HAL open science

# Un modèle mathématique des débuts de l'épidémie de coronavirus en France

Nicolas Bacaër

▶ **To cite this version:**

Nicolas Bacaër. Un modèle mathématique des débuts de l'épidémie de coronavirus en France. Mathematical Modelling of Natural Phenomena, 2020, 15, pp.29. 10.1051/mmnp/2020015 . hal-02509142v8

**HAL Id: hal-02509142**

**https://hal.science/hal-02509142v8**

Submitted on 9 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Un modèle mathématique des débuts de l'épidémie de coronavirus en France

Nicolas Bacaër*

Math. Model. Nat. Phenom. 15 (2020) 29
doi : 10.1051/mmnp/2020015
html

**Résumé**

On étudie un modèle mathématique de type S-E-I-R à deux phases, inspiré de l'épidémie actuelle de coronavirus. Si les contacts sont réduits à zéro à partir d'une certaine date $T$ proche du début de l'épidémie, la taille finale de l'épidémie est proche de celle que l'on obtient en multipliant le nombre cumulé de cas $R(T)$ à cette date par la reproductivité $\mathcal{R}_0$ de l'épidémie. Plus généralement, si les contacts sont divisés au temps $T$ par $q > 1$ de sorte que $\mathcal{R}_0/q < 1$, alors la taille finale de l'épidémie est proche de $R(T)\,\mathcal{R}_0\,(1-1/q)/(1-\mathcal{R}_0/q)$. On ajuste approximativement les paramètres du modèle aux données relatives au coronavirus en France.

## 1 Le modèle

La figure 1[a] montre le nombre cumulé de cas confirmés de coronavirus en France entre le 25 février et le 29 mars 2020 ; ces données incluent à la fois les laboratoires de biologie médicale de ville et les patients hospitalisés [10]. Il faut distinguer la date du 15 mars à partir de laquelle des mesures drastiques ont été prises subitement pour arrêter l'épidémie (fermeture des écoles, des restaurants, etc.). Pour ces trois dates, le nombre cumulé de cas confirmés est passé de 13 à 5 423 puis à 40 174. La figure 1[b] montre les mêmes données en coordonnées logarithmiques ainsi que des droites de régression linéaire. On observe trois périodes : dans la première, qui va jusqu'au 6 mars, la croissance est rapide mais assez irrégulière ; dans la deuxième, qui va jusqu'au 15 mars, la croissance est un peu moins rapide mais régulière ; dans la troisième, à partir du 16 mars, la croissance est ralentie mais toujours régulière. Si l'on ajuste une droite sur l'ensemble des deux premières périodes, qui va du 25 février au 15 mars, on trouve que le nombre de cas croît comme $e^{\lambda t}$ avec un taux $\lambda \simeq 0{,}31$ par jour

*Institut de recherche pour le développement, Unité de modélisation mathématique et informatique des systèmes complexes, Les Cordeliers, Paris, France, nicolas.bacaer@ird.fr

(droite rouge). Le temps de doublement est $(\ln 2)/\lambda \simeq 2{,}2$ jours. Si en revanche on se limite à la deuxième période, avec des données qui sont particulièrement bien alignées en échelle logarithmique, on obtient $\lambda \simeq 0{,}225$ par jour et un temps de doublement de 3,1 jours (droite bleue). Comme les données du début de l'épidémie sont perturbées par une part importante de nouveaux cas importés et par des effets stochastiques, c'est probablement la deuxième estimation qui est la plus fiable. Pour la troisième période, après la mise en place de mesures drastiques, le temps de doublement passe à 4,9 jours.

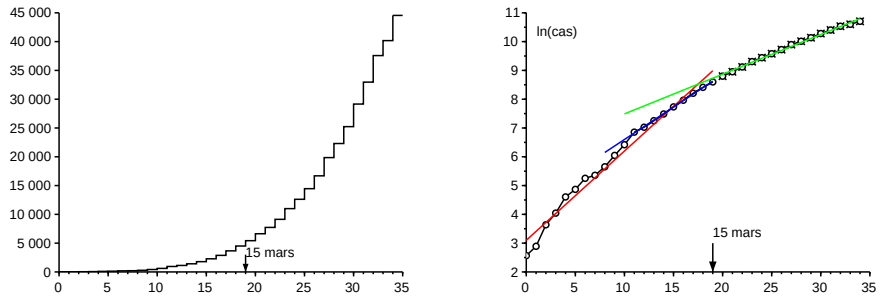


FIGURE 1 – a) Nombre cumulé de cas confirmés en France entre le 25 février et le 29 mars 2020, d'après Santé publique France et [11]. b) Logarithme népérien de ce nombre et droites de régression linéaire.

