

REVUE DE STATISTIQUE APPLIQUÉE

H. F. FREUNDLICH

La méthode du maximum de vraisemblance pour l'estimation des taux de survie à cinq ans

Revue de statistique appliquée, tome 10, n° 3 (1962), p. 59-74

http://www.numdam.org/item?id=RSA_1962__10_3_59_0

© Société française de statistique, 1962, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

LA MÉTHODE DU MAXIMUM DE VRAISEMBLANCE POUR L'ESTIMATION DES TAUX DE SURVIE A CINQ ANS (1)

H. F. FREUNDLICH

Service de Physique Médicale,
United Bristol Hospitals

Il est encore courant de comparer des taux de survie à cinq ans pour évaluer les mérites de différents traitements en thérapeutique du cancer. Mais ce procédé est assez long. On doit d'abord réunir un nombre de cas suffisant pour permettre une analyse statistique. Cela peut prendre quelques années. Ensuite, il faut encore attendre cinq ans avant de pouvoir estimer le taux de survie à 5 ans. Suivant l'importance du nombre de cas traités annuellement dans un Centre, on peut devoir attendre jusqu'à 8 ou 10 ans avant de pouvoir calculer un tel taux.

J.W. Boag [1] a publié en 1948 une méthode, basée sur celle du maximum de vraisemblance de R.A. Fisher [2], qui permet d'estimer le taux de survie à cinq ans bien avant la fin des cinq années. Bien que cette méthode permette de gagner un temps considérable pour juger une nouvelle méthode de traitement, elle n'a été que très peu utilisée depuis la publication de l'article de Boag. Ceci est probablement imputable à deux faits : d'une part le médecin moyen ne possède pas une formation mathématique suffisante pour mener à bien une telle estimation ; d'autre part, les spécialistes des centres de radiothérapie, qui eux devraient en être capables, sont employés à des travaux cliniques de routine, qui ne leur en laissent guère le temps. Une estimation moyenne par cette méthode peut demander une centaine d'heures ou plus, c'est-à-dire plusieurs semaines de travail, avec l'aide d'une machine à calculer ordinaire. L'utilisation d'un ordinateur semble s'imposer pour de si longs calculs : il existe maintenant un programme sur I.B.M. 704, ou sur calculateurs similaires, pour exécuter ce travail.

Le but de cet exposé est de donner quelques uns des résultats obtenus jusqu'ici et de montrer que les données nécessaires à une telle estimation peuvent être facilement rassemblées par un employé de bureau.

Bien qu'il soit préférable de lire l'exposé original de Boag [1], il est peut être utile de donner un bref aperçu de sa méthode d'estimation d'un taux de survie à cinq ans.

Si on considère, à différentes époques au cours d'une période de quelques années, des malades qui ont été traités pour un cancer, on peut les répartir en quatre groupes.

(1) Cet exposé a été présenté à la conférence : "Le rôle de l'ordinateur dans la Recherche Statistique en Biologie" (13-15 septembre 1961), organisée par le Centre d'Enseignement Européen d'I.B.M., Blaricum, Hollande.

1er groupe : ceux qui sont morts avec des signes de cancer, à l'époque t, après le début du traitement.

2ème groupe : ceux qui sont morts à l'époque t, après le début du traitement, mais pour d'autres raisons.

3ème groupe : ceux qui sont encore vivants et sans symptômes à l'époque t.

4ème groupe : ceux qui sont encore en vie, à l'époque t, mais avec un cancer qui s'aggrave.

Dans les estimations calculées sur des groupes cliniques de malades on a omis le 4ème groupe, car il est impossible de préciser la cause des symptômes résiduels : croissance persistante, simples effets de radiations, d'opérations, ou toute autre cause ? Ce groupe serait de toutes façons peu nombreux et sa sélection assez aléatoire.

(Boag [1], Gaddum [4] et d'autres ont montré que si l'on porte sur un graphique le nombre de malades morts avec un cancer en fonction du logarithme du temps de survie on obtient une courbe voisine de la courbe normale ou courbe d'erreur de Gauss. Ce phénomène est purement empirique mais semble largement valable (figure 1). Le modèle statistique d'une expérience clinique est indiqué dans la figure 2. Le carré (1) représente le groupe original de n malades. Après traitement et un certain temps d'observation, on peut partager ces malades en première approximation en deux groupes :

a) ceux qui sont morts avec des signes du cancer. (1er groupe)

Leur proportion (1-c) du groupe total est l'aire déterminée par la courbe normale.

b) ceux qui sont "guéris" (groupes 2 et 3)

