

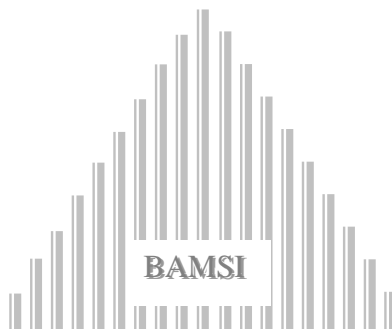
**BUREAU D'APPLICATION DES METHODES  
STATISTIQUES ET INFORMATIQUES**

**BAMSIREPRINT N° 03/2003**

**Le modèle logistique**

*Un peu de statistique et d'histoire*

*Samuel AMBAPOUR*



## **Le modèle logistique\***

*Un peu de statistique et d'histoire*

*Samuel AMBAPOUR\*\**

**Résumé :** Dans ce texte qui se veut essentiellement pédagogique, on présente le modèle logistique. Une description mathématique est d'abord introduite pour faciliter l'exposé du problème. Ensuite, on passe à l'estimation des paramètres de ce modèle ; on sait que dans ce cas, que la méthode des moindres carrés n'est pas directement applicable ; on a alors recours à des transformations diverses. Enfin, cette présentation est complétée par une page d'histoire, sans prétendre pour autant condenser en quelques lignes ce qu'un érudit peut désirer connaître sur le sujet.

**Mots clés :** modèle exponentiel, modèle logistique

\*Ce document a son origine dans un texte pédagogique pour un enseignement de statistique dispensé au Centre d'Application de la Statistique et de la Planification (CASP) en 1991. Nous remercions G. Ondongo (Université M. Ngouabi), G. Foumou (CASP) et Ch. Massamba (Ingénieur, ENSEA d'Abidjan) pour leurs commentaires sur une première version de ce texte.

\*\* Enseignant au CASP

*Ces documents de travail ne reflètent pas la position du BAMSI, mais n'engagent que ses auteurs.*

# Introduction

Dans ce document qui se veut essentiellement pédagogique, on présente le modèle logistique. Dans la première partie du texte, on donne une description mathématique de ce modèle afin de faciliter l'exposé du problème. On montre en fait, comment passer du modèle exponentiel (cas où la valeur d'une série croît rapidement) au modèle logistique (cas où la valeur de la série se stabilise et l'on admet qu'elle ne dépassera pas une certaine limite fixée). La deuxième partie du texte est consacrée à l'estimation des paramètres du modèle. Dans un premier temps, deux types de méthodes d'estimation sont proposés : tout d'abord, des méthodes dites « simplifiées » dont on ne peut porter un jugement en probabilité sur les résultats obtenus ; ensuite, des méthodes qui font appel aux techniques de régression. Et, dans un deuxième temps, un exemple illustratif est donné, ce qui permet de comparer les différentes estimations et tirer une conclusion d'ensemble. La troisième partie du texte donne un bref aperçu sur l'histoire du modèle basée sur les recherches de P. Verhulst, supposé être l'auteur à l'origine dudit modèle en déterminant les causes qui agissent sur l'accroissement d'une population, et sur les travaux de Pearl et Reed, sensés avoir trouvé une loi mathématique décrivant la croissance de la population des Etats-Unis.

## 1. Du modèle exponentiel au modèle logistique

Soit une chronique  $x_t$ . La valeur de cette dernière croît rapidement dans le cas d'une croissance exponentielle. Cependant, cette croissance peut être ralentie sous l'action de certaines contraintes et l'on admet dans ce cas qu'elle ne dépassera pas une certaine limite fixée, de telle sorte que la valeur de la chronique se stabilise. On obtient alors un modèle logistique. Dans ce qui va suivre, on va examiner le passage de l'un des modèles à l'autre.

## 1.1. Le modèle exponentiel

On montre que si le taux de variation  $\frac{\Delta x_t}{\Delta t}$  est proportionnel à  $x_t$ , le trend est exponentiel. En effet si  $r$  est le taux de variation relatif, on a par définition :

$$\frac{\frac{\Delta x_t}{\Delta t}}{x_t} = r ; \quad (1)$$

Soit dans le cas continu :

$$\frac{\frac{dx_t}{dt}}{x_t} = r \quad (1')$$

$$\Rightarrow \frac{dx_t}{x_t} = r dt \quad (2)$$

$$\Rightarrow \int \frac{dx_t}{x_t} = \int r dx_t$$

$$\ln x_t = rt + c$$

$$\Rightarrow x_t = e^{rt} + c \quad (3)$$

Où  $c$  est une constante. On obtient donc un trend exponentiel

## 1.2. Le modèle logistique

Ici, l'on suppose que l'existence de certaines contraintes techniques, économiques et sociales, entraîne le ralentissement de la croissance. On admet dans ce cas que cette croissance ne dépassera pas une valeur limite  $k$  ( $0 < x_t < k$ ). Considérons pour cela l'équation (2) et multiplions son membre de droite par  $\frac{k-x_t}{k}$  ; on obtient :