On va étudier un modèle mathématique inspiré de cette épidémie. Divisons la population en cinq compartiments, selon une variante du modèle classique S-E-I-R (voir par exemple [3, p. 61]) : susceptible d'être infecté ($S$), infecté en phase latente autrement dit non encore infectieux ($E$), infectieux sans protection ($I$), et retiré de la chaîne de transmission en étant comptabilisé parmi les cas confirmés ($R_1$) ou en ne l'étant pas ($R_2$). Chacun de ces deux derniers compartiments regroupe donc à la fois ceux qui sont encore infectieux mais isolés et ceux qui ne sont plus infectieux car guéris ou décédés. Certains malades ont des symptômes de faible gravité et restent chez eux sans se faire tester, d'autres habitent dans des maisons de retraites et n'ont pas non plus été testés malgré les complications voire le décès ; ce sont ces catégories que l'on retrouve dans le compartiment $R_2$. On peut évidemment raffiner à l'infini ce modèle pour le rendre plus réaliste mais on a essayé ici de limiter au maximum le nombre de paramètres inconnus ; on a aussi pour but principal d'obtenir un résultat théorique relatif à la taille finale de l'épidémie.

Notons $N$ la population totale (supposée grande) de sorte que $N = S(t) + E(t) + I(t) + R_1(t) + R_2(t)$. Notons $a$ le taux de contact effectif, $b$ le taux auquel les personnes infectées en phase latente deviennent infectieuses, et $c$ le taux moyen auquel les personnes infectieuses sont isolées et donc retirées de la chaîne de transmission. Notons $f$ la fraction d'individus infectieux qui sont comptabilisés parmi les cas confirmés au moment de l'isolement $(0 \leq f \leq 1)$ ;

cette fraction peut varier au fil du temps mais on supposera pour simplifier qu'elle est constante. Alors

$$\begin{aligned}
\frac{dS}{dt} &= -a\,S\,\frac{I}{N}\,, && (1)\\
\frac{dE}{dt} &= a\,S\,\frac{I}{N} - b\,E\,, && (2)\\
\frac{dI}{dt} &= b\,E - c\,I\,, && (3)\\
\frac{dR_1}{dt} &= f\,c\,I\,, && (4)\\
\frac{dR_2}{dt} &= (1-f)\,c\,I\,. && (5)
\end{aligned}$$

Pour faire le lien avec les données de la figure 1, $R_1(t)$ correspond au nombre cumulé de cas confirmés à l'instant $t$. Si l'on note $R(t) = R_1(t) + R_2(t)$, on remarque que

$$\frac{dR}{dt} = c\,I\,. \tag{6}$$

Avec $R_1(0) = R_2(0) = 0$, on en déduit que $R_1(t) = f\,R(t)$ et $R_2(t) = (1-f)R(t)$ pour tout $t \geq 0$.

Au début de l'épidémie, le nombre de cas reste très petit par rapport à la population totale de sorte que $S(t) \simeq N$, ce qui conduit à la linéarisation

$$\frac{dE}{dt} \simeq a\,I - b\,E, \qquad \frac{dI}{dt} = b\,E - c\,I.$$

Les compartiments $E$ et $I$ mais aussi les compartiments $R_1$ et $R_2$ tendent donc à croître exponentiellement comme $e^{\lambda t}$, où $\lambda$ est la plus grande valeur propre de la matrice

$$\begin{pmatrix} -b & a \\ b & -c \end{pmatrix}. \tag{7}$$

Le polynôme caractéristique est

$$\lambda^2 + (b+c)\lambda + b(c-a) = 0. \tag{8}$$

Donc

$$\lambda = \frac{-(b+c) + \sqrt{(b+c)^2 - 4b(c-a)}}{2} = \frac{-(b+c) + \sqrt{(b-c)^2 + 4ab}}{2}\,. \tag{9}$$

[9] indique que la durée d'incubation, c'est-à-dire la période avant apparition des symptômes, est de 5 à 6 jours. La période de latence peut être un peu plus courte puisqu'on peut devenir infectieux avant de montrer des symptômes. On fixe la durée moyenne $1/b$ dans la phase latente $E$ à 4 jours ; donc $b = 0{,}25$ par jour.

