



Contribution

Mais où sont les trous de balles ?

MARC-ANDRÉ DÉSAUTELS, DÉPARTEMENT DE MATHÉMATIQUES
CÉGEP DE SAINT-JEAN-SUR-RICHELIEU

Résumé

Abraham Wald, surtout connu pour son invention de l'analyse séquentielle, a été un des membres du Statistical Research Group (SRG) qui a été créé de 1942 à 1945 pour aider à l'effort de guerre. Nous discuterons du travail de Wald concernant le biais de l'armée américaine dans certaines de ses décisions et nous discuterons de la façon dont il s'y est pris pour estimer les probabilités qu'un avion de guerre criblé de balles revienne (ou pas !) du théâtre des opérations.

Mots clés : biais, probabilités

1 Introduction

This story, like many World War II stories, starts with the Nazis hounding a Jew out of Europe and ends with the Nazis regretting it.

Jordan Ellenberg

Dans cet article, nous introduisons quelques-uns des travaux moins connus de Abraham Wald. Durant les années 1943 à 1945, Wald travailla de concert avec l'armée américaine pour la conseiller sur le blindage de ses divers avions. Nous présenterons brièvement ses travaux dans la première section de cet article. Dans plusieurs mémos, Abraham Wald étudia la probabilité de survie des avions partant pour le front. Dans la seconde section de cet article, nous définirons le problème et le résoudrons pour un ensemble d'hypothèses simplifiées.

1.1 La jeunesse

Abraham Wald est né le 31 octobre 1902 dans la ville de Cluj-Napoca, en Roumanie. Au moment de son adolescence, la première guerre mondiale était terminée. Son talent en mathématiques a été découvert très tôt et en 1927, il entre à l'Université de Vienne, en Autriche, où il reçoit un doctorat en mathématiques, en 1931, sous la direction de Karl Menger.

À la fin de ses études, en raison des conditions économiques difficiles et de la discrimination de l'Autriche face aux Juifs, il ne parvient pas à trouver un emploi. Il se fait par contre offrir un emploi par Oskar Morgenstern (celui-ci émigre plus tard aux États-Unis et contribue à inventer la théorie des jeux). À cette époque, Morgenstern est directeur de l'Institut autrichien de recherches économiques. Il demande alors à Wald de venir faire de petits boulots de mathématiques pour lui. Malgré la situation politique qui continue de se dégrader en Autriche, Wald hésite à émigrer.

Au moment où les Nazis envahissent l'Autriche en 1938, Wald reçoit une bourse pour travailler pour la *Cowles Commission for Research in Economics*, un institut basé au Colorado qui travaille sur des questions économiques. Cette bourse lui est offerte en raison de son expérience acquise durant son travail pour Morgenstern. Abraham Wald décide donc enfin d'émigrer aux États-Unis avec sa famille.

Après quelques mois de vie au Colorado, on offre à Wald un emploi de professeur de statistiques à l'université Columbia, à New York. C'est à partir de ce moment qu'il commence à participer à l'effort de guerre. . .

Pour en apprendre davantage sur la vie d'Abraham Wald, la lectrice et le lecteur intéressés sont invités à consulter [2] et [3].

1.2 Le Statistical Research Group (SRG)

Le SRG était basé à l'Université Columbia et supporté par le *Applied Mathematics Panel (AMP)* du *National Defense Research Committee (NDRC)*, qui faisait partie du *Office of Scientific Research and Development (OSRD)*. Le SRG était un programme secret formé de statisticiens, un peu comme le projet Manhattan, mais fabriquant des équations plutôt que des bombes. Pour en apprendre davantage sur le SRG, la lectrice et le lecteur intéressés sont invités à consulter [4] et [5].

2 Le biais du survivant

Le premier problème que les officiers de l'armée américaine présentèrent à Wald était le suivant. Les avions qui revenaient du front étaient recouverts de trous produits par les balles ennemies. Vous pouvez voir un exemple à la Figure 1.

L'armée américaine ne voulant pas que ses avions soient abattus par des tirs ennemis, elle décide de leur ajouter du blindage.

Par contre,

- mettre trop de blindage est un problème,
- ne pas en mettre assez en est un aussi.