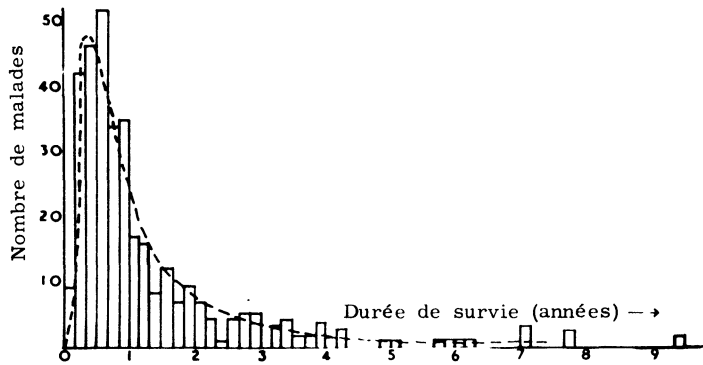
Leur proportion c étant représentée par la surface du petit carré. Comme la moyenne μ et l'écart-type σ déterminent complètement la loi normale c'est-à-dire la distribution des logarithmes des temps de survie, les trois paramètres μ , σ et c résument complètement les résultats d'une méthode de traitement.

Il est peut être utile de préciser ici que le taux de survie à cinq ans est la somme des proportions de ceux qui sont "guéris" (c'est-à-dire guéris ou morts d'une cause autre que le cancer) et de ceux qui mourront ultérieurement avec un cancer, (mais dont le temps de survie est supérieur à cinq ans). Il est facile de calculer ce dernier terme une fois connus les trois paramètres μ , σ et c.

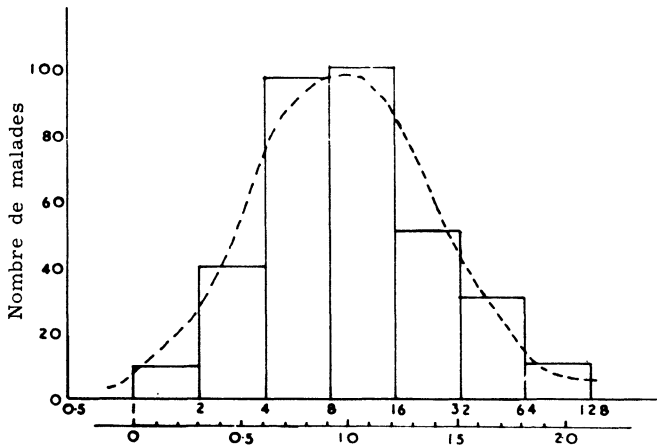
Le problème statistique est d'estimer de la meilleure façon possible ces trois paramètres, et cela bien avant la fin de la période de cinq ans. Ce problème a été résolu grâce à la méthode du maximum de vraisemblance de R.A. Fisher [2].

Le principe fondamental de cette méthode est de considérer une fonction L dite fonction de vraisemblance, fonction qui dépend des paramètres indépendants μ , σ et c.

Cette fonction L mesure la probabilité d'obtenir l'échantillon à partir d'un ensemble de valeurs données de μ , σ et c. On admet alors que les meilleures valeurs de μ , σ et c sont celles pour lesquelles la probabilité d'obtenir l'échantillon observé est maximum, c'est-à-dire celles qui maximisent la fonction L.



Distribution avec échelle des temps en années.



Même distribution avec échelle logarithmique des temps (esprimés en mois).

Figure 1 - Distributions des temps de survie et de leurs logarithmes.

Tableau 1

$$L = \sum_1 \log \left\{ \frac{(1 - c)z}{\sigma} \right\} + \sum_{2,3} \log \{c + (1 - c)q\} + \sum_4 \log(1 - c)q.$$

$$z = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} x^2}$$

$$x = \frac{(\log_e t - \mu)}{\sigma}$$

$$q = \int_x^\infty z \, dx$$

Le maximum de cette fonction :

$$L = f(\mu, \sigma, c)$$

peut être évalué par les méthodes mathématiques usuelles, en résolvant les équations aux dérivées partielles suivantes :

$$\frac{\partial L}{\partial \mu} = 0 \quad ; \quad \frac{\partial L}{\partial \sigma} = 0 \quad ; \quad \frac{\partial L}{\partial c} = 0$$

Malheureusement ces trois équations aux dérivées partielles ne peuvent pas être résolues rigoureusement ; Boag a surmonté la difficulté par une méthode d'approximations successives. Elle est exposée dans le tableau 2.