La durée moyenne dans le compartiment $I$ avant isolement, qui vaut $1/c$, est plus difficile à estimer car elle dépend de nombreux facteurs. Elle dépend des

caractéristiques biologiques du virus, des caractéristiques des individus telles que leur âge, mais aussi de la promptitude avec laquelle les cas sont isolés, ce qui varie d'un pays à l'autre. L'épidémie en France a lieu alors que les habitants sont déjà bien au courant de l'existence de la pandémie ; les malades ne tardent pas trop à être isolés. Certains ne seront pas du tout infectieux, d'autre le seront plusieurs jours avant d'être isolés. Supposons que la moyenne soit de l'ordre de 1 jour, la forme du modèle sous-entendant que la distribution est exponentielle. On aurait une moyenne de cet ordre dans un modèle plus raffiné si par exemple 80% des infectés restaient 0 jour infectieux et si 20% restaient 5 jours infectieux avant d'être isolés. En résumé, on a choisi $c = 1$ par jour.

On déduit de la formule (9) que

$$a = \frac{(2\lambda + b + c)^2 - (b - c)^2}{4b} = (\lambda + c)\left(1 + \frac{\lambda}{b}\right). \tag{10}$$

ce qui permettrait de calculer numériquement le taux de contact effectif $a$ à partir du taux de croissance observé $\lambda$.

Imaginons que des mesures de santé publique puissent diviser le taux de contact effectif par un nombre $k$ qui soit supérieur à 1. Combien doit valoir au minimum $k$ pour arrêter l'épidémie ? Cette valeur de $k$, traditionnellement notée $\mathcal{R}_0$ à la suite de Lotka et appelée par lui « reproductivité » [6, p. 102], s'obtient simplement en remarquant que lorsque $a$ est remplacé par $a' = a/\mathcal{R}_0$, le nouveau taux de croissance de l'épidémie $\lambda'$ doit être nul, ce qui d'après l'équation (8) conduit à $c - a/\mathcal{R}_0 = 0$ et à

$$\mathcal{R}_0 = \frac{a}{c} = \left(1 + \frac{\lambda}{b}\right)\left(1 + \frac{\lambda}{c}\right) \simeq 2{,}33$$

si l'on utilise la valeur numérique $\lambda \simeq 0{,}225$ par jour suggérée par la courbe épidémique de la figure 1 . Vues les incertitudes sur les paramètres $b$ et $c$, ceci ne peut être qu'une valeur approchée [1].

Revenons au modèle S-E-I-R non linéaire (1)-(6). Rappelons comment déterminer la taille finale de l'épidémie en l'absence complète d'intervention ; c'est une adaptation facile et d'ailleurs bien connue de la méthode utilisée pour le modèle S-I-R (voir par exemple [4, p. 76]). L'équation (1) montre que

$$\frac{d}{dt} \ln S = -\frac{a}{N} I(t).$$

Donc en intégrant de $t = 0$ à $t = +\infty$,

$$\ln S(\infty) - \ln S(0) = -\frac{a}{N} \int_0^\infty I(t)\, dt\,, \tag{11}$$

1. De manière plus technique (voir par exemple [7]), on aurait pu remarquer que $\mathcal{R}_0$ était aussi le rayon spectral de la matrice

$$\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} \begin{pmatrix} b & 0 \\ -b & c \end{pmatrix}^{-1}.$$

où $S(\infty)$ désigne la limite quand $t \to +\infty$ de la fonction $S(t)$ qui est décroissante et positive. Comme $R(0) = 0$, l'équation (6) montre que

$$R(\infty) = c \int_0^\infty I(t)\, dt. \tag{12}$$

Par ailleurs, on a à tout instant $S(t) + E(t) + I(t) + R(t) = N$. Quand $t \to +\infty$, l'épidémie finit par s'arrêter de sorte que $E(t)$ et $I(t)$ tendent vers 0. À la limite, il ne reste donc que les personnes qui ont échappé à l'épidémie et celles qui ont été infectées mais qui sont passées dans les compartiments $R$ :

$$S(\infty) + R(\infty) = N. \tag{13}$$

En combinant (11), (12) et (13), on voit que

$$N - R(\infty) = S(0)\, \exp\left(-\frac{a}{c}\frac{R(\infty)}{N}\right).$$

Au début de l'épidémie, il n'y a que quelques personnes infectées dans la population, donc $S(0) \simeq N$. L'équation implicite pour la taille finale de l'épidémie peut s'écrire comme

$$1 - \frac{R(\infty)}{N} \simeq \exp\left(-\mathcal{R}_0\, \frac{R(\infty)}{N}\right), \tag{14}$$

qui se trouve avoir la même forme que pour le modèle S-I-R [4]. Avec $\mathcal{R}_0 \simeq 2{,}33$, on trouve numériquement $R(\infty)/N \simeq 87\%$. Seule une fraction $f$ de ces cas est recensée.

