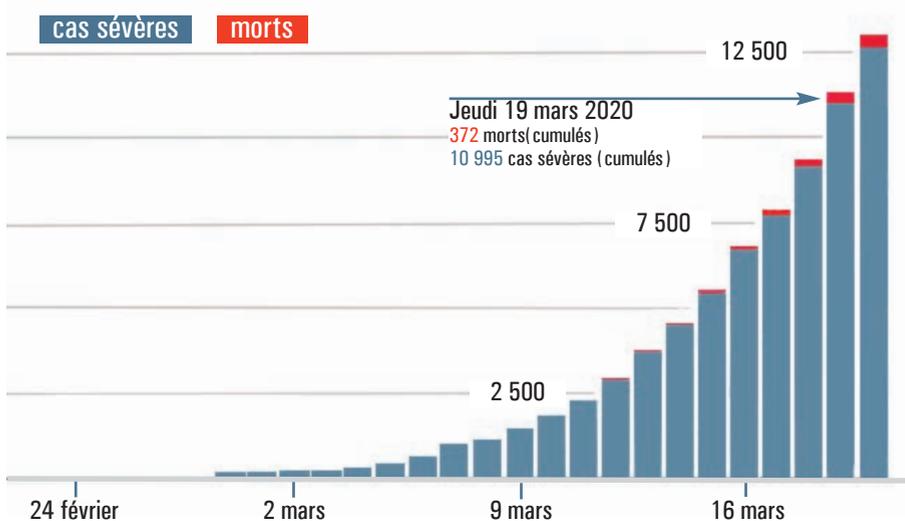


Les modèles mathématiques des épidémies*

Étienne Pardoux**

Professeur émérite à Aix-Marseille Université

L'année 2020 restera marquée par une pandémie mondiale inédite, qui s'est répandue à partir de la Chine dans le monde entier par le transport aérien, et se transforme en épidémies nationales une fois les frontières verrouillées. Les modèles mathématiques des épidémies permettent de répondre à des questions du type : jusqu'où la courbe des infectés et des décès va-t-elle grimper? quelle est l'efficacité des mesures de confinement décidées dans la plupart des pays?



La propagation du Covid-19 en France lors des premières semaines.

© CIJM, d'après Le Monde

* Cet article est une version complétée et remise à jour de la vidéo de l'auteur *Modèles mathématiques des épidémies comme celle du Covid-19*, disponible en ligne (https://youtu.be/fu92X74MS_M), elle-même inspirée par celle (en anglais) de Tom Britton, professeur à l'université de Stockholm (<https://www.youtube.com/watch?v=gSqIwXl6IjQ>). C'est également une adaptation de l'article de l'auteur *Modèles mathématiques des épidémies* paru dans le journal en ligne *The Conversation*.

**L'auteur remercie le professeur Tran Viet Chi, de l'université Paris-Est-Marne-la-Vallée, pour l'avoir aidé à s'initier en accéléré à l'utilisation du logiciel R pour les simulations.

Le modèle SIR et le nombre de reproduction de base R_0

L'un des plus vieux modèles mathématiques des épidémies est celui de Reed–Frost (1929). Il est simpliste mais il permet d'introduire des notions essentielles et d'obtenir une formule mathématique importante. Les individus sont de trois types : S (comme «susceptibles» d'être infectés), I (comme «infectés et infectieux», capables d'infecter un individu susceptible), et R (comme «remis» ou «retiré», soit guéri, soit mort). Dans ce modèle, le temps t est discret et représente par exemple un nombre de semaines. La taille n de la population est supposée «grande». Au début du modèle, la population comporte $n-1$ individus de type S, 1 de type I et 0 de type R. Un individu qui est infecté une semaine infecte chaque susceptible avec la probabilité p la semaine suivante, puis guérit. L'épidémie se poursuit tant qu'il y a des infectés, puis elle s'arrête. Pour simplifier, on néglige la phase d'incubation, et on suppose qu'un individu de type R, s'il n'est pas mort (ce qui heureusement est le cas de l'immense majorité des R), est immunisé.

Le nombre de reproduction de base R_0 est le nombre moyen de susceptibles qu'un individu I infecte «au début de l'épidémie», lorsque presque toute la population est susceptible. Combien vaut R_0 ? Prenons $n=1\,000$ et $p=0,0025$. Le premier infecté a, autour de lui, $n-1$ (soit environ 1 000) individus susceptibles. Puisqu'il infecte chacun d'eux avec la probabilité p , R_0 vaut (environ) $n \times p$, soit 2,5.