Tableau 2

Développement limité des dérivées partielles de L

$$\frac{\partial L}{\partial m_1} + \frac{\partial}{\partial m_1} \left(\frac{\partial L}{\partial m_1} \right) \delta m + \frac{\partial}{\partial s_1} \left(\frac{\partial L}{\partial m_1} \right) \delta s + \frac{\partial}{\partial c_1} \left(\frac{\partial L}{\partial m_1} \right) \delta c = 0$$

$$\frac{\partial L}{\partial s_1} + \frac{\partial}{\partial m_1} \left(\frac{\partial L}{\partial s_1} \right) \delta m + \frac{\partial}{\partial s_1} \left(\frac{\partial L}{\partial s_1} \right) \delta s + \frac{\partial}{\partial c_1} \left(\frac{\partial L}{\partial s_1} \right) \delta c = 0$$

$$\frac{\partial L}{\partial c_1} + \frac{\partial}{\partial m_1} \left(\frac{\partial L}{\partial c_1} \right) \delta m + \frac{\partial}{\partial s_1} \left(\frac{\partial L}{\partial c_1} \right) \delta s + \frac{\partial}{\partial c_1} \left(\frac{\partial L}{\partial c_1} \right) \delta c = 0$$

$$s_1 A - D \delta m - G \delta s - H s_1 \delta c = 0$$

$$s_1 B - G \delta m - E \delta s - K s_1 \delta c = 0$$

$$s_1 C - H \delta m - K \delta s - F s_1 \delta c = 0$$

Si m_1 , s_1 et c_1 constituent une première estimation approximative des paramètres, les estimations améliorées peuvent être obtenues en ajoutant les corrections δm , δs , δc . De plus amples détails sur ces calculs sont donnés dans le tableau 3 et les solutions de ces équations pour δm , δc , δs dans le tableau 4.

Dans la table ci-après :

$$\Phi = \frac{z}{(q + c/1 - c)} \quad \Psi = \frac{1}{(c + q/1 - q)} \quad z = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} x^2}$$

$$x = \frac{(u - \mu)}{\sigma} \quad q = \int_x^\infty Z dx \quad u = \log_e t$$

t = durée de survie.

Tableau 3

Dérivées du logarithme de la fonction de vraisemblance.

Quantité	Symbole	Contribution d'un malade du groupe 1	Contribution d'un malade des groupes 2 ou 3	Contribution d'un malade du groupe 4
$\sigma \frac{\partial L}{\partial \mu}$	A	x	Φ	z/q
$\sigma \frac{\partial L}{\partial \sigma}$	B	$x^2 - 1$	$x\Phi$	$x \frac{z}{q}$
$\frac{\partial L}{\partial c}$	C	$-\frac{1}{(1-c)}$	Ψ	$-\frac{1}{(1-c)}$
$-\sigma^2 \frac{\partial^2 L}{\partial \mu^2}$	D	1	$-x\Phi + \Phi^2$	$-x \frac{z}{q} + \left(\frac{z}{q}\right)^2$
$-\sigma^2 \frac{\partial^2 L}{\partial \sigma^2}$	E	$3x^2 - 1$	$2x\Phi - x^3\Phi + x^2\Phi^2$	$2x \frac{z}{q} - x^3 \frac{z}{q} + x^2 \left(\frac{z}{q}\right)^2$
$-\frac{\partial^2 L}{\partial c^2}$	F	$\frac{1}{(1-c)^2}$	Ψ^2	$\frac{1}{(1-c)^2}$
$-\sigma^2 \frac{\partial^2 L}{\partial \sigma \partial \mu}$	G	2x	$\Phi - x^2\Phi + x\Phi^2$	$\frac{z}{q} - x^2 \frac{z}{q} + x \left(\frac{z}{q}\right)^2$
$-\sigma \frac{\partial^2 L}{\partial c \partial \mu}$	H	0	$\frac{\Phi^2}{z(1-c)^2}$	0
$-\sigma \frac{\partial^2 L}{\partial c \partial \sigma}$	K	0	$\frac{x\Phi^2}{z(1-c)^2}$	0

Tableau 4

Solutions des équations du tableau 2, pour δm , δs , δc

$$\delta m = \frac{s_1}{\Delta} \begin{vmatrix} A & G & H \\ B & E & K \\ C & K & F \end{vmatrix} \quad \delta s = \frac{s_1}{\Delta} \begin{vmatrix} D & A & H \\ G & B & K \\ H & C & F \end{vmatrix}$$

$$\delta c = \frac{1}{\Delta} \begin{vmatrix} D & G & A \\ G & E & B \\ H & K & C \end{vmatrix} \quad \Delta = \begin{vmatrix} D & G & H \\ G & E & K \\ H & K & F \end{vmatrix}$$

En pratique, ces calculs sont répétés par approximations successives jusqu'à ce que les corrections δm , δs et δc ne soient plus significatives.