## 2 Deuxième phase avec intervention drastique

Imaginons qu'à une certaine date $T$, des mesures drastiques soient prises de sorte que le nouveau taux de contact effectif soit réduit à 0 alors qu'il y a $R_1(T)$ cas confirmés cumulés. Par exemple, il y avait en France 5 423 cas confirmés cumulés au 15 mars, date à laquelle sont entrées en vigueur les mesures concernant les écoles et les lieux publics. Peut-on alors prévoir quelle aurait été sous ces hypothèses idéales la nouvelle taille finale de l'épidémie $R(\infty)$, ou du moins celle confirmée $R_1(\infty)$ ?

Vers la fin de la phase exponentielle où $t \leq T$ et où le nombre total de cas représente encore une part infime de la population totale, on a

$$E(t) \simeq u\, e^{\lambda t}, \quad I(t) \simeq v\, e^{\lambda t}, \quad R(t) \simeq w\, e^{\lambda t},$$

où $(u, v)$ est un vecteur propre associé à la plus grande valeur propre $\lambda$ de la matrice (7). Ainsi, $-b\, u + a\, v = \lambda\, u$. Avec l'équation (10), on trouve que

$$u = \frac{a\, v}{\lambda + b} = \frac{\lambda + c}{b}\, v\,.$$

Comme $dR/dt \simeq \lambda R$ pour $t < T$ lorsque $t$ n'est pas trop proche de 0, on a

$$I(T) = \frac{1}{c}\frac{dR}{dt}(T) \simeq \frac{\lambda}{c}\,R(T).$$

Mais $I(T) \simeq v\,e^{\lambda T}$, donc

$$E(T) \simeq u\,e^{\lambda T} = \frac{\lambda + c}{b}\,v\,e^{\lambda T} \simeq \frac{\lambda + c}{b}\,I(T) \simeq \left(\frac{\lambda^2}{bc} + \frac{\lambda}{b}\right) R(T)\,.$$

Les contacts étant supposés réduits à zéro, on a pour $t > T$

$$\frac{dS}{dt} = 0, \quad \frac{dE}{dt} = -b\,E, \tag{15}$$

tandis que les autres équations (3), (4) et (5) restent identiques. Sans avoir à résoudre ce système, il est clair que la taille finale de l'épidémie sera $R(\infty) = R(T) + E(T) + I(T)$, puisqu'il y a $E(T) + I(T)$ individus infectés qui ne sont pas encore dans les compartiments $R$ au temps $T$. Ainsi

$$R(\infty) \simeq R(T)\left(1 + \frac{\lambda^2}{bc} + \frac{\lambda}{b} + \frac{\lambda}{c}\right) = R(T)\left(1 + \frac{\lambda}{b}\right)\left(1 + \frac{\lambda}{c}\right) = \mathcal{R}_0\,R(T)\,.$$

Puisque à tout instant $R_1(t) = f\,R(t)$, on en déduit aussi que

$$R_1(\infty) \simeq \mathcal{R}_0\,R_1(T).$$

Ainsi, si les contacts sont réduits à zéro à partir d'une certaine date proche du début de l'épidémie (assez proche pour que l'approximation linéaire soit encore valable mais pas trop proche pour que le système linéarisé ait eu le temps de converger vers le vecteur propre associé à la première valeur propre), alors la taille finale (confirmée ou totale) de l'épidémie est proche de celle que l'on obtient en multipliant le nombre cumulé de cas (confirmés ou au total) à cette date par la reproductivité $\mathcal{R}_0$ de l'épidémie. Un résultat semblable s'obtient de la même manière pour un modèle S-I-R. En annexe, on remarque cependant que ce n'est plus $\mathcal{R}_0$ qui détermine le rapport $R(\infty)/R(T)$ dans les modèles avec une période infectieuse qui n'est pas distribuée exponentiellement, mais une expression plus compliquée.

Avec $R_1(T) = 5\,423$ et $\mathcal{R}_0 \simeq 2{,}33$, cela donne $R_1(\infty) \simeq 12\,600$. Soulignons encore une fois l'incertitude autour des paramètres $b$ et $c$, qui se retrouve dans la valeur de $R_1(\infty)$.

Notons au passage l'analogie avec le concept de « potentiel d'accroissement » des populations en démographie [8, p. 176]. C'est le rapport entre la population finale stationnaire et la population à un certain instant si la fertilité est divisée subitement à cet instant par la reproductivité $\mathcal{R}_0$, de sorte que la population se retrouve avec un taux asymptotique de croissance nul. Comme dans notre calcul, c'est en supposant que la population à cet instant est « stable » au sens de Lotka (c'est-à-dire donnée par le premier vecteur propre) que Keyfitz a obtenu

une formule relativement simple pour le potentiel d'accroissement, formule qui fait aussi intervenir $\mathcal{R}_0$ quoique de manière plus compliquée que pour notre modèle S-E-I-R [8, p. 179].

