Cours numéro 5 :

équations différentielles du premier ordre, 3

l'équation logistique

1 Introduction : l'exemple fondamental, évolution de populations

On s'intéresse à l'évolution d'une population en fonction du temps. Le mot population est à prendre ici en un sens très large. Il peut aussi bien s'agir d'une population humaine, qu'animale, des victimes d'une épidémie, d'un ensemble de molécules, de particules, etc.

1.1 Euler, Malthus

Les premiers travaux connus sur ces questions sont ceux d'Euler (1707-1783) *(Recherches générales sur la mortalité et la multiplication du genre humain)*. Dans ce travail, qui date de 1760, Euler calcule notamment la population d'une ville ou d'une province pour une certaine année. On trouvera en annexe 1 quelques détails sur cet article. Retenons simplement qu'il trouve pour la population p_n à l'année n une relation de récurrence $p_{n+1} = \lambda p_n$ qui conduit à une suite géométrique, *a priori* plutôt avec une raison $\lambda > 1$ ce qui donne une croissance rapide.

On peut aussi donner un modèle en temps continu du phénomène en postulant que l'accroissement de population $p'(t)$ est proportionnel à $p(t)$. On obtient une équation différentielle $p'(t) = \lambda p(t)$ qui mène à une fonction exponentielle $p(t) = Ce^{\lambda t}$. Qualitativement, il n'y a pas une différence essentielle entre le modèle discret et le modèle continu.

L'idée d'un accroissement exponentiel de la population est reprise et développée en 1798 par Thomas Robert Malthus (1766-1834). Son analyse conduit à modéliser la population humaine comme une suite géométrique, tandis que la capacité de production se comporterait, elle, comme une suite arithmétique. La distorsion entre les deux, à terme, le conduit à une proposition de limitation des naissances qui est depuis attachée à son nom.

1.2 Verhulst

Le modèle malthusien est remis en cause vers 1840 par Pierre François Verhulst (1804-1849) qui propose un modèle dit logistique qui prend en compte la limitation de la population. Le principe est simple : l'accroissement de la population n'est proportionnel à la population que pour les petites valeurs de celle-ci. Lorsqu'elle croît, des facteurs limitants apparaissent [1] (place ou quantité de nourriture disponible, etc.) qui font qu'il y a une population maximale m. Verhulst postule alors que l'accroissement y' de la population y est proportionnel à la quantité $y(m-y)$. Ce modèle lui permet de donner en 1837 une prévision de la population de la France en 1930 de 40 millions, prévision somme toute raisonnable puisque la population de la France, en réalité, est de 41,5 millions en 1931.

1.3 Aujourd'hui

Le modèle logistique de Verhulst est encore utilisé aujourd'hui dans nombre de questions (en démographie, biologie, médecine, etc.) et il est assez probant. On utilise d'ailleurs en statistique une régression [2] logistique, analogue à la régression linéaire pour modéliser les évolutions de populations qui suivent des courbes “en S”. Ce modèle est important car il permet, avec une fiabilité relativement bonne, de prévoir l'évolution future de la population.

Comme dans le cas exponentiel, il y a deux modèles logistiques, selon que l'on fait varier le temps de manière continue ou discrète. Le modèle à temps continu conduit à une équation différentielle très simple, on en connaît parfaitement les solutions et on sait les interpréter. Nous allons l'étudier ici. En revanche, le modèle discret peut mener, pour certaines valeurs des paramètres, à des comportements beaucoup plus compliqués, voire chaotiques. Vous le verrez avec Pascal Gamblin au cours du projet *Modélisation*.

1. Le point de départ de mon intérêt pour la question est un exercice proposé par une étudiante de CAPES, dans lequel une population de canards en Scandinavie suivait une loi en 6^n. Il est clair que ce type de modèle est absurde. Le seul contre-exemple que je connaisse est celui de la population des lapins en Australie. En effet, 27 lapins furent introduits en 1859 dans le pays. N'ayant pas de prédateurs et une place quasiment infinie, ils se sont développés de manière exponentielle (ils étaient 22 millions 6 ans plus tard), dévorant une bonne partie de la végétation. L'introduction de renards comme prédateurs n'ayant pas été concluante, la seule façon d'enrayer cette invasion a été de répandre la myxomatose, avec un résultat foudroyant. Cette méthode est évidemment à employer avec précaution, un médecin français d'Eure-et-Loir l'ayant utilisée en 1952 pour se débarrasser des lapins qui ravageaient son jardin a répandu l'épidémie dans toute la France.

2. Ou plutôt plusieurs, voir Annexe 3.

2 L'équation logistique

On a donc à résoudre l'équation différentielle $y' = ay(m-y)$ où a, m sont des constantes positives et y une fonction inconnue de la variable t. Il s'agit d'une équation non linéaire, mais facile à intégrer cependant.

2.1 Le théorème de Cauchy

Rappelons qu'une solution de l'équation est une fonction $t \mapsto y(t)$, définie **sur un intervalle** de $\mathbf{R}$, dérivable et dont la dérivée vérifie $y' = ay(m-y)$.

Le second membre de l'équation est une fonction polynomiale, donc C^1, de y (et n'est pas fonction de t). Le théorème de Cauchy-Lipschitz s'applique donc et assure l'existence de solutions (locales mais aussi de solutions maximales, mais attention, pas nécessairement définies sur $\mathbf{R}$ tout entier comme on le verra).

Le point principal de Cauchy-Lipschitz, ici, est l'assertion d'unicité : par un point (t_0, y_0) donné il passe une seule courbe intégrale (voir annexe 2 pour l'idée d'une preuve). Or, il y a deux solutions particulières évidentes : les fonctions constantes $y = 0$ et $y = m$ (en termes de populations : la population nulle, et la population maximale). En vertu du théorème de Cauchy, toute autre solution ne peut donc prendre ces valeurs (les solutions évidentes jouent le rôle de barrières). On peut donc maintenant supposer $y(t) \neq 0$ et $y(t) \neq m$ pour tout t.

2.2 Méthode 1 : changement de fonction

C'est la méthode utilisée en terminale.

Comme y ne s'annule pas, on peut poser $z(t) = \dfrac{1}{y(t)}\cdot$ On obtient l'équation différentielle $z' + amz = a$ dont les solutions sont de la forme $\frac{1}{m} + Ce^{-amt}$ et on a donc pour y la solution $y(t) = \dfrac{m}{1 + Ke^{-amt}}$ (où K est égal à mC). On retrouve la solution constante $y = m$ en prenant $K = 0$; la solution $y = 0$ correspondrait au cas $K = \infty$).

2.3 Méthode 2 : éléments simples

Comme y est distinct de 0 et de m on écrit $\dfrac{y'}{y(m-y)} = a$, on décompose la fraction en éléments simples : $\dfrac{y'}{y} + \dfrac{y'}{m-y} = am$. On prend alors des primitives des deux membres et on a (en supposant par exemple $0 < y < m$)

$\ln \dfrac{y}{m-y} = amt + C$, d'où $\dfrac{y}{m-y} = K'e^{amt}$. On trouve la même forme de solutions (avec $K = 1/K'$).

2.4 Allure des solutions

Lorsque K est positif, la solution est définie sur $\mathbf{R}$ tout entier et elle est comprise strictement entre 0 et m. On obtient ainsi les solutions qui ont un sens "physique". Ces solutions sont croissantes. Inutile de calculer explicitement la dérivée pour le voir car on a l'équation différentielle $y' = ay(m-y)$! Les solutions ont des limites 0 et m en $-\infty$ et $+\infty$. Elles ont aussi une inflexion en laquelle y vaut $m/2$. Là encore, inutile de calculer la dérivée seconde. En dérivant l'équation on a $y'' = ay'(m-2y)$ et on voit que y'' est positif tant que y est plus petit que $m/2$ et négatif ensuite. Avec l'expression de $y(t)$ on voit que le point d'inflexion est en $t = \frac{\ln K}{am}$.