$$\frac{dx_t}{dt} = \frac{rx_t(k-x_t)}{k} \quad (4)$$

On peut remarquer que lorsque  $x_t$  est petit, la croissance est approximativement exponentielle. Et lorsque  $x_t \rightarrow k$ ,  $\frac{dx_t}{dt} \rightarrow 0$

$$\frac{dx_t}{dt} = \rho x_t(k-x_t) \text{ en posant que : } \frac{r}{k} = \rho$$

$$\frac{dx_t}{x_t(k-x_t)} = \rho dt$$

et

$$\int \frac{dx_t}{x_t(k-x_t)} = \int \rho dt$$

On obtient :

$$\frac{1}{k} \ln \left| \frac{x_t}{k-x_t} \right| = \rho t + c, \quad (5)$$

et

$$\ln \left| \frac{x_t}{k-x_t} \right| = k\rho t + kc \quad (6)$$

Compte tenu du fait que  $k > 0$  et  $(k-x_t) > 0$ , on peut écrire que :

$$\ln \left( \frac{x_t}{k-x_t} \right) = k\rho t + kc \quad (6')$$

Soit sous la forme exponentielle :

$$\begin{aligned} \frac{x_t}{k-x_t} &= e^{k\rho t + kc} \\ &= e^{k\rho t} + e^{kc} \end{aligned} \quad (7)$$

Si l'on remplace la constante positive  $e^{kc}$  par  $A$ , on a :

$$\frac{x_t}{k-x_t} = Ae^{k\rho t}$$

D'où :

$$\begin{aligned} x_t &= kAe^{k\rho t} - x_t Ae^{k\rho t} \\ \Rightarrow x_t (Ae^{k\rho t} + 1) &= kAe^{k\rho t} \end{aligned}$$

Donc

$$x_t = \frac{kAe^{k\rho t}}{Ae^{k\rho t} + 1} \quad (9)$$

En divisant le numérateur et le dénominateur de (9) par  $Ae^{k\rho t}$ , on a :

$$x_t = \frac{k}{1 + \frac{1}{Ae^{k\rho t}}} = \frac{k}{1 + \frac{1}{A} e^{-k\rho t}} \quad (10)$$

Et si l'on remplace  $\frac{1}{A}$  par  $b$  et  $k\rho$  par  $a$ , on a en définitive :

$$x_t = \frac{k}{1 + be^{-at}} \quad (11)$$

On obtient ainsi le modèle logistique où  $a$ ,  $b$  et  $k$  sont des paramètres à estimer.

D'après (11), on trouve que :  $x_t \rightarrow k$  lorsque  $t \rightarrow \infty$  ; et que  $x_t \rightarrow 0$  lorsque  $t \rightarrow -\infty$

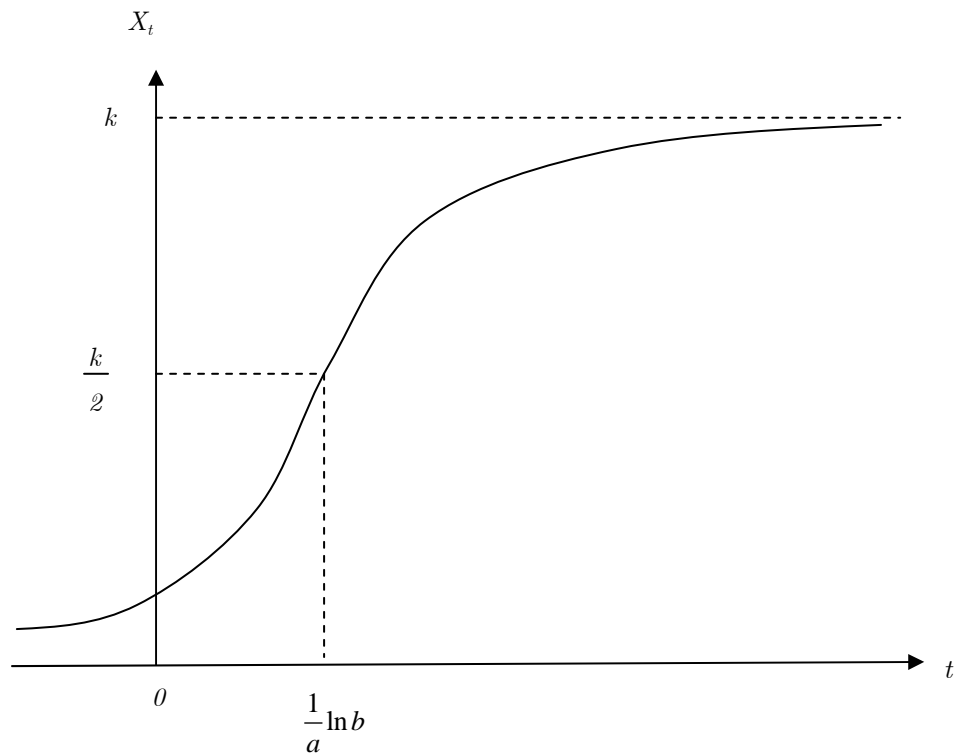
de la sorte  $k$  est la limite asymptotique supérieure et  $0$  la limite asymptotique inférieure. Sa dérivée est :

$$\frac{dx_t}{dt} = \frac{a}{k} x_t (k - x_t) \quad (12)$$

Cette dérivée est positive pour toute valeur de  $t$ . La dérivée seconde est :

$$\frac{d^2x_t}{dt^2} = \frac{a}{k}(k - 2x_t) \frac{dx_t}{dt} \quad (13)$$

Elle s'annule au point d'inflexion  $t = \frac{1}{a} \ln b$  ;  $x_t = \frac{k}{2}$ . Ainsi, le maximum du taux de croissance se produit au point ayant pour coordonnées :  $\left( \frac{1}{a} \ln b ; \frac{k}{2} \right)$