Notons aussi que l'estimation de $R(T) + E(T) + I(T)$ à partir de la donnée $R(T)$ seule est analogue au problème qui s'était posé aux débuts de l'épidémie de VIH pour estimer le nombre de personnes séropositives à partir du nombre de cas déclarés de SIDA.

La figure $2^a$ illustre ce modèle à deux phases. On a pris $N = 65 \times 10^6$ (la population totale de la France) et les conditions initiales

$$S(0) = N - 1, \quad E(0) = 1, \quad I(0) = 0, \quad R(0) = 0. \tag{16}$$

Le paramètre $a$ est donné par la formule (10) avec $\lambda = 0{,}225$ par jour, comme dans la figure 1. On dispose de peu d'information au sujet du paramètre $f$, si ce n'est rétrospectivement qu'un grand nombre de décès dus au virus dans les maisons de retraite n'étaient pas comptabilisés parmi les cas confirmés au début de l'épidémie ; fixons par exemple $f = 0{,}5$ pour l'illustration. On a pris $T = 43{,}2$ jours de sorte que $R_1(T) \simeq 5\,438$ soit proche de la donnée $5\,423$ du 15 mars. En poursuivant la simulation un peu plus longtemps que dans la figure, on trouve bien numériquement que $R_1(\infty)/R_1(T) \simeq 2{,}33 \simeq \mathcal{R}_0$.

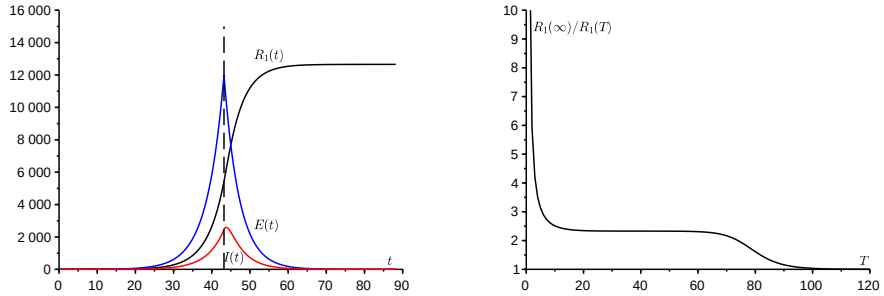


FIGURE 2 – a) Exemple de simulation du modèle à deux phases. b) Le rapport $R_1(\infty)/R_1(T)$ en fonction de $T$.

La figure $2^b$ montre comment le rapport $R_1(\infty)/R_1(T)$ varie en fonction de l'instant $T$ où le taux de contact est réduit à zéro. On observe effectivement un plateau où ce rapport est proche de $\mathcal{R}_0$. Quand $T \to 0$, on a $R_1(T) \to 0$ et $R_1(\infty) \to f(E(0) + I(0)) > 0$, donc le rapport $R_1(\infty)/R_1(T)$ tend vers l'infini. Le rapport se rapproche de $\mathcal{R}_0$ lorsque $T$ est de l'ordre de l'inverse de la différence entre les deux valeurs propres de la matrice (7). Quand au contraire $T \to \infty$, alors l'intervention intervient trop tard ; l'épidémie est déjà passée et $R_1(\infty)/R_1(T) \to 1$. On s'attend à ce que la largeur du plateau où $R_1(\infty)/R_1(T)$ est proche de $\mathcal{R}_0$ croisse comme $(\ln N)/\lambda$ lorsque $N \to +\infty$, puisque tel est par exemple le comportement du temps jusqu'au pic épidémique dans un modèle S-I-R à coefficients constants (voir [2] ou [5, p. 12]).

## 3 Généralisation

Dans la réalité, le taux de contact effectif n'est sûrement pas tout à fait nul pour $t > T$. La valeur obtenue pour $R(\infty)$ peut néanmoins être considérée comme une borne inférieure de la valeur réelle puisqu'il est certain que la taille finale de l'épidémie sera supérieure avec des contacts non nuls qu'avec des contacts nuls pour $t > T$. Rappelons cependant à ce sujet que les modèles épidémiques de type S-I-R ou S-E-I-R avec un taux de contact variable ne sont pas « monotones », dans le sens qu'une réduction du taux de contact peut parfois conduire à une taille finale de l'épidémie plus grande [1].