Voici l'allure de la courbe pour $K > 1$ (le point d'inflexion a un t positif).

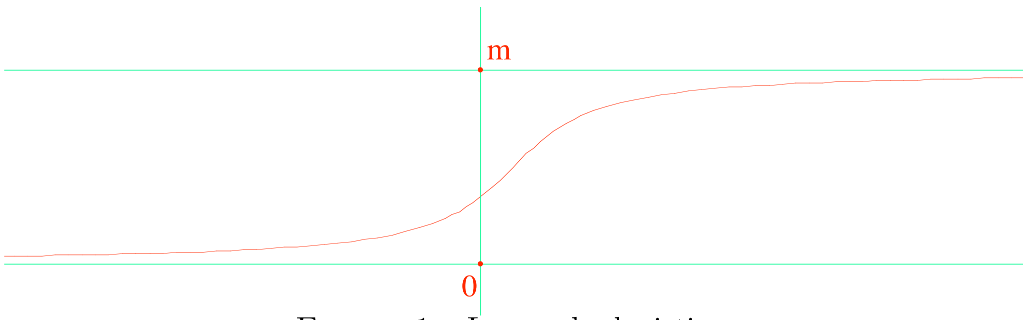

FIGURE 1 – La courbe logistique

Regardons le cas $K < 0$. Cette fois, la fonction $y(t) = \dfrac{m}{1 + Ke^{-amt}}$ n'est pas définie au point $t = \ln(-K)/am := \tau$. On a, en fait, deux solutions de l'équation, l'une définie sur $]-\infty, \tau[$ est négative et décroît de 0 à $-\infty$, l'autre, définie sur $]\tau, +\infty[$ est positive et décroît de $+\infty$ à m. On est dans le cas où les solutions maximales ne sont pas définies partout.

En regardant la variation des courbes intégrales selon le temps, on voit qu'elles s'éloignent de la solution nulle et se rapprochent de la solution m, ce qui signifie que la solution nulle est un équilibre instable, tandis que m est un équilibre stable. La raison est que, si l'on pose $H(y) = ay(m-y)$ on a $H'(y) = a(m-2y)$ et H' est positif en 0 mais négatif en m. Or, au voisinage d'une valeur d'équilibre y_0 (qui vérifie $H(y_0) = 0$), on a $H(y) = k(y-y_0) + o(y-y_0)$ avec $k = H'(y_0)$ et l'équation est "comme" l'équation linéaire $y' = k(y - y_0)$ dont les solutions sont $y = y_0 + e^{kt}$. Si k est positif, ces solutions s'écartent de y_0 tandis qu'elles s'en rapprochent si k est négatif.

3 La population des États-Unis

Ce qui suit est un exercice proposé en dossier de CAPES (à Orsay).

3.1 L'exercice

Le modèle de Verhulst pour décrire l'évolution d'une population (animale ou humaine) est le suivant. On appelle $p(t)$ la population au temps t (l'unité de temps est l'année, l'unité de population le millier d'individus). Pour des raisons de place disponible, de nourriture, etc. on postule que cette population admet une valeur maximum m et que l'accroissement de la population pendant un petit intervalle de temps est "proportionnel" à la fois à l'intervalle de temps, à la population $p(t)$ (plus il y a d'individus, plus il y a de naissances), mais aussi à l'écart entre population théorique maximum et population actuelle $m - p(t)$ (quand on approche du maximum de la population, la nourriture disponible devient rare et la mortalité augmente). Ce modèle conduit donc à l'équation différentielle (1) *:* $p' = a\,p\,(m - p)$ *où a est une constante positive.*

On suppose que $p(t)$ et $m - p(t)$ sont des nombres réels (et pas seulement des entiers) > 0.

1) On pose $g(t) = \dfrac{1}{p(t)}\cdot$ Montrer que g vérifie une équation différentielle (2) *de la forme $g' = \alpha g + \beta$ où α et β sont deux réels que l'on précisera.*

2) Résoudre l'équation (2) *et en déduire que les solutions de l'équation* (1) *sont de la forme :* $p(t) = \dfrac{m}{1 + \lambda e^{-amt}}$ *où λ est un réel > 0.*

3) Étudier les variations de la fonction p : signe de la dérivée p', variation de p' (on s'intéressera en particulier aux points où p'' s'annule). (Pour résoudre ces questions on pourra utiliser l'équation (1)*).*

Déterminer la limite de p quand t tend vers l'infini et donner l'allure du graphe de p.

4) La population $p(t)$ des États-Unis de 1790 *à* 1910 *a été étudiée par Pearl et Reed en* 1920 *au moyen de cette méthode. Avec l'origine des temps en* 1790*, les valeurs des paramètres sont alors données par $m = 197273$, $\lambda = 49,2$ et $a = 158 \times 10^{-9}$. Calculer avec la formule ci-dessus la population pour toutes les années multiples de* 10 *entre* 1790 *et* 1910*. Quand la population maximale sera-t-elle atteinte à un million près si l'on en croit ce modèle ?*

3.2 Questions complémentaires

Q2. La population réelle des États-Unis pour la période 1790 − 1910 (en milliers d'habitants) est donnée par le tableau ci-dessous. Commenter la différence entre les résultats calculés et les résultats réels. Comment ont été trouvés les paramètres[3] m, λ et a ?

1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910
3929	5308	7240	9638	12866	17069	23192	31443	38558	50156	62948	75995	91972

La population précise aux alentours de l'année 1914 est donnée par le tableau ci-dessous. En admettant que le modèle est bien adapté à la situation, comment peut-on prévoir la population maximale en examinant seulement ce tableau ?

1910	1911	1912	1913	1914	1915	1916	1917	1918
91972	93512	95055	96599	98144	99690	101235	102779	104320

Q3. La population américaine réelle pour les années 1920-2000 (en milliers d'habitants) est donnée par le tableau ci-dessous. La fonction $p(t)$ proposée à la question 4) de l'exercice est-elle encore plausible aujourd'hui ? En utilisant la fonction régression logistique de la calculatrice, donner une fonction qui prenne en compte l'évolution récente de la population. Quelle population peut-on attendre en 2060 ? Avec ce nouveau modèle, quand atteindra-t-on en principe la population maximale à 1 million d'habitants près ?

1920	1930	1940	1950	1960	1970	1980	1990	2000
106022	123203	132165	151326	179323	203302	226342	248710	281422

3.3 Corrigé

L'exercice reprend la résolution qui a été menée ci-dessus. Pour la question 4), il suffit de rentrer la fonction dans la calculatrice et de la tabuler. On a

3. Les paramètres donnés ci-dessus ont été arrondis. Voici les valeurs des paramètres trouvées par Pearl et Reed : $m = 197273,583363$, $\lambda = 49,2096165342$ et $a = 158,86333780255 \times 10^{-9}$.

les résultats suivants :

1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910
3929	5328	7205	9711	13028	17373	22987	30111	38949	49612	62050	76001	90978

On voit que les résultats sont très proches des valeurs réelles. Le maximum, à un million près, est atteint vers 2084 selon ce modèle. En réalité, il est déjà dépassé en 1970.

3.3.1 Réponse à la question Q2

La méthode pour estimer les trois paramètres m, λ et a consiste à utiliser trois valeurs de p en trois points régulièrement espacés. On prendra ici les temps $t = 0, 60, 120$ qui correspondent aux années 1790, 1850 et 1910. On a ainsi à résoudre un système (non linéaire) de trois équations en les trois inconnues m, λ et a ou plutôt m, λ et $u = e^{-60am}$:

$$\begin{cases} \frac{1+\lambda}{m} = \frac{1}{p(0)} = \alpha = 1/3929 \\ \frac{1+\lambda u}{m} = \frac{1}{p(60)} = \beta = 1/23192 \\ \frac{1+\lambda u^2}{m} = \frac{1}{p(120)} = \gamma = 1/91972 \end{cases}$$

L'astuce est de retrancher ces quantités, ce qui donne $u = \dfrac{\gamma - \beta}{\beta - \alpha}\cdot$ On en déduit λ en calculant α/β. On trouve $\lambda = \dfrac{\beta - \alpha}{\alpha u - \beta}$ et enfin on a m avec la formule $m = \dfrac{1+\lambda}{\alpha}\cdot$

En fait, une bonne calculatrice résout le système entier sans peine. On trouve les trois valeurs données (non arrondies), et, si on calcule avec ces valeurs non arrondies, on voit que les valeurs de $p(t)$ coïncident avec les données en les trois points choisis.

