x) = a + bx$ où a et b sont des coefficients réels.
Si on cherche une modélisation telle que la probabilité de guérir augmente avec la dose, quelle est la condition sur le coefficient b ?

Donner plusieurs exemples de valeurs pour a, b pour lesquels cette modélisation n'est pas adaptée.

7.2. La plupart du temps en médecine, les x mesurent log_{10}(dose) donc x peut prendre toutes les valeurs réelles.

a) Dans ce cas, montrer que les deux fonctions suivantes sont possibles pour une modélisation :

- Soit f définie par : $f(x) = \dfrac{e^{a+bx}}{1 + e^{a+bx}}$
- Soit f définie par : $f(x) = F_Z(a + bx)$ où F_Z est la fonction de répartition de la loi normale réduite centrée.

b) Soient x_1 et x_2, deux log-doses différentes de la “médication”. Pour x_1, on observe une proportion $\widehat{p_{x_1}}$ de personnes guéries et pour x_2, on observe une proportion $\widehat{p_{x_2}}$ de personnes guéries. D'après la “méthode numérique”, on peut penser que $\widehat{p_{x_1}}$ estime bien p_{x_1} et que $\widehat{p_{x_2}}$ estime bien p_{x_2}.

On décide d'utiliser la première modélisation :

$$f(x) = \frac{e^{a+bx}}{1 + e^{a+bx}}$$

Donner une méthode pour alors estimer les nombres inconnus a et b.

c) En déduire une solution au problème suivant :
Une dose de $d_1 = 2mg$ d'un remède est administrée à 20 patients. 12 sont guéris.
Une dose de $d_2 = 3mg$ du même remède est administrée à 25 autres patients. 18 sont guéris.
On ne peut augmenter la dose sans risques si elle est trop élevée.
Trouver la dose à administrer pour avoir une probabilité de guérison de 80%.

d) Qu'aurait dit Gavarret du résultat ?

7.3. Que feriez-vous si vous connaissiez les résultats $\widehat{p_{x_1}}$, $\widehat{p_{x_2}}$,...,$\widehat{p_{x_k}}$ pour k log-doses différentes x_1, x_2,...,x_k ?

Retour à l'article Indications pour cette activité

Activité 8

Compléter le tableau suivant :

Mortalité moyenne fournie par la statistique	STATISTIQUES de 950 cas	
	Répartition des malades	Erreur possible
	323 morts guéris	

Énoncer la conclusion comme l'aurait fait Gavarret.

Retour à l'article Indications pour cette activité

Activité 9

Appliquer la "nouvelle formule" $l = 2\sqrt{\left(\frac{1}{\mu} + \frac{1}{\mu'}\right)\left(\frac{2m^*n^*}{\mu^{*2}}\right)}$ aux données de la 1ère version et aux données de la 2ème version et énoncer la conclusion. Pourquoi obtient-on avec cette nouvelle formule la même limite l pour les deux versions ?

Retour à l'article   Indications pour cette activité

Activité 10

En utilisant la technique (erronée) qui consiste à calculer séparément pour p_1 et pour p_2 des intervalles de confiance de niveau $1 - 0{,}0047 = 0{,}9953$ et à décider que $p_1 \neq p_2$ si ces intervalles ne se chevauchent pas, quelles décisions auraient été prises concernant les trois versions de la page 23 du texte ? Comparer ces décisions à celles prises par Gavarret.

Retour à l'article Indications pour cette activité

Activité 11

Soit une population constituée de N personnes numérotées $\alpha = 1, 2, ..., N$. On note p_α la probabilité de contracter la maladie pour la personne α. Une mesure de santé publique est appliquée à l'ensemble de la population. On note p^*_α la probabilité de contracter la maladie pour la personne α après l'application de la mesure.

On pose $e_\alpha = p_\alpha - p^*_\alpha$, le gain de probabilité pour l'individu α.

Soient les deux situations suivantes avec $N = 4$ (données fictives ! ! !).

1ère situation :

p_α	0,9	0,6	0,8	0,9
p^*_α	0,6	0,5	0,4	0,9

2ème situation :

p_α	0,9	0,6	0,8	0,9
p^*_α	0,7	0,4	0,6	0,7

Calculer le gain en terme de “chance moyenne”, obtenu par l'application de la mesure et comparer les deux situations.

Trouver une formule générale pour calculer le gain dans le cas N quelconque.

Dans le cas où $e_\alpha = e$ pour tout α, où e est le gain de probabilité identique pour chaque individu, y a-t-il des exemples où ce modèle est mis en défaut ?

Retour à l'article Indications pour cette activité

Activité 12

Pour comparer la saison chaude et humide avec la saison chaude et sèche, Gavarret a pris comme "rapport des fréquences" $\dfrac{n_1}{n_1+n_2}$ où n_1 est le nombre correspondant à la deuxième ligne des données c'est-à-dire la saison chaude et humide mais pour comparer la saison chaude et humide avec la saison froide n_1 est le nombre correspondant à la première ligne des données.

Montrer qu'il est indifférent de prendre comme "rapport des fréquences" $\dfrac{n_1}{n_1+n_2}$ ou $\dfrac{n_2}{n_1+n_2}$.

Retour à l'article Indications pour cette activité

Activité 13

Justifier le test $H_0 : p = 4/7$ contre $H_1 : p \neq 4/7$.

Est-il équivalent à celui utilisé par Gavarret consistant dans l'artifice de multiplier n_1 par 3/4 et de tester ensuite $H_0 : p = 0{,}50$ contre $H_1 : p \neq 0{,}50$? Justifier mathématiquement la réponse.

Retour à l'article Indications pour cette activité

Activité 14

Calculer la valeur de la statistique de test et l'intervalle en utilisant l'artifice de multiplier n_1 par 3/4. Utiliser la "bonne méthode" n'utilisant pas l'artifice et conclure.

Retour à l'article   Indications pour cette activité

Activité 15

On utilise trois médicaments M_1, M_2 et M_3 pour guérir une maladie et chaque médicament est testé sur un échantillon de malades. On pose p_i la proportion inconnue de personnes guéries par le médicament M_i, pour $i = 1, 2, 3$. Les données sont fictives.

1er médicament : 350 sont guéris dans un échantillon de 1000 personnes.

2ème médicament : 45 sont guéris dans un échantillon de 100 personnes.

3ème médicament : 420 sont guéris dans un échantillon de 1000 personnes.

Tester en utilisant la formule de Poisson :

1er médicament : $H_0 : p_1 = p_2$ contre $H_1 : p_1 \neq p_2$.

2ère médicament : $H_0 : p_2 = p_3$ contre $H_1 : p_2 \neq p_3$.

3ère médicament : $H_0 : p_1 = p_3$ contre $H_1 : p_1 \neq p_3$.

Le résultat est-il contradictoire ?

Retour à l'article Indications pour cette activité

Activité 16

16.1. Dans l'exemple des individus atteints de dysenterie au Bengale, lors de la comparaison de la saison chaude et humide avec la saison chaude et sèche, il n'est pas indiqué le nombre d'habitants concernés pour chacune des saisons, mais on suppose cependant sans doute que c'est le même. En réalité, Gavarret teste l'égalité ou non de deux probabilités conditionnelles. En effet, on définit les événements suivants :

- S_1 : "être au Bengale pendant la saison chaude et humide" ;
- S_2 : "être au Bengale pendant la saison chaude et sèche" ;
- D : "être atteint de dysenterie".

On note :

- $p_1 = \mathbb{P}_D(S_1)$ probabilité d'être au Bengale dans la saison chaude et humide sachant qu'on est atteint de la dysenterie ;
- $p_2 = \mathbb{P}_D(S_2)$ probabilité d'être au Bengale dans la saison chaude et sèche sachant qu'on est atteint de la dysenterie ;
- $p_1^* = \mathbb{P}_{S_1}(D)$ probabilité d'être atteint de dysenterie sachant que l'on est au Bengale dans la saison chaude et humide ;
- $p_2^* = \mathbb{P}_{S_2}(D)$ probabilité d'être atteint de dysenterie sachant que l'on est au Bengale dans la saison chaude et sèche ;
- $p = \mathbb{P}(D)$ probabilité d'être atteint de dysenterie (cette probabilité est appelé par les médecins "prévalence de la dysenterie").

Le test effectué par Gavarret est donc en réalité le suivant :

$H_0 : p_1 = p_2$ contre $H_1 : p_1 \neq p_2$.
alors que le test intéressant est bien sûr :

$H_0 : p_1^* = p_2^*$ contre $H_1 : p_1^* \neq p_2^*$.
En effet, c'est celui qui permet de tester l'influence de la saison sur l'apparition de la dysenterie.

Montrer que $p_1 = p_2$ implique $p_1^* = p_2^*$ si et seulement si $\mathbb{P}(S_1) = \mathbb{P}(S_2)$.

16.2. Adopter la même démarche pour traiter les données de l'érysipèle et donner la conclusion.

Retour à l'article Indications pour cette activité

Indications sur l'activité 1.

1.1. Tirages possibles (sans remplacement) de numéros dans les deux populations (avant, après) :

Avant	14	2	6	9	19
Après	8	1	13	15	7

Tensions observées :

Avant :

i	$x_{1,i}$
1	16
2	13
3	13
4	15
5	17
$\overline{x_1}$	14,8

Après :

i	$x_{2,i}$
1	15
2	14
3	11
4	13
5	15
$\overline{x_2}$	13,5

Au vu de ces deux séries, puisque $\overline{x_1} = 14{,}8 > 13{,}5 = \overline{x_2}$, les tenants de la méthode numérique auraient conclu à l'efficacité du médicament.

Mais il y a d'autres tirages possibles, par exemple :

Avant	2	4	18	20	14
Après	19	3	7	1	12

Tensions observées :

Avant :

i	$x_{1,i}$
1	13
2	11
3	12
4	13
5	16
$\overline{x_1}$	13

Après :

i	$x_{2,i}$
1	15
2	18
3	15
4	14
5	15
$\overline{x_2}$	15,4

Au vu de ces deux nouvelles séries, puisque $\overline{x_1} = 13 < 15{,}4 = \overline{x_2}$, les tenants de la méthode numérique n'auraient pas conclu à l'efficacité du médicament.

1.2. Moyennes effectives avant et après : $\mu_1 = 14$ et $\mu_2 = 14$.

Écarts des tensions à la moyenne avant traitement :

α	1	2	3	4	5	6	7	8	9	10
$\epsilon_{1,\alpha}$	-2	-1	2	-3	0	-1	1	2	1	-2
α	11	12	13	14	15	16	17	18	19	20
$\epsilon_{1,\alpha}$	1	1	-2	2	1	0	0	-2	3	-1

Écarts des tensions à la moyenne après traitement :

α	1	2	3	4	5	6	7	8	9	10
$\epsilon_{2,\alpha}$	0	0	4	-2	-2	-2	1	1	-1	-2
α	11	12	13	14	15	16	17	18	19	20
$\epsilon_{2,\alpha}$	2	1	-3	4	-1	2	0	-3	1	0

Il est aisé de vérifier que : $\sum_{\alpha=1}^{20} \epsilon_{1,\alpha} = 0$ et $\sum_{\alpha=1}^{20} \epsilon_{2,\alpha} = 0$.

Ainsi donc, si on entend par "erreur" l'écart de la valeur de la tension à la moyenne des valeurs de **toute** la population, les "erreurs" se compensent. On pouvait prévoir le résultat car :

$$\sum_{\alpha=1}^{20} \epsilon_{1,\alpha,} = \sum_{\alpha=1}^{20}(\xi_{1,\alpha} - \mu_1) = (\sum_{\alpha=1}^{20} \xi_{1,\alpha}) - 20\mu_1 = 20\mu_1 - 20\mu_1 = 0.$$

Même raisonnement pour les écarts calculés après traitement. L'assertion de Louis semble donc vérifiée.

1.3. Dans le cas de la première série de tirages :

Avant :

i	$e_{1,i}$
1	2
2	-1
3	-1
4	1
5	3
$\sum_{i=1}^{5} e_{1,i}$	4

Après :

i	$e_{2,i}$
1	1
2	0
3	-3
4	-1
5	2
$\sum_{i=1}^{5} e_{2,i}$	-1

Ici, les "erreurs" ne se compensent pas.