Considérons maintenant le cas où le taux de contact n'est pas réduit à 0 mais simplement divisé par un nombre $q > 1$. La réduction à 0 correspond au cas limite où $q$ tend vers l'infini. On a pour $t > T$,

$$\frac{dS}{dt} = -\frac{a}{q}\, S\, \frac{I}{N}, \quad \frac{dE}{dt} = \frac{a}{q}\, S\, \frac{I}{N} - b\, E, \tag{17}$$

tandis que les équations (3), (4) et (5) restent identiques. Par le même raisonnement que dans la section 1, on a pour $t > T$

$$\frac{1}{S}\frac{dS}{dt} = -\frac{a}{q\, c\, N}\frac{dR}{dt}.$$

En intégrant entre $t = T$ et $t = +\infty$, on en déduit que

$$\ln \frac{S(\infty)}{S(T)} = -\frac{\mathcal{R}_0}{q}\, \frac{R(\infty) - R(T)}{N},$$

où $\mathcal{R}_0 = a/c > 1$. Comme $S(\infty) = N - R(\infty)$, on a donc

$$1 - \frac{R(\infty)}{N} = \frac{S(T)}{N} \exp\left(-\frac{\mathcal{R}_0}{q}\, \frac{R(\infty) - R(T)}{N}\right). \tag{18}$$

Supposons comme dans la section 2 que le temps $T$ ne soit ni trop petit ni trop grand, c'est-à-dire dans le plateau de la figure $2^b$. En première approximation, $S(T) \simeq N$ et $R(T)$ est encore petit devant $N$. Deux cas se présentent alors.

Si $1 < q < \mathcal{R}_0$, alors l'argument graphique classique [4] qui consiste à tracer les membres de gauche et de droite de l'équation (18) en fonction de $R(\infty)/N$ montre que la solution $R(\infty)/N$ n'est pas infinitésimale mais proche de la solution strictement positive de l'équation

$$1 - \frac{R(\infty)}{N} \simeq \exp\left(-\frac{\mathcal{R}_0}{q}\, \frac{R(\infty)}{N}\right). \tag{19}$$

Si au contraire $q > \mathcal{R}_0$, alors la solution $R(\infty)/N$ de l'équation (18) est petite. Comme $S(T) = N - E(T) - I(T) - R(T)$ et $E(T) + I(T) + R(T) \simeq \mathcal{R}_0\, R(T)$, un développement à l'ordre 1 de l'exponentielle dans l'équation (18) conduit à

$$1 - \frac{R(\infty)}{N} \simeq \left[1 - \frac{\mathcal{R}_0\, R(T)}{N}\right] \left[1 - \frac{\mathcal{R}_0}{q}\, \frac{R(\infty) - R(T)}{N}\right].$$

En ne gardant que les termes d'ordre inférieur ou égal à 1, on trouve que

$$1 - \frac{R(\infty)}{N} \simeq 1 - \frac{\mathcal{R}_0\, R(T)}{N} - \frac{\mathcal{R}_0}{q}\frac{R(\infty) - R(T)}{N}.$$

Finalement,

$$R(\infty) \simeq R(T)\,\mathcal{R}_0\,\frac{1 - 1/q}{1 - \mathcal{R}_0/q}. \tag{20}$$

Lorsque $q \to +\infty$, on retrouve bien que $R(\infty) \simeq R(T)\,\mathcal{R}_0$. On remarque aussi que $(1 - 1/q)/(1 - \mathcal{R}_0/q) > 1$, comme il se doit. Une formule identique à (20) lie $R_1(\infty)$ et $R_1(T)$.

La figure 3 montre avec une ligne continue, en fonction du paramètre $q$, la taille finale de l'épidémie en échelle logarithmique, $\ln(R(\infty)/N)$, obtenue par simulation numérique du système (1)-(6) pour $t < T$ avec les conditions initiales (16) et du système (17) pour $t > T$. Comme dans la figure 2[a], la population totale est $N = 65 \times 10^6$ et le paramètre $a$ est donné par la formule (10) avec $\lambda = 0{,}225$ par jour ; on a encore pris $f = 0{,}5$ et $T = 43{,}2$ jours de sorte que $R_1(T) \simeq 5\,438$. La figure montre aussi avec des petits ronds ce que donne la formule (20) pour $q > \mathcal{R}_0$. Elle montre enfin avec de petits losanges la solution strictement positive de l'équation (19) pour $q < \mathcal{R}_0$. On voit que les deux approximations cessent d'être valables au voisinage de $q = \mathcal{R}_0$.