Pour prévoir la population maximale (ou le nombre total de morts dans une épidémie, ou tout autre calcul de ce type), il y a une méthode évidente lorsque le phénomène relève de la loi logistique : on repère le point d'inflexion, on sait qu'en ce point la valeur de p est $m/2$ et on déduit m. Pour déterminer le point d'inflexion, il suffit d'examiner la variation de p et de repérer le maximum de la dérivée ou – comme dirait un journaliste – le moment où l'augmentation diminue. Sur l'exemple donné, si on tabule les différences $p(n) - p(n-1)$ aux alentours de 1914, on obtient :

1910	1911	1912	1913	1914	1915	1916	1917	1918
	1540	1543	1544	1545	1546	1545	1544	1541

On voit que le maximum est atteint en 1915. Comme la valeur de p à cette date est 99690, on peut s'attendre à ce que m soit de l'ordre de $2 \times 99690 = 199380$ (en réalité $m = 197273$ selon ce modèle).

3.3.2 Réponse à la question Q3

On voit sur le tableau donné que les prévisions du modèle sont correctes jusque vers les années 1940, mais très fausses au-delà. Les valeurs de p sont par exemple de 159230 (au lieu de 179323 en réalité) en 1970) et de 184677 (au lieu de 281422) en 2000. Il y a sans doute de bonnes raisons (démographiques, politiques, économiques) qui expliquent cette augmentation de la population après la seconde guerre mondiale, mais j'ignore lesquelles[4]. Toujours est-il qu'il faut réviser le modèle en intégrant les nouvelles données. Plutôt que de se livrer à un calcul comme ci-dessus, on peut utiliser ce qu'on appelle les *régressions logistiques* qui fonctionnent sur le même principe que la régression linéaire : on rentre les données et on récupère les valeurs des paramètres.

Sur la calculatrice TI Voyage, avec les données de 1790 à 2000 on a : $m = 440880$, $\lambda = 56,484$ et $a = 4,9 \times 10^{-8}$. Avec ces valeurs, le maximum est atteint à un million près vers 2260 et la population estimée en 2060 est 378280 (en milliers d'habitants). L'avenir dira[5] si cette prévision est correcte.

En fait, voir Annexe 3, il y a plusieurs méthodes pour faire ces régressions et elles ne sont pas vraiment convergentes. Dans le cas d'une population humaine, il faut donc utiliser ce modèle avec la plus grande circonspection.

4. J'imagine une forte immigration, mais je n'ai pas d'élément objectif pour le dire.

5. *Je serai sous la terre, et fantôme sans os*
Par les ombres myrteux je prendrai mon repos.
Je ne pourrai donc pas vérifier, mais vous pourrez le faire, j'espère.

4 Bernoulli et la petite vérole

4.1 Introduction

Il s'agit d'un travail [6] de Daniel Bernoulli (1700-1782), qui date de 1760 et qui propose une modélisation d'une épidémie de variole (on disait à l'époque "petite vérole", de nos jours cette maladie a pratiquement disparu, mais au cours du XVIII-ième siècle elle a tué 50 millions d'européens ; on rappelle que la population de l'Europe était à cette époque d'environ 110 millions). Le but est de savoir si la technique d'inoculation [7] de la maladie présentait pour la population plus d'avantages que de risques.

Cette technique était très controversée et le restera même après le papier de Bernoulli.

4.2 Les hypothèses de Bernoulli

Il y a un point qu'il faut bien comprendre pour mesurer l'intérêt du travail de Bernoulli. Comme il le dit, il s'agit simplement de mesurer l'impact de la variole sur la mortalité. Bien entendu, il suffirait pour cela d'avoir des statistiques correctement tenues sur cette mortalité, qui donneraient évidemment le résultat. Le travail de Bernoulli ne prend son sens que **parce qu'il n'existe pas, à l'époque [8], de telles statistiques**.

La seule chose dont dispose Bernoulli pour son étude est une table de survie construite par Halley (celui de la comète) pour la ville de Breslau (aujourd'hui Wrocław) sur une population de 1300 personnes. Cette table comporte deux colonnes : une pour le temps t et une pour le nombre de survivants au temps t (voir ci-dessous).

On observe une population de $N = 1300$ personnes de la naissance ($t = 0$) à 24 ans ($t = 24$). Bernoulli introduit deux nombres : $N(t)$ est le nombre de survivants parmi cette population à l'instant t (ce nombre est connu, donné

6. *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir.*

7. Cette technique consiste à inoculer volontairement la maladie, qu'on ne peut attraper deux fois. Elle provenait de Chine et avait été popularisée en Europe par Lady Montagu qui l'avait rencontrée dans l'empire ottoman. Attention, il ne s'agit pas d'une vaccination qui consiste à inoculer une forme affaiblie ou voisine de la maladie. L'idée de vaccination est un peu plus récente. Elle est due à Jenner (1796) qui avait constaté que les filles de ferme qui trayaient les vaches ne contractaient jamais la variole. En effet, elles étaient immunisées par leur contact avec une forme bénigne de la maladie qui touchait les vaches : la vaccine ! Le mot vaccin vient de là et il date de 1880 avec Pasteur.

8. Bernoulli dit qu'il va en exister bientôt à Londres et qu'il critiquera lui-même son travail à la lueur de ces statistiques.

par la table, en particulier on a $N(0) = N$) et $x(t)$ est le nombre de ceux parmi les $N(t)$ qui n'ont pas encore eu la variole à l'instant t. On a aussi $x(0) = N$, mais ensuite, ce nombre est **inconnu**, faute de statistiques. Parmi les gens qui n'ont pas encore eu la variole on suppose que le taux d'infection par cette maladie (c'est-à-dire la proportion de gens qui vont la contracter dans l'année qui suit) est a et qu'il ne dépend pas de l'âge, voir la note ci-dessous. Ensuite, parmi les personnes atteintes par la variole, on suppose que la proportion de celles qui vont mourir dans l'année est b. Dans son travail, Bernoulli prend, à la fois, $a = b = 1/8$ (on notera que a et b ne dépendent pas de t, voir en note ce que dit Bernoulli [9]). Enfin il suppose qu'il y a une fonction inconnue $m(t)$ qui désigne la proportion, sur toute la population, des morts par d'autres causes que la variole entre t et $t+1$. Cette fois, cette fonction dépend de t : plus on vit, plus on se rapproche de sa mort ...