Mais par exemple, pour les tirages suivants :

Avant	8	15	4	15	2
Après	10	3	4	19	9

les "erreurs" se compensent.

Gavarret avait donc raison en émettant des doutes sur le fait que les "erreurs" puissent se compenser. Cette compensation est vraie si c'est l'ensemble de la population qui est prise en compte mais elle ne l'est plus obligatoirement si c'est seulement une partie de la population étudiée qui est examinée. En effet, $\sum_{\alpha=1}^{20} \epsilon_{1,\alpha} = 0$ n'implique pas obligatoirement $\sum_{\alpha\in A} \epsilon_{1,\alpha} = 0$ pour $A \subset \{1, 2, ..., 19, 20\}$.

Si les "erreurs" se compensent, alors $\overline{x_1} - \overline{x_2} = (\mu_1 + \sum_{i=1}^{5} e_{1,i}) - (\mu_2 + \sum_{i=1}^{5} e_{2,i}) = \mu_1 - \mu_2$ et $\overline{x_1} > \overline{x_2}$ entraîne $\mu_1 > \mu_2$. Gavarret a bien compris qu'à partir où les observations se font sur une partie seulement de la population, la possibilité d'étendre les conclusions issues de ces observations à la population toute entière n'allait pas de soi.

1.4. Il semble plus pertinent de travailler sur les différences entre la tension après et la tension avant et de procéder ainsi :

Le médecin tire au hasard 5 numéros parmi les 20. Pour chacun des patients, il examine la tension avant et après. En effet, ce type d'expérimentation tiendra davantage compte de l'effet "patient" sur le médicament. Si un patient a une tension très élevée avant la prise du médicament et que celui-ci a un effet, la tension restera cependant élevée mais c'est la différence qu'il faudra prendre en compte.

Commentaire : Le but de cette activité est de faire prendre conscience de la différence entre la population étudiée et l'échantillon observé, et de la non-évidence à étendre à toute la population des conclusions obtenues à la suite d' observations issues d' un échantillon.

Retour à l'article   Retour à l'énoncé de l'activité

Indications sur l'activité 2.1

2.1.1. Exemple de possibilité d'obtention d'un tirage de 900 blanches et 100 noires avec $K < H$.

Si le tirage a lieu **avec remplacement**, il suffit de prendre N quelconque ($N > 2$ et $K < \frac{N}{2}$).
Par exemple, $N = 3$, $K = 1$ et $H = 2$. Ou encore $N = 1000$, $K = 499$ et $H = 501$.

Si le tirage a lieu **sans remplacement**, il faut prendre $K \geq 900$ et $N > 2K$.
Par exemple $N = 1801$, $K = 900$ et $H = 901$ conviennent mais aussi $N = 2000$, $K = 999$ et $H = 1001$, ou $N = 2000$, $K = 950$ et $H = 1050$.

2.1.2. Exemple de possibilité d'obtention d'un tirage de 900 blanches et 100 noires avec $K = H$.

Si le tirage a lieu **avec remplacement**, il suffit de prendre K quelconque et $N = 2K$.
Par exemple, $N = 2$, $K = 1$ et $H = 1$. Ou encore $N = 1000$, $K = 500$ et $H = 500$ mais aussi $N = 2000$, $K = 1000$ et $H = 1000$.

Si le tirage a lieu **sans remplacement**, il faut prendre $K \geq 900$ et $N = 2K$.
Par exemple, $N = 1800$, $K = 900$ et $H = 900$ conviennent.

Commentaire : Il s'agit, par des exemples, de montrer que l'observation d'un beaucoup plus grand nombre de blanches que de noires n'implique pas qu'il y ait plus de blanches que de noires dans l'urne.

Retour au sommaire des activités 2 Retour à l'énoncé de l'activité

Indications sur l'activité 2.2

2.2.1. En réalité, Gavarret énonce la conclusion $K > H$ à partir de l'observation suivant laquelle le nombre de boules blanches, dans un tirage de 1000 boules, est très supérieur à un nombre donné (ici il propose 900). On voit mal comment il concluerait $K > H$ en observant 900 blanches mais concluerait le contraire en observant 901 blanches, ..., ou 902 blanches ! Pour choisir entre les deux hypothèses : $K > H$ et $K \leq H$, la règle de décision est implicitement la suivante : "Soit X la variable qui associe à un tirage de 1000 boules le nombre x de blanches. On décide $K > H$ si $x \geq A$, où A est un nombre tel que, si en réalité on avait $K \leq H$ la probabilité de l'événement $(X \geq A)$ serait très petite" (ici, Gavarret prend $A = 900$).

En effet, dans **2.1.**, il est montré que, bien que x soit supérieur ou égal à 900, on ne peut exclure que $K \leq H$. Il faut seulement, dans ce cas, que la probabilité de se tromper en décidant $K > H$ soit très faible. Vérifions-le sur quelques exemples, où $K \leq H$.

2.2.2. Dans le cas d'un tirage **avec remplacement**, X suit une loi binomiale de paramètres $n = 1000$ et $p = \dfrac{K}{N}$.

$$\mathbb{P}(X = x) = \binom{1000}{x} p^x (1-p)^{1000-x}$$

Dans le cas d'un tirage **sans remplacement**, X suit une loi hypergéométrique :

$\mathbb{P}(X = x) = \dfrac{\dbinom{K}{x}\dbinom{H}{1000-x}}{\dbinom{N}{1000}}$ pour $0 \leq x \leq K$ et $0 \leq 1000 - x \leq H$, et $\mathbb{P}(X = x) = 0$ sinon.

- **Cas avec remplacement et $N = 3$, $K = 1$ et $H = 2$.**
  X suit une loi binomiale de paramètres $n = 1000$ et $p = \dfrac{1}{3}$. On utilise l'approximation normale c'est-à-dire $X \sim \mathcal{N}(np, \sqrt{np(1-p)})$.

  $$\mathbb{P}(X \geq 900) = 1 - \mathbb{P}(X < 900) = 1 - \mathbb{P}(X \leq 899) = 1 - \mathbb{P}\left(\frac{X - \frac{1000}{3}}{\sqrt{1000 \times \frac{1}{3} \times \frac{2}{3}}} \leq \frac{899 - \frac{1000}{3}}{\sqrt{1000 \times \frac{1}{3} \times \frac{2}{3}}}\right)$$

  $= 1 - \mathbb{P}(Z \leq 37{,}94) = 1 - F_Z(37{,}94)$ où F_Z est la fonction de répartition de la loi normale centrée réduite. Or $F_Z(37{,}94) \simeq 1$ donc $\mathbb{P}(X \geq 900) \simeq 0$

- **Cas avec remplacement et $N = 1000$, $K = 499$ et $H = 501$.**
  X suit une loi binomiale de paramètres $n = 1000$ et $p = \dfrac{499}{1000}$.
  Le même raisonnement montre que : $\mathbb{P}(X \geq 900) = 1 - F_Z(25{,}09) \simeq 0$.

- **Cas avec remplacement et $N = 2$, $K = 1$ et $H = 1$.** Ou encore $N = 1000$, $K = 500$ et $H = 500$ mais aussi $N = 2000$, $K = 1000$ et $H = 1000$.
  X suit une loi binomiale de paramètres $n = 1000$ et $p = 0{,}50$.
  Le même raisonnement montre que $\mathbb{P}(X \geq 900) = 1 - \mathbb{P}(Z \leq 25{,}23) \simeq 0$.

– **Cas sans remplacement avec** $N = 1801$**,** $K = 900$ **et** $H = 901$**.**

Ici $\mathbb{P}(X \geq 900) = \mathbb{P}(X = 900) = \dfrac{\dbinom{900}{900}\dbinom{901}{100}}{\dbinom{1801}{1000}} \simeq 0.$

– **Cas sans remplacement avec** $N = 2000$**,** $K = 999$ **et** $H = 1001$. Ou $N = 2000$, $K = 950$ et $H = 1050$, ou encore $N = 1800$, $K = 900$ et $H = 900$.
On a toujours : $\mathbb{P}(X \geq 900) \simeq 0$.

Commentaire : La question 2.2.1. vise à faire comprendre que dans un test où on souhaite décider entre $H_0 : K \leq H$ et $H_1 : K > H$, après l'observation de x blanches, la zone de rejet de l'hypothèse nulle est de la forme $X \geq A$ et non $X = A$, où A est un nombre "assez grand". Même si ici la forme de la zone de rejet paraît évidente, la détermination de celle-ci dans d'autres problèmes de test reste un sujet mathématiquement délicat. La question 2.2.2. permet de faire travailler sur les tables de la fonction de répartition de la loi normale centrée réduite et de montrer, en liaison avec le 2.1., qu'un événement de probabilité nulle n'est pas un événement impossible.

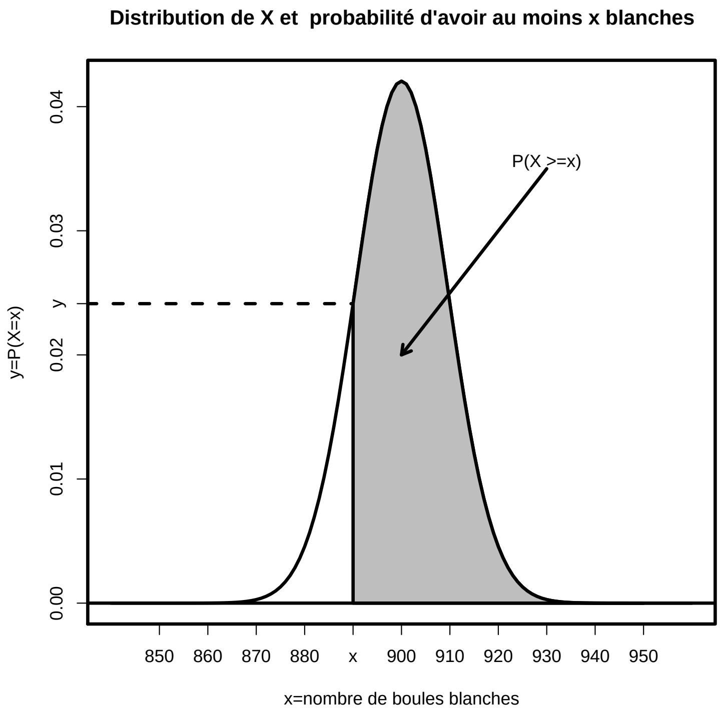
Retour au sommaire des activités 2 Retour à l'énoncé de l'activité

Indications sur l'activité 2.3

a) Soit X la variable aléatoire qui donne le nombre de blanches en supposant que la proportion de boules blanches dans l'urne est 90%. Alors X suit une loi binomiale de paramètre $n_2 = 1000$ et $p = 0{,}90$. Obtenir plus de blanches que de noires se traduit par la réalisation de $(X > 500)$ ou $(X \geq 501)$.

En utilisant l'approximation normale, on obtient :
$$\mathbb{P}(X \geq 501) = 1 - F_Z\left(\frac{500 - 1000 \times 0{,}90}{\sqrt{1000 \times 0{,}90 \times 0{,}10}}\right) = 1 - F_Z(-4{,}216) = 1 - 0{,}000\,012 = 0{,}999\,988.$$
Cette probabilité très proche de 1 semble donner raison à Gavarret.

b) Si la proportion de boules blanches dans l'urne est 90%, on a $\mathbb{P}(X \geq 950) \simeq 0$! ! ! ! Donc une probabilité presque nulle d'avoir "beaucoup plus de blanches que de noires" (si on considère que "plus de 950 " est équivalent à "beaucoup plus de blanches que de noires"). Ce paradoxe s'explique par le fait que, si la proportion de boules blanches dans l'urne est 90%, la distribution de X se concentre autour de son espérance c'est-à-dire 900 et la probabilité d'obtenir des valeurs éloignées de cette espérance est faible.



Commentaire : Puisque dans la suite de son texte, Gavarret semble considérer comme "absolument certain" [G, p. 257] tout événement dont la probabilité est supérieure ou égale à $0{,}9953$, Gavarret aurait raison d'affirmer qu'on obtiendrait plus de blanches que de noires. Mais peut-il dire qu' "on obtiendrait **beaucoup** plus de blanches que de noires" (c'est nous qui soulignons) ? C'est toute l'ambiguïté du mot "beaucoup" et qui se retrouvera plus loin dans le texte, et qui peut donner des résultats apparemment paradoxaux comme le montrera 2.4.5.