**Fig 1 : La fonction logistique**

## 2. Méthodes d'estimation des paramètres du modèle logistique

Ici, on va examiner quelques méthodes d'estimation des paramètres du modèle logistique. Deux types de méthodes sont envisagés. Dans le premier type, il s'agit des méthodes simplifiées qui donnent des résultats approximatifs. Il est donc impossible dans ce cas, de porter un jugement en probabilité sur les résultats obtenus. Il s'agira principalement de la méthode des « trois sommes » et de la méthode des « trois points ». Dans le deuxième type de méthodes, on fait appel aux techniques de régression. On sait que dans l'estimation des courbes non linéaires, la méthode des

moindres carrés n'est pas directement applicable ; on a alors recours à des transformations d'ordre divers. A ce sujet quatre méthodes de régression seront proposées.

## 2.1. Les méthodes simplifiées

### 2.1.1. La méthode des « trois sommes »

La chronique (11) est divisée en trois parties égales. Pour chaque partie on calcule la somme des inverses des valeurs de  $x_t$ . Et l'on note ces trois sommes par :

$$\begin{aligned} \sum_1 \frac{1}{x_t} &= \frac{1}{k}(n+bc) ; \\ \sum_2 \frac{1}{x_t} &= \frac{1}{k}(n+bce^{-na}) ; \\ \sum_3 \frac{1}{x_t} &= \frac{1}{k}(n+bce^{-2na}) \end{aligned} \quad (14)$$

On tire  $c$

$$c = \frac{(1-e^{-na})}{(1-e^{-a})}$$

Déterminons maintenant les deux différences ci-après :

$$\begin{aligned} D_1 &= \sum_1 \frac{1}{x_t} - \sum_2 \frac{1}{x_t} = \frac{bc}{k}(1-e^{-na}) ; \\ D_2 &= \sum_2 \frac{1}{x_t} - \sum_3 \frac{1}{x_t} = \frac{bc}{k}e^{-na}(1-e^{-na}). \end{aligned}$$

Le rapport entre les deux différences donne :

$$\frac{D_2}{D_1} = e^{-na}$$

On tire  $a$

$$a = \frac{1}{n}(\ln D_1 - \ln D_2) \quad (15)$$

Par ailleurs, on peut montrer que

$$\frac{D_1^2}{D_1 - D_2} = \frac{bc}{k}$$

Donc :

$$b = \frac{k}{c} \frac{D_1^2}{D_1 - D_2} \quad (16)$$

De la première équation du système (14), on trouve :

$$\frac{bc}{k} = \sum_1 \frac{1}{x_t} - \frac{n}{k}$$

et donc

$$k = \frac{n}{\left( \sum_1 \frac{1}{x_t} - \frac{D_1^2}{D_1 - D_2} \right)} \quad (17)$$

### 2.1.2. La méthode des « trois points »

On fait passer la courbe par trois points équidistants  $x_0$ ,  $x_1$  et  $x_2$  et dont les expressions sont respectivement :

$$\begin{aligned} \frac{1}{x_0} &= \frac{1}{k}(1+b) \\ \frac{1}{x_1} &= \frac{1}{k}(1+be^{-na}) \\ \frac{1}{x_2} &= \frac{1}{k}(1+be^{-2na}) \end{aligned} \quad (18)$$

On calcule les différences suivantes :

$$\begin{aligned} D_1 &= \frac{1}{x_0} - \frac{1}{x_1} = \frac{b}{k}(1 - e^{-na}) ; \\ D_2 &= \frac{1}{x_1} - \frac{1}{x_2} = \frac{b}{k}e^{-na}(1 - e^{-na}). \end{aligned}$$

Leur rapport donne :

$$\frac{D_2}{D_1} = e^{-na}.$$

On tire  $a$

$$a = \frac{1}{n}(\ln D_1 - \ln D_2) \quad (19)$$

Par ailleurs, on peut montrer que

$$\frac{D_1^2}{D_1 - D_2} = \frac{b}{k} = \frac{1}{x_0} - \frac{1}{k}$$

D'où

$$\frac{1}{k} = \frac{1}{x_0} - \frac{D_1^2}{D_1 - D_2}$$

et

$$k = \frac{1}{\left( \frac{1}{x_0} - \frac{D_1^2}{D_1 - D_2} \right)} \quad (20)$$

