JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

KRUMMEICH

# Contribution à l'étude du mouvement de la population (suite)

*Journal de la société statistique de Paris*, tome 68 (1927), p. 157-175
<http://www.numdam.org/item?id=JSFS_1927__68__157_0>

## III

# CONTRIBUTION A L'ÉTUDE DU MOUVEMENT DE LA POPULATION (*Suite*)

DEUXIÈME PARTIE

## PROCÉDÉS D'AJUSTEMENT ET EXEMPLES POUR L'ANGLETERRE ET LES ÉTATS-UNIS

7° *Procédés d'ajustement de la formule de Verhulst.* — 1° *Procédé des trois ordonnées équidistantes.*

Prenons trois recensements équidistants, pour temps 0 l'époque du premier des trois recensements, et pour unité de temps l'intervalle qui sépare deux recensements successifs. Nous avons en adoptant la forme utilisée par M. Yule :

$Y_2$, $Y_1$, $Y_0$, étant les trois ordonnées correspondant aux trois recensements.

$$\left\{\begin{aligned} \frac{1}{Y_0} &= \frac{1}{K}\left(1+e^{\frac{\beta}{\alpha}}\right) \\ \frac{1}{Y_1} &= \frac{1}{K}\left(1+e^{\frac{\beta-1}{\alpha}}\right) \\ \frac{1}{Y_2} &= \frac{1}{K}\left(1+e^{\frac{\beta-2}{\alpha}}\right). \end{aligned}\right. \tag{1}$$

Posons :

$$\begin{aligned} d_1 &= \frac{1}{Y_0} - \frac{1}{Y_1} = \frac{1}{K} e^{\frac{\beta}{\alpha}}\left(1 - e^{-\frac{1}{\alpha}}\right) \\ d_2 &= \frac{1}{Y_1} - \frac{1}{Y_2} = \frac{1}{K} e^{\frac{\beta-1}{\alpha}}\left(1 - e^{-\frac{1}{\alpha}}\right). \end{aligned} \tag{2}$$

En divisant membre à membre les deux expressions (2), on obtient l'intervalle type.

$$e^{\frac{1}{\alpha}} = \frac{d_1}{d_2}. \tag{3}$$

Nous avons par ailleurs :

$$d^2_1 = \frac{1}{K^2} e^{\frac{2\beta}{\alpha}}\left(1 - e^{-\frac{1}{\alpha}}\right)^2$$

$$d_1 - d_2 = \frac{1}{K} e^{\frac{\beta}{\alpha}}\left(1 - e^{-\frac{1}{\alpha}}\right)^2$$

$$\frac{d^2_1}{d_1 - d_2} = \frac{1}{K} e^{\frac{\beta}{\alpha}}. \tag{4}$$

En portant la valeur de $e^{\frac{\beta}{\alpha}}$ tirée de cette expression dans la première des équations (1), on obtient :

$$K = Y_0 \frac{(d_1 - d_2)}{d_1 - d_2 - d^1_2 Y_0}$$

$$\frac{1}{K} = \frac{1}{Y_0} - \frac{d^2_1}{d_1 - d_2}. \tag{5}$$

Ayant calculé $\alpha$ et K, l'équation (4) donne $\beta$, distance du point d'inflexion au point 0, l'unité de temps étant l'intervalle des recensements.

Nous verrons plus loin une application pratique de ce procédé à la population française.

2° *Procédé de la somme des inverses.* — Ce procédé dû à M. Yule, présente l'avantage d'utiliser toutes les données, pourvu que celles-ci soient en nombre multiple de trois.

Prenons les inverses des populations sous la forme des équations (1), l'unité de temps étant toujours l'intervalle entre deux recensements successifs. Le nombre des données étant 3 $r$, formons trois groupes successifs de $r$ chacun, $S_1$, $S_2$ et $S_3$.

$$S_1 = \frac{r}{K} + \frac{1}{K}\left(e^{\frac{\beta}{\alpha}} + e^{\frac{\beta-1}{\alpha}} + \ldots + e^{\frac{\beta-r+1}{\alpha}}\right)$$

$$= \frac{r}{K} + \frac{e^{\frac{\beta}{\alpha}}}{K} \frac{1 - e^{\frac{-r}{\alpha}}}{1 - e^{\frac{-1}{\alpha}}}.$$

Posons :

$$c = \frac{1 - e^{\frac{-r}{\alpha}}}{1 - e^{\frac{-1}{\alpha}}}.$$

Nous pouvons alors écrire :

$$\left\{\begin{aligned} S_1 &= \frac{r}{K} + \frac{c}{K} e^{\frac{\beta}{\alpha}} \\ S_2 &= \frac{r}{K} + \frac{c}{K} e^{\frac{\beta - r}{\alpha}} \\ S_3 &= \frac{r}{K} + \frac{c}{K} e^{\frac{\beta - 2r}{\alpha}}. \end{aligned}\right. \tag{6}$$

Posons :

$$\begin{aligned} D_1 = S_1 - S_2 &= \frac{c}{K} e^{\frac{\beta}{\alpha}} \left(1 - e^{-\frac{r}{\alpha}}\right) \\ D_2 = S_2 - S_3 &= \frac{c}{K} e^{\frac{\beta - r}{\alpha}} \left(1 - e^{-\frac{r}{\alpha}}\right). \end{aligned} \tag{7}$$

On obtient par division membre à membre :

$$e^{\frac{r}{\alpha}} = \frac{D_1}{D_2} \tag{8}$$

$$D^2_1 = \frac{C^2}{K^2} e^{\frac{2\beta}{\alpha}} \left(1 - e^{-\frac{r}{\alpha}}\right)^2$$

$$D_1 - D_2 = \frac{c}{K} e^{\frac{\beta}{\alpha}} \left(1 - e^{-\frac{r}{\alpha}}\right)^2$$

$$\frac{D^2_1}{D_1 - D_2} = \frac{c}{K} e^{\frac{\beta}{\alpha}}.$$

En portant cette valeur dans la première des équations (6), on a :

$$\frac{r}{K} = S_1 - \frac{D^2_1}{D_1 - D_2}.$$

On achève ces calculs comme précédemment, nous verrons plus loin dans quelles conditions on peut appliquer ce procédé à la population française.

3° *Méthode des moindres carrés.* — Cette méthode, utilisée par M. Yule, donne de médiocres résultats, par contre, elle a un grand intérêt théorique car elle est fondée sur l'équation.

$$\frac{Yt + h - Yt}{Yt} = 1 - \frac{Yt + h}{K} \left(e^{\frac{h}{\alpha}} - 1\right).$$

On fait porter l'ajustement par moindres carrés sur les accroissements successifs pendant les intervalles $h$ égaux, la courbe ajustée étant une droite. K est l'abscisse du point où cette droite coupe l'axe des temps. On a ensuite $\alpha$. $\beta$ est obtenu à l'aide de la première équation (6).