Retour au sommaire des activités 2 Retour à l'énoncé de l'activité

Indications sur l'activité 2.4

2.4.1. a) Déjà montré en **2.2.2.** : $\mathbb{P}(X \geq 900) = 1 - \mathbb{P}(Z \leq 25{,}23) \simeq 0 < 0{,}0047$, donc ce cas est "impossible au sens de Gavarret".

Calcul de la probabilité des événements suivants :

– "tirer au moins 546 blanches". $\mathbb{P}(X \geq 546) = 1 - F_Z\left(\dfrac{545 - 1000 \times 0{,}50}{\sqrt{1000 \times 0{,}50 \times 0{,}50}}\right) =$
$= 1 - \mathbb{P}(Z \leq 2{,}846) = 0{,}002\,213 < 0{,}0047$. Événement "impossible au sens de Gavarret" ;

– "tirer au moins 543 blanches". $\mathbb{P}(X \geq 543) = 1 - \mathbb{P}(Z \leq 2{,}656) = 0{,}00395 < 0{,}0047$.
Événement impossible "impossible au sens de Gavarret" ;

– "tirer au moins 542 blanches". $\mathbb{P}(X \geq 542) = 1 - \mathbb{P}(Z \leq 2{,}593) = 0{,}00475 > 0{,}0047$;

– " tirer au moins 541 blanches". $\mathbb{P}(X \geq 541) = 1 - \mathbb{P}(Z \leq 2{,}530) = 0{,}00570 > 0{,}0047$.

b) Il est aisé de démontrer que la fonction u définie ainsi :
$u(x) = \dfrac{(x-1) - 1000 \times 0{,}50}{\sqrt{1000 \times 0{,}50 \times 0{,}50}}$ est strictement croissante et puisque la fonction F_Z est aussi strictement croissante, la fonction g définie ainsi :
$g(x) = \mathbb{P}(X \geq x) = 1 - F_Z(u(x))$ est alors strictement décroissante.

Remarque : L'utilisation des tables montre que $F_Z(2{,}597) = 0{,}9953$ et donc que si Z est une loi normale centrée réduite, alors $\mathbb{P}(Z \geq 2{,}597) = 0{,}0047$. On peut trouver A en cherchant le plus petit entier x tel que $(x-1) - 1000 \times 0{,}50 \geq 2{,}597\sqrt{1000 \times 0{,}50 \times 0{,}50}$. Puisque $501 + 2{,}597\sqrt{1000 \times 0{,}50 \times 0{,}50} = 542{,}062$, on trouve bien $A = 543$.

Ainsi donc, si dans l'urne il y a autant de blanches que de noires, $A = 543$ est le plus petit des entiers x tels que, l'événement "on a tiré au moins x boules blanches à la suite de 1000 tirages" est impossible "au sens de Gavarret".

Commentaire : Ici, après avoir en 2.2.1. déterminé la forme : $X \geq A$ de la zone de rejet de l'hypothèse nulle, il s'agit de déterminer A de telle sorte que, quand si l'hypothèse nulle est vraie, alors la probabilité que le résultat de l'expérience tombe dans cette zone de rejet soit inférieure à $0{,}0047$. Ici le calcul explicite de A s'effectue en supposant que la composition de l'urne est telle que $K = H$, c'est-à-dire quand $p = \dfrac{K}{N} = 0{,}50$. Pour être rigoureux, il faudrait prouver aussi que la probabilité que le résultat de l'expérience tombe dans cette zone de rejet est inférieure à $0{,}0047$, non seulement quand $K = H$, mais aussi quand $K < H$ ("moins de blanches que de noires"). C'est l'objet de la question suivante.

2.4.2. a) Si la proportion de blanches dans l'urne est 49%, la probabilité de tirer au moins 543 blanches est égale à :

$$\mathbb{P}(X \leq 543) = 1 - F_Z\left(\frac{542 - 1000 \times 0{,}49}{\sqrt{1000 \times 0{,}49 \times (1 - 0{,}49)}}\right) = 1 - \mathbb{P}(Z \leq 3{,}289) = 0{,}000\,502 < 0{,}0047.$$

Si la proportion de blanches dans l'urne est 48%, la probabilité de tirer au moins 543 blanches est égale à :

$$\mathbb{P}(X \leq 543) = 1 - F_Z\left(\frac{542 - 1000 \times 0{,}48}{\sqrt{1000 \times 0{,}48 \times (1 - 0{,}48)}}\right) = 1 - \mathbb{P}(Z \leq 3{,}924) = 0{,}000\,043 < 0{,}0047.$$

$\mathbb{P}_{0,48}(X \geq 543) \leq \mathbb{P}_{0,49}(X \geq 543) \leq \mathbb{P}_{0,50}(X \geq 543) < 0{,}0047$.

b) On pose h la fonction définie par :
$h(p) = \mathbb{P}_p(X \geq 543) = 1 - F_Z\Big(\dfrac{542 - 1000 \times p}{\sqrt{1000 \times p \times (1-p)}}\Big)$.

Calcul de la dérivée :

$$h^{'}(p) = \frac{1000(542 - p(2 \times 542 - 1000))}{2 \times (1000p(1-p))^{\frac{3}{2}}} \frac{1}{\sqrt{2\pi}} e^{-\dfrac{(542-1000p)^2}{2 \times 1000p(1-p)}}$$

or $542 - p(2 \times 542 - 1000) > 0$ car $p < 1 < \dfrac{542}{2 \times 542 - 1000} = 6{,}452$, donc $h^{'}(p) > 0$ pour $0 < p < 1$ et la fonction h est croissante.

Remarque : Cette propriété est vraie pour tout A tel que $\dfrac{n}{2} \leq A \leq n$ car $\dfrac{A}{2A - n} \geq 1 \geq p$.

On peut donc généraliser.

Soit $0 < \alpha < 1$. On pose $\mathbb{P}_p(X \geq x)$ la probabilité d'avoir au moins x boules blanches parmi n tirées dans une urne contenant une proportion p de blanches et A le plus petit entier x tel que $\mathbb{P}_{0.50}(X \geq x) \leq \alpha$ alors si c'est l'hypothèse $K < H$ qui est vraie, on a donc encore $\mathbb{P}_p(X \geq A) < \mathbb{P}_{0{,}50}(X \geq A) \leq \alpha$.

c) Dans l'exemple de Gavarret, si l'hypothèse $K < H$ est vérifiée, $p < 0{,}50$ donc $P_p(X \geq 543) < P_{0{,}50}(X \geq 543) = 0{,}00395 < 0{,}0047$.

Commentaire : Le problème de test peut se ramener ici à celui-ci : choisir entre l'hypothèse nulle $H_0 : p \leq 0{,}50$ et l'hypothèse alternative $H_1 : p > 0{,}50$, avec un risque de se tromper $\alpha = 0{,}0047$ en décidant H_1 et un échantillon de taille $n = 1000$. On vient de montrer que pour calculer explicitement A, il suffisait de trouver le plus petit entier x tel que $\mathbb{P}_p(X \geq x) \leq \alpha$, mais en se restreignant à $p = 0{,}50$ c'est-à-dire la "frontière" entre les deux zones possibles pour le paramètre inconnu.

2.4.3. Montrons que les deux règles de décision sont équivalentes.

Si $x \geq 543$ alors $(X \geq x) \subset (X \geq 543)$ donc $\mathbb{P}_{0{,}50}(X \geq x) \leq \mathbb{P}_{0{,}50}(X \geq 543)$.

Or 543 a été choisi tel que $\mathbb{P}_{0{,}50}(X \geq 543) \leq \alpha$, donc $\mathbb{P}_{0{,}50}(X \geq x) \leq \alpha$.

Si $\mathbb{P}_{0{,}50}(X \geq x) \leq \alpha$, $x < 543$ est impossible car, par construction 543 est le plus petit entier x tel que $\mathbb{P}_{0{,}50}(X \geq x) \leq \alpha$. Donc $x \geq 543$.

Commentaire : Aujourd'hui, tous les logiciels et les calculettes utilisent la notion de **p-value** notée souvent **P** dans les calculettes. Ici, il s'agit de **p-value = P =** $\mathbb{P}_{0{,}50}(X \geq x)$ où x est le résultat (ici le nombre de blanches) obtenu après l'expérience. Pour conclure un test, il suffira de comparer cette **p-value** au risque de première espèce choisi α. Si **p-value** $\leq \alpha$, on rejette l'hypothèse nulle et si **p-value** $> \alpha$, on ne rejette pas l'hypothèse nulle.

Il est maintenant d'usage d'accompagner les conclusions d'un test de sa "signification" suivant les valeurs de la **p-value**.

	Résultat	Notation
$0{,}05 <$ **p-value**	"pas significatif"	
$0{,}01 \leq$ **p-value** $< 0{,}05$	"significatif"	*
$0{,}001 \leq$ **p-value** $< 0{,}01$	"très significatif"	**
p-value $< 0{,}001$	"hautement significatif"	***

2.4.4.

a) Calcul de $\beta(p) = \mathbb{P}_p(X < 543) = \mathbb{P}_p(X \leq 542)$ pour :

– Pour $p = 0{,}51$, $\beta(p) = \mathbb{P}_p(X < 543) = 0{,}978\,528$;
– Pour $p = 0{,}52$, $\beta(p) = \mathbb{P}_p(X < 543) = 0{,}918\,117$;
– Pour $p = 0{,}53$, $\beta(p) = \mathbb{P}_p(X < 543) = 0{,}776\,467$;
– Pour $p = 0{,}54$, $\beta(p) = \mathbb{P}_p(X < 543) = 0{,}550\,489$;
– Pour $p = 0{,}55$, $\beta(p) = \mathbb{P}_p(X < 543) = 0{,}305\,546$;
– Pour $p = 0{,}60$, $\beta(p) = \mathbb{P}_p(X < 543) = 0{,}000\,090$;

$\beta(p) = 1 - h(p)$ or h est croissante donc β est décroissante.

b) **Calcul avec un risque** $\alpha = 0{,}05$. Cas classique.

D'après 2.4.2., pour trouver A^* il suffit de trouver le plus petit x tel que $\mathbb{P}_{0,50}(X \geq x) < 0{,}05$ ou encore le plus petit x tel que $P_{0,50}(X \leq x-1) > 0{,}95$.

$$\mathbb{P}_{0,50}(X \leq x-1) = \mathbb{P}_{0,50}\left(\frac{X - 1000 \times 0.50}{\sqrt{1000 \times 0{,}50 \times 0{,}50}} \leq \frac{(x-1) - 1000 \times 0{,}50}{\sqrt{1000 \times 0{,}50 \times 0{,}50}}\right) = F_Z\left(\frac{(x-1) - 1000 \times 0{,}50}{\sqrt{1000 \times 0{,}50 \times 0{,}50}}\right)$$

En utilisant une table ou une calculette, on trouve $F_Z(1{,}65) = 0{,}95$.

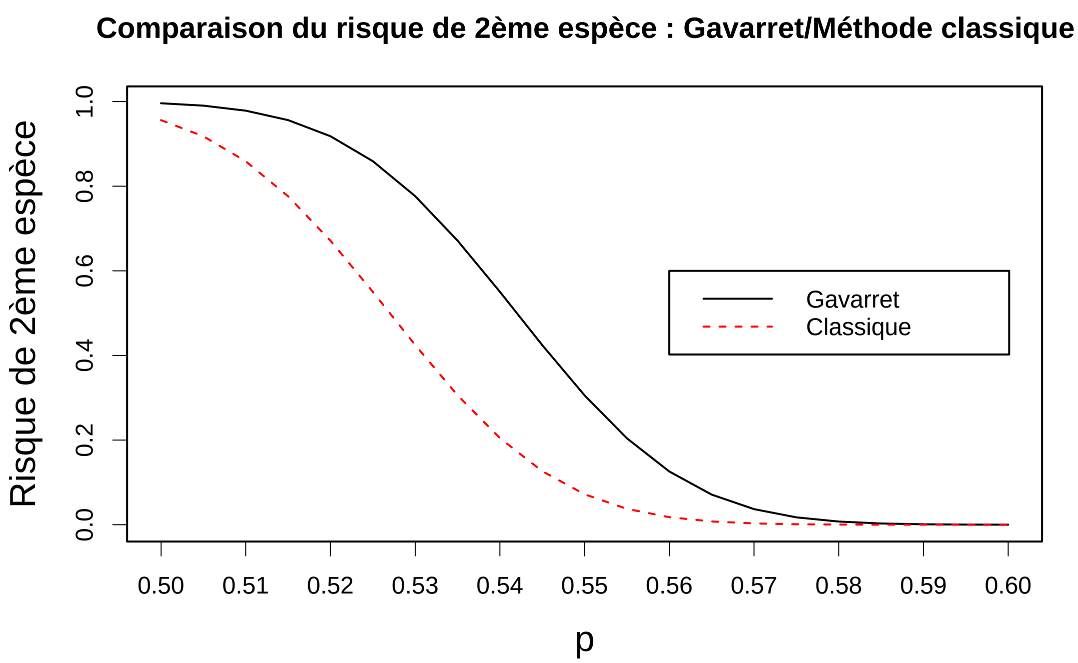
Il suffit alors de prendre pour A^* le plus petit entier x tel que : $\dfrac{(x-1) - 1000 \times 0{,}50}{\sqrt{1000 \times 0{,}50 \times 0.50}} > 1{,}65$.