9. Voilà ce que dit Bernoulli : *Quel est le risque annuel à différents âges d'être surpris par la petite vérole, pour ceux qui ne l'ont pas eue, et quel est celui d'en mourir pour ceux qui en sont attaqués ? Il est vrai que nous ne sommes pas directement informés sur ces deux éléments, mais d'autres connaissances m'ont paru y suppléer avec beaucoup de vraisemblance ... Nous voyons que la petite vérole n'attaque guère que les enfants et les jeunes gens, nous sommes d'abord portés à juger que la seule jeunesse y est exposée par sa constitution ... mais un peu plus de réflexion nous fait bientôt revenir sur cette erreur. S'il est rare que la petite vérole attaque les adultes, c'est qu'il est rare que les adultes ne l'aient pas eue, et qu'elle n'attaque jamais, ou presque jamais, deux fois la même personne. C'est ici le caractère essentiel de cette maladie ... Il est donc vraisemblable que les vieillards qui n'ont pas eu la petite vérole, courent le même risque de l'avoir que les jeunes gens. Pour peu que ce risque diminuât avec l'âge, ce devrait être une chose sans exemple d'avoir la petite vérole à l'âge de 70 ans, et on en connaît plusieurs. Je n'ai donc plus hésité d'adopter mon premier principe, qui est que tant qu'on a pas eu la petite vérole, on court continuellement le même risque de l'avoir. Nous n'avons encore aucune observation qui nous oblige à renoncer à cette supposition, et les lois de la Nature les plus simples sont toujours les plus vraisemblables ... Quant au risque annuel d'être attaqué par la petite vérole, pour ceux qui ne l'ont pas eue, j'ai cru ne pouvoir satisfaire aux notions générales que nous avons sur cette maladie, qu'en la supposant d'un huitième, ce rapport de 1 sur 8 étant supposé constant ... pour peu qu'on voulût changer ledit rapport de 1 sur 8, l'effet qui en rejaillirait sur les adultes et sur les vieillards serait trop sensible, et peut-être manifestement faux ... Disons encore un mot sur le risque de la petite vérole pour ceux qui en sont attaqués : la plupart l'ont fait d'un septième ; je l'ai un peu diminué, en le faisant d'un huitième : deux raisons m'y ont engagé, la première est qu'on apprend exactement tous ceux qui en meurent, et qu'on ne saurait apprendre si exactement tous ceux qui ont la maladie ; la seconde, est que le rapport de 1 sur 7 ferait la mortalité variolique trop grande par rapport à la mortalité entière, pendant que celui de 1 sur 8 est entièrement conforme à l'observation la mieux constatée, qui est que la petite vérole enlève la treizième partie du total des morts...* ≫

4.3 Mise en équation

On commence par une mise en équation discrète pour passer de t à $t+1$ (en années). On obtient le système :

$$\begin{cases} x(t+1) - x(t) = -ax(t) - m(t)x(t), \\ N(t+1) - N(t) = -abx(t) - m(t)N(t). \end{cases}$$

Pour la première équation, le nombre de gens qui n'ont pas encore eu la variole diminue entre t et $t+1$. La diminution provient de deux sources : ceux qui ont attrapé la variole, et on sait que la proportion en est a, et ceux qui sont morts d'autre chose, avec une proportion $m(t)$. Pour la seconde, c'est pareil, sauf que seuls b parmi les $ax(t)$ ayant contracté la variole en décèdent.

On fait ensuite le même raisonnement en fractionnant l'année en intervalles de largeur Δt et en supposant que les évolutions sont régulières tout au long de l'année :

$$\begin{cases} x(t+\Delta t) - x(t) = (-ax(t) - m(t)x(t))\Delta t, \\ N(t+\Delta t) - N(t) = (-abx(t) - m(t)N(t))\Delta t. \end{cases}$$

En divisant par Δt et en faisant tendre cette quantité vers 0 on obtient un système différentiel :

$$\begin{cases} x'(t) = -ax(t) - m(t)x(t), \\ N'(t) = -abx(t) - m(t)N(t). \end{cases}$$

4.4 Résolution

On pose $f(t) = \dfrac{x(t)}{N(t)}\cdot$ On a $f(0) = 1$ et $f'(t) = -af(t) + abf(t)^2$ (on note que le terme inconnu $m(t)$ disparaît). On reconnaît une équation du type logistique, que l'on intègre grâce au changement de fonction [10] $g = 1/f$. On tombe sur l'équation $g' = ag - ab$ dont les solutions sont de la forme $g(t) = b + Ce^{at}$. En tenant compte de la valeur initiale de f, on a $f(t) =$

10. On suppose $f(t) \neq 0$ ou encore $x(t) \neq 0$: tout le monde n'a pas la variole !

$\dfrac{1}{b+(1-b)e^{at}}\cdot$ En remplaçant f par x/N et en tenant compte des valeurs numériques, on a $x(t) = \dfrac{8}{1+7e^{t/8}}\, N(t)$.

Comme on connaît $N(t)$, on en déduit $x(t)$. C'est ce que fait Bernoulli, et il en déduit la Table 1 ci-dessous. Dans cette table figurent donc t (de 0 à 24), puis $N(t)$ (nombre de survivants au temps t, donné par la table de Halley), puis $x(t)$ (ceux parmi les $N(t)$ qui n'ont pas encore eu la variole, calculé avec la formule ci-dessus), puis $N(t) - x(t)$ (ceux qui l'ont eue). On trouve ensuite le nombre $v(t) = \dfrac{x(t-1)+x(t)}{16}\cdot$ C'est le nombre de ceux qui vont attraper la variole pendant l'année t. Ici Bernoulli ne prend pas $x(t)/8$ comme on s'y attendrait, mais une moyenne prise entre l'année en cours et l'année précédente. La colonne suivante est $m_v(t) = v(t)/8$: nombre de morts dûs à la variole dans l'année t, et la suivante est le cumul de ces morts $s(t) = \sum_{i=0}^{t} m_v(i)$. Enfin, la dernière colonne est le nombre de morts $m_{nv}(t)$ relevant d'une autre cause (c'est donc $m_{nv}(t) = N(t-1) - N(t) - m_v(t))$ ou aussi, avec les notations précédentes, $m(t-1)N(t-1)$. On note qu'on récupère ainsi la fonction inconnue m.

Pour la lisibilité, j'ai reproduit les tables de Bernoulli ci-dessous. Le lecteur trouvera en annexe 4 les fac-similés des originaux.

Cette table, et en particulier la colonne $m_{nv}(t)$, permet à Bernoulli d'estimer quelle serait la population $y(t)$ à l'année t si personne ne mourait [11] de la variole. Au départ, il y a 1300 personnes. Il n'y en a plus que [12] 1000 à l'année 1, dont 17,1 morts de la variole. Si personne ne mourait de la variole, il y aurait donc $y(1) = 1017,1$ survivants à l'année 1. Pour calculer les suivants, Bernoulli utilise la fonction $m(t)$ calculée ci-dessus. La proportion de morts par autre chose que la variole entre t et $t+1$ étant $m(t)$, si celle-ci est éradiquée, on a la relation $y(t+1) = y(t) - y(t)m(t) = y(t) \times \big(1 - \dfrac{m_{nv}(t+1)}{N(t)}\big)\cdot$ C'est ainsi que Bernoulli obtient la Table 2 ci-dessous, dans laquelle il fait apparaître le gain $y(t) - N(t)$ que constituerait une éradication de la variole. Sur la population de 1300 personnes, entre 10 et 24 ans, ce serait chaque année une moyenne de plus de 70 vies qui seraient épargnées.

11. *Mais si l'on ne mourait plus,*
J'crèverais de faim sur mon talus,
J'suis un pauvre fossoyeur.