4° *Calcul des constantes de :*

$$Y = \frac{K}{1 + me^{a_1 x + a_2 x^2 + a_3 x^3}}$$

Ce calcul est exposé comme suit par MM. Pearl et Reed :

$$a_1 x + a_2 x^2 + a_3 x^3 = \log (K - y) - \log m - \log y.$$

Si l'on pose :

$$\log m = a_0,$$

Il vient :

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 = \log \frac{K - y}{y}.$$

Faisons passer la courbe par cinq points observés :

$$x_0 = 0,\ x_1,\ x_2,\ x_3,\ x_4.$$

On a :

$$a_0 - \log \frac{K - Y_0}{Y_0} = 0 \qquad (9)$$

$$a_0 + a_1 x_1 + A_2 x_1^2 + a_3 x_1^3 - \log \frac{K - Y_1}{Y_1} = 0 \qquad (10)$$

$$a_0 + a_1 x_4 + a_2 x_4^2 + a_3 x_4^3 - \log \frac{K - Y_4}{Y_4} = 0 \qquad (13)$$

En supposant les coordonnées régulièrement espacées le long de l'axe des $x$, K est donné par l'équation du huitième degré en K :

$$Y_1^4\, Y_3^4\, (K - Y_0)\, (K - y_2)^6\, (K - Y_4) = Y_0\, Y_2^6\, Y_4\, (K - Y_1)^4\, (K - Y_3)^4 = 0.$$

Pour obtenir les $a$, on pose :

$$B_1 = \log \frac{K - Y_1}{Y_1} 1 - \log \frac{K - Y_0}{Y_0}$$

$$B_2 = \log \frac{K - Y_2}{Y_2} - \log \frac{K - Y_0}{Y_0}$$

$$B_3 = \log \frac{K - Y_3}{Y_3} - \log \frac{K - Y_0}{Y_0}.$$

En portant ces valeurs dans les équations (9) à (13) on a :

$$a_0 = \log \frac{K - Y_0}{Y_0}$$

$$a_1 = \frac{18\, B_1 - 9\, B_2 + 2 B_3}{6\, x_1}$$

$$a_2 = \frac{4\, B_2 - 5\, B_1 - B_3}{2 x_1^2}$$

$$a_3 = \frac{B_3 + 3\, B_1 - 3\, B_2}{6\, x_1^3}$$

8° *Calcul des constantes de la forme généralisée :*

$$Y = p + \frac{K}{1 + e^{\frac{\beta - t}{\alpha}}}$$

MM. Pearl et Reed ne disent pas par quel procédé ils effectuent le calcul de ces constantes. Plusieurs artifices de caclul permettent du reste de le faire facilement :

Prenons quatre recensements équidistants, le temps 0 étant l'époque du premier, et l'unité de temps étant l'intervalle entre deux recensements successifs.

On doit avoir :

$$Y_0 = p + \frac{K}{1 + e^{\frac{\beta}{\alpha}}}$$
$$Y_1 = p + \frac{K}{1 + e^{\frac{\beta - 1}{\alpha}}} \tag{14}$$
$$Y_2 = p + \frac{K}{1 + e^{\frac{\beta - 2}{\alpha}}}$$
$$Y_3 = p + \frac{K}{1 + e^{\frac{b - 3}{\alpha}}}.$$

Posons :

$$Z_0 = Y_0 - p = \frac{K}{1 + e^{\frac{\beta}{\alpha}}}$$
$$Z_1 = Y_1 - p = \frac{K}{1 + e^{\frac{\beta - 1}{\alpha}}} \tag{15}$$
$$Z_2 = Y_2 - p = \frac{K}{1 + e^{\frac{\beta - 2}{\alpha}}}$$
$$Z_3 = Y_3 - p = \frac{K}{1 + e^{\frac{\beta - 3}{\alpha}}}.$$

Formons les expressions :

$$\delta_1 = \frac{1}{Z_0} - \frac{1}{Z_1} = \frac{1}{K} e^{\frac{\beta}{\alpha}} \left(1 - e^{-\frac{1}{\alpha}}\right)$$
$$\delta_2 = \frac{1}{Z_1} - \frac{1}{Z_2} = \frac{1}{K} e^{\frac{\beta - 1}{\alpha}} \left(1 - e^{-\frac{1}{\alpha}}\right)$$
$$\delta_3 = \frac{1}{Z_2} - \frac{1}{Z_3} = \frac{1}{K} e^{\frac{\beta - 2}{\alpha}} \left(1 - e^{-\frac{1}{\alpha}}\right)$$
$$\frac{\delta_1}{\delta_2} = \frac{e^{\frac{\beta}{\alpha}}}{e^{\frac{\beta - 1}{\alpha}}} = e^{-\frac{1}{\alpha}}$$
$$\frac{\delta_2}{\delta_3} = \frac{e^{\frac{\beta - 1}{\alpha}}}{e^{\frac{b - 2}{\alpha}}} = e^{-\frac{1}{\alpha}}.$$

On en déduit :

$$\delta_1 \delta_3 = \delta_2^2$$

Ou :

$$\frac{(Z_1 - Z_0)(Z_3 - Z_2)}{Z_0 Z_3} = \frac{(Z_2 - Z_1')^2}{Z_1 Z_2}$$
$$(y_2 - p)(y_1 - y_0)(y_3 - y_2)(y_1 - p) = (y_2 - y_1)^2 (y_0 - p)(y_3 - p).$$

On peut tirer $p$ de cette équation du second degré, et n'y a plus qu'à appliquer une des méthodes précédentes en utilisant les Z comme données.

Il nous semble préférable d'employer le procédé suivant : Les quatre valeurs expérimentales de $y$ nécessaires pour déterminer nos quatre constantes sont encore fournies par les équations (14).

Posons :

$$\begin{aligned}
Y_1 &= y_1 - y_0 = K\left(\frac{1}{1+e^{\frac{\beta-1}{\alpha}}} - \frac{1}{1+e^{\frac{\beta}{\alpha}}}\right) \\
Y_2 &= y_2 - y_0 = K\left(\frac{1}{1+e^{\frac{\beta-2}{\alpha}}} - \frac{1}{1+e^{\frac{\beta}{\alpha}}}\right) \\
Y_3 &= y_3 - y_0 = K\left(\frac{1}{1+e^{\frac{\beta-3}{\alpha}}} - \frac{1}{1+e^{\frac{\beta}{\alpha}}}\right) \quad (16) \\
Y_4 &= y_2 - y_1 = K\left(\frac{1}{1+e^{\frac{\beta-2}{\alpha}}} - \frac{1}{1+e^{\frac{\beta-1}{\alpha}}}\right) \\
Y_5 &= y_3 - y_1 = K\left(\frac{1}{1+e^{\frac{\beta-3}{\alpha}}} - \frac{1}{1+e^{\frac{\beta-1}{\alpha}}}\right)
\end{aligned}$$

Formons ensuite les expressions :

$$\begin{aligned}
d_1 &= \frac{1}{Y_1} - \frac{1}{Y_2} \\
d_2 &= \frac{1}{Y_2} - \frac{1}{Y_3} \\
d_3 &= \frac{1}{Y_4} - \frac{1}{Y_5}
\end{aligned}$$

on trouve immédiatement :

$$\begin{aligned}
d_1 &= \frac{\left(1+e^{\frac{\beta}{\alpha}}\right)^2}{Ke^{\frac{\beta}{\alpha}}} \; \frac{e^{\frac{-1}{\alpha}}}{1-e^{\frac{-2}{\alpha}}} \\
d_2 &= \frac{\left(1+e^{\frac{\beta}{\alpha}}\right)^2}{Ke^{\frac{\beta}{\alpha}}} \; \frac{e^{\frac{-2}{\alpha}}\left(1-e^{\frac{-1}{\alpha}}\right)}{\left(1-e^{\frac{-2}{\alpha}}\right)\left(1-e^{\frac{-3}{\alpha}}\right)} \qquad (17) \\
d_3 &= \frac{\left(1+e^{\frac{\beta-1}{\alpha}}\right)^2}{Ke^{\frac{\beta}{\alpha}}} \; \frac{1}{1-e^{\frac{-2}{\alpha}}}
\end{aligned}$$