FIGURE 1 – Les avions revenaient du combat couverts de trous de balles, mais les trous n'étaient pas uniformément distribués. Source : <https://tomtunguz.com/clandestine-statisticians/>

En effet, lorsque l'avion est trop blindé, il perd de la manoeuvrabilité et il consomme davantage de carburant. Au contraire, ne pas mettre suffisamment de blindage rend les avions trop vulnérables. Les officiers de l'armée voulaient obtenir la même protection avec moins de blindage, en se concentrant sur les endroits en ayant davantage besoin. Ils demandèrent donc à Wald quelle quantité de blindage ajouter aux diverses parties de l'avion.

Lorsque Wald commença à étudier le problème, il se rendit compte que les avions revenaient du combat couverts de trous de balles, mais que les trous n'étaient pas uniformément distribués, un peu comme nous pouvons le voir à la Figure 1. Le tableau 1 présente les informations sur lesquelles se basait Wald pour tirer ses conclusions.

Les militaires étaient enclins à ajouter de la protection aux parties qui, sur les avions revenant du front, montraient davantage de trous. Quant à lui, Wald conclut que le blindage ne devait pas aller aux endroits où il y avait des trous mais là où il n'y en avait pas.

TABLEAU 1 – Un exemple des informations que recevait Wald à propos des avions qui revenaient du combat recouverts de trous de balles.

Section de l'avion	Nombre de trous par pied carré
Moteur	1,11
Fuselage (partie principale de l'avion)	1,73
Système de carburant	1,55
Reste de l'avion	1,8

Wald s'était posé la question suivante : où sont les trous manquants ? Il avait supposé, en se basant sur de bonnes informations, que les coups reçus par un avion au combat se devaient d'être uniformément distribués sur tout l'avion. Les trous sur les avions de retour n'étaient par contre pas uniformément distribués. Les trous manquants devaient donc être logiquement sur les avions qui ne sont pas revenus.

La raison pour laquelle les avions de retour avaient moins de trous dans la région des moteurs (Tableau 1) était que les avions ayant été touchés aux moteurs ne revenaient pas. Il s'ensuit donc que les coups sur les parties les plus vulnérables avaient moins de chance d'être trouvés sur les avions revenants que les coups portants sur les parties les moins vulnérables, puisque les avions recevant des coups sur les parties les plus vulnérables avaient moins de chance de revenir pour offrir des données. À partir de ces hypothèses, Wald inventa diverses méthodes pour estimer la vulnérabilité de diverses parties de l'avion [1].

Comment est-ce que Wald a pu voir ce que les officiers expérimentés de l'armée ne pouvaient pas voir ? Un statisticien se pose toujours les questions suivantes : Quelles sont vos hypothèses ? Sont-elles justifiées ? Les officiers de l'armée étaient sous l'impression que les avions qui revenaient formaient un échantillon aléatoire de tous les avions. C'était bien sûr faux. Une fois que vous avez compris que c'est l'hypothèse sous laquelle vous travaillez, ça prend un petit moment pour comprendre que vous faites fausse route.

L'exemple étudié par Wald est un exemple de ce qu'on appelle le biais du survivant. C'est une forme de biais de sélection consistant à surévaluer les chances de succès d'une initiative en concentrant l'attention sur les sujets ayant réussi mais qui sont des exceptions statistiques (des « survivants ») plutôt que des cas représentatifs.

Les exemples de biais du survivant sont légion :

- En finance, le biais du survivant survient lors de l'exclusion des études de performance des entreprises ayant fait faillite.
- En production industrielle, on entend souvent « les produits étaient mieux fabriqués avant ». Bien sûr, on oublie tous les objets de la vie courante que nous avons dû mettre à la poubelle.
- En architecture et en construction, les bâtiments de plus de 100 ans encore debout. Encore une fois, on oublie les bâtiments qui se sont écroulés et qui étaient mal bâtis.
- Dans les carrières hyper-compétitives, comme les athlètes d'élite.
- Une étude réalisée en 1987 montre que les chats qui chutent d'une hauteur équivalente à moins de six étages et qui survivent ont des blessures plus graves que les chats tombant d'une hauteur plus élevée (Whitney WO, Mehlhaff C.J. (1987) [12]). La situation pourrait cependant s'expliquer par un biais du survivant. Il est beaucoup moins probable que les chats qui meurent à la suite d'une chute soient emmenés chez le vétérinaire par rapport aux chats blessés par une chute. Les chats morts à la suite d'une chute de plus de six étages sont donc retirés des statistiques, ne laissant la place qu'aux chats « survivants ».