Enfin de la première équation du système (18), on obtient :

$$b = \frac{k - x_0}{x_0} \quad (21)$$

## 2.2. Méthodes de régression

### 2.2.1. Méthode de Fisher

Considérons l'expression (12), que l'on peut réécrire comme suit :

$$\frac{dx_t}{dt} = ax_t \left(1 - \frac{x_t}{k}\right) \quad (22)$$

Divisons les deux membres de (22) par  $x_t$  :

$$\frac{dx_t}{dt} \frac{1}{x_t} = a \left(1 - \frac{x_t}{k}\right) \quad (23)$$

Après quelques arrangements, on peut écrire (23) comme une fonction linéaire du temps de paramètre  $a$  et  $\frac{a}{k}$  :

$$\tau_t = \frac{dx_t}{dt} \frac{1}{x_t} = a - \frac{a}{k} x_t \quad (24)$$

Ici, l'estimation directe des paramètres n'est pas possible. La série ne permet de déterminer  $dx_t/dt$  et donc  $\tau_t$ . Fisher propose d'approximer  $\tau_t$  par l'expression suivante :

$$\tau_t \cong \frac{1}{2} \ln \frac{x_{t+1}}{x_{t-1}}, \quad t = 2, 3, \dots, n-1. \quad (25)$$

Ainsi disposant de  $n$  observations, les paramètres de (24) peuvent être estimés en résolvant le système :

$$\begin{cases} \sum \tau_t = a(n-2) + \frac{a}{k} \sum x_t \\ \sum \tau_t x_t = a \sum x_t + \frac{a}{k} \sum x_t^2, \quad t = 2, \dots, n-1 \end{cases} \quad (26)$$

### 2.2.1. Méthode de Jull

Cette méthode est basée sur la propriété selon laquelle, le taux d'accroissement est une fonction linéaire du niveau de la série :

$$\frac{x_{t+1} - x_t}{x_t} = (e^a - 1) \left(1 - \frac{1}{k} x_{t+1}\right),$$

Où encore

$$\frac{x_{t+1} - x_t}{x_t} = (e^a - 1) - \frac{e^a - 1}{k} x_{t+1} \quad (27)$$

De cette façon, on cherche à régresser  $\frac{x_{t+1} - x_t}{x_t}$  sur  $x_{t+1}$ . Le système d'équations normales prend alors la forme la suivante :

$$\begin{cases} \sum \tau_t = \alpha(n-1) + \frac{\alpha}{k} \sum x_t; \\ \sum \tau_t x_{t+1} = \alpha \sum x_t + \frac{\alpha}{k} \sum x_t^2, \quad t = 1, \dots, n-1 \end{cases} \quad (28)$$

Où

$$\tau_t = \frac{x_{t+1} - x_t}{x_t}, \quad \alpha = e^a - 1$$

### 2.2.2. Méthode de Rhodes

Considérons deux membres successifs de la série. On peut écrire la différence de leurs inverses de la manière suivante :

$$\frac{1}{x_{t+1}} - \frac{1}{x_t} = \frac{e^a - 1}{e^a + 1} \left[ \frac{2}{k} - \left( \frac{1}{x_t} + \frac{1}{x_{t+1}} \right) \right]. \quad (29)$$

De là :

$$\frac{1}{x_{t+1}} = \frac{1 - e^{-a}}{k} + e^{-a} \frac{1}{x_t} \quad (30)$$

Rhodes a proposé d'estimer  $a$  et  $k$  en se basant sur l'équation (30). En d'autres termes, on régresse  $\frac{1}{x_{t+1}}$  sur  $\frac{1}{x_t}$  avec pour coefficients  $e^{-a}$  et  $\frac{(1 - e^a)}{k}$ . Le système d'équations normales est le suivant :

$$\begin{cases} \sum \frac{1}{x_{t+1}} = \left( \frac{1 - e^{-a}}{k} \right) (n-1) + e^{-a} \sum \frac{1}{x_t}; \\ \sum \frac{1}{x_{t+1}} \frac{1}{x_t} = \left( \frac{1 - e^{-a}}{k} \right) \sum \frac{1}{x_t} + e^{-a} \sum \left( \frac{1}{x_t} \right)^2, \quad t = 1, \dots, n-1 \end{cases} \quad (31)$$

### 2.2.3. Méthode de Nair

Se basant sur la relation (29), Nair a proposé de régresser  $\frac{1}{x_{t+1}} - \frac{1}{x_t}$  sur  $\frac{1}{x_{t+1}} + \frac{1}{x_t}$ . Si l'on

note par :

$$\gamma_t = \frac{1}{x_{t+1}} - \frac{1}{x_t};$$

$$\lambda_t = \frac{1}{x_{t+1}} + \frac{1}{x_t};$$

$$\nu = \frac{e^a - 1}{e^a + 1}.$$

Dans ce cas, (29) s'écrit :

$$\gamma_t = \nu \frac{2}{k} - \nu \lambda_t \quad (32)$$

Les paramètres  $\nu$  et  $\nu \frac{2}{k}$  peuvent être estimés en résolvant le système (33):

$$\begin{cases} \sum \gamma_t = \nu \frac{2}{k} (n-1) - \nu \sum \lambda_t ; \\ \sum \gamma_t \lambda_t = \nu \frac{2}{k} \sum \lambda_t - \nu \sum \lambda_t^2, \quad t=1, \dots, n-1 \end{cases} \quad (33)$$