Divisons membre à membre les deux premières équations (17). Il vient :

$$\frac{d_1}{d_2} = e^{\frac{1}{\alpha}} + 1 + e^{\frac{-1}{\alpha}}$$

Pour résoudre cette équation réciproque qui admet pour racines $e^{\frac{1}{\alpha}}$ et $e^{-\frac{1}{\alpha}}$, posons :

$$\frac{d_1}{d_2} - 1 = A$$

$$e^{\pm\frac{1}{\alpha}} = \frac{A \pm \sqrt{A^2 - 4}}{2} \qquad (18)$$

Divisons membre à membre les deux dernières équations (17) :

$$\frac{d_1}{d_3} = \left(\frac{1 + e^{\frac{\beta}{\alpha}}}{1 + e^{\frac{\beta - 1}{\alpha}}}\right)^2 e^{-\frac{1}{\alpha}}$$

d'où :

$$e^{\frac{\beta - 1}{\alpha}} = \frac{1 - \sqrt{\frac{d_1}{d_3} e^{\frac{1}{\alpha}}}}{\sqrt{\frac{d_1}{d_3} e^{\frac{1}{\alpha}}} - e^{\frac{1}{\alpha}}}$$

On en tire facilement $\beta$.

En portant les valeurs de $\alpha$ et $\beta$ dans deux quelconques des équations (14), on a immédiatement les valeurs de $p$ et de K.

C'est ce procédé que nous utiliserons plus loin pour ajouter la population française.

9° *Expériences de M. Pearl et M^me^ Parker sur les drosophiles.* — La loi de Verhulst étant une loi biologique, on doit pouvoir la vérifier à l'aide d'expériences biologiques; en les réglant convenablement, on peut éliminer des facteurs parasitaires comme l'émigration et l'immigration, la variation des conditions économiques; on peut aussi avoir des données aussi nombreuses qu'on le désire, ce que ne fournissent pas les statistiques démographiques. Dans une des nombreuses expériences faites sur la drosophila melanogaster par le Dr Pearl et Mme S. Parker, on isola dans une bouteille de lait d'une demi-pinte un couple de ces mouches; tous les trois jours on procéda à un recensement de la population résultante vivant dans un milieu parfaitement limité. Au bout d'un mois, leur nombre dépassa 300, on dressa la courbe de leur accroissement. Elle suivait une courbe de Verhulst d'aussi près que l'on pouvait l'attendre d'un nombre aussi restreint d'individus soumis à l'examen; par contre, les derniers points, nettement au-dessous semblent indiquer une asymétrie de la courbe.

10° *Ajustement de la population anglaise.* — Nous donnons ci-après à titre d'exemples, quelques ajustements effectués par M. Yule, pour les populations anglaises et américaine.

TABLEAU

*Population de l'Angleterre et du pays de Galles.*

| ANNÉE | POPULATION (en millions) | VALEUR AJUSTÉE DE LA POPULATION (1) | (2) | (3) |
|---|---|---|---|---|
| 1801 | 8,89 | 8,89 | 8,93 | 9,98 |
| 1811 | 10,16 | 10,29 | 10,32 | 10,35 |
| 1821 | 12,00 | 11,88 | 11,89 | 11,91 |
| 1831 | 13,90 | 13,68 | 13,66 | 13,67 |
| 1841 | 15,91 | 15,69 | 15,65 | 15,64 |
| 1851 | 17,93 | 17,93 | 17,87 | 17,83 |
| 1861 | 20,07 | 20,40 | 20,33 | 20,26 |
| 1871 | 22,71 | 23,11 | 23,02 | 22,93 |
| 1881 | 25,97 | 26,05 | 25,95 | 25,83 |
| 1891 | 29,00 | 29,19 | 29,11 | 28,96 |
| 1901 | 32,53 | 32,53 | 32,47 | 32,31 |
| 1911 | 36,07 | 36,03 | 36,02 | 35,85 |
| Sommes des erreurs positives |  | +1,13 | + 0,88 | + 0,69 |
| Sommes des erreurs négatives |  | —0,60 | — 0,80 | — 1,31 |
| Sommes des valeurs absolues des erreurs |  | 1,73 | 1,68 | 2,00 |

La courbe (1) est déterminée par les recensements de 1801, 1851 et 1901.
La courbe (2) est déterminée à l'aide de la somme des inverses.
La courbe (3) est déterminée par la méthode des moindres carrés.
Pour chacune des ces trois courbes, les constates sont
(1) K = 91,211 millions, $\alpha$ = 61,134 ans, $t_0$ = 1936,065.
(2) K = 97,300 — $\alpha$ = 62,475 ans, $t_0$ = 1944,19.
(3) K = 100,054 — $\alpha$ = 63,435 ans, $t_0$ = 1947,96.
Le plus grand écart est de 400 000 habitants, produit par la courbe (1) en 1871. Les chiffres de la courbe (2) ne sont pas beaucoup plus satisfaisants que ceux de la courbe (1), mais les écarts y sont répartis à peu près également au-dessus et au-dessous de la courbe
C est la courbe (3) qui donne le plus mauvais ajustement

Il faut retenir de ce tableau : 1° que toutes les données se trouvent à gauche du point d'inflexion, et que par conséquent on ne peut fixer ce dernier avec précision, et encore moins avoir une idée, même approchée, de la population limite; 2° que les données ne couvrent que 1,8 intervalle type; d'une portion si restreinte de la courbe, on ne peut évidemment conclure que la courbe est bien celle de Verhulst.

11° *Ajustement de la population des États-Unis.* — Le tableau suivant donne un ajustement de la population des États-Unis à l'aide d'une parabole cubique calculée par le professeur Bowley par la méthode des moindres carrés.

$$y = 23{,}48 + 65{,}95\, t + 0{,}687\, t^2 + 0{,}00215\, t^3$$

Origine 1850, unité = 1 an (1).

Il donne à côté l'ajustement par la courbe de Verhulst.

$$y = \frac{199{,}790}{1 + e^{\frac{1915 - t}{31{,}93}}} \quad (2)$$

La courbe (2) est particulièrement intéressante car les données couvrent une grande partie de sa moitié gauche, et atteignent presque le point d'inflexion, grâce au peu de durée de l'intervalle-type.

TABLEAU

*Population des États-Unis (Unité = 10.000)*

| Année | Population | Courbe (1) | Courbe (2) de Verhulst |
|---|---|---|---|
| — | — | — | — |
| **1790** . . . . | 393 | 400 | 390 |
| **1800** . . . . | 531 | 499 | 530 |
| **1810** . . . . | 724 | 672 | 718 |
| **1820** . . . . | 964 | 930 | 968 |
| **1830** . . . . | 1.287 | 1.287 | 1.302 |
| **1840** . . . . | 1.707 | 1.755 | 1.739 |
| **1850** . . . . | 2.319 | 2.348 | 2.305 |
| **1860** . . . . | 3.144 | 3.078 | 3.025 |
| **1870** . . . . | 3.856 | 3.959 | 3.919 |
| **1880** . . . . | 5.016 | 5.003 | 5.000 |
| **1890** . . . . | 6.295 | 6.223 | 6.262 |
| **1900** . . . . | 7.599 | 7.632 | 7.680 |
| **1910** . . . . | 9 197 | 9.243 | 9.203 |
| *Erreurs* | | | |
| maxima | | 103 | 119 |
| moyenne | | 41 | 30 |
| quadratique moyenne | | 49 | 47 |

On voit qu'une parabole cubique ajuste la population américaine aussi bien qu'une courbe de Verhulst dont le seul intérêt ici est de ne pas conduire à des résultats absurdes si on veut extrapoler.