2.1 La navette Challenger



FIGURE 2 – L'explosion de la navette Challenger, 73 secondes après son décollage. La raison de l'explosion est une déféctuosité au niveau des joints d'étanchéité. Source : <https://www.cbsnews.com/pictures/challenger-shuttle-disaster/>

L'accident de la navette Challenger s'est produit le 28 janvier 1986. Une défaillance au niveau des joints d'étanchéité, en raison des températures inhabituellement froides au moment du décollage, a causé la perte des sept astronautes (Figure 2).

L'accident choque énormément la société américaine, en partie en raison de la présence à bord de Christa McAuliffe, qui aurait été la première astronaute-enseignante, choisie par le projet *Teacher in Space*. Cet accident a pour conséquence la création de la *Rogers Commission* (formée par Neil A. Armstrong et le prix Nobel Richard P. Feynman, entre autres). Ce sont les membres de cette commission qui ont démontré la fragilité des joints d'étanchéité lors de températures inhabituellement froides.

Suite aux travaux de cette commission, on apprend que le problème des joints d'étanchéité était connu ! La nuit avant le décollage, une téléconférence avait eu lieu entre les ingénieurs et les directeurs de la NASA pour discuter des effets des basses températures sur le décollage et les joints d'étanchéité.

Temperature data [are] not conclusive on predicting primary O-ring blowby.

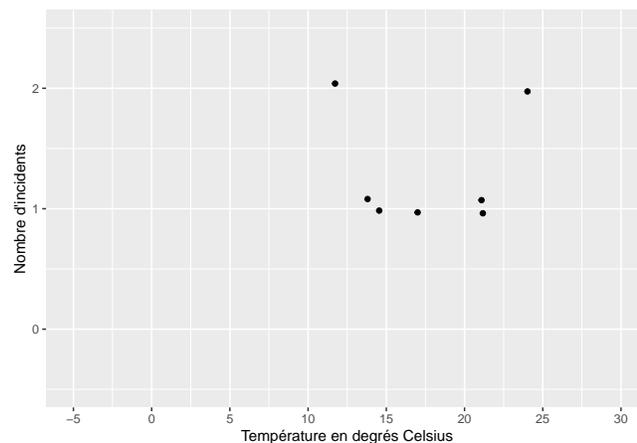


FIGURE 3 – Représentation du nombre d'incidents concernant les joints d'étanchéité, en fonction de la température en degrés Celsius. Une petite variation aléatoire verticale a été ajoutée pour bien distinguer les divers points.

La conclusion de cette conférence est que les données concernant les problèmes des joints d'étanchéité et la température ne permettent pas de conclure que les températures froides affectent les joints.

La conclusion de la téléconférence a été fortement influencée par la Figure 3. Avant le vol de la navette Challenger, 24 autres vols avaient été effectués et la NASA avait compilé le nombre d'incidents et la température pour chacun de ces vols.

La commission Rogers trouva une grave erreur dans la Figure 3 : **les vols sans incidents avaient été exclus** du graphique, car on croyait que ces vols ne donnaient aucune information sur l'effet de la température. C'est un exemple devenu maintenant « classique » de biais du survivant.

A careful analysis of the flight history of O-ring performance would have revealed the correlation of O-ring damage in low temperature

La Figure 4 présente tous les vols (avec ou sans incidents), qui ont précédé l'accident de la navette Challenger.

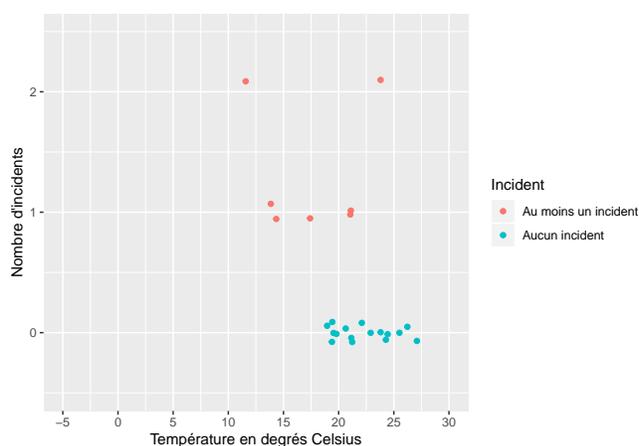


FIGURE 4 – Représentation du nombre d'incidents concernant les joints d'étanchéité, en fonction de la température en degrés Celsius, en ajoutant les vols sans incident. Une petite variation aléatoire verticale a été ajoutée pour bien distinguer les divers points.

La Figure 5 présente quant à elle les données concernant tous les vols précédant celui de la navette Challenger. Nous avons ajouté des barres verticales rouges qui représentent la température au moment du décollage de la navette. De plus, nous avons effectué une régression logistique de nos données, présentée en bleu.