On trouve : $A^* = 528$.

c) Calcul de $\beta^*(p) = \mathbb{P}_p(X < 528) = \mathbb{P}_p(X \leq 527)$.

– Pour $p = 0{,}51$, $\beta^*(p) = \mathbb{P}_p(X < 528) = 0{,}858\,900$;
– Pour $p = 0{,}52$, $\beta^*(p) = \mathbb{P}_p(X < 528) = 0{,}671\,144$;
– Pour $p = 0{,}53$, $\beta^*(p) = \mathbb{P}_p(X < 528) = 0{,}424\,623$;
– Pour $p = 0{,}54$, $\beta^*(p) = \mathbb{P}_p(X < 528) = 0{,}204\,732$;
– Pour $p = 0{,}55$, $\beta^*(p) = \mathbb{P}_p(X < 528) = 0{,}071\,874$;
– Pour $p = 0{,}60$, $\beta^*(p) = \mathbb{P}_p(X < 528) = 0{,}000\,001$.

On observe que pour $p = 0{,}51; 0{,}52; 0{,}53; 0{,}54; 0.55; 0.60$, on a $\beta^*(p) < \beta(p)$.
Puisque $(X < 528) \subset (X < 543)$, $\mathbb{P}_p(X < 528) < \mathbb{P}_p(X < 543)$ et ce résultat se généralise.



Commentaire :

1ère remarque : L'hypothèse nulle $H_0 : K \leq H$ correspond en fait à un ensemble possible de valeurs pour p qui est l'intervalle $[0\,;0{,}50]$. Pour un "risque de première espèce" donné, le "risque de deuxième espèce" est d'autant plus fort que la "vraie" valeur du paramètre p, tout en n'appartenant pas à cet ensemble, est "proche" de lui. C'est ce qu'avait bien perçu Gavarret quand il prend beaucoup de précautions pour énoncer la conclusion dans le cas où l'hypothèse nulle n'est pas rejetée (voir p. 23 de l'article et [G, p. 158]).

2ème remarque : Si on prend un risque $\alpha = 0{,}0047$, pour $p = 0{,}53$, le "risque de deuxième espèce" est égal à $\beta(0{,}53) = 0{,}776\,467$ alors que si on prend un risque $\alpha = 0{,}05 > 0{,}0047$, le "risque de deuxième espèce" est égal à $\beta^*(0{,}53) = 0{,}424\,623 < 0{,}776\,467 = \beta(0{,}53)$. De façon générale, pour une même valeur du paramètre inconnu, lorsque le "risque de première espèce" diminue, le "risque de deuxième espèce" augmente.

Il n'est pas malheureusement pas possible de "gagner sur les deux tableaux". Dans son ouvrage, Gavarret s'est limité à prendre de fait un "risque de première espèce" égal à $0{,}0047$ car il a conservé les formules de Poisson. Le "risque de seconde espèce" n'est jamais calculé mais Gavarret sent confusément que celui-ci n'est pas égal au "risque de première espèce" dans les conclusions qu'il donne à plusieurs de ses tests. Ce "risque de seconde espèce" ne pourrait l'être exactement que si on connaissait la véritable valeur de p, or celle-ci ne peut être qu'estimée.
Si ici, on l'estime par $\widehat{p} = \dfrac{900}{1000}$, il est facile de montrer que $\beta(\widehat{p}) \simeq 0$.

2.4.5. Soit $p = \dfrac{K}{K+H}$. On note x le nombre de blanches tirées.

Première interprétation : on teste $H_0 : p \leq 0{,}75$ contre $H_1 : p > 0{,}75$.

Règle de décision : on rejette $H_0 : p \leq 0{,}75$ si et seulement si $x \geq 765$ ou

Règle de décision : on rejette $H_0 : p \leq 0{,}75$ si et seulement si $P_{0{,}75}(X \geq x) < 0{,}0047$.

Ici, $x = 900$ donc on rejette $H_0 : p \leq 0{,}75$. Il y a "beaucoup plus de blanches que de noires". On évalue p par $\widehat{p} = 0{,}90$.

Le "risque de deuxième espèce" est évalué par : $\beta(0{,}90) = \mathbb{P}_{0{,}90}(X < 765) \simeq 0$

Deuxième interprétation : on teste $H_0 : p \leq 0{,}80$ contre $H_1 : p > 0{,}80$.

Règle de décision : on rejette $H_0 : p \leq 0{,}80$ si et seulement si $x \geq 814$ ou

Règle de décision : on rejette $H_0 : p \leq 0{,}80$ si et seulement si $P_{0{,}80}(X \geq x) < 0{,}0047$

Ici, $x = 900$ donc on rejette $H_0 : p \leq 0{,}80$. Il y a "beaucoup plus de blanches que de noires". On évalue p par $\widehat{p} = 0{,}90$.

Le "risque de deuxième espèce" est évalué par : $\beta(0{,}90) = \mathbb{P}_{0{,}90}(X < 814) \simeq 0$.

Troisième interprétation : on teste $H_0 : p \leq 0{,}90$ contre $H_1 : p > 0{,}90$.

Règle de décision : on rejette $H_0 : p \leq 0{,}80$ si et seulement si $x \geq 911$ ou

Règle de décision : on rejette $H_0 : p \leq 0{,}80$ si et seulement si $P_{0.90}(X \geq x) < 0{,}0047$

Ici, $x = 900$ donc on ne rejette pas $H_0 : p \leq 0{,}80$. On ne peut pas décider : "il y a beaucoup plus de blanches que de noires". On évalue p par $\widehat{p} = 0{,}90$.

Le "risque de deuxième espèce" est évalué par : $\beta(0{,}90) = P_{0.90}(X < 911) = 0{,}854$.

Ainsi donc, puisqu'on ne rejette pas H_0, on est tenté de décider qu' "il n'y a pas beaucoup plus de blanches que de noires". Mais dans ce cas on a $85{,}4\%$ de chances de se tromper.

Commentaire : Cette activité illustre la nécessité d'être précis dans la définition de l'alternative qui se pose au statisticien.

Retour au sommaire des activités 2

Retour à l'énoncé de l'activité

Indications sur l'activité 2.5

2.5.1 D'après la théorie des intervalles de confiance pour une proportion inconnue :
$\mathbb{P}(\widehat{p}-\epsilon < p < \widehat{p}+\epsilon) = 1-\alpha$ avec $\epsilon = z_\alpha\sqrt{\dfrac{\widehat{p}(1-\widehat{p})}{n}}$
où z_α tel que $\mathbb{P}(-z_\alpha \leq Z \leq z_\alpha) = 1-\alpha$. Ici pour $\alpha = 0{,}0047$ on a $z_\alpha \simeq 2\sqrt{2}$ donc :
$\epsilon = 2\sqrt{\dfrac{2\times 0{,}90\times(1-0{,}90)}{1000}} = 0{,}0268$ soit $2{,}68\%$.
L'intervalle de confiance pour p est donc $[0{,}8732\,;0{,}9268]$, ou, en pourcentage, $[87{,}32\%\,;92{,}68\%]$.

Commentaire : Il s'agit simplement du calcul d'un intervalle de confiance avec un niveau différent de celui utilisé dans les classes de B.T.S. et de lycée.

2.5.2. Puisque l'on a $\left(|p-\widehat{p}| \leq 2\sqrt{\dfrac{2p(1-p)}{n}}\right) \subset \left(|p-\widehat{p}| \leq \sqrt{\dfrac{2}{n}}\right)$ et
$\mathbb{P}((|p-\widehat{p}| \leq 2\sqrt{\dfrac{2p(1-p)}{n}}) = 1-0{,}0047$, alors $\mathbb{P}(|p-\widehat{p}| \leq \sqrt{\dfrac{2}{n}}) \geq 0{,}9953$.
Il suffit de prendre n tel que $\epsilon = \sqrt{\dfrac{2}{n}}$ d'où $n \simeq \dfrac{2}{\epsilon^2}$.

Commentaire : Il existe des méthodes plus fines pour optimiser la taille de l'échantillon connaissant l'"erreur" que l'on accepte entre la "réalité" et la proportion observée, mais celle-ci fournit une première estimation de cette taille. Si, avec le niveau de confiance implicite pris par Gavarret, on accepte une "erreur" de 1% autour du pourcentage observé, il faut effectuer un sondage sur $20\,000$ personnes !

Retour au sommaire des activités 2 Retour à l'énoncé de l'activité

Indications sur l'activité 2.6

2.6.1. $\binom{20}{20} = 1,\qquad \binom{20}{19} = 20,\qquad \binom{20}{18} = 190,\qquad \binom{20}{17} = 1140,\qquad \binom{20}{16} = 4845$

2.6.2.

- $\mathbb{P}_{0,50}(Y \geq 16) = 0{,}005\,91$, $\mathbb{P}_{0,50}(Y \geq 17) = 0{,}001\,29$, $\mathbb{P}_{0,50}(Y \geq 18) = 0{,}000\,20$, $\mathbb{P}_{0,50}(Y \geq 19) = 0{,}000\,02$, $\mathbb{P}_{0,50}(Y \geq 20) \simeq 0$
- $\mathbb{P}_{0,75}(Y \geq 16) = 0{,}414\,84$, $\mathbb{P}_{0,75}(Y \geq 17) = 0{,}225\,15$, $\mathbb{P}_{0,75}(Y \geq 18) = 0{,}091\,26$, $\mathbb{P}_{0,75}(Y \geq 19) = 0{,}024\,31$, $\mathbb{P}_{0,75}(Y \geq 20) = 0{,}003\,17$
- $\mathbb{P}_{0,80}(Y \geq 16) = 0{,}629\,64$ $\mathbb{P}_{0,80}(Y \geq 17) = 0{,}411\,45$, $\mathbb{P}_{0,80}(Y \geq 18) = 0{,}206\,08$, $\mathbb{P}_{0,80}(Y \geq 19) = 0{,}069\,17$, $\mathbb{P}_{0,80}(Y \geq 20) = 0{,}011\,53$
- $\mathbb{P}_{0,90}(Y \geq 16) = 0{,}956\,82$, $\mathbb{P}_{0,90}(Y \geq 17) = 0{,}867\,04$, $\mathbb{P}_{0,90}(Y \geq 18) = 0{,}676\,93$, $\mathbb{P}_{0,90}(Y \geq 19) = 0{,}391\,75$, $\mathbb{P}_{0,90}(Y \geq 20) = 0{,}121\,57$.

2.6.3.

a)

1er cas : *Règle de décision 1 :* on rejette l'hypothèse nulle (et donc on accepte l'hypothèse alternative "il y a plus de blanches que de noires") si et seulement si $y \geq 17$.

Règle de décision 2 : on rejette l'hypothèse nulle (et donc on accepte l'hypothèse alternative "il y a plus de blanches que de noires") si et seulement $\mathbb{P}_{0,50}(Y \geq y) \leq 0{,}0047$.

Puisque $y = 18$, la décision sera donc : "il y a plus de blanches que de noires".

Le "risque de deuxième espèce" pour ce test est évalué par $\mathbb{P}_{0,90}(Y < 17) = 0{,}132\,95$.

2ème cas : *Règle de décision 1 :* on rejette l'hypothèse nulle (et donc on accepte l'hypothèse alternative "il y a plus de blanches que de noires") si et seulement si $y = 20$.