Il convient du reste de formuler des réserves sur la courbe (2), car les États-Unis ne répondent guère aux conditions fondamentales posées au début de ce travail. Toute la population blanche était à l'origine groupée le long du littoral de l'Atlantique et du Pacifique. L'occupation des régions centrales fut davantage une colonisation qu'une extension, et il ne peut guère être question d'étendue bornée, au moins dans la première moitié du siècle dernier. D'autre part, on peut avoir quelques doutes sur la valeur des premiers recensements dont on se représente mal l'opération dans un pays à l'époque en partie habité par des indigènes peu sociables.

---

# TROISIÈME PARTIE

## APPLICATION A LA FRANCE DE LA FORMULE DE VERHULST

12° *Matériel statistique.* — Le matériel statistique dont on dispose pour appliquer à la France la formule de Verhulst ne peut être utilisé sans précautions. Des discontinuités ont été apportées à l'accroissement naturel de la population par l'acquisition de la Savoie et du Comté de Nice en 1860, par la perte de l'Alsace-Lorraine en 1871 et son recouvrement en 1918. Nous verrons plus loin comment on peut résoudre cette difficulté à l'aide d'évaluations ne nécessitant que des hypothèses très légitimes.

Quant aux chiffres fournis par les recensements, ils n'ont pas tous la même valeur. Le nombre des habitants vivant en 1801 sur le territoire de 1914 atteignait environ 27 millions. En 1806 le chiffre indiqué par les rapports des préfets atteignait presque 29 millions; la comparaison avec la balance des naissances et des décès accusait une immigration de 1.350.000 étrangers. Par contre, en comparant les chiffres de 1806 avec ceux de 1821, on trouve une émigration apparente de 1.200.000. On est donc amené à conclure que le chiffre de 1806 est trop élevé, et que le chiffre de 1801 est peut-être trop faible.

Le recensement de 1821 comporte encore une part d'évaluation.

L'opération devient régulière à partir de 1831, mais c'est seulement ed 1836 que la création effective de la liste nominative permit un relevé sûr. L'emploi du bulletin individuel à partir de 1876 constitue un perfectionnement important. Les recensements utilisables seront donc ceux de : 1801, 1821, 1831 à 1866, 1876 à 1911.

On ne peut se servir du recensement de 1872, fait dans de mauvaises conditions matérielles; on doit du reste le rejeter à canse des pertes de la guerre et d'un taux de mortalité énorme, et aussi parce que la détermination facile des constantes de la formule de Verhulst nécessite des données équidistantes dans le temps.

Une légère source d'erreur est constituée par le fait que les recensements ont été effectués à des périodes différentes de l'année; elle est négligeable pour les recensements récents.

Les chiffres que nous utiliserons dans la présente étude sont ceux fournis par le recensement de 1911 pour la population vivant sur le territoire à cette date. La correction nécessitée par les acquisitions et pertes territoriales a été effectuée à l'aide de la balance des naissances et des décès dans ces territoires.

M. Yule s'est servi dans le même but d'une hypothèse qui donne des résultats comparables aux précédents et qui consiste à admettre que la population ajoutée ou retranchée a suivi dans le passé la même loi d'évolution que la population totale. Voici comment on peut par exemple tenir compte de l'acquisition de Nice et de la Savoie en 1860 :

Si nous appelons $p$ et P la population française en 1861, avant et après les acquisitions $\varpi$, on a évidemment :

$$P = p + \varpi$$

Posons :

$$P = \alpha p$$

et supposons que par le passé $\varpi$ ait varié comme $p$. Si à l'époque $t_\beta$ les valeurs de $p$ et $\varpi$ étaient des fractions $\beta$ de ce qu'elles étaient en 1861, on a :

$$P_\beta = \beta p + \beta \varpi = \beta (p + \varpi)$$
$$= \beta \alpha p = \alpha p_\beta$$

On aura donc les valeurs corrigées de P en multipliant les valeurs brutes de $p$ par la constante $\frac{P}{p}$ calculée en 1861:

Le tableau ci-joint fournit les chiffres bruts du recensement, les chiffres corrigés fournis par la statistique générale de la France, les chiffres corrigés, par la méthode de M. Yule et la comparaison des deux dernières séries de chiffres. On voit que, seuls les recensements douteux de 1801 et 1821 donnent des résultats nettement différents. Toutefois, aux chiffres fournis par la méthode de M. Yule il faut préférer ceux fournis par la balance des naissances et des décès et qui proviennent de registres d'état-civil bien tenus.

*Population des territoires français occupés avant* **1918**
*(Nice et la Savoie compris. Alsace-Lorraine exclue.)*

| ANNÉES | POPULATION légale fournie par le recensement | CHIFFRE CORRIGÉ DE LA POPULATION Donné par la méthode de M. Yule | Donné par la Statistique générale de la France | DIFFÉRENCE (en milliers) des 2 colonnes précédentes |
|---|---|---|---|---|
| 1801 | 27,349 | 26,694 | 27,000 | — 306 |
| 1821 | 30,462 | 29,732 | 29,871 | — 139 |
| 1831 | 32,569 | 31,790 | 31,788 | + 2 |
| 1836 | 33,541 | 32,738 | 32,760 | — 22 |
| 1841 | 34,230 | 33,411 | 33,401 | + 10 |
| 1846 | 35,402 | 34,554 | 34,547 | + 7 |
| 1851 | 35,783 | 34,927 | 34,902 | + 25 |
| 1856 | 36,039 | 35,177 | 35,174 | + 3 |
| 1861 | 36,717 (1) | » | » | |
| 1861 | 37,386 (2) | 35,838 | 35,835 | + 3 |
| 1866 | 38,067 | 36,491 | 36,485 | + 6 |
| 1872 | 37,662 (3) | | » | |
| 1872 | 36,103 (4) | | 36,103 | |
| 1876 | 36,906 | | 36,906 | |
| 1881 | 37,672 | | 37,672 | |
| 1886 | 38,219 | | 38,219 | |
| 1891 | 38,343 | | 38,343 | |
| 1896 | 38,518 | | 38,518 | |
| 1901 | 38,962 | | 38,962 | |
| 1906 | 39,252 | | 39,252 | |
| 1911 | 39,602 | | 39,602 | |

Les chiffres de la population sont exprimés en millions
(1) Nice et la Savoie non compris.
(2) Nice et la Savoie compris.
(3) Alsace-Lorraine comprise.
(4) Alsace Lorraine non comprise.

**13°** *Ajustement de la population française par la formule de Verhulst.* — 1° Procédé des trois ordonnées équidistantes.