12. On notera les progrès accomplis face à la mortalité infantile !

Table 1

t	$N(t)$	$x(t)$	$N(t)-x(t)$	$v(t)$	$m_v(t)$	$s(t)$	$m_{nv}(t)$
0	1300	1300	0				
1	1000	896	104	137	17,1	17,1	283
2	855	685	170	99	12,4	29,5	133
3	798	571	227	78	9,7	39,2	47
4	760	485	275	66	8,3	47,5	30
5	732	416	316	56	7	54,5	21
6	710	359	351	48	6	60,5	16
7	692	311	381	42	5,2	65,7	12,8
8	680	272	408	36	4,5	70,2	7,5
9	670	237	433	32	4	74,2	6
10	661	208	453	28	3,5	77,7	5,5
11	653	182	471	54,4	3	80,7	5
12	646	160	486	21,4	2,7	83,4	4,3
13	640	140	500	18,7	2,3	85,7	3,7
14	634	123	511	16,6	2,1	87,8	3,9
15	628	108	520	14,4	1,8	89,6	4,2
16	622	94	528	12,6	1,6	91,2	4,4
17	616	83	533	11	1,4	92,6	4,6
18	610	72	538	9,7	1,2	93,8	4,8
19	604	63	541	8,4	1	94,8	5
20	598	56	542	7,4	0,9	95,7	5,1
21	592	48,5	543	6,5	0,8	96,5	5,2
22	586	42,5	543	5,6	0,7	97,2	5,3
23	579	37	542	5	0,6	97,8	6,4
24	572	32,4	540	4,4	0,5	98,3	6,5

Table 2 (La valeur $y(13) = 741,1$ est erronée : $y(13) = 714,1$.)

t	$N(t)$	$y(t)$	$y(t) - N(t)$		t	$N(t)$	$y(t)$	$y(t) - N(t)$
0	1300	1300	0		13	640	741,1	74,1
1	1000	1017,1	17,1		14	634	709,7	75,7
2	855	881,8	26,8		15	628	705	77
3	798	833,3	35,3		16	622	700,1	78,1
4	760	802	42		17	616	695	79
5	732	779,8	47,8		18	610	689,6	79,6
6	710	762,8	52,8		19	604	684	80
7	692	749,1	57,2		20	598	678,2	80,2
8	680	740,9	60,9		21	592	672,3	80,3
9	670	734,4	64,4		22	586	666,3	80,3
10	661	728,4	67,4		23	579	659	80
11	653	722,9	69,9		24	572	651,7	79,7
12	646	718,2	72,2		25	565	644,3	79,3

4.5 Analyse des résultats

Voici les mots de Bernoulli : *je me suis attaché surtout, à exposer dans une même Table les deux états de l'humanité, l'un tel qu'il est effectivement, et l'autre tel qu'il serait si on pouvait affranchir de la petite vérole tout le genre humain. J'ai pensé que le parallèle de ces deux états en expliquerait mieux la différence et le contraste, que ne ferait le plus ample commentaire...*

Il montre aussi que l'espérance de vie passerait de 26 ans et 7 mois à 29 ans et 9 mois si la variole était totalement éradiquée. Pour cela, il ajoute tous les nombres de la colonne $N(t)$ (de 1 jusqu'à 84 ans), ce qui lui donne la somme des durées de vie des 1300 personnes, il trouve 34605 et l'espérance de vie s'obtient en divisant ce nombre par 1300. En effet, l'espérance, *a priori*, est

$$\frac{1}{1300}\sum_{t\geq 1} t(N(t)-N(t+1)), \text{ mais on peut la récrire } \frac{1}{1300}\sum_{t\geq 1} N(t)(t-(t-1)).$$
(Dans la figure ci-dessous, cela revient à passer de la somme en lignes à la somme en colonnes.) Pour l'espérance dans un monde sans variole, il suffit d'ajouter tous les termes de la colonne $y(t)$ au lieu de $N(t)$. On trouve 38513 et on obtient le résultat.

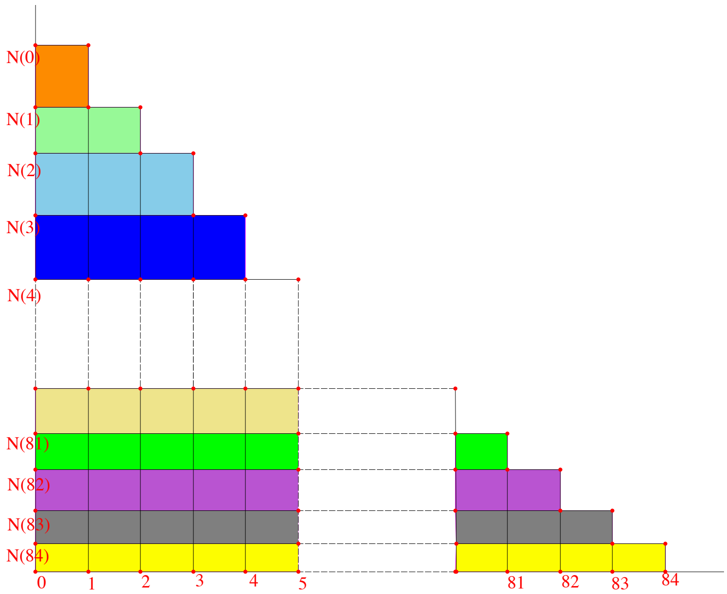

FIGURE 2 – Calcul de l'espérance

4.5.1 En faveur de l'inoculation

Comme on l'a dit, l'intérêt essentiel du travail de Bernoulli est de suppléer l'absence de statistiques fiables concernant la mortalité due à la variole. C'est ce que nous avons vu ci-dessus. Ce qui est plus intéressant est d'estimer l'effet de l'inoculation, ce que fait Bernoulli ensuite. Le défaut de cette technique d'inoculation est qu'elle peut être mortelle, et c'est l'argument principal de ses opposants. D. Bernoulli postule qu'on a une chance sur 200 de mourir de la variole à cause de l'inoculation (les tables démographiques semblent indiquer que ce rapport était un maximum [13]). Il calcule alors ce que deviendrait l'espérance de vie si l'on procédait à une inoculation généralisée dans la première année. Comme il y a au départ 1300 personnes et qu'il en succombe 1/200-ième par l'inoculation, cela fait environ 6, 5 morts (à comparer avec les

13. Ce taux est de 1 sur 600 à Londres en 1755.

79, 3 vies gagnées, voir ci-dessus). Comme l'espérance de vie de ces personnes aurait été de 29, 75 ans, cela fait une perte totale de 193 années sur les 38513. Il reste 38320 qui donne, en divisant par 1300, une espérance de vie de 29 ans 7 mois et 10 jours, chiffre qui cette fois tient compte des risques de l'inoculation. Cela permet d'évaluer le tribut payé à l'inoculation à 1 mois et 20 jours.

Ce calcul conduisit Bernoulli à prendre parti pour l'inoculation préventive en dépit du risque individuel que cette mesure comportait. Un long débat mathématique et philosophique, auquel D'Alembert (1717-1783) prit une part active, s'ensuivit. C'est l'aspect statistique lui-même qui posait problème à D'Alembert, ce dernier estimant que le calcul du risque n'a pas le même sens vu dans la masse et vu par une mère et son enfant. D'Alembert écrit : *Je suppose avec monsieur Bernoulli que le risque de mourir de l'inoculation soit de 1 sur 200. Cela posé, il me semble que pour apprécier l'avantage de l'inoculation, il faut comparer, non la vie moyenne de 34 ans à la vie moyenne de 30, mais le risque de 1 sur 200 auquel on s'expose de mourir en un mois par l'inoculation (et cela à l'âge de trente ans, dans la force de la santé et de la jeunesse) à l'avantage éloigné de vivre quatre ans de plus au bout de 60 ans lorsqu'on sera beaucoup moins en état de jouir de la vie ... Voilà, il n'en faut point douter, ce qui rend tant de personnes, et surtout tant de mères, peu favorables parmi nous à l'inoculation.*

Le texte qui suit donne une idée de l'ambiance de l'époque. Dans l'Encyclopédie de Diderot (Denis, 1713-1784), on trouve l'article "Inoculation" suivant écrit par le médecin Théodore Tronchin (1709-1781) :

Ce nom synonyme d'insertion a prévalu pour désigner l'opération par laquelle on communique artificiellement la petite vérole, dans la vue de prévenir le danger et les ravages de cette maladie contractée naturellement. [...] Quand toute la France serait persuadée de l'importance et de l'utilité de cette pratique, elle ne peut s'introduire parmi nous sans la faveur du gouvernement ; et le gouvernement se déterminera-t-il jamais à la favoriser sans consulter les témoignages les plus décisifs en pareille matière ? C'est donc aux facultés de théologie et de médecine, c'est aux académies, c'est aux chefs de la magistrature, aux savants, aux gens de lettres, qu'il appartient de bannir des scrupules fomentés par l'ignorance, et de faire sentir au peuple que son utilité propre, que la charité chrétienne, que le bien de l'État, que la conservation des hommes sont intéressés à l'établissement de l'inoculation. Quand il s'agit du bien public, il est du devoir de la partie pensante de la nation d'éclairer ceux qui sont susceptibles de lumière, et d'entraîner par le poids de l'autorité cette foule sur qui l'évidence n'a point de prise.