Règle de décision 2 : on rejette l'hypothèse nulle (et donc on accepte l'hypothèse alternative "il y a plus de blanches que de noires") si et seulement $\mathbb{P}_{0,75}(Y \geq y) \leq 0{,}0047$.

Puisque $y = 18$, la décision sera donc : "il n'y a pas trois fois plus de blanches que de noires". Le "risque de deuxième espèce" pour ce test est évalué par $\mathbb{P}_{0,90}(Y < 20) = 0{,}878\,42$.

3ème cas : puisque $\mathbb{P}_{0,80}(Y = 20) = 0{,}011\,53 > 0{,}0047$, il n'est pas possible de construire un test ayant $\alpha = 0{,}0047$ comme "risque de première espèce".

4ème cas : puisque $\mathbb{P}_{0,90}(Y = 20) = 0{,}12157 > 0,0047$, il n'est pas possible de construire un test ayant $\alpha = 0{,}0047$ comme "risque de première espèce".

Dans tous les cas, que ce soit avec un tirage de 1000 boules ou avec un tirage de 20 boules, la proportion p est estimée par $0{,}90$.

b) Pour le premier cas, avec un tirage de 1000 boules, le "risque de deuxième espèce" est évalué par $\mathbb{P}_{0,90}(X < 543) \simeq 0$ alors qu'ici, pour le même cas, mais avec un tirage de 20 boules, ce risque est évalué à $\mathbb{P}_{0,90}(Y < 17) = 0{,}1329$.

Pour le deuxième cas, avec un tirage de 1000 boules, le "risque de deuxième espèce" est évalué par $\mathbb{P}_{0,90}(X < 765) \simeq 0$ (voir 2.4.4.) alors qu'ici, pour le même cas, mais avec un tirage de 20 boules, ce risque est évalué à $\mathbb{P}_{0,90}(Y < 20) = 0{,}8784$.

Cet exemple permet donc d'observer que le "risque de deuxième espèce" dépend de n. Plus n est grand, plus le "risque de deuxième espèce" diminue.

c) Soit le problème de test : $H_0 : K \leq H$ contre $H_1 : K > H$ avec un "risque de premier espèce" $\alpha = 0{,}0047$.

Posons $p_0 = \dfrac{K}{N}$ et soit A_n le plus petit entier y tel que $\mathbb{P}_{p_0}(Y \geq y) \leq \alpha$.

Rappelons que $F_Z(2{,}597) = 1 - 0{,}0047$. On a vu que $A_n \sim np_0 + 2{,}597\sqrt{np_0(1-p_0)} + 1$

Notons le "risque de deuxième espèce" $\beta_n(p) = \mathbb{P}_p(Y < A_n)$.

$$\begin{aligned}\beta_n(p) &= \mathbb{P}_p\left(Y \leq np_0 + 2{,}597\sqrt{np_0(1-p_0)}\right)\\ &= \mathbb{P}_p\left(\frac{Y-np}{\sqrt{np(1-p)}} \leq \frac{np_0 + 2{,}597\sqrt{np_0(1-p_0)} - np}{\sqrt{np(1-p)}}\right)\\ &= F_Z\left(\frac{\sqrt{n}(p_0-p)}{\sqrt{p(1-p)}} + 2{,}597\frac{\sqrt{p_0(1-p_0)}}{\sqrt{p(1-p)}}\right)\end{aligned}$$

Pour $p > p_0$, $\dfrac{\sqrt{n}(p_0-p)}{\sqrt{p(1-p)}} + 2{,}597\dfrac{\sqrt{p_0(1-p_0)}}{\sqrt{p(1-p)}}$ tend vers $-\infty$ quand n tend vers l'infini.

Donc $\beta_n(p)$ tend vers 0 quand n tend vers l'infini.

La puissance du test $l_n(p) = 1 - \beta_n(p)$, probabilité d'avoir confiance dans la décision quand celle-ci est le rejet de l'hypothèse nulle, converge donc vers 1 quand n tend vers l'infini.

2.6.4.

1er problème : Aujourd'hui, avec le risque $\alpha = 0{,}05$, puisque $x = 532 > 528$, on rejetterait l'hypothèse nulle donc, on conclurait : "il y a plus de blanches que de noires dans l'urne".

Parce que $x = 532 < 543$, avec le risque pris par Gavarret, on ne pourrait rejeter l'hypothèse nulle et on ne conclurait pas : "il y a plus de blanches que de noires dans l'urne".

En utilisant la méthode de la **p-value**, on obtiendrait la même conclusion.

En effet, ici **p-value** $= \mathbb{P}(X \geq 532) = 0{,}0249$. Puisque **p-value** $= 0{,}0249 < 0{,}05$, aujourd'hui, on rejetterait l'hypothèse nulle et, avec le risque pris par Gavarret, puisque **p-value**$=0{,}0249 > 0{,}0047$, on ne rejetterait pas l'hypothèse nulle.

Avec les conventions actuelles rappelées dans le commentaire du 2.4.3., on conclurait que le nombre de blanches dans l'urne est **significativement** plus élevé que le nombre de noires, car $0{,}01 < 0{,}0249 < 0{,}05$.

2ème problème : Soit T la variable aléatoire qui, à un tirage de 2000 boules dans une urne contenant un proportion p de blanches associe le nombre de boules blanches obtenu. $T \sim \mathcal{B}(2000, p)$.

Pour utiliser la deuxième règle de décision, il suffit de calculer $\mathbb{P}_{0{,}50}(T \geq 1064)$. On utilise l'approximation normale.

p-value $= \mathbb{P}_{0{,}50}(T \geq 1064) = 1 - \mathbb{P}_{0{,}50}(T < 1064) = 1 - \mathbb{P}_{0{,}50}(T \leq 1063)$

$$= 1 - F_Z\left(\frac{1063 - 2000 \times 0{,}50}{\sqrt{2000 \times 0{,}50 \times 0{,}50}}\right) = 1 - F_Z(2{,}817) = 0{,}0024$$

Puisque **p-value** < $0{,}05$, aujourd'hui on rejetterait l'hypothèse nulle donc, on conclurait : "il y a plus de blanches que de noires dans l'urne".

Puisque **p-value** < $0{,}0047$, avec le risque pris par Gavarret, on rejetterait l'hypothèse nulle donc on conclurait : "il y a plus de blanches que de noires dans l'urne". Avec les conventions actuelles rappelées dans le commentaire du 2.4.3., on conclurait que le nombre de blanches dans l'urne est **très significativement** plus élevé que le nombre de noires, car $0{,}001 < 0{,}0024 < 0{,}01$.

Première remarque importante :

Dans les deux cas, si on estime la valeur du paramètre inconnu p (composition de l'urne) par la proportion observée, celle-ci est la même dans les deux problèmes c'est-à-dire $\widehat{p} = 0{,}532 = 53{,}2\%$.

On constate donc que, dans les deux cas, cette valeur est aussi éloignée de l'ensemble des valeurs correspondant à l'hypothèse nulle soit $[0\ ; 0{,}50]$. Or dans le premier problème, on énonce une conclusion "**significative**" alors que dans le deuxième problème cette conclusion est "**très significative**".

A l'écoute de ces conclusions différentes, il est souvent entendu que ceci impliquerait qu'il pourrait y avoir beaucoup plus de blanches dans l'urne pour le deuxième cas que dans le premier cas. Ceci n'est pas exact.

Le qualificatif de "**significatif**" ou "**très significatif**" ne concerne que le niveau de confiance portée à la conclusion qui dépend de la taille de l'échantillon.

C'est pourquoi lorsque deux laboratoires différents constatent pour l'un un effet "**significatif**" et dans l'autre un effet "**très significatif**", avant de conclure que l'effet est meilleur pour le dernier laboratoire, il faut s'assurer que les tailles des échantillons observés sont identiques. Cette remarque vaut bien sûr à fortiori quand un des laboratoires déclare une absence d'effet et l'autre un effet "**significatif**".

Deuxième remarque :

Au lieu de raisonner sur x nombre de boules blanches tirées dans un tirage de n boules, il est aussi possible de le faire sur la proportion observée $\widehat{p} = \dfrac{x}{n}$.

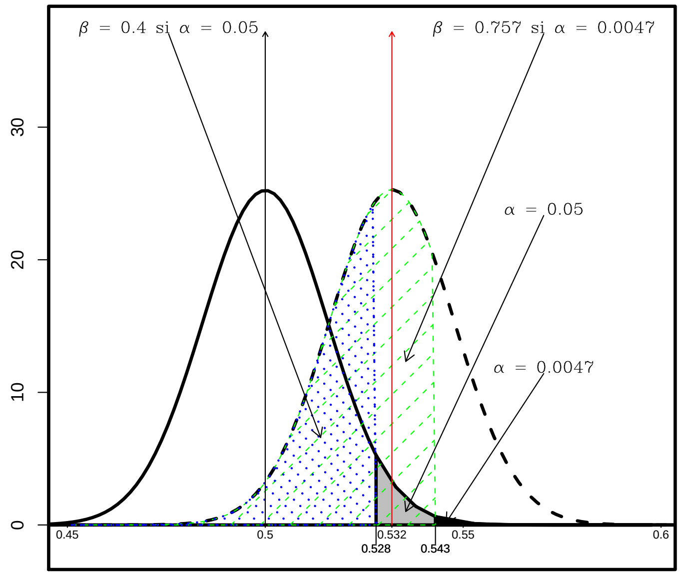
Pour connaître la proportion p^* telle que $\widehat{p} \geq p^*$ entraîne le rejet de $H_0 : p \leq 0{,}50$ avec une probabilité inférieure à α, il suffit de trouver A tel que $x \geq A$ entraîne le rejet de $H_0 : p \leq 0{,}50$ avec une probabilité inférieure à α.

Pour exemple, avec le risque $\alpha = 0{,}0047$ et un tirage de $n = 1000$ boules, on avait trouvé $A = 543$. Donc $p^* = 0{,}543$, et si la proportion observée est supérieure ou égale à $0{,}543$, la décision sera le rejet de $H_0 : p \leq 0{,}50$.

De même, avec le risque $\alpha = 0{,}0047$ et un tirage de $n = 2000$ boules, on trouve $A = 1060$. Donc $p^* = 0{,}530$, et si la proportion observée est supérieure ou égale à $0{,}530$, la décision sera le rejet de $H_0 : p \leq 0{,}50$.

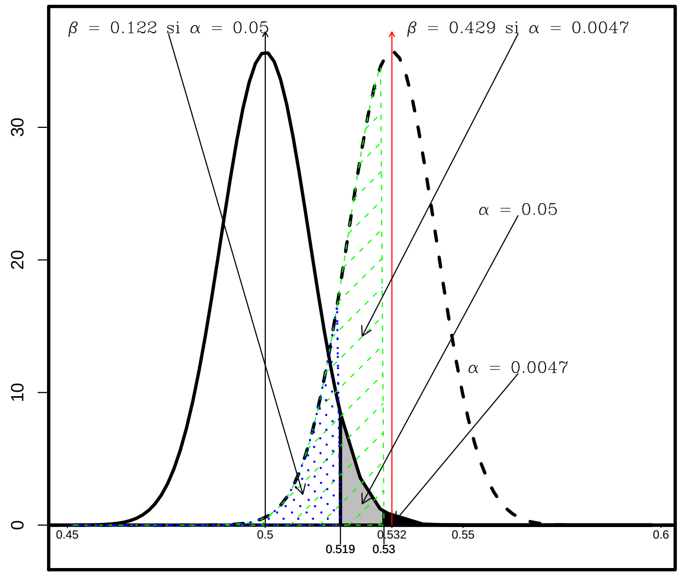
Les deux figures (à la même échelle) de la page suivante illustrent le cas où, avec des tailles d'échantillons différentes, mais avec une même proportion observée $\widehat{p} = 0{,}532$, les décisions peuvent être différentes suivant le "risque de première espèce" choisi et la taille de l'échantillon.