#### 2.2.4. Détermination du paramètre $b$

Ayant estimé les paramètres  $a$  et  $k$  par une des méthodes de régression, il reste enfin à déterminer le paramètre  $b$  de la fonction logistique. On peut réécrire cette dernière comme suit :

$$\frac{k}{x_t} - 1 = b e^{-at}.$$

Et en prenant le logarithme de cette expression, on obtient :

$$\ln \left( \frac{k}{x_t} - 1 \right) = \ln b - at$$

D'où :

$$\ln b = at + \ln \left( \frac{k}{x_t} - 1 \right). \quad (34)$$

L'expression (34) peut être écrite pour chaque niveau de la série, c'est-à-dire :

$$\ln b = a + \ln \left( \frac{k}{x_1} - 1 \right), \text{ pour } t=1 ;$$

$$\ln b = 2a + \ln \left( \frac{k}{x_2} - 1 \right), \text{ pour } t=2 ;$$

⋮

$$\ln b = na + \ln \left( \frac{k}{x_n} - 1 \right) \text{ pour } t=n$$

En sommant toutes les observation on a :

$$n \ln b = a \sum t + \sum \ln \left( \frac{k}{x_t} - 1 \right)$$

En calculant la moyenne de cette expression et en tenant compte du fait que

$$\sum t = \frac{n(n+1)}{2}$$

On trouve :

$$\ln b = \frac{a(n+1)}{2} + \frac{1}{n} \sum \ln \left( \frac{k}{x_t} - 1 \right) \quad (35)$$

## 2.3. Comparaison des résultats des estimations des paramètres

Dans le tableau 1, est donnée l'évolution de la population des USA entre 1790 et 1960 (colonne 1, en millions). A partir de ce tableau, nous allons dans un premier temps estimer les paramètres du modèle logistique en se basant sur les différentes méthodes exposées ci-dessus. Dans un deuxième temps, on compare les résultats obtenus et on tire une conclusion d'ensemble.

### 2.3.1. Méthode des trois sommes

Considérons la colonne 5 du tableau 1 ; on a :

$$\sum_1 \frac{1}{x_t} = 0.266739 + 0.189861 + \dots + 0.058875 = 0.82799 ;$$

$$\sum_2 \frac{1}{x_t} = 0.043348 + 0.031971 + \dots + 0.0113204 = 0.150534 ;$$

$$\sum_3 \frac{1}{x_t} = 0.010892 + 0.009107 + \dots + 0.005603 = 0.047978.$$

On calcule :

$$D_1 = \sum_1 \frac{1}{x_t} - \sum_2 \frac{1}{x_t} = 0.82799 - 0.150534 = 0.677456 ;$$

$$D_2 = \sum_2 \frac{1}{x_t} - \sum_3 \frac{1}{x_t} = 0.150534 - 0.047978 = 0.102556.$$

On tire alors le paramètre  $a$  en appliquant la formule (15) ; soit :

$$a = \frac{1}{6} (\ln 0.677456 - \ln 0.102556) = 0.314656.$$

En utilisant l'expression (17), on calcule le paramètre  $k$ . On trouve :

$$k = \frac{6}{0.82799 - \frac{(0.677456)^2}{0.677456 - 0.102556}} = 202$$

Enfin, pour calculer  $b$  on se sert de (16). Dans cette expression, il faut d'abord déterminer le paramètre  $c$  ; soit :

$$c = \frac{(1 - e^{-na})}{(1 - e^{-a})} = \frac{(1 - e^{-6 \times 3.14656})}{(1 - e^{-3.14656})} = 1.3811 ;$$

et donc :

$$b = \frac{202}{1.3811} \times \frac{(0.677456)^2}{(0.677456 - 0.102556)} = 51$$

### 2.3.2. Méthode des trois points

On choisit trois points équidistants représentant les années 1800, 1880 et 1960 (voir colonne 5 du tableau 1). On trouve  $D_1 = 0.169831$  et  $D_2 = 0.014427$  ; d'où  $a = 0.352243$  en se servant de l'expression (19). On calcule ensuite le paramètre  $k = 234$  en appliquant la formule (20). Enfin, de la relation (21) on tire  $b = 43$ .