5 Exercices

5.1 Une épidémie

Le but de l'exercice est d'étudier le modèle épidémiologique SIS, qui s'applique notamment pour des maladies assez bénignes comme le rhume, dont l'issue est rarement fatale, mais où les individus guéris peuvent être infectés à nouveau. On note N la population totale, supposée constante, et $y(t)$ le nombre total d'individus infectés au temps t.

On suppose que la variation dy de y, rapportée à un temps infinitésimal dt provient de deux sources :

• les nouvelles contaminations que l'on suppose proportionnelles au produit $y(N-y)$ (avec un coefficient α positif),

• les guérisons que l'on suppose proportionnelles à y (avec un coefficient β positif, supposé constant).

1) Discuter les hypothèses ci-dessus et notamment la raison d'être du terme $y(N-y)$.

2) Montrer que l'équation différentielle en y qui traduit ces hypothèses est la suivante :

$$y' = (\alpha N - \beta)y - \alpha y^2.$$

3) Intégrer cette équation. Discuter le signe de $\alpha N - \beta$. Indiquer rapidement l'allure du graphe de y dans le cas $\alpha N - \beta > 0$.

4) Le tableau ci-dessous indique le nombre total de cas dans une petite ville pendant les mois de décembre-janvier en fonction du nombre de jours écoulés (ainsi, le 28-ième jour, il y avait eu, en tout, 1940 cas). En supposant que le modèle étudié ci-dessus est pertinent, utiliser ce tableau pour donner une prévision (approximative) du maximum de cas d'infection.

20	22	24	26	28	30	32	34	36	38	40	42	44
480	697	1000	1408	1940	2600	3366	4194	5014	5791	6452	6986	7396

On donne $\alpha = 2,382 \times 10^{-5}$ et $\beta = 0,1572$. Donner une estimation de la population de la ville.

Le modèle permet-il de rendre compte de ce qui se passera au printemps ?

5.2 Les lapins

Une population de lapins est introduite dans une petite île. On appelle $p(t)$ la population de lapins au temps t (l'unité de temps est l'année). On suppose que les lapins n'ont pas de prédateurs, mais, qu'en raison de la limitation de la quantité de nourriture disponible, il y a une quantité maximum m de lapins qui peut survivre sur l'île.

Le modèle proposé pour étudier l'évolution de cette population consiste à dire que l'accroissement de la population pendant un petit intervalle de temps est "proportionnel" à la fois à l'intervalle de temps, à la quantité de lapins présente $p(t)$ (plus il y a de lapins, plus ils peuvent faire de petits), mais aussi à l'écart entre population théorique maximum et population actuelle $m - p(t)$ (quand on approche du maximum de la population, la nourriture disponible devient rare et la mortalité augmente). Ce modèle conduit donc à l'équation différentielle (1) : $p' = a\,p\,(m - p)$ où a est une constante positive.

On suppose que $p(t)$ et $m - p(t)$ sont des nombres réels (et pas seulement des entiers) > 0.

1) On pose $g(t) = \dfrac{1}{p(t)}\cdot$ Montrer que g vérifie une équation différentielle (2) de la forme $g' = \alpha g + \beta$ où α et β sont deux réels que l'on précisera.

2) Résoudre l'équation (2) et en déduire que les solutions de l'équation (1) sont de la forme : $p(t) = \dfrac{m}{1 + \lambda e^{-amt}}$ où λ est un réel > 0.

3) Étudier les variations de la fonction p : signe de la dérivée p', variation de p' (on s'intéressera en particulier aux points où p'' s'annule), limite de p quand t tend vers l'infini (on pourra utiliser l'équation (1)). Donner l'allure du graphe de p.

4) Voici le tableau des valeurs de p aux alentours de la troisième année :

t	$2,5$	$2,6$	$2,7$	$2,8$	$2,9$	3	$3,1$	$3,2$	$3,3$
p(t)	347	377	408	440	473	507	539	570	600

Comment ce tableau permet-il, en supposant que le modèle est pertinent, d'avoir une idée approximative de la valeur maximum m ?

5) On suppose qu'on a $m = 1000$, $p(0) = 20$ et $p(1) = 70$. Calculer a et λ. Au bout de combien de temps aura-t'on atteint la moitié de la population limite ? la population limite à une unité près ?

6 Annexe 1 : Euler

Voici ce qu'Euler dit *loc. cit.* p. 157 : *Car si nous posons le nombre de tous les vivans l'année prochaine* $= nM$*, celui des vivans à présent étant* $= M$*, il faut tirer la valeur de* n *de l'équation trouvée*

$$\frac{M}{N} = 1 + \frac{(1)}{n} + \frac{(2)}{n^2} + \frac{(3)}{n^3} + \frac{(4)}{n^4} + \&c.$$

.

Avec des notations plus modernes [14] , on a la formule suivante :

$$p_n = q_n\big(1 + \frac{\pi_1}{\lambda} + \frac{\pi_2}{\lambda^2} + \cdots + \frac{\pi_{100}}{\lambda^{100}} + \cdots\big)$$

où p_n est la population de la ville à l'année n, q_n le nombre de naissances de cette même année, π_k la probabilité (calculée à partir de tables, voir [Euler] p. 152 ou ci-dessous) de vivre k années dans la région donnée et $\lambda = q_{n+1}/q_n$ le rapport entre le nombre de naissances de l'année $n+1$ et de l'année n, supposé indépendant de n. Le nombre de naissances à l'année $n - k$ est donc $\frac{q_n}{\lambda^k}$, ce qui donne la formule ci-dessus. Euler arrête généralement la sommation à 100, considérant qu'au-delà la probabilité de survie est négligeable. Avec cette formule, il est clair que le rapport entre la population à l'année $n + 1$ et celle à l'année n est égal au rapport des nombres de naissances donc à λ. Ainsi, on a la formule $p_{n+1} = \lambda p_n$, qui conduit à une suite géométrique [15].

14. Euler note (k) ce que je note π_k et n ce que je note λ.

15. Comme on l'a dit, on a donc un accroissement (ou une diminution) exponentiel selon la position de λ par rapport à 1.

La table de mortalité utilisée par Euler

152

(1) = 0,804	(31) = 0,499	(61) = 0,264	(91) = 0,006
(2) = 0,768	(32) = 0,490	(62) = 0,254	(92) = 0,004
(3) = 0,756	(33) = 0,482	(63) = 0,245	(93) = 0,003
(4) = 0,709	(34) = 0,475	(64) = 0,235	(94) = 0,002
(5) = 0,688	(35) = 0,468	(65) = 0,225	(95) = 0,001
(6) = 0,676	(36) = 0,461	(66) = 0,215	
(7) = 0,664	(37) = 0,454	(67) = 0,205	
(8) = 0,653	(38) = 0,446	(68) = 0,195	
(9) = 0,646	(39) = 0,439	(69) = 0,185	
(10) = 0,639	(40) = 0,432	(70) = 0,175	
(11) = 0,633	(41) = 0,426	(71) = 0,165	
(12) = 0,627	(42) = 0,420	(72) = 0,155	
(13) = 0,621	(43) = 0,413	(73) = 0,145	
(14) = 0,616	(44) = 0,406	(74) = 0,135	
(15) = 0,611	(45) = 0,400	(75) = 0,125	
(16) = 0,606	(46) = 0,393	(76) = 0,114	
(17) = 0,601	(47) = 0,386	(77) = 0,104	
(18) = 0,596	(48) = 0,378	(78) = 0,093	
(19) = 0,590	(49) = 0,370	(79) = 0,082	
(20) = 0,584	(50) = 0,362	(80) = 0,072	
(21) = 0,577	(51) = 0,354	(81) = 0,063	
(22) = 0,571	(52) = 0,345	(82) = 0,054	
(23) = 0,565	(53) = 0,336	(83) = 0,046	
(24) = 0,559	(54) = 0,327	(84) = 0,039	
(25) = 0,552	(55) = 0,319	(85) = 0,032	
(26) = 0,544	(56) = 0,310	(86) = 0,026	
(27) = 0,535	(57) = 0,301	(87) = 0,020	
(28) = 0,525	(58) = 0,291	(88) = 0,015	
(29) = 0,516	(59) = 0,282	(89) = 0,011	
(30) = 0,507	(60) = 0,273	(90) = 0,008	