Pour vérifier si l'on a bien trouvé un maximum de la fonction de vraisemblance L , on vérifie si les expressions :

$$\frac{\partial L}{\partial \mu} = A/\sigma \quad \frac{\partial L}{\partial \sigma} = B/\sigma \quad \frac{\partial L}{\partial c} = C$$

sont nulles ou au moins très petites.

Si l'on estime simultanément μ , σ et c (méthode A), les calculs sont particulièrement compliqués.

Cependant, Boag a montré, qu'en de nombreux cas il était très satisfaisant de supposer σ connu et d'estimer μ et c . Les valeurs de σ semblent ne varier qu'entre des limites assez étroites, pour des instituts, des lieux et des périodes de traitement différents (Méthode B).

De plus, si la valeur choisie pour σ est tant soit peu trop petite, la valeur de c correspondante sera trop grande et vice versa (figure 2).

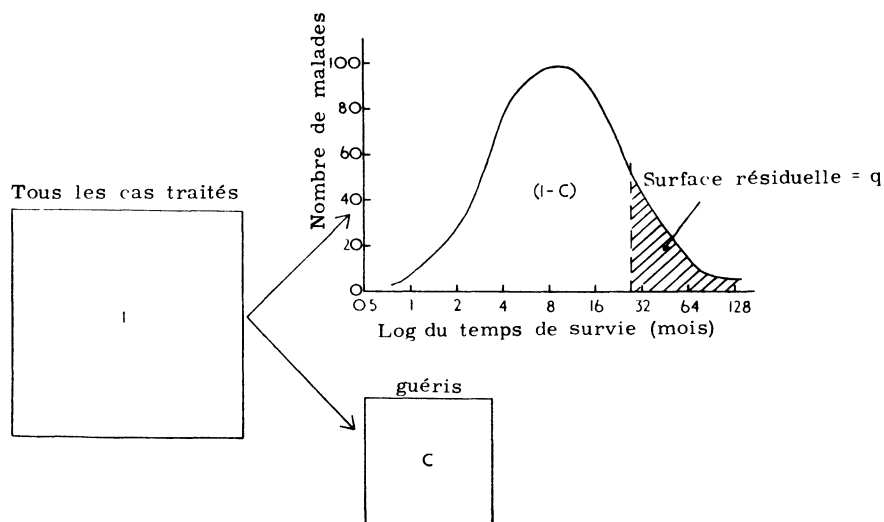


Figure 2 - Modèle statistique d'une expérience clinique (d'après Boag [1]).

Alors, que dans ce cas la valeur de c , le taux de guérison, ne peut plus être considérée comme la réponse correcte à notre problème, les taux estimés de survie sur cinq, dix ou même quinze ans seront beaucoup moins affectés puisque les deux erreurs tendent à se compenser quand on calcule le taux K de survie après n années d'après la formule appropriée :

$$K_n = c + (1 - c) q_n$$

Il existe dès maintenant un programme sur calculateur I.B.M. 704 pour les méthodes A et B. Ce programme, présenté par Fortran, nécessite la mémoire suivante :

Données	1 551 cases
Programme principal	1 406 cases
Entrée, sortie, logarithmes, racines carrées et divers	1 500 cases (environ)
Total	<u>4 500 cases (environ)</u>

Le temps de calcul T par programme est :

$$T = (5 + 5n) \text{ secondes,}$$

où n est le nombre d'itérations. Donc si l'on a besoin de six itérations, le temps nécessaire sera de 35 secondes. Le même calcul fait à la main nécessiterait probablement 12 à 14 jours.

Avant de présenter les résultats obtenus jusqu'ici à l'aide d'un ordinateur I.B.M., nous devons discuter un autre aspect de la méthode du maximum de vraisemblance qui est fondamental quand on organise une expérience clinique.

Si l'on ne possède que des données insuffisantes pour une estimation, l'ordinateur risque d'effectuer un grand nombre d'itérations avant d'atteindre des résultats qui sont encore relativement entachés d'erreur. Il peut même arriver que l'ordinateur soit incapable d'atteindre un résultat valable et continue indéfiniment s'il n'est pas arrêté en appuyant sur un bouton. Il est inutile de dire que dans un tel cas, les résultats calculés pour μ , σ et c n'ont aucun sens.

Il est donc souhaitable de s'assurer préalablement si les données sont suffisantes pour calculer une estimation avec succès. Quand on traite ce problème, le concept de quantité d'information, introduit à l'origine par R.A. Fisher est très pratique. La quantité d'information introduite par un paramètre quelconque est définie comme l'inverse de la variance de l'échantillon. Quand cette quantité croît, la précision de l'estimation augmente. Boag a montré dans son étude que l'expression de la quantité d'information relative au paramètre c est toujours $I_c = \frac{Ni}{c(1-c)}$ (N nombre de malades, c taux de guérison).