### 2.3.3. Méthodes de régression

#### a) Méthode de Fisher

Selon la méthode de Fisher, on effectue la régression de  $\tau_t$  (colonne 3 du tableau 1) sur  $x_t$  (colonne 1). On trouve l'équation suivante :  $\tau_t = 0.309615 - 0.001418x_t$ . Connaissant  $a$  et  $\frac{a}{k}$ , on en déduit que  $k = 217$

#### b) Méthode de Jull

On régresse  $\tau_t = \frac{x_{t+1} - x_t}{x_t}$  sur  $x_{t+1}$ . On trouve :  $\tau_t = 0.353261 - 0.001484x_{t+1}$ . Comme  $e^a = 0.353261$  et  $\frac{e^{-a} - 1}{k} = 0.001484$  donc  $k = 238$  ; et  $a = \ln 0.353261 = 0.302252$

#### c) Méthode de Rhodes

On effectue la régression entre  $\frac{1}{x_{t+1}}$  et  $\frac{1}{x_t}$ . On obtient :  $\frac{1}{x_{t+1}} = 0.001213 + 0.733808 \frac{1}{x_t}$  ; d'où  $e^{-a} = 0.733808 \Rightarrow a = \ln 0.733808 = 0.309508$ . D'autre, comme  $\frac{1 - e^{-a}}{k} = 0.001213 \Rightarrow k = \frac{1 - 0.733808}{0.001213} = 219$

#### d) Méthode de Nair

Enfin, par la méthode de Nair, on obtient l'équation de régression suivante :  $\frac{1}{x_{t+1}} - \frac{1}{x_t} = 0.001394 - 0.153466 \left( \frac{1}{x_{t+1}} + \frac{1}{x_t} \right)$ . Par conséquent on a les coefficients de régression :  $v \frac{2}{k} = 0.001394$  et  $v = 0.153466$ . A partir de l'expression  $v = \frac{e^a - 1}{e^a + 1}$ , on obtient :  $e^a = \frac{1 + 0.153466}{1 - 0.153466} = 1.326257$  et  $a = \ln 1.326257 = 0.282361$ . Comme  $v \frac{2}{k} = 0.001394$ , donc  $k = 0.153466 \frac{2}{0.001394} = 220$

e) *Calcul du paramètre b*

Le calcul du paramètre  $b$  pour les différentes méthodes de régression s'obtient en appliquant la formule (35). Par exemple par la méthode de Fisher, on a :

$$\ln b = 0.309615 \left( \frac{18+1}{2} \right) + \frac{1}{18} \sum \ln \left( \frac{217}{x_t} - 1 \right) = 2.329608 ; \Rightarrow b = 75.906969$$

**Tableau 1 : Calcul des paramètres du modèle logistique**

$t$	$x_t$	$\frac{1}{2} \ln \left( \frac{x_{t+1}}{x_{t-1}} \right)$	$\left( \frac{x_{t+1} - x_t}{x_{t-1}} \right)$	$\frac{1}{x_t}$	$\frac{1}{x_{t+1}} + \frac{1}{x_t}$	$\frac{1}{x_{t+1}} - \frac{1}{x_t}$
1	2	3	4	5	6	7
1790	3,895	-	0,352246	0,256739	0,446600	-0,066878
1800	5,267	0,305942	0,363584	0,189861	0,329098	-0,050624
1810	7,182	0,298377	0,331941	0,139237	0,244606	-0,033841
1820	9,566	0,290260	0,341626	0,105369	0,183287	-0,027451
1830	12,834	0,287058	0,323438	0,077918	0,136793	-0,019043
1840	16,985	0,293196	0,358198	0,058875	0,102223	-0,015527
1850	23,069	0,30529	0,355845	0,043348	0,075319	-0,011377
1860	31,278	0,25499	0,228211	0,031971	0,058001	-0,005941
1870	38,416	0,2337934	0,299563	0,026030	0,046060	-0,006000
1880	49,924	0,244880	0,255749	0,020030	0,035981	-0,004079
1890	62,692	0,208363	0,208033	0,015951	0,029155	-0,002747
1900	75,734	0,190755	0,212296	0,013204	0,024095	-0,0023120
1910	91,812	0,185744	0,195987	0,010892	0,019999	-0,001785
1920	109,806	0,145305	0,118108	0,009107	0,017252	-0,000962
1930	122,775	0,090788	0,072441	0,008145	0,015740	-0,000550
1940	131,669	0,102459	0,144514	0,007595	0,014231	-0,000959
1950	150,697	0,152048	0,184257	0,006636	0,012239	-0,001033
1960	178,464	-	-	0,005603	-	-

### 2.3.4. Vue d'ensemble

Comme on le constate (tableau 2), les méthodes étudiées donnent des résultats différents.

La meilleure estimation semble être la méthode de Nair, ensuite vient celle de Rhodes et cela, à en juger par le coefficient de corrélation entre les régressions correspondantes.

La méthode des trois points donnent des résultats approximatifs et ne doit être utilisée qu'en cas de manque d'informations complètes sur la série. On a constaté aussi que la longueur de la série, jouait beaucoup sur la valeur du paramètre  $k$  ; ce qui conduit parfois aux résultats absurdes.