On peut utiliser les recensements de 1801, 1856, 1911 séparés par des intervalles de cinquante-cinq ans. Nous donnons à titre d'exemple le détail des calculs.

| ANNÉE | POPULATION | INVERSES | DIFFÉRENCE 1re | DIFFÉRENCE 2e |
|---|---|---|---|---|
| 1801 | $y_0 = 27{,}000$ | 0,037037 | | |
| 1856 | $y_1 = 35{,}174$ | 0,028430 | $d_1 = 0{,}008607$ | |
| 1911 | $y_2 = 39{,}602$ | 0,025251 | $d_2 = 0{,}003179$ | $d_1 - d_2 = 0{,}005428$ |

$$e^{\frac{1}{\alpha}} = \frac{d_1}{d_2} = \frac{0{,}008607}{0{,}003179} = 2{,}707509$$

$$\alpha = \frac{\log e}{\log 2{,}707509} = \frac{0{,}434.2945}{0{,}432.5699} = 1{,}003.9869.$$

L'unité de temps étant cinquante-cinq ans, la valeur de $\alpha$ en années est : $\alpha = 1{,}003.9869 \times 55 = 55{,}219$ ans.

$$\frac{d_1^2}{d_1 - d_2} = \frac{0{,}000.0741}{0{,}005.4280} = 0{,}013.6475$$

$$\frac{1}{K} = \frac{1}{y_0} - \frac{d_1^2}{d_1 - d_2} = 0{,}037037 - 0{,}013648 = 0{,}023389$$

$$K = \frac{1}{0{,}023389} = 42{,}754 \text{ millions.}$$

C'est la population limite.

$$e^{\frac{\beta}{\alpha}} = \frac{K - y_0}{y_0} = \frac{15,754}{27,000} = 0,58349$$

$$\frac{\beta}{\alpha} = \frac{\log 0,58349}{0,4342945} = - 0,5387$$

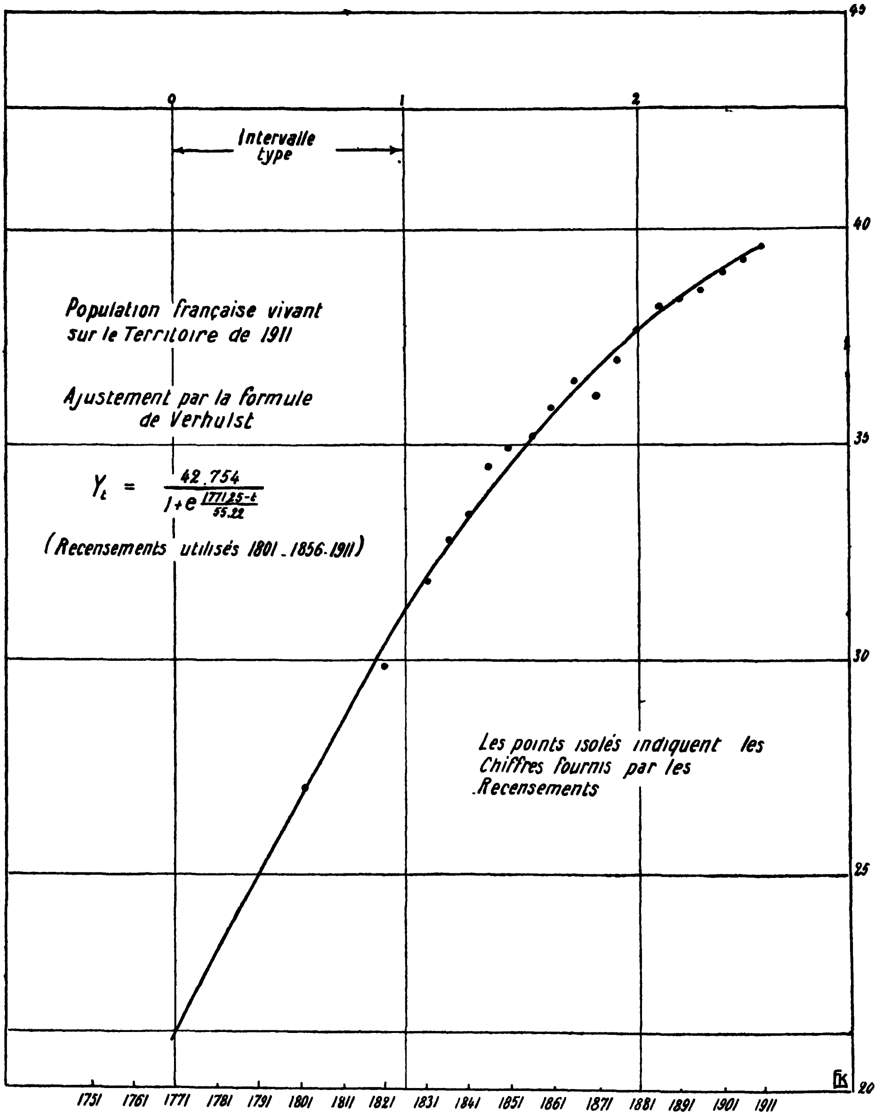
$$\beta = - 0,5387 \times 55,219 = - 29,748.$$

L'abscisse du point d'inflexion est donc :

$$1801 - 29,748 = 1.771,252$$

Finalement la formule est :

$$y_t = = \frac{42,754}{1 + e^{\frac{1771,252 - t}{55,219}}}.$$




2° *Procédé de la somme des inverses.* — L'absence de plusieurs recensements, et le décalage d'une année de celui de 1872, d'ailleurs à rejeter, empêchent l'utilisation de tout l'ensemble, en particulier des recensements les plus anciens. On ne peut guère employer que le groupement suivant, l'intervalle entre deux recensements successifs étant 15 ans.

| ANNÉE | INVERSE de la population | ANNÉE | INVERSE de la population | ANNÉE | INVERSE de la population |
|---|---|---|---|---|---|
| 1836 . . . . . | 0,030.5250 | 1866 . . . . | 0,027.4085 | 1896 . . . . . | 0,025.9619 |
| 1851 . . . . | 0,028.6517 | 1881 . . . . . | 0,026.5449 | 1911 . . . . . | 0,025.2513 |
| $S_1$ . . . . . . | 0,059.1767 | $S_2$ . . . . . . | 0,053.9534 | $S_3$ . . . . . . | 0,051 2132 |

Posons :

$$D_1 = S_1 - S_2 = 0{,}005\ 2232$$

$$D_2 = S_2 - S_3 = 0{,}002.7403$$

$$\frac{2}{e\alpha} = \frac{D_1}{D_2} = \frac{0{,}005.2232}{0{,}002.7403} = 1{,}906.0833$$

$$\alpha = 2 \times \frac{0{,}434.2945}{\log 1{,}9060833} = 3{,}100.5334.$$

L'unité étant une période de 15 ans, la valeur de $\alpha$ en années est :

$$3{,}100.5334 \times 15 = 46{,}508 \text{ ans.}$$

$$\frac{D_1^2}{D_1 - D_2} = 0{,}010.9879$$

$$\frac{r}{K} = \frac{2}{K} = S_1 - \frac{D_1^2}{D_1 - D_2} = 0{,}059.1767 - 0{,}010.9879 = 0{,}048.1888$$

$$K = \frac{2}{0{,}048.1888} = 41{,}503 \text{ millions}$$

$$e = \frac{1 - e^{-\frac{2}{\alpha}}}{1 - e^{-\frac{1}{\alpha}}} = 1 + e^{-\frac{1}{\alpha}}$$

$$e^{-\frac{1}{\alpha}} = \frac{1}{\sqrt{e^{\frac{2}{\alpha}}}} = 0{,}724.3176. \text{ d'où}$$

$$C = 1{,}724.3176$$

$$e^{\frac{\beta}{\alpha}} = \frac{K}{c} \frac{D_1^2}{D_1 - D_2} = 0{,}264473$$

$$\beta = \alpha \frac{\log 0{,}264473}{\log e} \times (15 \text{ ans}) = -\ 61{,}856.$$