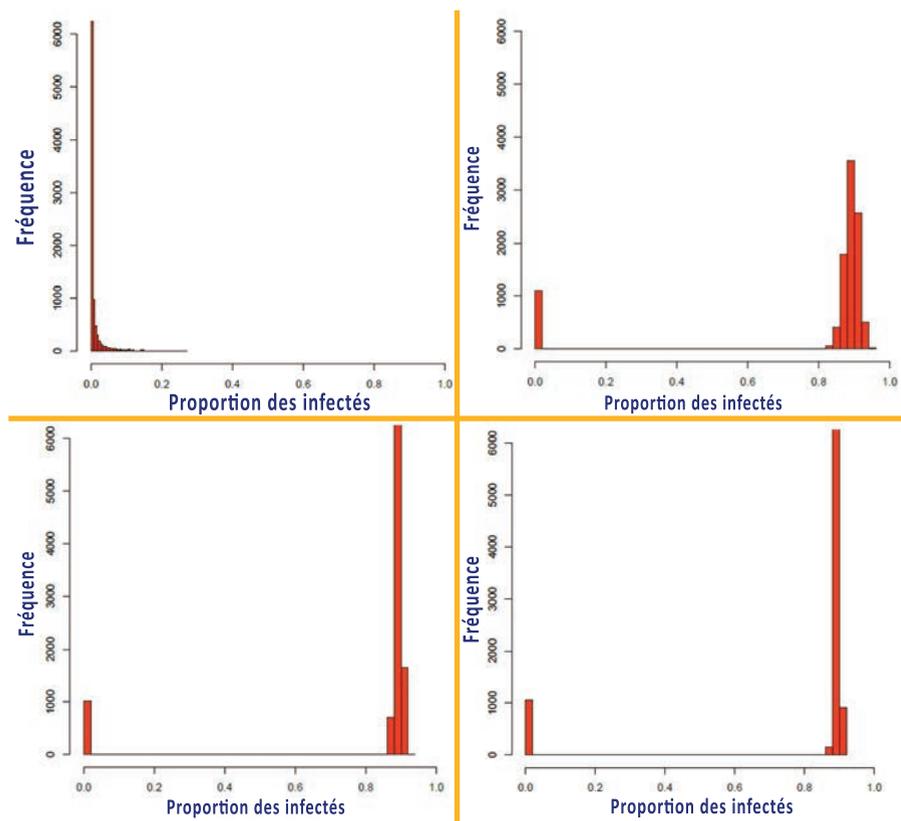
Si $R_0 < 1$, il n'y aura pas d'épidémie majeure avec un petit nombre d'infectés initiaux. De même si $R_0 = 1$. Par contre, si $R_0 > 1$, un seul infecté initial peut déclencher une épidémie majeure (qui touche une fraction importante de la population).

La fraction de la population qui sera touchée

Une question importante est d'estimer la fraction de la population totale qui sera touchée par l'épidémie. Si l'on admet qu'une personne guérie est immunisée, l'épidémie va s'arrêter tôt ou tard, au pire quand tout le monde aura été contaminé, et sera guéri et immunisé (un petit pourcentage étant mort). Mais dans la réalité tout le monde n'est pas touché. Lorsqu'une fraction de la population est immunisée, pour savoir combien de susceptibles un individu I infecte en moyenne, on multiplie p par le nombre de susceptibles. Bien avant que ce nombre ne s'annule, ce produit passe en dessous de 1, et alors l'épidémie s'arrête.

Effectuons dix mille simulations du modèle de Reed-Frost, dans les cas $R_0=0,95$ (lorsque $R_0 < 1$, aucune épidémie majeure n'a lieu) et $R_0=2,5$. La hauteur de chaque barre indique le nombre de simulations qui ont conduit à la proportion d'infectés indiquée sur l'axe des abscisses. Dans le cas $R_0=2,5$, une

certaine fraction des simulations (qui ne dépend pas de n) n'aboutissent pas à une épidémie majeure, tandis que la proportion d'individus infectés dans le cas d'une épidémie majeure se concentre quand n augmente autour d'une certaine valeur (laquelle augmente avec R_0).