**Tableau 2 : Comparaison des différentes méthodes**

Méthode	$a$	Paramètres		
		$k$	$b$	$r$ Coefficient de corrélation
Fisher	0,31	217	76	0,9525
Jull	0,35	238	82	0,8850
Rhodes	0,31	219	77	0,9987
Nair	0,28	220	59	0,9998
Trois sommes	0,31	202	51	-
Trois points	0,35	234	43	-

## 3. Aperçu historique

### 3.1. Les travaux de P. Verhulst (1804-1849)

Pour Verhulst, « de tous les problèmes que l'économie politique offre aux méditations des philosophes, l'un des plus intéressants est sans contredit, la connaissance de la loi qui règle le progrès de la population. Pour le résoudre avec exactitude, il faudrait pouvoir apprécier l'influence des causes nombreuses qui empêchent ou favorisent la multiplication de l'espèce humaine ». Ainsi dit, Verhulst va établir deux catégories de causes qui agissent sur l'accroissement de la population : les causes constantes et les

causes variables. Dans les causes constantes ou permanentes sont inclus : la fécondité, la salubrité du pays, les lois civiles et religieuses, ainsi que les mœurs de la nation. Quant aux causes variables, elles sont subdivisées selon leur caractère accidentel (épidémies, famines passagères) ou non (se résumant dans la difficulté de plus en plus grande que la population éprouve pour se procurer de subsistances, lorsqu'elle est devenue assez nombreuse et que toutes les bonnes terres se trouvent occupées).

Dans un premier temps pour déterminer la loi théorique de l'accroissement de la population, Verhulst va faire abstraction des causes aléatoires, en s'appuyant sur l'idée de Malthus selon laquelle la population croît selon une progression géométrique. Il va donc choisir la fonction suivante :

$$P = P_0 e^{rt} \quad (36)$$

Il considère le taux d'accroissement  $r$  comme « une mesure de l'énergie avec laquelle la population tend à croître lorsqu'elle n'est pas retenue par la crainte de manque de subsistance ». Cependant, et très rapidement, il va abandonner cette hypothèse de progression géométrique. Il ne la trouve réaliste que dans « les circonstances tout à fait exceptionnelles ».

C'est ainsi que dans un deuxième temps, il va se lancer dans différents types d'hypothèses pouvant conduire à une formulation qui tiennent compte des obstacles susceptibles de freiner l'accroissement de la population. Pour trouver la loi d'affaiblissement du taux d'accroissement, il va partir de l'idée suivante : lorsque toutes les « bonnes terres » sont cultivées, la population a atteint son niveau normale qu'elle ferait bien ne pas dépasser, car toute population excédentaire (population surabondante) sera responsable du ralentissement de la croissance au point de l'annuler en fin de compte. Il traduit sous forme mathématique, cette hypothèse qui consiste à considérer l'affaiblissement du taux d'accroissement comme proportionnel à l'augmentation de la population, à partir du moment où la difficulté de trouver les bonnes terres commence à se faire sentir par l'équation différentielle suivante :

$$\frac{1}{P} \frac{dP}{dt} = r - k(P - b) ; t \geq 0 \quad (37)$$

où  $b$  est la « population normale » qui correspond à l'époque remarquable où toutes les bonnes terres sont utilisées (l'origine du temps est pris à ce moment là) ;  $P - b$  est la « population surabondante » à une date  $t$  ;  $r$ ,  $k$  et  $b$  sont des paramètres. La

solution de cette équation différentielle conduit à la fonction ci-après à laquelle il donne le nom de « logistique » :

$$t = \frac{\ell}{m} \ln \left[ \frac{P(m - kb)}{b(m - kP)} \right] ; \quad m = r + kb, \quad t \geq 0 \quad (38)$$

Si  $t = 0$ ,  $P = b$  ; si  $t \rightarrow \infty$ ,  $P \rightarrow \frac{m}{k}$ , ce qui pour Verhulst représente « l'extrême limite de la population ».

### 3.1. Les expériences de Pearl et Reed.

En 1920, le biologiste américain Pearl, annonce avoir trouvé une loi mathématique décrivant la population des Etats-Unis. Celle-ci permettrait de faire des prévisions sur une période assez longue. En collaboration avec le mathématicien Read, il étudie les propriétés et justifie l'utilisation de ce modèle par une série d'expériences biologiques, notamment celle de la multiplication des mouches drosophiles. En effet entre avril et mai 1921, Pearl se sert de deux bouteilles pour élever dans chacune un couple de mouches drosophiles dans un environnement précis et dans des conditions bien déterminées. Sur la base des données obtenues, il trace un graphique de croissance des mouches ; sa ressemblance avec le modèle logistique fut éminemment étonnante (voir graphique 2).