Si un temps infini s'est écoulé, le facteur i devient égal à 1 :

$$(I_c)_\infty = \frac{N}{c(1-c)}$$

Pour n'importe quelle période de temps fini, i est toujours plus petit que 1.

La figure 3 montre comment i dépend de la variable $x = \frac{\log_e t - \mu}{\sigma}$. De plus, elle montre que le facteur i est plus grand, pour une valeur donnée de x dans le cas de la méthode B que dans celui de la méthode A. Cela veut dire qu'une plus grande précision peut être obtenue avec les mêmes données, si l'on n'estime que deux paramètres au lieu de trois.

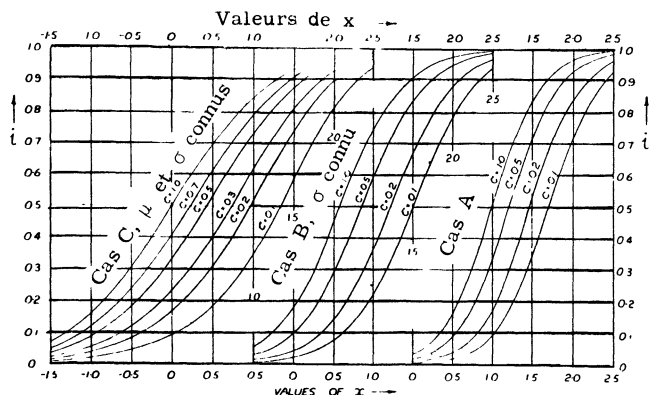


Figure 3 - Fraction de la quantité d'information disponible, utilisée pour estimer c , en fonction de x .

Les premiers calculs ont été effectués sur un groupe fictif de mille malades, que Boag [5] avait préparé pour analyser les aspects statistiques de la préparation d'une expérience clinique. Ce groupe fictif avait les caractéristiques suivantes :

- 1/ On considère une admission de 10 malades par mois.
- 2/ Le taux de guérison est de 0,30.
- 3/ La distribution du temps de survie de ceux des malades qui meurent est log-normale.

Pour réaliser la seconde condition une table de nombres aléatoires a été utilisée.

Quand le numéro d'un malade donné coïncidait avec un nombre aléatoire se terminant par 0, 1 ou 2, on admettait que ce malade avait un long temps de survie ou, pratiquement, qu'il était guéri.

La troisième condition a été réalisée à l'aide d'une planche de Galton (figure 4). On sait que les boules que l'on recueille dans les différentes cases au bas de la planche ont une distribution normale. Boag a utilisé 32 trous, 16 de chaque côté de la moyenne, et il a calculé pour chaque trou, l'écart x correspondant. Le temps de survie de chaque malade était alors déterminé en faisant passer une boule à travers les clous de la planche et en observant dans quel trou elle tombait. Ceci donnait un certain écart x , et bien évidemment on pouvait, à partir de ce résultat, calculer le temps de survie t puisque :

$$x = \frac{\log_e t - \mu}{\sigma},$$

t étant le temps de survie en mois.

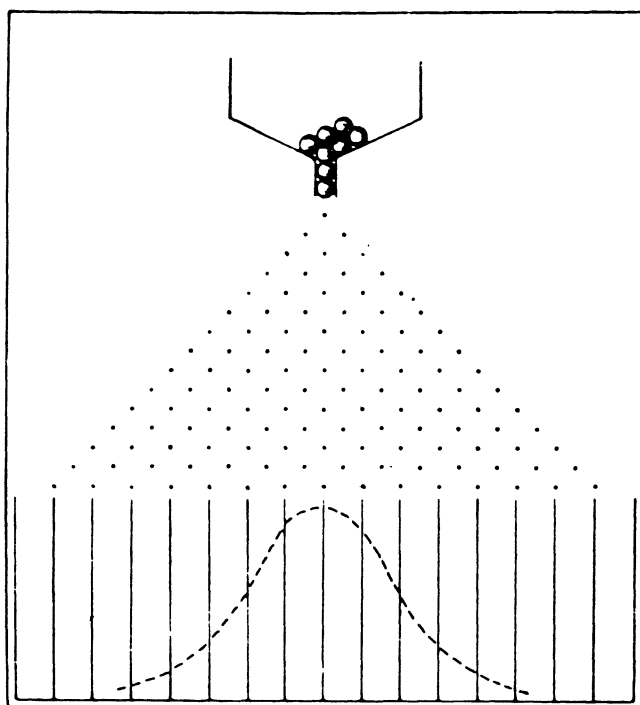


Figure 4 - Planche de Galton.