Test H0 : p<= 0.5 contre H1 : p> 0.5



Résultat : 532 blanches parmi 1000 boules tirées

Test H0 : p<= 0.5 contre H1 : p> 0.5



Résultat : 1064 blanches parmi 2000 boules tirées

	Décision "Gavarret"	Décision "aujourd'hui"
1000 boules	"$p \leq 0{,}50$" Risque de se tromper évalué à $0{,}757$	"$p > 0{,}50$" Risque de se tromper $\leq 0{,}05$
2000 boules	"$p > 0{,}50$" Risque de se tromper $\leq 0{,}0047$	"$p > 0{,}50$" Risque de se tromper $\leq 0{,}05$

Gavarret a bien sûr raison de ne pas se sentir éclairé par un nombre de tirages aussi faible mais son exigence d'une centaine d'observations doit être révisée.

2.6.5. : Pour montrer qu'il n'est pas équivalent d'effectuer le test $H_0 : K \leq H$ contre $H_1 : K > H$ et le test $H_0 : K \geq H$ contre $H_1 : K < H$, il faut déterminer la zone de rejet dans le cas de ce dernier test.

Pour choisir entre les deux hypothèses : $K \geq H$ et $K < H$, la règle de décision est implicitement la suivante : "Soit X la variable qui associe à un tirage de 1000 boules le nombre x de blanches. On décide $K < H$ si $x \leq A'$, où A' est un nombre tel que, si en réalité on avait $K \geq H$ la probabilité de l'événement $(X \leq A')$ serait très petite" (ici, inférieure à $0{,}0047$).

L'utilisation des tables montre que $F_Z(2{,}597) = 0{,}9953$. Puisque la symétrie de la densité de probabilité de Z, variable suivant une loi normale centrée réduite, entraîne $\mathbb{P}(Z \leq -2{,}597) = 0{,}0047$. On peut trouver A' en cherchant le plus grand entier x tel que $x - 1000 \times 0{,}50 \leq -2{,}597\sqrt{1000 \times 0{,}50 \times 0{,}50}$. Puisque $500 - 2{,}597\sqrt{1000 \times 0{,}50 \times 0{,}50} = 458{,}937$, on trouve $A' = 458$.

Pour ce nouveau problème de test, la règle de décision est donc la suivante :

On rejette $H_0 : K \geq H$ si $x \leq 458$.

Comparaison des décisions (entre guillemets) suivant le test et x nombre de boules blanches tirées dans une urne avec 1000 boules.
Rappel : Pour le test $H_0 : K \leq H$ contre $H_1 : K > H$, on rejette $H_0 : K \leq H$ si $x \geq 543$.

x	Test $H_0 : K \leq H$ contre $H_1 : K > H$	Test $H_0 : K \geq H$ contre $H_1 : K < H$
550	"$K > H$"	"$K \geq H$"
502	"$K \leq H$"	"$K \geq H$"
480	"$K \leq H$"	"$K \geq H$"
420	"$K \leq H$"	"$K < H$"

On remarque :

1. Pour un tirage de 502 boules blanches, les conclusions sont contradictoires. Il en est de même pour un tirage de 480 boules blanches.
2. Pour un tirage de 550 boules blanches, en décidant "$K > H$" avec le test $H_0 : K \leq H$ contre $H_1 : K > H$, le risque de se tromper est inférieur à $0{,}0047$ alors qu'avec le test $H_0 : K \geq H$ contre $H_1 : K < H$ le risque de se tromper est celui que l'on prend en ne rejetant pas l'hypothèse nulle (donc le risque de deuxième espèce que l'on ne peut pas contrôler).
3. Pour un tirage de 420 boules blanches, en décidant $H_0 : K \geq H$ contre $H_1 : K < H$ avec le test $H_0 : K \geq H$ contre $H_1 : K < H$, le risque de se tromper est inférieur à $0{,}0047$ alors qu'avec le test $H_0 : K \leq H$ contre $H_1 : K > H$ le risque de se tromper est celui que l'on prend en ne rejetant pas l'hypothèse nulle (donc le risque de deuxième espèce que l'on ne peut pas contrôler).

Les différences s'expliquent par l'absence de symétrie entre les deux hypothèses pour la construction de la zone de rejet. Entre deux hypothèses alternatives, il faut d'abord décider celle pour laquelle on veut contrôler le risque de se tromper en affirmant qu'elle est vraie. C'est cette hypothèse qui sera choisie comme hypothèse H_1 (et par conséquent l'autre sera choisie comme H_0).

Commentaire général sur les activités 2 :

Les conclusions tirées d'un test dépendent de plusieurs éléments dont entre autres :

- la taille de l'échantillon (voir 2.6.3.) ;
- la "distance" entre la valeur du paramètre inconnu et l'ensemble des valeurs du paramètres correspondant à l'hypothèse nulle. Plus cette distance est petite, plus le risque de se tromper en ne rejetant pas l'hypothèse nulle est grand (voir 2.4.4.) ;
- la précision de l'énoncé (voir 2.4.5.) ;
- le risque (choisi par l'expérimentateur) de se tromper en rejetant l'hypothèse nulle (voir 2.6.4.).

Gavarret avait très bien perçu l'importance des deux premiers éléments sans les formaliser mathématiquement. Quant à la construction d'intervalles de confiance, il a, dès le début de son ouvrage, posé les questions importantes qui sont encore celles posées aujourd'hui, en particulier par la pratique des sondages.

Retour au sommaire des activités 2 Retour à l'énoncé de l'activité

Indications sur l'activité 3.

On pose $p_\alpha = \frac{K_\alpha}{K_\alpha + H_\alpha}$.

La probabilité de l'événement : "l'individu numéro α meurt de la pathologie" est bien sûr p_α.

On note par p la probabilité de l'événement E : "un individu tiré au hasard dans la population meurt de la pathologie". Soit T la variable aléatoire qui associe au tirage au hasard le numéro de l'individu tirée.

D'après la formule des probabilités totales,
$\mathbb{P}(E) = \mathbb{P}(T = 1)\mathbb{P}_{(T=1)}(E) + \mathbb{P}(T = 2)\mathbb{P}_{(T=2)}(E) + \cdots + \mathbb{P}(T = \alpha)\mathbb{P}_{(T=\alpha)}(E) +$
$\cdots + \mathbb{P}(T = N)\mathbb{P}_{(T=N)}(E)$. Mais $\mathbb{P}_{(T=\alpha)}(E) = p_\alpha$ donc :

$$p = \frac{1}{N}\sum_{\alpha=1}^{N} p_\alpha.$$

Commentaire : Outre l'utilisation de la formule des probabilités totales, l'objectif est de faire comprendre la différence entre l'examen d'un malade particulier numéro α et l'examen d'un malade pris au hasard dans une population. Dans le deuxième cas, l'événement "mort de l'individu" est le résultat de deux hasards, d'abord celui résultant du tirage au hasard de l'individu dans la population et ensuite celui résultant de la "chance" de mourir de la pathologie pour cet individu, "chance" qui peut être différente suivant le malade.

Retour à l'article   Retour à l'énoncé de l'activité

Indications sur l'activité 4.

Série initiale	En ajoutant la série 1	En ajoutant la série 2
12 morts sur 30 malades	0,200	0,500
120 morts sur 300 malades	0,384	0,413
240 morts sur 600 malades	0,392	0,406
360 morts sur 900 malades	0,394	0,404
480 morts sur 1200 malades	0,396	0,403
600 morts sur 1500 malades	0,397	0,402
720 morts sur 1800 malades	0,397	0,402
840 morts sur 2100 malades	0,398	0,401

Le résultat pour la ligne $l \geq 2$ est donné :

– pour la première colonne par : $\dfrac{120(l-1)+3}{300(l-1)+20}$

– pour la deuxième colonne par : $\dfrac{120(l-1)+8}{300(l-1)+10}$

Il est aisé de montrer que chacun de ces résultats tend vers 0,40 quand l tend vers l'infini.

Commentaire : Il s'agit simplement de ne pas se contenter d'observer le rétrécissement de l'intervalle quand l'échantillon devient grand mais de le démontrer.

Retour à l'article

Retour à l'énoncé de l'activité

Indications sur l'activité 5.

Enoncé des résultats avec un niveau de confiance de $0{,}95$ pour chacun des trois cas.

1. L'erreur est égale à $0{,}0010$ d'où :

   Gavarret aurait conclu : au lieu de la proposition précédente, nous devons réellement admettre celle-ci :

   Sur $10\,000$ naissances, le nombre des garçons varie entre $\begin{cases} 5\,185 \\ 5\,165 \end{cases}$

2. L'erreur est égale à $0{,}0072$ d'où :

   Il aurait conclu : au lieu de la proposition précédente, nous devons réellement admettre celle-ci :

   Sur $10\,000$ naissances, le nombre des garçons varie entre $\begin{cases} 5\,140 \\ 4\,996 \end{cases}$

3. L'erreur est égale à $0{,}0102$ d'où :

   Il aurait conclu : au lieu de la proposition précédente, nous devons réellement admettre celle-ci :

   Sur $10\,000$ naissances, le nombre des garçons varie entre $\begin{cases} 5\,147 \\ 4\,943 \end{cases}$

On remarque que plus le degré de confiance diminue (on passe de $0{,}9953$ à $0{,}95$), plus la longueur de l'intervalle diminue et donc plus la précision augmente.

Avec la méthode des intervalles de confiance avec un niveau de confiance égal à $0{,}9953$, dans le premier exemple, l'hypothèse nulle aurait été rejetée donc on aurait énoncé la proposition suivante : "le nombre de garçons qui naissent est différent de celui du nombre de filles qui naissent". La probabilité de se tromper en énonçant ce résultat est inférieure à $0{,}0047$.

Avec la méthode des intervalles de confiance de niveau de confiance égal à $0,9953$, dans le deuxième et dans le troisième exemple, l'hypothèse nulle n'aurait pas été rejetée donc on aurait pu énoncer la proposition suivante : "lorsqu'un enfant naît, il y a autant de chance que ce soit un garçon qu'une fille" mais avec un "risque de deuxième espèce" qu'il convient d'estimer.

Ces conclusions restent les mêmes dans le cas d'un "risque de premier espèce" égal à $\alpha = 0{,}05$ (ou un niveau de confiance égal à $0{,}95$).

Commentaire : Ici encore, il s'agit de faire comprendre combien le niveau de confiance de l'intervalle est déterminant pour la construction de l'intervalle. Une question pourrait être posée : quel intervalle pour un niveau de confiance égal à 1 ?

Retour à l'article

Retour à l'énoncé de l'activité

Indications sur l'activité 6.

$\widehat{p_1} = 0{,}516\,431, \quad \widehat{p_2} = 0{,}517\,526$ donc $d = |\widehat{p_1} - \widehat{p_2}| = 0{,}001\,095.$

$$l = 2\sqrt{\frac{2 \times 417\,490 \times 441\,488}{912\,978^3} + \frac{2 \times 468\,151 \times 436\,443}{904\,594^3}} = 0{,}002\,036.$$

Donc $d < l$. La différence reste bien inférieure à la limite, ce qui justifie la décision de Gavarret.

Commentaire : Ici, il s'agit simplement de faire manipuler la formule de Poisson.

Retour à l'article Retour à l'énoncé de l'activité

Indications sur l'activité 7.

7.1. Il faut bien sûr que $b > 0$ pour que f soit croissante.

Deux exemples qui ne permettent pas une telle modélisation :

Exemple 1 : $a = 0$ et $b = 0{,}2$. Pour $x > 5$, $f(x) > 1$. Or une probabilité, par définition, ne peut être strictement supérieure à 1. Donc cette modélisation ne convient pas.

Exemple 2 : $a = -0{,}5$ et $b = 0{,}1$. Pour $x < 5$, $f(x) < 0$. Or une probabilité, par définition, ne peut être strictement négative. Donc cette modélisation ne convient pas.

7.2. a) Soit f définie par : $f(x) = \dfrac{e^{a+bx}}{1+e^{a+bx}}$

Une étude de la fonction montre que $0 < f(x) < 1$ et que, si $b > 0$, cette fonction est croissante.

Soit f définie par : $f(x) = F_Z(a + bx)$ où F_Z est la fonction de répartition de la loi normale réduite centrée.

Puisque $0 < F_Z(z) < 0$ pour tout z réel, alors $0 < f(x) < 1$. F_Z est croissante donc f aussi.

Il suffit de calculer h fonction réciproque de g où $g(u) = \dfrac{e^u}{1+e^u}$. On trouve facilement h par :

$h(v) = \ln\left(\dfrac{v}{1-v}\right)$. Or $f(x) = g(a + bx)$ donc si $y = f(x)$ alors $a + bx = \ln\left(\dfrac{y}{1-y}\right)$.