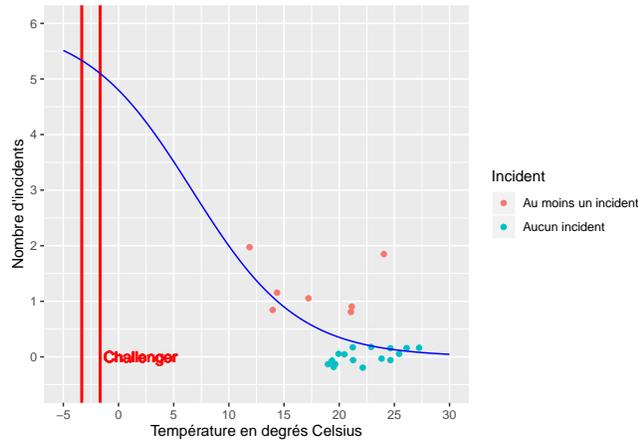


FIGURE 5 – Représentation du nombre d’incidents concernant les joints d’étanchéité, en fonction de la température en degrés Celsius, en ajoutant les vols sans aucun incident. Une petite variation aléatoire verticale a été ajoutée pour bien distinguer les divers points.

3 Les trous de balles

Durant ces années au SRG, Wald produit huit mémos ([1] et [6]) portant sur l’estimation des probabilités de survie d’un avion, sachant que celui-ci a déjà été touché. L’aspect le plus étonnant de ses travaux est qu’ils permettent d’estimer le dommage sur des avions qui ne sont jamais revenus du combat ! Nous nous intéresserons dans cet article au premier mémo de Wald.

Dans ce premier mémo, Wald a accès aux données concernant le nombre de trous sur les avions revenant du front. Dans l’exemple traité par Wald ([1]) et aussi par Mangel et Samaniego ([6]), les données sont celles que nous pouvons trouver au tableau 2.

À partir de ce type de données, que pouvons-nous dire sur les probabilités qu’un avion survive à un nombre donné de coups ? Nous ne savons rien des avions qui ne sont pas revenus... sauf qu’ils ne sont pas revenus. Plusieurs raisons possibles peuvent expliquer pourquoi ces avions sont disparus : ils peuvent manquer de carburant, ils peuvent avoir subi une avarie mécanique, etc. La réponse à la question posée à Wald est donc extrêmement complexe et elle est encore plus difficile à résoudre en raison de toutes les données manquantes. Wald devait donc rester très prudent dans les hypothèses qu’il utilisait.

TABLEAU 2 – Un exemple des informations que recevait Wald à propos des avions qui revenaient du combat recouverts de trous de balles. La colonne des fractions est calculée en divisant les éléments de la colonne Nombre par le nombre d’avions partis en mission, dans ce cas-ci 400.

	Nombre	Fraction
Avions partis en mission	400	1,000
Avions qui reviennent	380	0,950
Avions abattus	20	0,050
Avions revenus sans aucun coup au but	320	0,800
Avions revenus avec 1 coup au but	32	0,080
Avions revenus avec 2 coups au but	20	0,050
Avions revenus avec 3 coups au but	4	0,010
Avions revenus avec 4 coups au but	2	0,005
Avions revenus avec 5 coups au but	2	0,005

3.1 Quelques notations

Avant de débiter, le tableau 3 décrit les notations dont nous allons avoir besoin pour résoudre ce problème.

TABLEAU 3 – Les notations utilisées pour résoudre le problème de Wald.

Notation	Définition
N	le nombre total d’avions
S	le nombre d’avions survivants
D	le nombre d’avions disparus
X_i	le nombre d’avions touchés i fois
$x_i = \frac{X_i}{N}$	la fraction du nombre d’avions touchés i fois
P_i	la probabilité qu’un avion soit abattu par i coups au but
Q_i	la probabilité de l’événement contraire
p_i	la probabilité conditionnelle qu’un avion soit abattu par le i ème coup au but, étant donné que les premiers $i - 1$ coups ne l’ont pas abattu
q_i	la probabilité de l’événement contraire

3.2 Les hypothèses

À l'aide des notations du tableau 3, nous pouvons écrire les hypothèses suivantes :

Hypothèse 1 : Nous connaissons N , le nombre total d'avions .

Hypothèse 2 : Nous connaissons pour tout i ($i = 0, 1, 2, \dots$) les nombres S_i , c'est-à-dire le nombre d'avions ayant survécu à i coups.