Ce groupe de mille malades a été divisé en cinq groupes de 200 malades (Groupes A, B, C, D et E). Pour chaque groupe, on a effectué deux estimations, l'une 10 mois, l'autre 30 mois après la fin du traitement. Les calculs sur ce groupe fictif ont été réalisés tout d'abord, pour voir expérimentalement comment cette méthode était applicable, par exemple pour trouver le nombre d'itérations nécessaires pour une quantité d'information donnée, pour tester la stabilité des calculs, etc.

Les estimations calculées pour c sont données dans le tableau 6.

Comme il était prévisible, les estimations après 10 mois montrent encore de grands écarts par rapport à la vraie valeur de c . De même les écarts-types sont encore importants. Les estimations après 30 mois sont considérablement améliorées. En fait, on peut probablement énoncer simplement que, dans ce cas, 10 mois est une période encore trop courte alors que 30 mois est une période convenable pour entreprendre ces estimations.

Les estimations pour μ et σ que nous n'avons pas données conduisent à des résultats similaires.

Tableau 5

Estimation de c à l'aide d'un calculateur I/B.M. 704, pour
des groupes fictifs

(Méthode A)

Groupe	Nombre de malades	Estimation de c		Vraie valeur de c dans la population totale
		Après 10 mois	Après 30 mois	
A	200	0.183	0.31 0.04	0.30 0.03
B	200	0.324 0.07	0.33 0.04	0.30
C	200	0.407 0.06	0.35 0.04	0.30
D	200	0.181 0.11	0.30 0.04	0.30
E	200	0.384 0.05	0.31 0.04	0.30

Tableau 6

Echantillons expérimentaux
Nombre d'itérations et Quantité d'information

(Méthode A)

Groupe	Après 10 mois		Après 30 mois	
	Nombre d'itérations	Quantité d'information	Nombre d'itérations	Quantité d'information
A	7	-	2	-
B	22	0.21	3	0.64
C	4	0.28	4	0.64
D	7	0.08	2	0.64
E	4	0.41	2	0.64

Le tableau 6 donne pour chaque estimation le nombre d'itérations effectuées par le calculateur, ainsi que le facteur i qui est la fraction de la quantité d'information disponible dans chaque cas. Bien que le nombre d'expériences soit encore trop petit pour établir une corrélation, il semble à première vue que le facteur i doive être au moins de l'ordre de 0,25, et de préférence plus grand, pour pouvoir effectuer une estimation.

Tableau 7
 Comparaison des Méthodes A et B
 Estimation de c

Groupe	Après 10 mois		Après 30 mois	
	A	B	A	B
A	0.18	0.40 0.05	0.31 0.04	0.30 0.04
B	0.32 0.07	0.29 0.06	0.33 0.04	0.32 0.04
C	0.41 0.06	0.33 0.06	0.35 0.04	0.32 0.04
D	0.18 0.11	0.21 0.07	0.30 0.04	0.29 0.04
E	0.38 0.05	0.32 0.05	0.31 0.04	0.30 0.04

Le tableau 7 montre une comparaison entre la méthode A et la méthode B. Les remarques faites dans les deux derniers paragraphes sont également valables pour la méthode B. Les expériences entreprises jusqu'ici ne montrent aucune différence significative entre les deux méthodes. Des expériences plus nombreuses et plus rapprochées sont nécessaires pour mettre en évidence quelques différences significatives. Le seul résultat obtenu jusqu'ici est que le nombre d'itérations est quelque peu plus petit dans la méthode B que dans la méthode A, comme le montre le tableau 8.

Tableau 8
 Comparaison des méthodes A et B
 Nombre d'itérations

Groupe	Nombre de malades	Après 10 mois		Après 30 mois	
		A	B	A	B
A	200	7	1	2	2
B	200	22	2	3	2
C	200	4	2	4	2
D	200	7	2	2	2
E	200	4	4	2	2

Les estimations ultérieures ont été effectuées sur des groupes de malades atteintes du cancer du col de l'utérus. Les données concernant ces malades ont été relevées dans le rapport que Sorensen [6] a publié sur 713 malades atteintes de cancer du col de l'utérus et traitées entre 1922 et 1929.

Toutes ces malades ont été suivies jusqu'en 1954. Pour chaque cas est indiqué clairement l'année du traitement, la durée de survie, la date de décès, la cause du décès (soit le cancer, soit une autre maladie) et de nombreux autres détails cliniques. Cinq groupes différents, soignés cinq années consécutives (1925-1929) ont été choisis. On a supposé que le nombre de malades pris en charge chaque mois de l'année considérée était le même.