D'où la relation :

$$a + bx = \ln\left(\frac{p_x}{1-p_x}\right)$$

b) Ici on estime p_{x_1} par $\widehat{p_{x_1}}$ et p_{x_2} par $\widehat{p_{x_2}}$.

D'où 2 équations à 2 inconnues a et b :

$$\begin{cases} a + bx_1 = \ln\left(\dfrac{\widehat{p_{x_1}}}{1-\widehat{p_{x_1}}}\right) \\ a + bx_2 = \ln\left(\dfrac{\widehat{p_{x_2}}}{1-\widehat{p_{x_2}}}\right) \end{cases}$$

c) En appliquant à $x_1 = log_{10}(2) = 0{,}30103$ et $x_2 = log_{10}(3) = 0{,}477\,121$, on obtient $\widehat{p_{x_1}} = \dfrac{12}{20} = 0{,}60$ et $\widehat{p_{x_2}} = \dfrac{18}{25} = 0{,}72$. Et ensuite : $a = -0{,}516$ et $b = 3{,}061$.

D'où les relations : $p_x = \dfrac{e^{-0{,}516+3{,}061x}}{1+e^{-0{,}516+3{,}061x}}$ et $-0{,}516 + 3{,}061x = \ln\left(\dfrac{p_x}{1-p_x}\right)$.

En remplaçant p_x par $0{,}80$, on obtient $x = 0{,}621\,461$ d'où la dose = $10^{0{,}621\,461} = 4{,}18mg$.

d) Gavarret aurait certainement contesté ce résultat au vu du peu de malades testés pour chacune des doses.

7.3. Dans le cas d'une suite de données : $\widehat{p_{x_1}}$, $\widehat{p_{x_2}}$,...,$\widehat{p_{x_k}}$ pour k log-doses différentes x_1, x_2,...,x_k, une technique souvent utilisée est celle de du calcul de a et b à l'aide d'une régression linéaire simple, mais cette technique comporte des défauts, en particulier celui de provoquer un biais sur l'estimation de a et de b.

Commentaire : L'objectif visé par cette activité, outre l'étude de fonctions et la résolution d'un système de deux équations à deux inconnues, est surtout de montrer que le problème fondamental en médecine est de relier la probabilité d'apparition d'un phénomène à un facteur que l'on peut contrôler ou non (ici la dose administrée). Ceci nécessite une modélisation qui n'est pas unique.

Retour à l'article

Retour à l'énoncé de l'activité

Indications sur l'activité 8.

Mortalité moyenne fournie par la statistique	STATISTIQUES de 950 cas	
	Répartition des malades	Erreur possible
0,340 000	323 morts 627 guéris	0,043 471

Gavarret aurait conclu : il en résulte que la véritable mortalité cherchée est comprise entre les nombres : $0{,}340\,000$ plus $0{,}043\,471$, c'est-à-dire $0{,}383\,471$ et $0{,}340\,000$ moins $0{,}043\,471$, c'est-à-dire $0{,}296\,529$.

Commentaire : Ici, il ne s'agit que de faire comprendre le fonctionnement de la table et sa construction.

Retour à l'article

Retour à l'énoncé de l'activité

Indications sur l'activité 9.

Calcul de l avec la méthode actuelle.

– Pour la 1ère version, $l = 0{,}053\,231$. On a toujours $0{,}06 = d > l$, d'où la même conclusion que Gavarret.
– Pour la 2ème version, $l = 0{,}053\,231$. On a toujours $0{,}02 = d < l$, d'où la même conclusion que Gavarret.

On obtient la même limite car dans les deux cas, les tailles des deux échantillons restent les mêmes et la proportion commune p^* est estimée par la même quantité : $p^* = \dfrac{460}{1540} = 0{,}298\,701$.

> *Commentaire :* La raison principale qui amène aujourd'hui à utiliser la formule $l = 2\sqrt{\left(\frac{1}{\mu} + \frac{1}{\mu'}\right)\left(\frac{2m^*n^*}{\mu^{*2}}\right)}$ plutôt que celle de Poisson est la suivante. On a vu dans l'activité 2 que la construction d'un test s'opérait en considérant au départ que l'hypothèse nulle était vraie. Or dans le test $H_0 : p_1 = p_2$ contre $H_1 : p_2 \neq p_2$, sous l'hypothèse nulle, $p_1 = p_2$ donc il est préférable d'utiliser cette information et plutôt que d'estimer séparement p_1 et p_2, il vaut mieux estimer la probabilité commune p^*. On peut démontrer mathématiquement que cette méthode est plus "puissante" que celle de Poisson car le "risque de deuxième espèce" est plus petit avec la nouvelle formule qu'avec la formule de Poisson.

Retour à l'article Retour à l'énoncé de l'activité

Indications sur l'activité 10.

On calcule les intervalles de confiance donc de niveau $0{,}9953$ pour p_1 probabilité de mort avec la 1ère médication et pour p_2 probabilité de mort avec la 2ème médication, grâce à la formule de Poisson.

- Pour la 1ère version, l'intervalle de confiance pour p_1 est $[0{,}1642\ ; 0{,}2358]$ et l'intervalle de confiance pour p_2 est $[0{,}2208\ ; 0{,}2992]$. Ces intervalles se chevauchent, donc d'après la règle de décision, on n'aurait pas repoussé l'hypothèse nulle, **contrairement** à la décision de Gavarret.
- Pour la 2ème version, l'intervalle de confiance pour p_1 est $[0{,}1829\ ; 0{,}2570]$ et l'intervalle de confiance pour p_2 est $[0{,}2018\ ; 0{,}2782]$. Ces intervalles se chevauchent, donc d'après la règle de décision, on n'aurait pas repoussé l'hypothèse nulle, **conformément** à la décision de Gavarret.
- Pour la 3ème version, l'intervalle de confiance pour p_1 est $[0{,}2113\ ; 0{,}2887]$ et l'intervalle de confiance pour p_2 est $[0{,}1549\ ; 0{,}2251]$. Ces intervalles se chevauchent, donc d'après la règle de décision, on n'aurait pas repoussé l'hypothèse nulle, **contrairement** à la décision de Gavarret.

Commentaire : Cette activité permet de "tordre le cou" à une pratique souvent répandue et qui n'a pas de justification mathématique.

Retour à l'article Retour à l'énoncé de l'activité

Indications sur l'activité 11.

Dans les deux situations, la "chance moyenne" est égale à $p = 0{,}80$ avant et elle est égale à $0{,}60$ après l'application de la mesure donc un gain de $0{,}20$. Dans la deuxième situation, la probabilité de contracter la maladie a baissé de $0{,}20$ pour chaque individu ce qui n'est pas le cas dans la première situation.

Cas général.

Soit $p = \dfrac{1}{N}\sum_{\alpha=1}^{N} p_\alpha$ la "chance moyenne" de contracter la maladie avant la mesure.

Soit $p^* = \dfrac{1}{N}\sum_{\alpha=1}^{N} p^*_\alpha$ la "chance moyenne" de contracter la maladie après la mesure.

Le gain en terme de chance moyenne est égal à $\dfrac{1}{N}\sum_{\alpha=1}^{N} e_\alpha$.

Ce gain peut cacher des disparités importantes entre individus. Or, la position du médecin qui se trouve devant l'individu α est de faire baisser la probabilité qu'il contracte la maladie. Pour savoir si une mesure est capable d'assurer un gain individuel qui soit égal au gain global sur toute la population, il est nécessaire d'introduire dans l'analyse un modèle dit "modèle mixte" où interviennent au moins deux facteurs de variabilité : variabilité des probabilités entre individus et, pour un individu donné, variabilité du résultat "a contracté la maladie" suivant une loi déterminée par la probabilité affectée à l'individu.

Si pour chaque individu $e_\alpha = e$, alors le gain global sera égal à e.

Le modèle $e_\alpha = e$ est mis en défaut s'il existe α tel que $p_\alpha - e < 0$. Puisque pour cet individu α, on a $p^*_\alpha = p_\alpha - e$ et donc $p^*_\alpha < 0$. Or une probabilité ne peut être négative !

Commentaire : Cette activité peut être une des explications au peu d'intérêt accordé à l'époque par les médecins aux travaux de Gavarret. En effet, un bénéfice en terme de santé publique ne se répercute pas automatiquement de la même façon sur chacun des individus concernés. Gavarret avait prévenu l'objection en disant que la statistique ne pouvait prédire un événement pour un individu particulier.

Retour à l'article Retour à l'énoncé de l'activité

Indications sur l'activité 12.

En effet, si $0{,}5-\sqrt{\frac{2}{\mu}} \leq \frac{n_1}{n_1+n_2} \leq 0{,}5+\sqrt{\frac{2}{\mu}}$, alors $1-(0{,}5-\sqrt{\frac{2}{\mu}}) \geq 1-\frac{n_1}{n_1+n_2} \geq 1-(0{,}5+\sqrt{\frac{2}{\mu}})$ d'où $0{,}5-\sqrt{\frac{2}{\mu}} \leq \frac{n_2}{n_1+n_2} \leq 0{,}5+\sqrt{\frac{2}{\mu}}$.

Commentaire : Cette "indifférence" ne vaut ici que dans le cas où on veut tester $H_0 : p = 0{,}50$ contre $H_1 : p \neq 0{,}50$ mais dans des cas où la valeur de référence n'est plus $0{,}50$, des précautions doivent être prises.

Retour à l'article

Retour à l'énoncé de l'activité

Indications sur l'activité 13.

Soit Y la variable aléatoire qui associe à un individu, le jour de la semaine où il est rentré à l'hôpital.

Soit $J_1 \subset \{1,2,3,4,5,6,7\}$, un sous-ensemble de k_1 jours et $J_2 \subset \{1,2,3,4,5,6,7\}$ un sous-ensemble de $k_2 = 7 - k_1$ jours avec J_1 et J_2 disjoints.

Si le résultat est uniquement dû au hasard, alors $\mathbb{P}(Y \in J_1) = \dfrac{k_1}{7}$. Si le résultat n'est pas uniquement dû au hasard, alors $\mathbb{P}(Y \in J_1) \neq \dfrac{k_1}{7}$. Si $J = J_1 \cup J_2 = \{1,2,3,4,5,6,7\}$ numérote les jours de la semaine, alors en prenant $J_1 = \{1,2,4,5\}$ et en posant $p = \mathbb{P}(Y \in J_1)$, le problème de test est donc :

$$H_0 : p = 4/7 \text{ contre } H_1 : p \neq 4/7$$

On peut utiliser le test énoncé par Gavarret [G, p. 289-290] (voir article p. 28-29).

Le test utilisant l'artifice de se ramener au même nombre de jours n'est pas équivalent à celui qui devrait être utilisé car

$$0{,}50 - \sqrt{\frac{2}{\frac{3n_1}{4} + n_2}} \leq \frac{\frac{3n_1}{4}}{\frac{3n_1}{4} + n_2} \leq 0{,}50 + \sqrt{\frac{2}{\frac{3n_1}{4} + n_2}}$$

n'est pas équivalent à

$$\frac{4}{7} - 2\sqrt{\frac{2 \times \frac{4}{7}\frac{3}{7}}{n_1 + n_2}} \leq \frac{n_1}{n_1 + n_2} \leq \frac{4}{7} + 2\sqrt{\frac{2 \times \frac{4}{7}\frac{3}{7}}{n_1 + n_2}}$$

La démonstration n'est pas immédiate. On pose $n = n_1 + n_2$ et $x = n_1$. L'artifice de Gavarret consiste, dans la formule de Poisson, à remplacer n_1 par $\frac{3}{4}n_1 = \frac{3}{4}x$ et n par $n' = \frac{3}{4}n_1 + n_2 = \frac{3}{4}x + (n - x) = n - \frac{x}{4}$.

Problème posé :

Existe-t-il $n \in \mathbb{N}$ et $x \in \mathbb{N}$ avec $x \leq n$ tels que les deux propositions suivantes ne soient pas simultanément vraies ?