Hypothèse 3 : Nous supposons que tous les avions disparus le sont en raison de tirs ennemis et donc $D_0 = 0$. Ceci implique que nous supposons qu'aucun avion ne peut être porté disparu en raison de problèmes mécaniques.

Hypothèse 4 : Le nombre de tirs sur un avion est borné, c'est-à-dire que j étant le rang du tir considéré, $D_j = 0$ pour j plus grand qu'un certain entier m .

Pour résoudre le problème, nous suivrons les étapes décrites dans [1] et [6]. Par contre, pour la lectrice ou le lecteur intéressé, on peut trouver dans [7] une façon différente de résoudre ce même problème.

3.3 Quelques résultats « évidents »

Résultat 1 : Le nombre total d'avions envoyés au combat est égal à la somme du nombre d'avions survivants et de celui d'avions disparus, c'est-à-dire que $N = S + D$.

Résultat 2 : De façon similaire, en étudiant les avions touchés au but i fois, nous avons $N_i = S_i + D_i$.

Résultat 3 : La probabilité qu'un avion ne soit pas abattu par i coups au but est en fait la probabilité de ne pas être abattu par 1 coup et de ne pas être abattu par 2 coups (en ayant survécu au coup précédent), ... et de ne pas être abattu par i coups (en ayant survécu aux coups précédents). Nous avons donc $Q_i = q_1 q_2 \dots q_i$.

Résultat 4 : La probabilité d'être abattu par i coups au but est donc $P_i = 1 - Q_i = 1 - q_1 q_2 \dots q_i$.

Puisque p_i représente la probabilité conditionnelle qu'un avion soit abattu par le i ème coup au but, étant donné que les premiers $i - 1$ coups ne l'ont pas abattu. Nous pouvons donc écrire :

$$p_i = \frac{\text{Nombre d'avions disparus avec } i \text{ coups au but}}{\text{Nombre d'avions ayant reçu } i \text{ coups au but ou plus}} = \frac{D_i}{\sum_{j \geq i} N_j} \quad (3.1)$$

À partir de l'équation (3.1), nous pouvons obtenir :

$$\begin{aligned}
D_i &= p_i \left(\sum_{j \geq i} N_j \right) \\
&= p_i \left(N - \sum_{j < i} N_j \right) \\
&= p_i \left(N - \sum_{j < i} S_j - \sum_{j < i} D_j \right). \tag{3.2}
\end{aligned}$$

Nous pouvons résoudre l'équation (3.2) par récurrence sur i car nous avons émis l'hypothèse :

$$\begin{aligned}
D_0 &= 0. \\
D_1 &= p_1 \left(N - \sum_{j < 1} S_j - \sum_{j < 1} D_j \right) \\
&= p_1 (N - S_0) \\
D_2 &= p_2 \left(N - \sum_{j < 2} S_j - \sum_{j < 2} D_j \right) \\
&= p_2 (N - S_0 - S_1 - D_1) \\
&\dots
\end{aligned} \tag{3.3}$$

Les équations (3.3) nous permettent donc de trouver le nombre d'avions disparus en fonction du nombre de trous de balles $i \dots$ si nous connaissons les probabilités p_i . Pour trouver ces probabilités, Wald utilise l'équation suivante :

$$\sum_{j=1}^m \frac{S_j}{q_1 q_2 \dots q_j} = N - S_0. \tag{3.4}$$

Cette équation ne semble pas au premier abord avoir un lien avec le problème de Wald. Il y figure de nombreuses inconnues et elle a donc plusieurs solutions. Les travaux de Wald portaient en fait sur les bonnes hypothèses à poser pour la résoudre.

3.4 De fausses munitions

L'équation (3.4) peut être retrouvée en utilisant, entre autres, le fait que le nombre de tirs sur un avion est borné. Par contre, Wald offre dans [1] une façon alternative de retrouver l'équation (3.4), et c'est celle-ci que nous proposons, tant sa façon de résoudre le problème est originale.

Considérons les avions ayant reçu des tirs effectués avec de fausses munitions (munitions de petit calibre ne comportant pas de projectile) et posons F_i le nombre d'avions ayant été touchés

i fois par de fausses munitions. Ceux-ci ne sont pas endommagés, même s'ils reçoivent un coup. Il est donc clair que le nombre d'avions survivants touchés i fois par de *fausses* munitions est supérieur ou égal au nombre d'avions survivants touchés i fois par de *vraies* munitions. Autrement dit $F_i \geq S_i$. De plus, si nous additionnons tous les avions touchés i fois par de fausses munitions, nous obtiendrons le nombre total d'avions envoyés au combat, c'est-à-dire $\sum_{j=0}^n F_j = N$. Puisque $F_i \geq S_i$, nous allons utiliser une nouvelle variable Y_i telle que $Y_i = F_i - S_i$.