L'abscisse du point d'inflexion est donc :

$$1836 - 61{,}856 = 1774{,}144$$

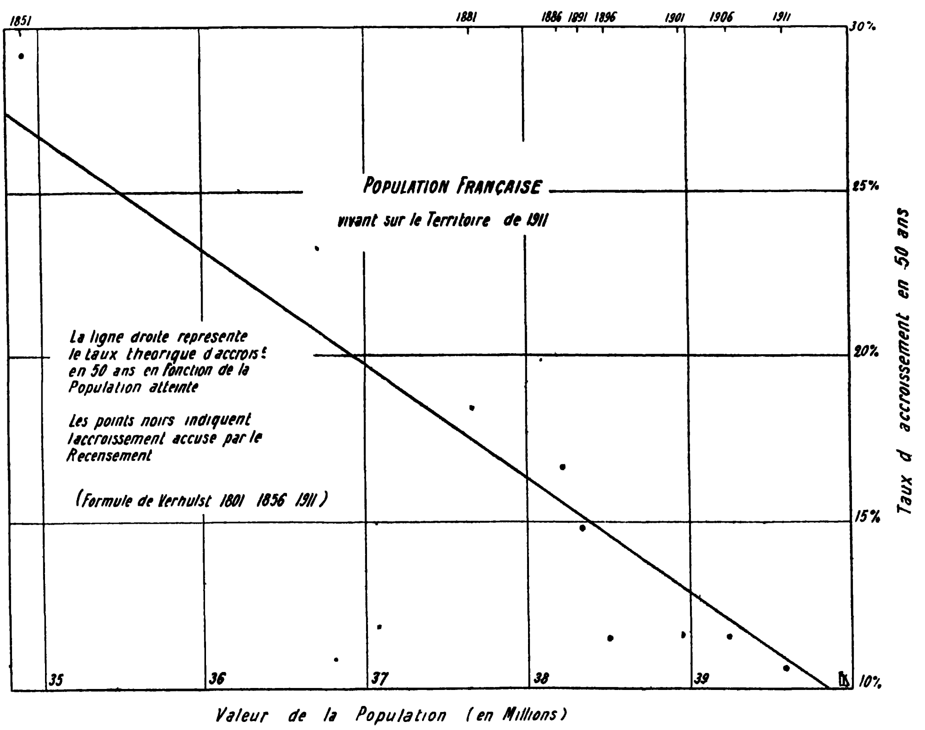
et l'équation cherchée est finalement :

$$y_t = \frac{41{,}503}{1 + e^{\frac{1774{,}144 - t}{46{,}508}}}$$

Dans le tableau qui suit, nous comparons la population enregistrée, avec la population ajustée à l'aide de ces deux formules.

*Comparaison d'ajustements de la population française à l'aide de diverses formules.*

| ANNÉE | POPULATION enregistrée | VALEURS AJUSTÉES DE LA POPULATION — (1) Formule de Verhulst — Popul. | Écarts en + | Écarts en — | (2) Formule de Verhulst — Popul. | Écarts en + | Écarts en — | (3) Formule de Pearl — Popul. | Écarts en + | Écarts en — | (4) Formule du Dr Bowley — Popul. | Écarts en + | Écarts en — |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1801 | 27,000 | 27,000 | | | 26,582 | | 418 | 27,000 | | | 27,090 | 90 | |
| 1821 | 29,871 | 30,404 | 533 | | 30,403 | 532 | | 30,197 | 326 | | 30,440 | 570 | |
| 1831 | 31,788 | 31,932 | 144 | | 32,061 | 273 | | 31,788 | | | 31,940 | 150 | |
| 1836 | 32,760 | 32,648 | | 112 | 32,823 | 63 | | 32,556 | | 204 | 32,630 | | 130 |
| 1841 | 33,401 | 33,330 | | 71 | 33,538 | 137 | | 33,287 | | 114 | 33,300 | | 100 |
| 1846 | 34,547 | 33,978 | | 569 | 34,207 | | 340 | 33,987 | | 560 | 33,950 | | 600 |
| 1851 | 34,902 | 34,592 | | 310 | 34,831 | | 71 | 34,646 | | 256 | 34,560 | | 340 |
| 1856 | 35,174 | 35,174 | | | 35,411 | 237 | | 35,263 | 89 | | 35,140 | | 30 |
| 1861 | 35,835 | 35,722 | | 113 | 35,960 | 125 | | 35,835 | | | 35,690 | | 150 |
| 1866 | 36,485 | 36,238 | | 247 | 36,446 | | 39 | 36,361 | | 124 | 36,210 | | 270 |
| 1872 | 36,103 | 36,816 | 713 | | 36,992 | 889 | | 36,932 | 829 | | 36,790 | 690 | |
| 1876 | 36,906 | 37,177 | 271 | | 37,326 | 420 | | 37,278 | 372 | | 37,150 | 240 | |
| 1881 | 37,672 | 37,601 | | 71 | 37,713 | 41 | | 37,672 | | | 37,600 | | 70 |
| 1886 | 38,219 | 37,998 | | 221 | 38,068 | | 151 | 38,026 | | 193 | 37,990 | | 230 |
| 1891 | 38,343 | 38,367 | 24 | | 38,392 | 49 | | 38,343 | | | 38,360 | 20 | |
| 1896 | 38,518 | 38,711 | 193 | | 38,687 | 169 | | 38,624 | 106 | | 38,710 | 190 | |
| 1901 | 38,962 | 39,031 | 69 | | 38,957 | | 5 | 38,874 | | 88 | 39,020 | 60 | |
| 1906 | 39,252 | 39,327 | 75 | | 39,202 | | 50 | 39,095 | | 157 | 39,300 | 50 | |
| 1911 | 39,602 | 39,602 | | | 39,425 | | 177 | 39,290 | | 312 | 39,550 | | 50 |
| | Somme des écarts | | 2 022 | 1.714 | | 2.935 | 1.251 | | 1 722 | 2.008 | | 2.060 | 1.970 |
| | | | 3.736 | | | 4.186 | | | 3.730 | | | 4,030 | |
| | Écart moyen. | | 196 | | | 220 | | | 196 | | | 212 | |




Il résulte immédiatement de ce tableau que les courbes calculées par les deux procédés donnent des résultats presque identiques. Le point d'inflexion se place entre 1771 et 1774, et la valeur limite de la population varie entre 41,5 millions et 42,8 millions; par contre, l'écart entre les deux valeurs de l'intervalle type atteint environ un cinquième de ces valeurs qu'on doit fixer à environ 50 ans.

Si l'on effectue la somme des écarts à la population enregistrée, on voit que pour la courbe (1) calculée à l'aide des trois ordonnées équidistantes 1801-1856-1911, les écarts positifs et négatifs ont à peu près la même importance, leur total est 3,736, ce qui donne un écart moyen de 0,196. Avec la courbe (2) établie à l'aide de la somme des inverses, on a une somme d'écarts positifs près de trois fois plus grande que celle des écarts négatifs; le total des écarts est 4,186, ce qui donne un écart moyen de 0,220, il semblerait donc que cet ajustement soit moins bon que le précédent; en réalité, si on laisse de côté les écarts correspondant à la région extrapolée de la courbe (1801 1831), l'écart moyen s'abaisse à 0,185, mais les écarts positifs restent toujours beaucoup plus importants que les écarts négatifs; il faut attribuer ceci en grande partie aux deux recensements exceptionnels de 1872 et 1876 qui ensemble provoquent un total d'écarts positifs de 1,309 et faussent la comparaison, comparaison faussée pour la même raison dans la courbe (1).