Proposition 1 :

$$0{,}50 - \sqrt{\frac{2}{n - \frac{x}{4}}} \leq \frac{\frac{3x}{4}}{n - \frac{x}{4}} \leq 0{,}50 + \sqrt{\frac{2}{n - \frac{x}{4}}} \qquad (1)$$

Proposition 2 :

$$\frac{4}{7} - 2\sqrt{\frac{2 \times \frac{4}{7}\frac{3}{7}}{n}} \leq \frac{x}{n} \leq \frac{4}{7} + 2\sqrt{\frac{2 \times \frac{4}{7}\frac{3}{7}}{n}} \qquad (2)$$

Résolution du problème :

Dans la suite de la démonstration, nous poserons $c = (\frac{4}{7})^2 = \frac{16}{49}$.

Remarquons d'abord que, en multipliant les termes des inégalités de la proposition 2 par n et en posant :

$I_2^{\pm} = \dfrac{4n}{7} \pm \sqrt{6cn}$,

la proposition 2 est équivalente à :

$x \in I_2 = [I_2^- \,; I_2^+]$ (3)

Démontrons d'abord que la proposition 1 est équivalente à dire que $x \in I_1 = [I_1^- \,; I_1^+]$, où I_1^- et I_1^+ seront des réels à déterminer.

(1) est équivalent à :

$$\frac{1}{2}-\sqrt{\frac{8}{4n-x}} \leq \frac{3x}{4n-x} \leq \frac{1}{2}+\sqrt{\frac{8}{4n-x}} \tag{4}$$

ou encore à :

$$\left|\frac{1}{2}-\frac{3x}{4n-x}\right| \leq \sqrt{\frac{8}{4n-x}} \tag{5}$$

En élevant au carré :

$$\left(\frac{1}{2}-\frac{3x}{4n-x}\right)^2 \leq \frac{8}{4n-x} \tag{6}$$

En réduisant au même dénominateur, on obtient l'inéquation :

$$49x^2+32x-56nx-128n+16n^2 \leq 0 \tag{7}$$

Le calcul du discriminant réduit Δ' donne :

$\Delta' = \sqrt{(16-28n)^2-49(16n^2-128n)} = \sqrt{5356n+256}$ d'où les 2 racines du polynôme en x :

$\alpha_1 = \frac{-(16-28n)-\sqrt{5356n+256}}{49}$ et $\alpha_2 = \frac{-(16-28n)-\sqrt{5356n+256}}{49}$

En simplifiant et en utilisant $c = \frac{16}{49}$, l'inéquation (6) est équivalente à :

$$\frac{4n}{7}-c-c\sqrt{1+21n} \leq x \leq \frac{4n}{7}-c+c\sqrt{1+21n} \tag{8}$$

En posant :

$I_1^{\pm} = \dfrac{4n}{7}-c \pm c\sqrt{1+21n},$

on prouve donc que la proposition 1 est équivalente à

$$x \in I_1 = [I_1^- \,; I_1^+] \tag{9}$$

Pour démontrer qu'il existe $n \in \mathbb{N}$ et $x \in \mathbb{N}$ $(x \leq n)$ ne vérifiant pas simultanément les deux propositions, il suffit de trouver $n \in \mathbb{N}$ et $x \in \mathbb{N}$ $(x \leq n)$ tels que $x \in I_1$ et $x \notin I_2$ soit vrai, ou bien $x \notin I_1$ et $x \in I_2$ soit vrai.

Remarque 1 : Soient a et b deux nombres réels avec $a < b$. Si $b-a > 1$, il existe $x \in \mathbb{N}$ tel que $a \leq x < b$.

Il suffit de prendre $x = \lfloor b \rfloor$ où $\lfloor b \rfloor$ désigne la partie entière de b.

Remarque 2 : $(\frac{4n}{7}-\sqrt{6cn})-(\frac{4n}{7}-c-c\sqrt{1+21n}) > 0$.

Montrons d'abord $c\sqrt{1+21n}-\sqrt{6cn} > 0$. En effet, $\dfrac{c\sqrt{21}}{\sqrt{6c}} = \sqrt{\dfrac{3\times 7\times 4\times 4}{2\times 3\times 7\times 7}} = \sqrt{\dfrac{8}{7}} > 1$.

On en déduit : $c\sqrt{21} > \sqrt{6c}$ et $c\sqrt{1+21n}-\sqrt{6cn} > c\sqrt{21n}-\sqrt{6cn} = \sqrt{n}(c\sqrt{21}-\sqrt{6c}) > 0$.

Il suffit de remarquer que cette dernière inégalité entraîne :

$(\frac{4n}{7}-\sqrt{6cn})-(\frac{4n}{7}-c-c\sqrt{1+21n}) = c+(c\sqrt{1+21n}-\sqrt{6cn}) > c > 0$.

En prenant $b = I_2^- = \dfrac{4n}{7}-\sqrt{6cn}$ et $a = I_1^- = \dfrac{4n}{7}-c-c\sqrt{1+21n}$, la remarque 1 amène à chercher $n \in \mathbb{N}$ tel que :

$(\dfrac{4n}{7}-\sqrt{6cn})-(\dfrac{4n}{7}-c-c\sqrt{1+21n}) > 1$ ou encore tel que :

$$c+c\sqrt{1+21n}-\sqrt{6cn}-1 > 0 \tag{10}$$

Mais, puisque $c+c\sqrt{1+21n}-\sqrt{6cn}-1 > c+c\sqrt{21n}-\sqrt{6cn}-1$, il suffira de trouver $n \in \mathbb{N}$ tel que :

$c+c\sqrt{21n}-\sqrt{6cn}-1 > 0.$

En prenant :

$\sqrt{n} > \dfrac{1-c}{c\sqrt{21}-\sqrt{6c}}$ c'est-à-dire $n > 48{,}56199$ ou encore $n \geq 49$, on obtient le résultat recherché.

Pour $n = 49$, $I_1 = [17{,}19\,;38{,}15]$ et $I_2 = [18{,}20\,;37{,}79]$. Il suffit de prendre $x = 18$ car alors $x \in I_1$ mais $x \notin I_2$. Dans ce cas, ceci revient donc à prendre $n_1 = 18$ et $n_2 = n - x = 31$.

La valeur de la fonction de test utilisée par Gavarret, pour ces valeurs de n_1 et de n_2, est égale à : $\varphi = \dfrac{\frac{3n_1}{4}}{\frac{3n_1}{4}+n_2} = 0{,}3033$. Puisque $\varphi \in [0{,}50 - \sqrt{\dfrac{2}{\frac{3n_1}{4}+n_2}}\,;0{,}50 + \sqrt{\dfrac{2}{\frac{3n_1}{4}+n_2}}] = [0{,}288\,;0{,}712]$, l'hypothèse nulle est acceptée avec ce test.

La valeur de la fonction de test dans le cas du test adéquat, pour ces valeurs de n_1 et de n_2, est égale à : $\varphi = \dfrac{n_1}{n_1+n_2} = 0{,}367$. Puisque $\varphi \notin [\dfrac{4}{7} - 2\sqrt{\dfrac{2\times\frac{4}{7}\frac{3}{7}}{n}}\,;\dfrac{4}{7} + 2\sqrt{\dfrac{2\times\frac{4}{7}\frac{3}{7}}{n}}] = [0{,}371\,;0{,}771]$, l'hypothèse nulle est rejetée avec ce test.

Commentaire : L'artifice utilisé par Gavarret est encore, sous d'autres formes, en vigueur de nos jours, en particulier pour augmenter le nombre de données. Dans beaucoup de cas, il ne peut se justifier mathématiquement.

Retour à l'article

Retour à l'énoncé de l'activité

Indications sur l'activité 14.

En utilisant l'artifice, la valeur de la statistique de test est égale à $0{,}50\,977$ et se trouve bien dans l'intervalle qui est $[0, 48\,697\,;0{,}51\,303]$.

En utilisant la "bonne méthode", la valeur de la statistique de test est $0{,}58\,096$ et se trouve encore dans l'intervalle qui est $[0{,}55\,951\,;0{,}58\,335]$.

Commentaire : Il s'agit simplement de mettre en oeuvre l'artifice de Gavarret et de comparer les résultats avec une méthode plus rigoureuse.

Retour à l'article Retour à l'énoncé de l'activité

Indications sur l'activité 15.

- Test $H_0 : p_1 = p_2$ contre $H_1 : p_1 \neq p_2$.
  Ici $d = 0{,}10 < 0{,}1470 = l$. On ne rejette pas H_0. Décision : $p_1 = p_2$.
- Test $H_0 : p_2 = p_3$ contre $H_1 : p_2 \neq p_3$.
  Ici $d = 0{,}03 < 0{,}1474 = l$. On ne rejette pas H_0. Décision : $p_2 = p_3$.
- Test $H_0 : p_1 = p_3$ contre $H_1 : p_1 \neq p_3$.
  Ici $d = 0{,}07 > 0{,}0614 = l$. On rejette H_0. Décision : $p_1 \neq p_3$.

  En toute logique $p_1 = p_2$ et $p_2 = p_3$ devrait impliquer $p_1 = p_3$. Mais ici, $p_1 = p_2$ et $p_2 = p_3$ ne sont que des décisions prises après une expérience aléatoire et donc assorties d'une probabilité d'être vraie. On a vu que dans le cas d'une décision du type $p_1 = p_2$, la probabilité de se tromper pouvait être importante en particulier si la différence entre p_1 et p_2 est "petite". Les règles de la logique ne s'appliquent pas à des décisions qui ne sont pas "certaines".

Commentaire : Aujourd'hui, ce sont des méthodes du type analyse de la variance ou méthode du chi-deux qui s'appliquent pour tester : $H_0 : p_1 = p_2 = p_3$ contre H_1 : il existe au moins deux p_i différents. Cependant, lorsque l'hypothèse nulle est rejetée, se pose ensuite le problème de la comparaison des p_i et, dans l'état actuel des connaissances statistiques, il n'existe pas de test idéal. La tâche que s'était donné Gavarret de classer plusieurs "médications" n'est pas achevée !

Retour à l'article

Retour à l'énoncé de l'activité

Indications sur l'activité 16.

16.1. D'après la formule de Bayes : $\mathbb{P}_{S_1}(D) = \dfrac{\mathbb{P}(D)\mathbb{P}_D(S_1)}{\mathbb{P}(S_1)}$ et $\mathbb{P}_{S_2}(D) = \dfrac{\mathbb{P}(D)\mathbb{P}_D(S_2)}{\mathbb{P}(S_2)}$.

Si $\mathbb{P}(S_1) = \mathbb{P}(S_2) = p^*$, alors $p_1^* = \dfrac{p \times p_1}{p^*}$ et $p_2^* = \dfrac{p \times p_2}{p^*}$.
Il s'ensuit que $p_1 = p_2$ entraine $p_1^* = p_2^*$ et réciproquement.

16.2. Pour traiter les données de l'érysipèle, on peut poser les événements suivants :
- S_1 : "être un homme" ;
- S_2 : "être une femme" ;
- E : "être atteint de d'érysipèle".

On note :
- $p_1 = \mathbb{P}_E(S_1)$ probabilité d'être un homme sachant qu'on est atteint d'érysipèle ;
- $p_2 = \mathbb{P}_E(S_2)$ probabilité d'être une femme sachant qu'on est atteint d'érysipèle ;
- $p_1^* = \mathbb{P}_{S_1}(E)$ probabilité d'être atteint d'érysipèle sachant que l'on est un homme ;
- $p_2^* = \mathbb{P}_{S_2}(E)$ probabilité d'être atteint d'érysipèle sachant que l'on est une femme ;
- $p = \mathbb{P}(E)$ probabilité d'être atteint d'érysipèle.

Il s'agit de tester :
$H_0 : p_1^* = p_2^*$ contre $H_1 : p_1^* \neq p_2^*$.
On peut supposer que $\mathbb{P}(S_1) = \mathbb{P}(S_2)$ et donc tester suivant la méthode de Gavarret :
$H_0 : p_1 = p_2$ contre $H_1 : p_1 \neq p_2$
La valeur de la statistique de test est $\dfrac{n_1}{n_1 + n_2} = 0{,}477\,0115$. Elle se trouve dans l'intervalle $[0{,}446\,394\,;0.553\,605$
On ne peut donc pas affirmer que le genre a une influence sur le fait d'attraper ou non l'érysipèle.

> *Commentaire :* Cette activité permet l'utilisation de la formule de Bayes. Elle amène à mieux formaliser les hypothèses en présence.

Retour à l'article   Retour à l'énoncé de l'activité