Il est clair que le nombre d'avions non touchés par de fausses munitions est égal au nombre d'avions survivants n'ayant jamais été touchés, ce qui implique que $F_0 = S_0$ et que $Y_0 = 0$. Pour trouver le nombre d'avions Y_i , nous pouvons multiplier le nombre d'avions touchés i fois par de fausses munitions, par la probabilité qu'un avion soit abattu par i coups au but. C'est-à-dire :

$$\begin{aligned}
 Y_i &= P_i F_i \\
 &= P_i (S_i + Y_i) \quad (\text{pour } i = 1, 2, \dots, n) \\
 (1 - P_i) Y_i &= P_i S_i \\
 Y_i &= \frac{P_i}{Q_i} S_i \\
 &= \frac{1 - q_1 q_2 \dots q_i}{q_1 q_2 \dots q_i} S_i \\
 Y_i &= \frac{S_i}{q_1 q_2 \dots q_i} - S_i. \tag{3.5}
 \end{aligned}$$

L'équation (3.5) est valide pour tous $i = 1, 2, \dots, n$ et nous pouvons donc en trouver la somme.

$$\begin{aligned}
 \sum_{j=1}^m Y_j &= \sum_{j=1}^m \frac{S_j}{q_1 q_2 \dots q_j} - \sum_{j=1}^m S_j \\
 \sum_{j=1}^m \frac{S_j}{q_1 q_2 \dots q_j} &= \sum_{j=1}^m (Y_j + S_j) \\
 &= \sum_{j=1}^m F_j \\
 \sum_{j=1}^m \frac{S_j}{q_1 q_2 \dots q_j} &= N - S_0. \tag{3.6}
 \end{aligned}$$

L'équation (3.6) nous redonne donc bien l'équation (3.4) cherchée.

Il semble raisonnable de supposer que les tirs ennemis diminuent les chances pour un avion de survivre, ce qui implique que $q_1 \geq q_2 \geq \dots$. Si nous supposons que les probabilités q_i sont toutes égales, c'est-à-dire $q_1 = q_2 = \dots = q_n$, alors notre problème est résolu complètement et

nous pouvons trouver la valeur de q telle que $q_1 = q_2 = \dots = q_n$ en trouvant la solution de l'équation (3.4) telle que

$$\sum_{j=1}^m \frac{S_j}{q^j} = N - S_0. \quad (3.7)$$

L'équation (3.7) ne possède qu'une seule solution entre 0 et 1.

Pour sa part, Wald trouve plusieurs façons d'estimer les probabilités avec des hypothèses moins restrictives [1].

3.5 Un premier exemple

Nous allons résoudre le problème de Wald en utilisant les données du tableau 2. En utilisant l'équation (3.7) et les données du tableau 2, nous obtenons :

$$\begin{aligned} \sum_{j=1}^m \frac{S_j}{q^j} &= N - S_0 \\ \frac{S_1}{q} + \frac{S_2}{q^2} + \frac{S_3}{q^3} + \frac{S_4}{q^4} + \frac{S_5}{q^5} &= N - S_0 \\ \frac{32}{q} + \frac{20}{q^2} + \frac{4}{q^3} + \frac{2}{q^4} + \frac{2}{q^5} &= 400 - 320 \\ 80q^5 - 32q^4 - 20q^3 - 4q^2 - 2q - 2 &= 0. \end{aligned} \quad (3.8)$$

Si nous représentons le polynôme précédent (équation (3.8)) à la Figure 6, nous pouvons voir la valeur de q qui résout l'équation (3.7). En trouvant la racine du polynôme se trouvant dans l'intervalle $[0; 1]$, nous obtenons que la valeur de q est 0.8510246.

À partir de la valeur de q obtenue, nous avons que $p = 1 - q$ est égal à 0.1489754, qui représente la probabilité qu'un avion soit abattu au i -ème coup, en sachant que les $i - 1$ coups précédents ne l'ont pas abattu.