7 Annexe 2 : une idée de preuve de l'unicité dans le théorème de Cauchy-Lipschitz

Nous nous contenterons du résultat suivant, qui ne concerne en apparence que l'équation logistique et sa solution nulle, mais contient en germe l'unicité de Cauchy-Lipschitz :

7.1 Proposition. *Soit p une fonction dérivable sur un intervalle ouvert I de $\mathbf{R}$, à valeurs réelles, vérifiant l'équation différentielle $p' = ap(m-p)$ où a et m sont des constantes positives. On suppose qu'il existe $x_0 \in I$ tel que $p(x_0) = 0$. Alors p est identiquement nulle sur I.*

Démonstration. Soit F l'ensemble (non vide) des points de I où p s'annule. Comme p est continue, F est fermé. Comme I est connexe, il suffit donc de montrer que F est aussi ouvert. Pour cela, choisissons a et b dans I vérifiant $a < x_0 < b$. Sur le segment $[a, b]$, la fonction $|m-p|$ est majorée par un nombre $M > 0$. On en déduit qu'on a, pour tout $x \in [a, b]$, $|p'(x)| \leq aM|p(x)|$. Soit $\epsilon > 0$ vérifiant $aM\epsilon < 1/2$ et tel que $J = [x_0 - \epsilon, x_0 + \epsilon]$ soit contenu dans $[a, b]$. Nous allons montrer que p est nulle sur J. Soit K le maximum de $|p|$ sur J. Il est atteint en un point x_1. L'égalité des accroissements finis s'écrit $p(x_1) = p(x_1)-p(x_0) = (x_1-x_0)p'(c)$ avec $c \in]x_0, x_1[$. La majoration ci-dessus donne alors $K = |p(x_1)| \leq aM|p(c)|\epsilon \leq \frac{1}{2}K$. Cela impose $K = 0$.

8 Annexe 3 : régressions logistiques

Sur ce sujet, ma science est toute récente et je me suis inspiré d'un petit article de Samuel Ambapour, *Le modèle logistique, un peu de statistique et d'histoire*, document du BAMSI (Brazzaville), que l'on trouve sur Internet.

On suppose qu'on a un phénomène régi par l'équation différentielle logistique $y' = ay(m-y)$ dont les solutions sont de la forme $y(t) = \dfrac{m}{1+\lambda e^{-amt}}$. On dispose de données statistiques et on cherche à évaluer les paramètres. Pour plus de commodité, on pose $c = am$. Il y a plusieurs méthodes, qui toutes consistent à se ramener au cas d'une régression linéaire en établissant des formules plus ou moins compliquées. Quitte à changer l'origine des temps et l'unité de temps [16], on peut supposer qu'on a les valeurs de y pour les temps $1, 2, \ldots, n$, à savoir $y_1, y_2, \ldots, y_n$. On a donc, si le modèle est correct,

16. Par exemple on prendra l'origine en 1790 et l'unité de temps égale à 10 ans.

$y_i = \dfrac{m}{1+\lambda e^{-ci}}\cdot$ On va déterminer c et m par des méthodes de régression linéaire. Ensuite, pour calculer λ on utilise la formule :

$$\lambda = \big(\frac{m}{y_i} - 1\big)e^{ci}.$$

Si la donnée était vraiment logistique, un seul calcul suffirait, mais comme ce n'est pas le cas, on calcule λ en prenant la moyenne de toutes les valeurs trouvées.

8.1 Les méthodes exactes

Le lecteur vérifiera sans peine les formules annoncées ci-dessous.

8.1.1 La méthode de Rhodes

C'est peut-être la plus simple de toutes. Elle repose sur le principe que $1/y$ est plus simple que y. Précisément, on a la relation :

$$\frac{1}{y_{i+1}} = \frac{1-e^{-c}}{m} + \frac{e^{-c}}{y_i}$$

et on fait la régression linéaire de $Y = \frac{1}{y_{i+1}}$ par rapport à $X = \frac{1}{y_i}\cdot$ On en déduit les coefficients de $Y = AX + B$: $A = e^{-c}$ et $B = \dfrac{1-e^{-c}}{m}$ d'où c et m.

8.1.2 La méthode de Jull

Elle repose sur la formule (exacte) :

$$\frac{y_{i+1}}{y_i} = e^c - \frac{e^c-1}{m}y_{i+1}$$

et il suffit de faire la régression linéaire de $Y = \frac{y_{i+1}}{y_i}$ par rapport à $X = y_{i+1}$. On obtient une estimation des coefficients de $Y = AX + B$: $A = -\frac{e^c-1}{m}$ et $B = e^c$ dont on déduit aussitôt c et m.

8.1.3 La méthode de Nair

On a $\dfrac{1}{y_i} = \dfrac{1}{m} + \dfrac{\lambda}{m}e^{-ci}$. Un petit calcul donne la formule :

$$\frac{1}{y_i} - \frac{1}{y_{i+1}} = -\frac{e^c-1}{e^c+1}\left(\frac{1}{y_i} + \frac{1}{y_{i+1}}\right) + \frac{2}{m}\frac{e^c-1}{e^c+1}.$$

On effectue une régression linéaire avec comme X les $\frac{1}{y_i} + \frac{1}{y_{i+1}}$ et comme Y les $\frac{1}{y_i} - \frac{1}{y_{i+1}}$. On estime les coefficients $A = \frac{e^c-1}{e^c+1}$ et $B = \frac{2}{m}\frac{e^c-1}{e^c+1}$ de la relation $Y = -aX + B$ et on en déduit les valeurs de m et c :

$$m = \frac{2A}{B} \qquad \text{et} \qquad c = \ln \frac{1+A}{1-A}.$$

8.2 Les méthodes approchées

Cette fois on utilise des formules approchées.

8.2.1 La méthode de Perrin

On part de $y' = ay(m-y)$, qui donne $\dfrac{y'}{y} = -ay + c$ (avec $c = am$). Si l'on a les y_i, il s'agit de calculer les y'_i. On peut considérer, grossièrement [17], qu'on a, pour un accroissement de 1, $y'_i = y_{i+1} - y_i$. Cela donne pour y'_i/y_i la valeur $\dfrac{y_{i+1}}{y_i} - 1$. La régression linéaire de $Y = \dfrac{y_{i+1}}{y_i}$ par rapport à $X = y_i$ a donc pour coefficients $Y = -aX + c + 1$. On en déduit donc a et c, donc m. On notera que cette méthode est essentiellement la même que celle de Jull et les résultats sont analogues en remplaçant e^c par $c+1$.

8.2.2 La méthode de Fisher

Le principe est le même, mais avec une autre estimation de y'_i/y_i par $\dfrac{1}{2}\ln\dfrac{y_{i+1}}{y_{i-1}} = \dfrac{1}{2}(\ln(y_{i+1}) - \ln(y_{i-1}))$. La justification est la suivante : y'/y est la dérivée de $z = \ln y$ et on approche cette dérivée par $\frac{1}{2}(z(x+1) - z(x-1))$.

8.3 Résultats

Il est fascinant de constater combien les résultats sont différents selon les méthodes. Voici certains de ces résultats appliquées aux chiffres de la population des USA de 1790 à 1910. Je donne seulement la valeur de m (en millions d'habitants). Par la méthode de Jull on trouve 199, avec celle de Perrin, 163, avec celle de Rhodes ou de Nair 217, avec celle de Fisher 190, etc.