Deux estimations du taux de survie à cinq ans ont été faites pour chaque groupe, l'un avec les chiffres disponibles après deux ans et l'autre après trois ans (Tableau 9). Le résultat le plus intéressant est visible sur le tableau 10 qui donne le nombre d'itérations, ainsi que le facteur i .

Tableau 9

Estimation du taux de survie à cinq ans
cinq groupes de malades atteintes du cancer du col de l'utérus
(Sorenson Acta Rad. 1958 - suppl. n° 169)

Année du traitement	Nombre de malades	Estimation d'un taux de survie à cinq ans (Méthode A)		Taux net de survie à cinq ans
		après 2 années	après 3 années	
1925	82	19 %	19 %	15 %
1926	122	22 %	20 %	21 %
1927	105	31 %	31 %	28 %
1928	143	34 %	33 %	32 %
1929	130	34 %	34 %	32 %

Tableau 10

Rapport Sørensen
Nombre d'itérations et quantité d'information

Groupe	Nombre de malades	après 2 années		après 3 années	
		Nombre d'itérations	i	Nombre d'itérations	i
1925	82	5	0.31	3	0.61
1926	122	27	0.11	19	0.28
1927	105	2	0.39	3	0.50
1928	143	8	0.11	6	0.31
1929	130	4	0.74	3	0.90

On peut noter que dans le cas du groupe de 1926, le calculateur a effectué 27 itérations pour l'estimation faite au bout de 2 ans et 19 itérations pour celle faite au bout de 3 ans. Un examen serré des résultats révéla que le

calculateur oscillait entre des valeurs supérieures et inférieures au résultat final. Cet effet mérite des recherches plus approfondies et on pourrait envisager d'introduire une sorte d'amortissement dans le programme.

Comme dernier exemple, nous prendrons des estimations faites sur des malades atteints de cancer du larynx. Comme cet exemple illustre la valeur pratique d'une telle estimation, il semble utile de donner brièvement les raisons qui nous ont poussé à faire ces estimations [7].

Les deux principales méthodes utilisées pour traiter le cancer sont la chirurgie et la radiothérapie, c'est-à-dire le traitement par les rayons X. L'équipement normal, pour rayons X, qui est encore le plus largement utilisé, est l'appareil à rayons X de 200 kV.

Il y a environ 10 ou 12 ans, de nombreux articles publiés suggéraient que des radiations plus dures ou plus pénétrantes seraient plus efficaces que les rayons X de 200 kV utilisés dans le passé pour traiter le cancer. Pour cette raison, deux unités de Co^{60} ont été installées au centre de Bristol en 1955. La radiation gamma émise par le Co^{60} est à peu près équivalente à des rayons X de 3 MeV. Très vite les cancers du larynx traités aux radiations de Co^{60} montrèrent cliniquement de meilleurs résultats. Pour confirmer cette observation clinique, on décida d'effectuer une estimation par le maximum de vraisemblance. Le tableau 11 montre deux groupes de malades atteints du cancer du larynx. Le premier groupe a été soigné entre 1948 et 1953. 38,5 % des cas ont été opérés et 61,5 % des cas ont été traités aux rayons X ; à cette époque, les cas les plus favorables étaient opérés, alors que les autres, plus avancés étaient laissés à la radiothérapie.

Tableau 11
Cancer du Larynx
Méthodes de traitement avant et après l'emploi du Co^{60}

Traitement	Groupe 1955-1958		Groupe 1948-1953	
	Effectif traité	Pourcentage du total	Effectif traité	Pourcentage du total
Co^{60}	49	79 %		
Chirurgie	5	8,1 %	25	38,5 %
Rayons X	5	8,1 %	40	61,5 %
Médication Palliative	3	4,8 %		
Total	62	100 %	65	100 %

Avec l'introduction du Co^{60} , le schéma a été complètement changé. Le traitement au Co^{60} est devenu le traitement de choix, ainsi qu'il est clairement indiqué par le chiffre de 79 % de cas soumis au Co^{60} . Les deux groupes ont été comparés dans leur ensemble, car les effectifs étaient trop petits pour pouvoir être subdivisés.

On peut simplement espérer que la répartition des malades se présentant à l'hôpital n'a pas changé de manière significative de 1948 à 1955 en ce qui concerne aussi bien la durée d'évolution déjà acquise par la maladie que le type pathologique ou clinique de l'affection.

On a fait une estimation sur le premier groupe avec les chiffres disponibles au 1/6/1954 (Tableau 12).

Tableau 12
Taux de survie à cinq ans.
Cancer du larynx

Date de l'estimation	Groupe 1948-1953	Groupe 1955-1958	
		1.6.54	1.6.59
Taux de survie estimé	36 %	53 %	51 %
Valeur réelle du taux de survie	32 %	?	