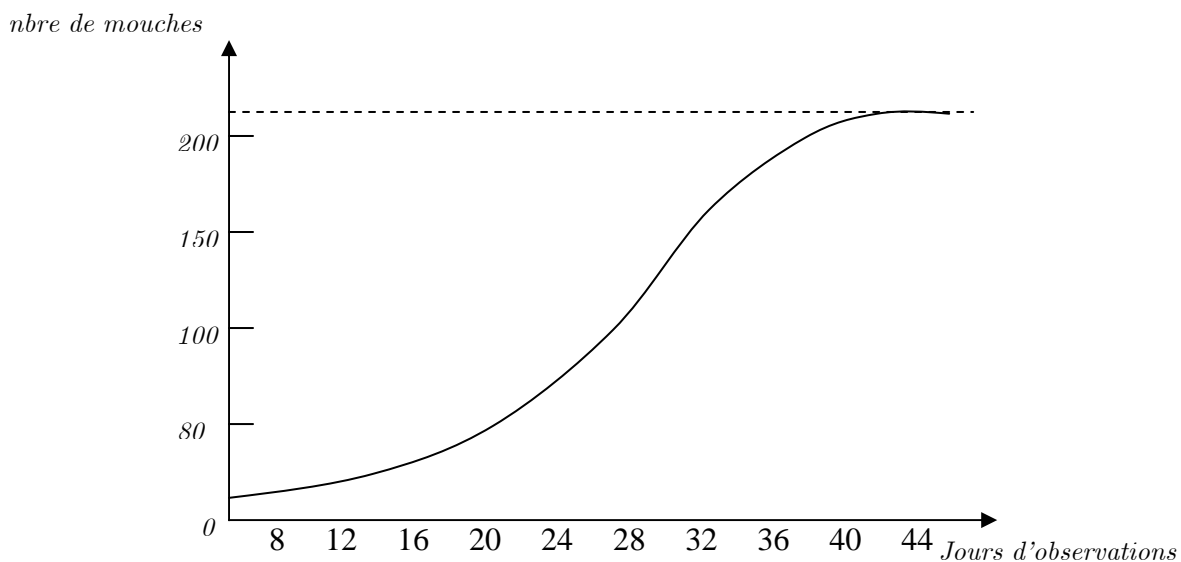


Fig 2 : Les expériences de Pearl et Reed

Par la suite, sur la base des données de la population des Etats-Unis de 1790-1910, Pearl et Reed vont calculer les paramètres du modèle logistique. Ils ont trouvé l'équation suivante :

$$P = \frac{197.27}{1 + 67.32^{-0.13137t}}$$

On peut comparer les résultats des prévisions de Pearl et Reed sur la période allant de 1920 à 1980 avec les données réellement observées

**Tableau 3 : Les prévisions de Pearl et Reed**

Années	Prévisions	Données observées
1920	107,4	109,8
1930	122,4	122,8
1940	136,3	131,7
1950	148,7	150,7
1960	159,2	178,4
1970	167,9	204,3
1980	174,9	-

## Conclusion : champ actuel

Aujourd'hui, le modèle logistique est appliqué avec succès dans l'étude des problèmes de diffusion et d'innovations :

- lancement d'un nouveau produit sur le marché ;
- mise en place d'un appareil de production supplémentaire ;
- ouverture d'un nouveau marché ;
- la demande d'un bien de consommation.

Dans le cas de la demande d'un bien de consommation, on remarque que la demande pour un produit augmente avec le revenu, mais que cette demande est inégalement intense suivant les produits. L'analyse des séries longues, montre que l'élasticité de la demande par rapport au revenu est loin d'être constante. Elle a tendance à décroître pour certains produits. Pour d'autre au contraire, elle a tendance à augmenter puis à diminuer. Ainsi, la variation de l'élasticité peut être le résultat :

- d'une modification du goût des consommateurs, indépendante de leur revenu ;
- d'une tendance à la saturation, qui explique l'élasticité décroissante. Dans ce cas, le phénomène de saturation peut se représenter par le modèle logistique suivant :

$$D = \frac{k_D}{1 + bc^{-aR}}$$

où  $D$  est la demande pour le bien considéré,  $R$  le revenu et  $k_D$  la limite supérieure de  $D$ .

## Bibliographie

- Ansion, G.**, (1990). *Les méthodes de prévision à court terme*, Armand Colin.
- Dupaquier, J.**, (1980). Avez-vous lu Malthus ? *Population*, Vol 35, 2.
- Guittou, H.**, (1964). *Statistique et économétrie*, Dalloz.
- Johnson, J.**, (1985). *Méthodes économétriques*, Economica.
- Jull, G.V.**, (1925). The growth of population and the factors which control it. *The Journal of the Royal Statistical Society*, Vol 88.
- Horvath, R.A.**, (1980). De Christophe Bernoulli à Achille Guillard : les tentatives de création d'une discipline démographique autonome au milieu du XIX-ème siècle, *Population*, Vol 35, 4-5.
- Hotteling, H.**, (1927). Differential equations subject to error and population estimates, *The Journal of American Statistical Association*, Vol 22.
- Malthus, P.T.**, (1980). *Essai sur le principe de population*, Ined, Puf.
- Nair, K. R.**, (1954). The fitting of growth curves, In *Statistics and Mathematics in Biology*, Jova.
- Rhodes, E. C.**, (1940). Population mathematics, *The Journal of the Royal Statistical Society*, Vol 103.
- Schtickzelle, M.**, (1981). Pierre Verhulst (1804-1849). La première découverte de la fonction logistique, *Population*, Vol 36,3.
- Tchitirkin, E.**, (1977). *Statisticheski metodi prognozirovanie*, Statistika, Moskva.
- Trebici, V.**, (1975). *Mica enciclopedie de demografie*, Bucuresti
- Ourlanis, B.**, (1940). *Rost nacilienie ve evropie*, Statistika, Moskva.