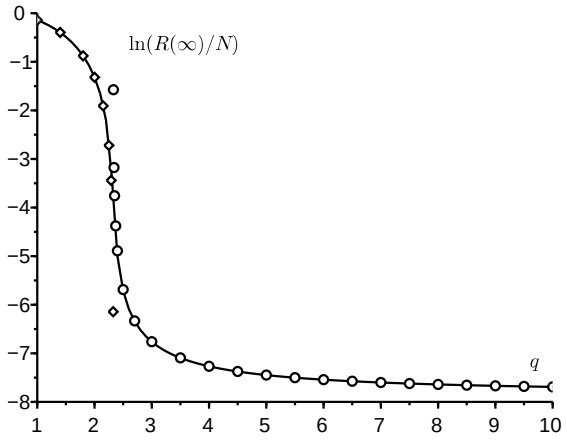


FIGURE 3 – $\ln(R(\infty)/N)$ en fonction de $q$ (ligne continue), comparé avec la formule (20) (petits ronds) valable pour $q > \mathcal{R}_0$ et avec la solution de l'équation (19) (petits losanges) valable pour $q < \mathcal{R}_0$.

On notera que la taille finale de l'épidémie varie de plusieurs ordres de grandeurs lorsque le paramètre $q$ est proche de $\mathcal{R}_0$. Comme il est difficile de le quantifier, la prédiction de la taille finale de l'épidémie est également difficile

dans cette zone. Il n'y a que si le paramètre $q$ est nettement supérieur à $\mathcal{R}_0$ que la prévision avec la formule (20) devient moins sensible à la valeur de $q$.

# 4 Tentative d'estimation du paramètre $q$

Essayons d'estimer le paramètre $q$ en ajustant une simulation du modèle aux données postérieures au 15 mars, y compris celles jusqu'au 15 avril qui ne figuraient pas dans la figure 1. [10] avertit néanmoins que « le nombre de cas confirmés en France ne reflète plus de manière satisfaisante la dynamique de l'épidémie », étant donné que « les patients présentant des signes de COVID-19 ne sont plus systématiquement confirmés par un test biologique ». Pour cela, on part de la donnée $R_1(T) = 5\,423$ et des relations $R(T) = R_1(T)/f$ et $R_2(T) = (1-f)R(T)$. Comme les données des 8 jours qui précèdent sont particulièrement bien alignées, on démarre la simulation de notre modèle avec $R(T-\tau) = e^{-\lambda\tau}R(T)$ où $\lambda = 0{,}225$ par jour et $\tau = 8$ jours, et avec les estimations correspondantes $I(T-\tau) \simeq \frac{\lambda}{c}\,R(T-\tau)$, $E(T-\tau) \simeq (\frac{\lambda^2}{bc} + \frac{\lambda}{b})R(T-\tau)$ et $S(T-\tau) = N - E(T-\tau) - I(T-\tau) - R(T-\tau)$.

Pour $t > T$, le taux de contact effectif est $a/q$ et l'on essaie d'ajuster $R_1(t)$ aux données jusqu'au 15 avril. Le meilleur ajustement se trouve aux alentours de $q = 1{,}7$. Comme cette valeur est inférieure à $\mathcal{R}_0$, il semblerait que les mesures de confinement soient encore insuffisantes. Mais les tous derniers points de la figure montrent que l'écart avec le modèle grandit dans le sens d'un ralentissement de l'épidémie réelle. Il se peut que la valeur de $f$ choisie ne soit pas appropriée ou qu'elle ait varié au cours de l'épidémie. Ou alors le modèle est peut-être un peu trop simpliste ; on s'attend notamment à ce qu'une distribution non exponentielle des temps passés dans les différents compartiments influence le moment où la courbe commence à s'infléchir.

En conclusion, on a exploré un scénario à deux phases où le taux de contact est réduit à partir d'un certaine date. On a trouvé une formule approchée simple pour la taille finale de l'épidémie en fonction du nombre de cas détectés au moment de la réduction. Il reste néanmoins à énoncer et à démontrer plus rigoureusement ce résultat, probablement en le faisant apparaître comme un résultat asymptotique lorsque $N \to +\infty$.

**Remerciements**.

On remercie Hisashi Inaba, Ali Moussaoui et Frédéric Hamelin pour leurs commentaires sur le manuscrit.

# Annexe

Considérons un modèle S-I-R avec une période infectieuse qui n'est pas nécessairement distribuée exponentiellement. Soit $I(t,x)$ la densité de personnes

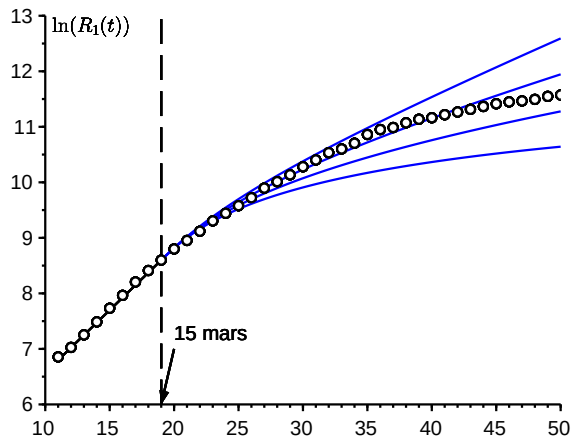


FIGURE 4 – Logarithme népérien du nombre de cas recensés entre le 7 mars et le 15 avril (petits ronds, données de Santé publique France [11]) et $\ln(R_1(t))$ en fonction du temps $t$ dans 4 simulations avec de haut en bas $q \in \{1{,}5;\ 1{,}7;\ 2;\ 2{,}5\}$.