Cette estimation donna un taux de survie à cinq ans de 36 %. La valeur réelle du taux, que l'on connaît maintenant, est de 32 %. La concordance entre ces deux chiffres serait encore meilleure si l'on tenait compte des cas de mort par d'autres maladies. L'estimation pour le groupe 1955/58 fut de 53 % le 1/6/59 et 51 % le 1/6/61. On peut donc affirmer raisonnablement qu'un cas sur deux peut être guéri, contre un sur trois avant 1955. De plus, ce travail a confirmé statistiquement la supériorité du traitement par le Co⁶⁰ (1).

Comme nous l'avons déjà signalé, il serait très souhaitable d'établir préalablement si les données sont adéquates pour l'obtention par le calculateur d'une estimation valable.

De plus, bien souvent, il serait important de savoir combien de temps il est nécessaire d'attendre avant de pouvoir valablement entreprendre une estimation du taux de survie à cinq ans à partir d'un nombre mensuel donné de malades entrant à l'hôpital.

(1) Monsieur Schwartz, Directeur de l'Unité de Recherches Statistiques de l'Institut National d'Hygiène, à qui l'article de Monsieur H.F. Freundlich avait été communiqué, nous fait part de l'observation suivante :

"La méthode de calcul des taux de survie, exposée par Monsieur Freundlich est ingénieuse et on doit le féliciter de l'avoir exposée en détail. Je voudrais seulement exprimer une réserve à propos de l'exemple final qui est utilisé pour l'illustrer. L'auteur compare en effet les résultats du cobalt et de la radiothérapie classique sur deux groupes de malades et, d'après les résultats, conclut qu'on a "confirmé statistiquement la supériorité du cobalt". Indépendamment du fait que la différence des taux de survie entre les deux groupes ne semble guère significative (le test n'est pas fait), les deux groupes de malades ont été soignés à des époques différentes et rien ne permet d'affirmer qu'ils soient comparables. La conclusion énoncée ne me paraît donc pas justifiée. Cette critique ne retire naturellement rien à la méthode de calcul des taux de survie".

On trouvera dans le tableau 13 un exemple de préparation d'une expérience clinique sur le cancer du poumon. Les valeurs retenues pour μ , σ et c proviennent des données de R.J. Guttman [8]. On voit dans ce tableau qu'il faut attendre trois ans et demi et quatre ans avant de pouvoir faire une estimation correcte.

Tableau 13

Préparation d'une expérience clinique sur le cancer du poumon

$$I + \frac{1}{T} \int_0^T \left[c \Psi - \frac{c \varphi^2}{(1-c)^2 [1-q+z(\varphi-x)]} \right] dt \quad (\text{méthode B})$$

$$c = 0.08 \quad \mu = 1.00 \quad \sigma = 0.33$$

t (mois)	i	N	Erreur type de t
18	0.04	75	0.15
24	0.08	100	0.10
30	0.14	125	0.07
36	0.20	150	0.05
42	0.26	175	0.04
48	0.32	200	0.03

Prise en charge de 50 malades par an.

En conclusion, on peut affirmer qu'un taux de survie à cinq ans, estimé par la méthode de Boag, peut être calculé en quelques minutes, et que les données nécessaires pour mener à bien une telle estimation peuvent être rassemblées par un employé sans compétence spéciale.

REMERCIEMENTS -

J'exprime ici ma profonde reconnaissance envers le Dr. J.W. Boag, pour ses précieux commentaires et observations, lors de la préparation de cet exposé.

Je tiens aussi à remercier la Compagnie I.B.M. pour l'aide matérielle, prise sur leurs fonds de Recherche, qui a permis de mener à bien ces calculs.

BIBLIOGRAPHIE

- [1] J.W. BOAG - Brit. J. Rad. Vol.21 (1948), p.128 et p.189. J. Roy. Stat. Soc. Series B, vol. XI (1949), p.15.
- [2] R.A. FISHER - Phil. Trans. Roy. Soc. A 222 (1922), p.309.
- [3] J.W. BOAG, H.F. FREUNDLICH et B.J. TURNER - Manuscrit in preparation.

- [4] J.H. GADDUM - Nature 156 (1945), p.463.
- [5] J.W. BOAG - Acta, Rad. Vol.54 (1960) p.289.
- [6] Bent SØRENSEN - Acta Rad. Suppl. 169, (1958).
- [7] R.C. TUDWAY, H.F. FREUNDLICH - Br. J. Rad. Vol.33 (1960), p.98.
- [8] Ruth J. GUTTMANN - Radiology vol.76 (1961), p.83.