(On peut noter ici que les perturbations provoquées par les guerres des deux Empires et les épidémies de choléra qui affligèrent la France au cours du XIXe siècle, n'ont apporté que des écarts de peu de durée à la loi de Verhulst, conformément au principe formulé par Botero et énoncé au début de ce travail.)

Dans les deux cas, les données ne couvrent que la portion de la courbe à droite du point d'inflexion, sur une longueur d'environ 2,4 intervalles-types. Elles permettent donc une bonne évaluation de la population limite; par contre, comme elles n'atteignent pas le point d'inflexion, on ne peut déterminer avec exactitude l'intervalle-type, c'est-à-dire l'allongement plus ou moins grand de la courbe. On ne peut donc conclure de ces chiffres que la population française suit un cycle de Verhulst et que la courbe est symétrique autour de son point d'inflexion.

Malgré l'absence de recensements avant 1801, on peut cependant faire des évaluations plus ou moins exactes de notre population avant cette date (on relève déjà un recensement des feux dans tout le pays sous Charles IX). Elles donnent toutes des valeurs très supérieures à celles qu'accorde la loi de Verhulst. Ainsi, celle-ci donne à la population vers 1600 une valeur d'environ 2 millions, chiffre grossièrement faux.

Les données sont à peine suffisantes pour vérifier la loi au cours du XIXe siècle; tout au plus peut-on affirmer qu'elle permet de suivre de près l'évolution de notre population pendant cette période; les quelques données plus anciennes que nous possédons montrent qu'elle ne s'applique pas du tout au XVIIIe siècle. Nous ne pouvons guère invoquer en sa faveur que l'exemple un peu douteux des États-Unis, partis au début du siècle dernier d'une population très faible, et ayant atteint leur point d'inflexion vers 1915.

Cette grave objection d'un ajustement défectueux du passé disparaît si l'on admet l'hypothèse d'additivité des cycles de Pearl et Reed.

14° *Ajustement de la population française par la formule généralisée.*

$$y_t = p + \frac{K}{1 + e^{\frac{b-t}{\alpha}}}.$$

Les lacunes dans les recensements et la guerre de 1870-1871 nous empêchent de faire un ajustement par la formule de Pearl couvrant toute la période 1801-1911. Les seules séries de quatre recensements équidistants nous permettant un ajustement de quelque étendue, sont :

1801 — 1831 — 1861 — 1891
1836 — 1861 — 1886 — 1911

Exposons, à titre d'exemple, le détail des calculs de la première qui remontant au plus ancien recensement est évidemment celle qui nous fournira la meilleure détermination de $p$.

| ANNÉE | POPULATION | | |
|---|---|---|---|
| 1801 | $y_0 = 27,000$ | $Y_1 = y_1 - y_0 = 4,788$ | $\frac{1}{Y_1} = 0,208.8555$ |
| 1831 | $y_1 = 31,788$ | $Y_2 = y_2 - y_0 = 8,835$ | $\frac{1}{Y_2} = 0,113.1862$ |
| 1861 | $y_2 = 35,835$ | $Y_3 = y_3 - y_0 = 11,343$ | $\frac{1}{Y_3} = 0,088.1601$ |
| 1891 | $y_3 = 38,343$ | $Y_4 = y_2 - y_1 = 4,047$ | $\frac{1}{Y_4} = 0,247.0966$ |
| | | $Y_5 = y_3 - y_1 = 6,555$ | $\frac{1}{Y_5} = 0,152.5553$ |

$$d_1 = \frac{1}{Y_1} - \frac{1}{Y_2} = 0,095.6693$$

$$d_2 = \frac{1}{Y_2} - \frac{1}{Y_3} = 0,025.0261$$

$$d_3 = \frac{1}{Y_4} - \frac{1}{Y_5} = 0,094.5413$$

$$\frac{d_1}{d_2} = 3,822.7810$$

$$\frac{d_1}{d_3} = 1,011.9313$$

$$A = \frac{d_1}{d_2} - 1 = 2,822781$$

$$e^{\frac{1}{\alpha}} = \frac{A + \sqrt{A^2 - 4}}{2} = 2,407394$$

$$e^{-\frac{1}{\alpha}} = \frac{A - \sqrt{A^2 - 4}}{2} = 0,415387$$

$$\alpha = \frac{\log e}{\log 2,407394} = 1,1382458.$$

l'unité de temps étant 30 ans. La valeur en années est donc :

$$\alpha = 34,147374.$$

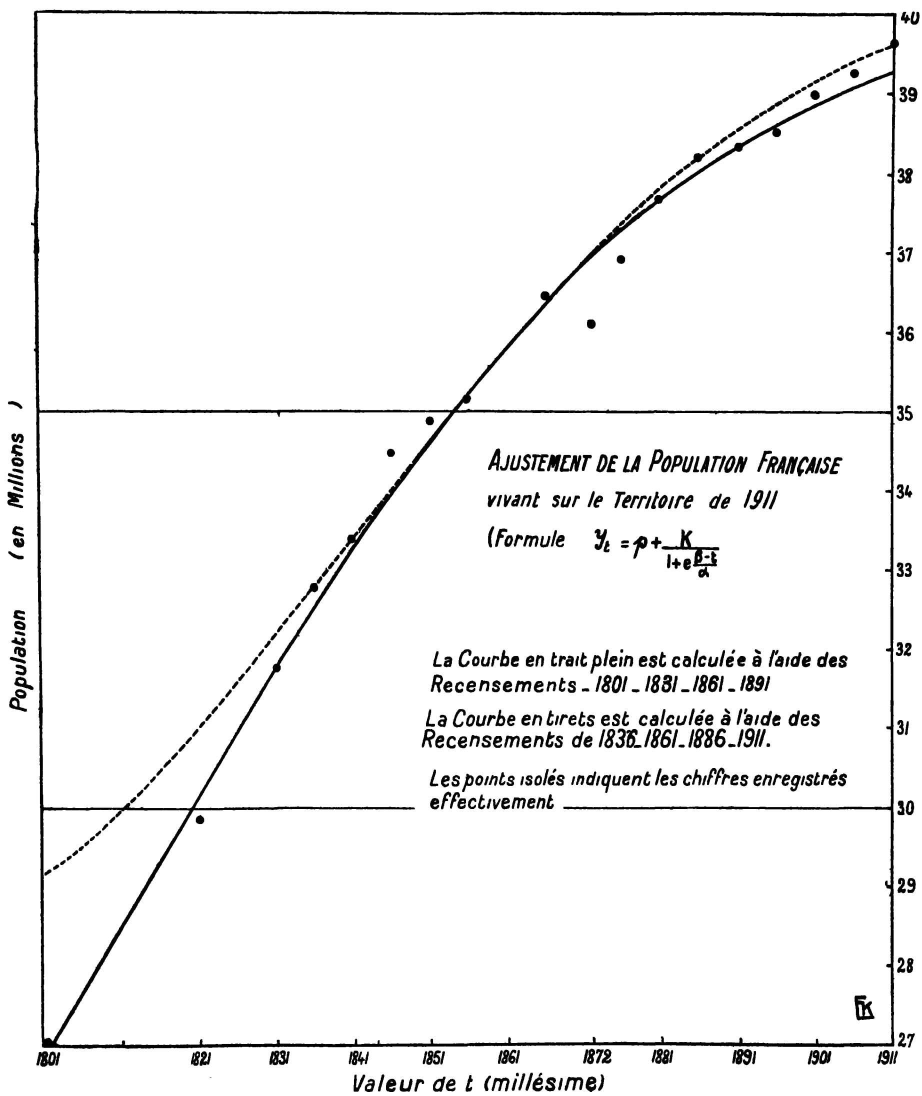
Les derniers chiffres significatifs n'ont évidemment aucune valeur et ne sont conservés qu'afin de retrouver, les constantes calculées, les mêmes chiffres calculés que les chiffres enregistrés pour les quatre recensements utilisés.