infectées depuis $x$ unités de temps au temps $t$. Soit $a(x)$ le taux de contact effectif et $b(x)$ le taux auquel les personnes infectées cessent de transmettre l'infection. On a au début de l'épidémie

$$\begin{aligned} I(t,0) &\simeq \int_0^\infty a(x)\, I(t,x)\, dx \\ \frac{\partial I}{\partial t} + \frac{\partial I}{\partial x} &= -b(x)\, I(t,x) \\ \frac{dR}{dt} &= \int_0^\infty b(x)\, I(t,x)\, dx \end{aligned}$$

On en déduit, comme dans la théorie des populations stables de Lotka [6] que

$$I(t,x) \simeq k\, e^{\lambda t}\, e^{-\lambda x - \int_0^x b(y)\, dy}$$

où $k$ est une constante et le taux de croissance $\lambda$ est l'unique solution de l'équation

$$1 = \int_0^\infty a(x) e^{-\lambda x - \int_0^x b(y)\, dy}\, dx.$$

Si l'on pose $I(t) = \int_0^\infty I(t,x)\, dx$, le problème est d'estimer $I(T) + R(T)$ à partir de $R(T)$. Or

$$\lambda R(T) \simeq \frac{dR}{dt}(T) = \int_0^\infty b(x)\, I(T,x)\, dx \simeq \int_0^\infty b(x)\, k\, e^{\lambda T}\ e^{-\lambda x - \int_0^x b(y)\, dy}\, dx.$$

On en déduit que

$$k \simeq \frac{\lambda R(T) e^{-\lambda T}}{\int_0^\infty b(x)\, e^{-\lambda x - \int_0^x b(y)\, dy}\, dx}.$$

Finalement,

$$\frac{I(T) + R(T)}{R(T)} \simeq \frac{\lambda \int_0^\infty e^{-\lambda x - \int_0^x b(y)\, dy}}{\int_0^\infty b(x)\, e^{-\lambda x - \int_0^x b(y)\, dy}} + 1.$$

On voit que le membre de droite n'a pas de raison particulière de coïncider avec $\mathcal{R}_0 = \int_0^\infty a(x)\, e^{-\int_0^x b(y)\, dy}\, dx$. Dans le cas spécial où les taux sont constants, avec $a(x) \equiv a$ et $b(x) \equiv b$, on a cependant $\lambda = a - b$ et donc $(I(T) + R(T))/R(T) \simeq \frac{\lambda}{b} + 1 = \frac{a}{b} = \mathcal{R}_0$.

# Références


[1] N. Bacaër et M.G.M. Gomes, Sur la taille finale des épidémies avec saisonnalité, *Bull. Math. Biol.* **71** (2009) 1954–1966. `https://hal.archives-ouvertes.fr/hal-01299608`

[2] N. Bacaër, Sur le pic épidémique dans un modèle S-I-R, *Quadrature* **117** (2020) 1-4. `https://hal.archives-ouvertes.fr/hal-02518993`

[3] M. Corlosquet-Habart, J. Janssen et R. Manca (2012), Modélisation stochastique du risque de pandémie : stratégies de couverture et d'assurance. Lavoisier, Cachan (2012).

[4] A. Hillion, Les Théories mathématiques des populations. Presses Universitaires de France, Paris (1986).

[5] H.A. Lauwerier, Mathematical Models of Epidemics. Mathematisch Centrum, Amsterdam (1984).

[6] A.J. Lotka, Théorie analytique des associations biologiques, 2e partie. Hermann, Paris (1939).

[7] L. Nkague Nkamba, Robustesse des seuils en épidémiologie et stabilité asymptotique d'un modèle à infectivité et susceptibilité différentielle. Thèse, Université de Lorraine et Université Gaston Berger (2012).

[8] R. Pressat, Éléments de démographie mathématique. AIDELF, Paris (1995).

[9] Ph. Sansonetti, Covid-19 ou la chronique d'une émergence annoncée. Exposé, Collège de France, 18 mars 2020.

[10] Santé publique France, Covid-19, point épidémiologique hebdomadaire du 09 avril 2020. www.santepubliquefrance.fr

[11] `https://fr.wikipedia.org/wiki/Pand%C3%A9mie_de_Covid-19_en_France`