Si l'on prend les chiffres de 1790 à 2000 c'est encore bien pire : les méthodes de Nair et Rhodes donnent des valeurs de m inférieures à certaines données ! (Et donc des coefficients imaginaires pour λ.)

17. La méthode est grossière, mais je l'ai inventée tout seul !

La morale de tout cela c'est qu'il est bien hasardeux de penser que la population des États-Unis suit une loi logistique. Dans les populations humaines, il y a beaucoup de paramètres qui interviennent et qui rendent difficiles les prévisions. Par ailleurs, le caractère essentiel de la fonction logistique est l'existence du point d'inflexion à mi-course. Or, ici, il n'est pas évident qu'il existe sur les données (surtout si l'on va jusqu'en 2000), de sorte que la loi est très mal déterminée par les valeurs détenues.

Dans la figure ci-dessous les croix correspondent aux populations (en millions) observées entre la décade 0 (1790) et la décade 21 (2000) tandis que la courbe est obtenue avec la régression de Jull. On voit bien qu'elle est sujette à caution.

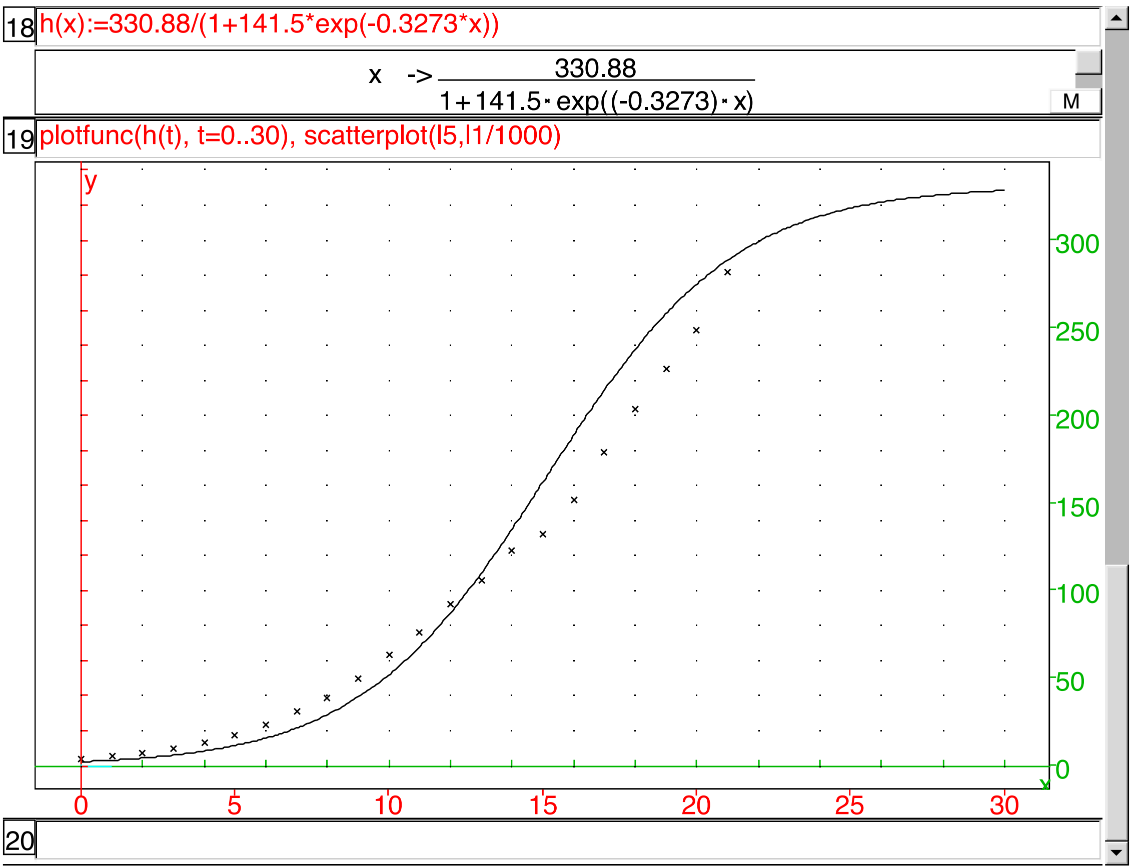

Figure 3 – Les données et la régression logistique de Jull de 1790 à 2000

Annexe 4 : les tables de Bernoulli

TABLE I.

Ages par années.	Survivans selon M. Halley.	N'ayant pas eu la pet. vérole.	Ayant eu la pet. vérol.	Prenant la pet. vérole pendant ch. année.	Morts de la pet. vérole pendant chaq. ann.	Somme des morts de la pet. vérole.	Morts par d'autres maladies pend. chaq. année.
0	1300	1300	0				
1	1000	896	104	137	17,1	17,1	283
2	855	685	170	99	12,4	29,5	133
3	798	571	227	78	9,7	39,2	47
4	760	485	275	66	8,3	47,5	30
5	732	416	316	56	7,0	54,5	21
6	710	359	351	48	6,0	60,5	16
7	692	311	381	42	5,2	65,7	12,8
8	680	272	408	36	4,5	70,2	7,5
9	670	237	433	32	4,0	74,2	6
10	661	208	453	28	3,5	77,7	5,5
11	653	182	471	24.4	3.0	80.7	5
12	646	160	486	21,4	2,7	83,4	4,3
13	640	140	500	18,7	2,3	85,7	3,7
14	634	123	511	16,6	2,1	87,8	3,9
15	628	108	520	14.4	1.8	89,6	4.2
16	622	94	528	12,6	1,6	91,2	4,4
17	616	83	533	11,0	1,4	92,6	4,6
18	610	72	538	9,7	1,2	93,8	4,8
19	604	63	541	8,4	1,0	94,8	5
20	598	56	542	7,4	0,9	95,7	5,1
21	592	48,5	543	6,5	0,8	96,5	5,2
22	586	42.5	543	5.6	0.7	97.2	5,3
23	579	37	542	5,0	0,6	97,8	6,4
24	572	32.4	540	4,4	0,5	98,3	6,5

Dans l'ordre des colonnes : l'année t, $N(t)$, $x(t)$, $N(t)-x(t)$, $v(t) = x(t)/8$ (ou plutôt une moyenne avec $x(t)$ et $x(t-1)$), $v(t)/8$ (les morts par variole), le cumul des morts par variole, les autres morts de l'année $t-1$ à t.

La deuxième table de Bernoulli

TABLE II.

Ages par années.	État naturel & variolique.	État non-varioliq.	Differ. ou gains.	Ages par années.	État naturel & variolique.	État non-varioliq.	Differ. ou gains.
0	1300	1300	0	13	640	741.1	74.1
1	1000	1017,1	17,1	14	634	709,7	75,7
2	855	881,8	26,8	15	628	705,0	77,0
3	798	833.3	35.3	16	622	700,1	78,1
4	760	802,0	42,0	17	616	695,0	79,0
5	732	779,8	47,8	18	610	689,6	79,6
6	710	762,8	52,8	19	604	684,0	80,0
7	692	749,2	57,2	20	598	678,2	80,2
8	680	740.9	60.9	21	592	672.3	80.3
9	670	734.4	64.4	22	586	666,3	80,3
10	661	728.4	67.4	23	579	659,0	80.0
11	653	722,9	69,9	24	572	651,7	79,7
12	646	718,2	72,2	25	565	644,3	79,3

Cette Table fait voir d'un coup d'œil, combien sur 1300 enfans, supposés nés en même temps, il en resteroit de vivans d'année en année jusqu'à l'âge de vingt-cinq ans, en les supposant tous sujets à la petite vérole : & combien il en resteroit s'ils étoient tous exempts de cette maladie, avec la comparaison & la différence des deux états.