Résultat des simulations dans les cas $R_0 = 0,95$ (en haut à gauche), $R_0 = 2,5$ et $n = 500$ (en haut à droite), $R_0 = 2,5$ et $n = 2\,500$ (en bas à gauche) et enfin $R_0 = 2,5$ et $n = 5\,000$ (en bas à droite).

© É. Pardoux, 2020

Supposons donc $R_0 > 1$. Dans la réalité, il n'y a pas un seul infecté initial, il y en a eu suffisamment dans chaque pays pour que l'épidémie majeure soit inévitable. Supposons que l'épidémie touche au total une fraction τ de la population. Alors la fraction $1 - \tau$ de la population qui n'est pas touchée par l'épidémie est égale à la probabilité qu'un individu ne soit infecté par aucun des infectieux, car tous les individus dans la population ont la même probabilité d'être infecté. Ce qui est égal, par définition, à $(1 - p)^\tau$, ou encore

à $(1 - R_0/n)^{n\tau}$, car les « choix des victimes » par les divers infectés sont indépendants. Cette quantité vaut approximativement $\exp(-R_0\tau)$ pour n « grand ». De même que la proportion de « piles » dans un « grand » nombre de lancers successifs d'une pièce équilibrée tend vers $1/2$, la proportion aléatoire τ tend vers une valeur donnée τ^* quand n tend vers l'infini. Cette valeur τ^* est donc la solution de l'équation $\tau = 1 - \exp(R_0\tau)$.

0 est toujours solution de cette équation, et c'est la seule solution si $R_0 \leq 1$. Par contre, si $R_0 > 1$, alors il y a une seconde solution $0 < \tau^* < 1$, qui ne dépend que de R_0 . C'est elle qui nous intéresse.

Un ordinateur fournit une valeur approchée de τ^* pour chaque valeur donnée de R_0 . On trouve en particulier, pour $R_0 = 1,5, 2, 2,5$ et 3 , les valeurs respectives de τ suivantes : 58 %, 79 %, 89 % et 94 %.

Trois hypothèses essentielles

Les résultats obtenus avec le modèle de Reed-Frost peuvent en fait être établis dans un cadre plus général, respectant les trois hypothèses suivantes.

(H1) Pas d'immunité au départ (naturelle ou par vaccination). (H1) est vraie pour la maladie à coronavirus 2019 (Covid-19), pas pour la grippe saisonnière.

(H2) Une communauté homogène. (H2) n'est pas vraie pour le Covid-19, ni pour aucune maladie ! Les hétérogénéités qu'il convient de prendre en compte dépendent de chaque maladie et de son mode de transmission. En situation réelle, on considère qu'il faut réduire le nombre d'infectés d'environ 10 % à 20 % par rapport aux prédictions du modèle homogène.

(H3) Le comportement des individus ne change pas au cours de l'épidémie. Les mesures prises par les autorités (fermeture des écoles et des lieux publics, confinement...) changent évidemment la donne.

Réduire R_0 : confinement, mesures barrières, vaccination...

Comment réduire R_0 ? Factorisons ce nombre ainsi pour voir comment il est construit : $R_0 = p \times c \times u$, avec p la probabilité qu'un contact produise une infection, c le nombre moyen de « contacts » par jour et u la durée (en nombre de jours) de la période d'infection. Les mesures de prévention visent à réduire R_0 en diminuant p (port de masques, lavage des mains...), c (confinement, interdiction des regroupements...) et u (diagnostic rapide, mise à l'isolement des infectés...).

Maintenant, si une proportion v de la population est vaccinée, alors R_0 est remplacé par $R_v = (1-v)R_0$ car chaque «tentative d'infecter» ne réussit qu'avec la probabilité $1-v$. On en déduit que $R_v < 1$ dès que $v > 1 - 1/R_0$. Ainsi, R_0 nous indique la fraction de la population qu'il faut vacciner si l'on veut être sûr qu'il n'y aura pas d'épidémie majeure (si $R_0 = 2,5$, il faut vacciner 60 % de la population).

Estimer R_0 est vraiment crucial ! Mais, observer la courbe d'incidence (la première figure de l'article) ne suffit pas. Le taux de croissance r de la courbe dépend en effet de deux facteurs : R_0 , et le *temps de génération* G (la durée entre le moment où l'on est infecté et celui où l'on infecte). Le taux r croît avec R_0 , décroît avec G . Les mathématiciens ont établi des formules qui relient r , R_0 et la loi de probabilité de G . Par une méthode de *suivi des contacts* (recherche de qui a infecté qui et quand), on peut obtenir des informations sur la loi de G . En les combinant avec la lecture de r sur les courbes d'incidence, on peut estimer R_0 !