$$\sqrt{\frac{d_1}{d_3}\, e^{\frac{1}{\alpha}}} = \sqrt{2,4361173} = \pm\, 1,560807.$$

Seule, la valeur positive est à conserver, car seule elle rend positif $e^{\frac{b-1}{\alpha}}$

$$e^{\frac{b-1}{\alpha}} = \frac{1 - \sqrt{\frac{d_1}{d_3}\, e^{\frac{1}{\alpha}}}}{\sqrt{\frac{d_1}{d_3}\, e^{\frac{1}{\alpha}}} - e^{\frac{1}{\alpha}}} = \frac{1 - 1,560807}{1,560807 - 2,407394} = 0,6624328$$

$$\frac{b-1}{\alpha} \log e = \log 0,6624328 = -\, 0,1788582\cdot$$



L'unité de temps étant 30 ans, on a en années :

$$b - 30 = 34{,}147374 \times (-\,0{,}4118362) = -\,14{,}063125$$

$$b = 15{,}936875$$

$$e^{\frac{b-1}{a}} = 0{,}6624328$$

$$e^{\frac{b}{a}} = 1{,}5947367.$$

Le point d'inflexion a lieu en :

$$1801 + 15{,}936875 = 1816{,}936875.$$

Les deux premiers recensements nous donnent d'autre part :

$$y_0 = 27{,}000 = p + \frac{K}{1 + 1{,}5947367}$$

$$y_1 = 31{,}788 = p + \frac{K}{1 + 0{,}6624328}.$$

D'où :

$$y_1 - y_0 = 4{,}788 = K \times 0{,}2161325, \text{ ou :}$$

$$K = 22{,}153073$$

$$p = 18{,}462304.$$

La population limite est $K + p = 40{,}615$ millions.

On a finalement la formule :

$$y_t = 18{,}462 + \frac{22{,}153}{1 + e^{\frac{1816{,}937 - t}{34{,}147}}}.$$

Si au lieu d'utiliser les recensements de 1801, 1831, 1861, 1891 on utilise ceux de 1836, 1861, 1886, 1911, on obtient la formule :

$$y_t = 26{,}545 + \frac{14{,}243}{1 + e^{\frac{1843{,}230 - t}{28{,}252}}}.$$

Ces deux formules, à première vue très différentes, nous donnent cependant des valeurs presque identiques de la population limite $p + K$ atteinte à la fin du cycle (40,615 et 40,788), ce qui s'explique fort bien en admettant que la France soit dans la dernière phase du présent cycle. La différence entre les intervalles-types atteint environ un cinquième de leur valeur. La détermination du point d'inflexion varie de 26 ans; quant à celle de la constante $p$ correspondant à l'accroissement des précédents cycles, elle est franchement mauvaise, l'une des valeurs étant les deux tiers de l'autre. Il est évident que c'est la détermination 18.462 qui se rapproche le plus de la vérité, puisqu'elle est obtenue à l'aide de recensements plus anciens que l'autre, mais il faut bien en conclure que de tels chiffres sont illusoires puisque la possession de renseignements antérieurs de 35 ans suffisent à modifier aussi profondément la physionomie de la courbe. En fait, l'examen des points fournis par les chiffres des recensements montre deux inflexions assez nettes : l'une vers 1845 et l'autre vers 1817.

La détermination d'un cycle ne peut donc présenter quelque valeur que si les données en couvrent la plus grande partie.

Ceci devrait suffire à mettre en garde contre les extrapolations, nous verrons plus loin d'autres raisons qui les rendent particulièrement dangereuses lorsqu'une population suit la formule de Pearl.

Dans l'incertitude où nous sommes de la véritable forme de la loi démographique, il est indispensable de vérifier la répartition des écarts entre les chiffres bruts et les chiffres ajustés. Comme les observations ont porté sur un nombre de têtes peu variables, il est inutile de leur affecter un poids.

Le tableau suivant donne la distribution des écarts comparée à la distribution normale, les valeurs de l'écart type $\sigma$ fourni par la racine carrée de la moyenne des carrés des écarts étant calculées :

1° Pour la formule de Verhulst de 1836 à 1911 (procédé de la somme des inv.);
2° Comme ci-dessus en négligeant les écarts de 1872 et 1876;
3° Pour la formule de Pearl 1801-1891 extrapolée jusqu'en 1911;
4° Comme ci-dessus en négligeant les écarts de 1872 à 1876.

| | VALEUR supérieure des écarts | | | | | DISTRIBUTION théorique des écarts | | | | DISTRIBUTION observée des écarts | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Verhulst | | Pearl | | | Verhulst | | Pearl | | Verhulst | | Pearl | |
| | 1° | 2° | 3° | 4° | | 1° | 2° | 3° | 4° | 1° | 2° | 3° | 4° |
| $\pm \frac{1}{3}\sigma$ | 0,094 | 0,050 | 0,096 | 0,070 | 26 % | 4,2 | 3,6 | 4,9 | 4,4 | 7 | 4 | 7 | 5 |
| $\pm \frac{2}{3}\sigma$ | 0,188 | 0,099 | 0,191 | 0,139 | 50 % | 8 | 7 | 9,5 | 8,5 | 12 | 7 | 11 | 10 |
| $\pm \sigma$ | 0,283 | 0,149 | 0,287 | 0,209 | 68 % | 10,9 | 9,5 | 12,9 | 11,6 | 13 | 9 | 14 | 13 |
| $\pm \frac{4}{3}\sigma$ | 0,376 | 0,198 | 0,384 | 0,279 | 82 % | 13,1 | 11,5 | 15,6 | 13,9 | 14 | 12 | 17 | 14 |
| $\pm 2\sigma$ | 0,566 | 0,300 | 0,574 | 0,418 | 95 % | 15,2 | 13,3 | 18,1 | 16,2 | 15 | 13 | 18 | 16 |
| $\pm 3\sigma$ | 0,849 | 0,447 | 0,861 | 0,627 | 100 % | 16 | 14 | 19 | 17 | 15 | 14 | 19 | 17 |

On voit que la distribution des écarts est sensiblement conforme à la distribution théorique si on élimine les écarts de 1872 et 1876. Elle est aussi bonne pour la formule simple de Verhulst que pour celle de Pearl. On ne saurait guère en exiger une meilleure pour un nombre d'observations variant de 14 à 19.

(*A suivre*). KRUMMEICH.