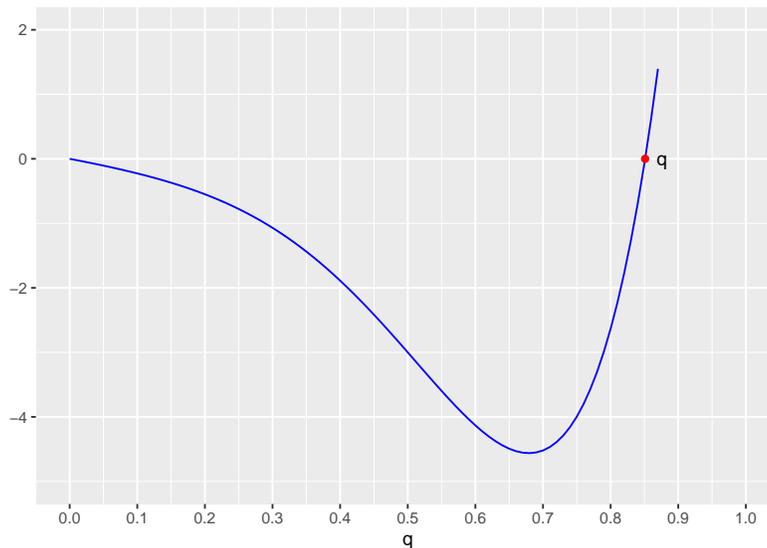


FIGURE 6 – Le polynôme de degré cinq obtenu à l'équation 3.8.

En utilisant les équations (3.3) et $p = 0.1489754$, nous avons :

$$D_0 = 0$$

$$D_1 = p(N - S_0) = 11,9180343$$

$$D_2 = p(N - S_0 - S_1 - D_1) = 5,3753263$$

$$D_3 = p(N - S_0 - S_1 - S_2 - D_1 - D_2) = 1,5950262$$

$$D_4 = p(N - S_0 - S_1 - S_2 - S_3 - D_1 - D_2 - D_3) = 0,7615048$$

$$D_5 = p(N - S_0 - S_1 - S_2 - S_3 - S_4 - D_1 - D_2 - D_3 - D_4) = 0,3501084.$$

Nous présentons ces résultats au tableau 4. Nous ajoutons également une colonne représentant la fraction d'avions disparus D_i sur le nombre total d'avions, qui est de 400 dans cet exemple.

3.6 Un second exemple

Dans son article, Wald étudie un autre exemple de données artificielles, pour comparer les valeurs de q obtenues ainsi que les valeurs D_i obtenues. Les données se trouvent au tableau 5.

En utilisant une méthode similaire à celle utilisée à la section 3.5, nous obtenons que la valeur

TABLEAU 4 – Nous trouvons dans ce tableau les valeurs D_i (dans la colonne Nombre) trouvées à l'aide des formules de Wald ainsi que les valeurs d_i (dans la colonne Fraction), pour le premier exemple.

Di	Nombre	Fraction
0	0.0000000	0.0000000
1	11.9180343	0.0297951
2	5.3753263	0.0134383
3	1.5950262	0.0039876
4	0.7615048	0.0019038
5	0.3501084	0.0008753

TABLEAU 5 – Un second exemple des informations que recevait Wald à propos des avions qui revenaient du combat recouverts de trous de balles.

	Nombre	Fraction
Avions partis en mission	400	1,000
Avions qui reviennent	320	0,800
Avions abattus	80	0,200
Avions revenus sans aucun coup au but	120	0,300
Avions revenus avec 1 coup au but	80	0,200
Avions revenus avec 2 coups au but	40	0,100
Avions revenus avec 3 coups au but	40	0,100
Avions revenus avec 4 coups au but	20	0,050
Avions revenus avec 5 coups au but	20	0,050

de q est 0.8707043. À partir de cette valeur de q , nous obtenons que $p = 1 - q$ est égal à 0.1292957, qui représente la probabilité qu'un avion soit abattu par le i -ème coup, en sachant que les $i - 1$ coups précédents ne l'ont pas abattu.

Nous présentons ces résultats au tableau 6. Nous ajoutons également une colonne représentant la fraction D_i du nombre total d'avions (400 dans cet exemple) qui ont disparu. De plus, nous mettons côte à côte les résultats du premier et du second exemple de Wald.

TABLEAU 6 – Nous trouvons dans ce tableau les valeurs D_i (dans la colonne Nombre) trouvées à l'aide des formules de Wald ainsi que les valeurs d_i (dans la colonne fraction), pour le premier et le second exemple.