Un modèle déterministe pour la dynamique de l'épidémie

Pour décrire l'évolution de l'épidémie, dans le cas d'une « grande » population, et partant d'une situation où une fraction significative de cette population est déjà touchée, on va utiliser un modèle déterministe.

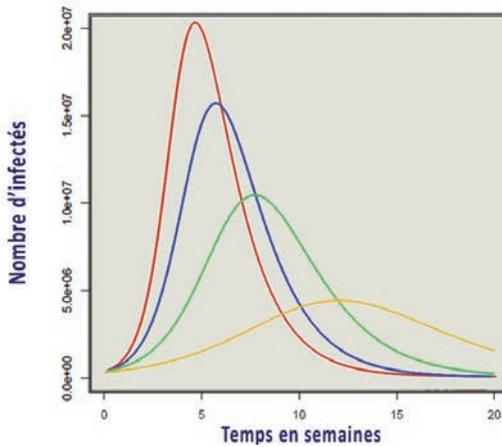
Le modèle d'évolution

Le modèle déterministe SIR pour décrire l'évolution de l'épidémie est le suivant. On appelle $s(t)$ la fraction d'individus susceptibles à l'instant t , et $i(t)$ celle des infectés. Le modèle s'écrit sous la forme d'un système d'équations différentielles :

$$\begin{cases} \frac{ds}{dt}(t) = -\alpha s(t)i(t), \\ \frac{di}{dt}(t) = \alpha s(t)i(t) - \beta i(t), \end{cases}$$

où α est la force de l'infection et β l'inverse de la durée moyenne d'infection. Au début de l'épidémie, $s(t)$ est « proche » de 1, et chaque individu I infecte au taux α pendant une durée de moyenne $1/\beta$. Donc $R_0 = \alpha/\beta$.

Le système d'équations différentielles qui régit le modèle est non linéaire. Si l'on fixe la condition initiale, ce système a une solution unique. Sur les quatre premières courbes ci-dessous, plus R_0 est grand, plus la vague des infectés arrive tôt et monte haut. En outre, lorsque l'on réduit R_0 , la réduction de la hauteur du pic est plus marquée que la réduction de la taille totale de l'épidémie.



L'évolution du nombre d'infectés pour des valeurs de R_0 égales à 3 (rouge), 2,5 (bleu), 2 (vert) et 1,5 (jaune).

© É. Pardoux, 2020

Concernant le Covid-19 en France, alors même que l'épidémie était déjà à un stade très avancé, beaucoup d'incertitudes subsistaient quant à la valeur de R_0 : les estimations variaient entre 2,5 et 5. Si le gouvernement n'avait pas pris de mesures fortes, entre 60 et 80 % de la population aurait probablement été touchée. Surtout, la « vague » serait arrivée très vite et montée très haut, submergeant le système de santé. D'autres facteurs ont contribué à rendre les modèles particulièrement délicats avec le Covid-19 : on ne dispose pas à ce jour d'une bonne estimation du nombre de personnes en France qui ont été touchées par le virus. Cette maladie est en effet bénigne, et même indécélable, dans beaucoup de cas. Il est donc difficile de savoir quelle fraction de la population est immunisée !

Les modèles mathématiques sont indispensables pour les décideurs. Il convient de poursuivre leur étude pour obtenir, à l'avenir, des modèles plus fins, plus proches des conditions dans lesquelles une nouvelle épidémie pourrait survenir. Voici quelques problèmes sur lesquels travaillent les mathématiciens et statisticiens qui s'intéressent aux modèles des épidémies : prise en compte de l'hétérogénéité de la population (maisonnées et lieux de travail, répartition spatiale, voyages...) et analyser ses effets ; comparer l'intérêt des modèles les plus « réalistes » et des modèles plus « simples » ; mesurer l'effet de diverses mesures préventives ; délimiter précisément ce que les données disponibles permettent d'obtenir comme conclusions ; identifier les données supplémentaires qui pourraient être utiles... Le chantier est vaste. Donc si vous voulez sauver des vies, vous pouvez aussi faire des maths !

É. P.