Di	Nombre 1	Nombre 2	Fraction 1	Fraction 2
0	0.0000000	0.0000000	0.0000000	0.0000000
1	11.9180343	36.202792	0.0297951	0.0905070
2	5.3753263	21.178272	0.0134383	0.0529457
3	1.5950262	13.268186	0.0039876	0.0331705
4	0.7615048	6.380839	0.0019038	0.0159521
5	0.3501084	2.969910	0.0008753	0.0074248

Rappelons que la valeur de q trouvée dans le premier exemple (0.8510246) est presque la même que celle trouvée dans le second exemple (0.8707043), malgré le fait que les valeurs S_i sont très différentes. La différence entre les valeurs de S_i semble être due au fait que la probabilité qu'un avion reçoive un coup est beaucoup plus petite dans le premier que dans le second exemple. La probabilité qu'un avion reçoive un coup n'est pas en lien avec la probabilité qu'un avion s'écrase s'il reçoit un coup.

4 Conclusion

Le biais du survivant, comme présenté dans cet article, est présent un peu partout autour de nous. Le problème du blindage des avions devrait agir comme une petite sonnette d'alarme pour toujours avoir en tête les moments où nous prenons une décision en ne tenant pas compte de toutes les données existantes, mais seulement des données présentes.

Quant au problème d'étudier la probabilité que les avions reviennent du front, la lectrice et le lecteur intéressés pourront découvrir en [1], plusieurs moyens d'utiliser des hypothèses autres que $q_1 = q_2 = \dots = q_n$. Dans les mémos 87 et 88, Wald étudie respectivement les bornes exactes et approximatives des valeurs de $1 - \prod_{j=1}^i q_j$. Le mémo 89 consiste quant à lui à étudier les bornes de $1 - \prod_{j=1}^i q_j$ sous l'hypothèse que $\lambda_1 q_j \leq q_{j+1} \leq \lambda_2 q_j$. Les mémos suivants servent ensuite à partager l'avion en zones de vulnérabilités variées et à étudier les erreurs possibles

d'échantillonnage. Le dernier mémo porte sur une variété d'autres sujets reliés au sujet d'origine. L'article [7] est aussi très pertinent pour sa présentation du problème de Wald en utilisant une autre méthode de résolution.

Enfin, nous invitons également les lecteurs à lire [4] pour une version « non mathématique » des travaux du SRG, et par le fait même des travaux de Wald.

Références

- [1] Wald, Abraham (1980). A Method of Estimating Plane Vulnerability Based on Damage of Survivors. *CRC 432*, Juillet 1980.
- [2] Morgenstern, Oskar (1951). Abraham Wald, 1902-1950. *Econometrica*, vol. 19, no. 4, pages 361-367. JSTOR, www.jstor.org/stable/1907462.
- [3] Wolfowitz, J. (1952). Abraham Wald, 1902-1950. *The Annals of Mathematical Statistics*, vol. 23, no. 1, pages 1-13. JSTOR, www.jstor.org/stable/2236396.
- [4] Wallis, W. Allen (1980). The Statistical Research Group, 1942-1945. *Journal of the American Statistical Association*, vol. 75, no. 370, pages 320-330. JSTOR, www.jstor.org/stable/2287451.
- [5] Wallis, W. Allen (1980). The Statistical Research Group, 1942-1945 : Rejoinder. *Journal of the American Statistical Association*, vol. 75, no. 370, pages 334-335. JSTOR, www.jstor.org/stable/2287454.
- [6] Mangel, Marc et Samaniego, F.J.. (1984). Abraham Wald's Work on Aircraft Survivability. *Journal of The American Statistical Association*, vol. 79, no. 386, pages 259-267. JSTOR, <https://www.jstor.org/stable/2288257>.
- [7] Berger, James O. (1984). Abraham Wald's Work on Aircraft Survivability : Comment. *Journal of the American Statistical Association*, vol. 79, no. 386, pages 267-269. JSTOR, www.jstor.org/stable/2288258.
- [8] <https://tomtunguz.com/clandestine-statisticians/>
- [9] https://en.wikipedia.org/wiki/Space_Shuttle_Challenger_disaster
- [10] <http://www.ams.org/publicoutreach/feature-column/fc-2016-06>
- [11] Ellenberg, Jordan (2014). How Not to Be Wrong : The Power of Mathematical Thinking. *Penguin Press*, 480 pages.
- [12] Whitney WO, Mehlhaff CJ. (1987). *Journal of the American Veterinary Medical Association*, 191 (11), pages 1399-1403.