

# Naviguer au travers d'une épidémie

Janvier 2020 : une épidémie éclate.

Le nombre d'infectés explose comme un feu de broussailles.

Les données sont incomplètes et peu fiables.

La santé publique veut sauver des vies. Que faire?

**Christiane Rousseau**  
**Yvan Saint-Aubin**  
Université de Montréal

Le principe des modèles mathématiques de phénomènes naturels, tels ceux capturant le réchauffement global, la croissance des populations, les épidémies, est de simplifier le problème pour se concentrer sur ses caractéristiques essentielles. Ainsi, dans l'étude du climat, on veut éliminer les extrêmes météo pour se concentrer sur les tendances à long terme : on pourra utiliser des moyennes, dans le temps et dans l'espace, pour le faire. On fera de même pour modéliser la propagation d'une épidémie, tout en précisant les hypothèses simplificatrices.

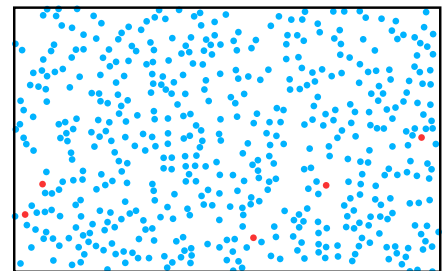
## Le modèle le plus simple

Pour ce premier modèle, la population sera compartimentée en personnes *susceptibles*, c'est-à-dire qui n'ont pas encore contracté le virus, et en personnes *infectées*, c'est-à-dire celles qui portent le virus et peuvent le transmettre. Certaines personnes infectées seront en contact avec beaucoup de personnes susceptibles et répandront beaucoup la maladie, alors que d'autres, plus isolées, ne contamineront presque personne. L'hypothèse simplificatrice ignore ces variations :

## Hypothèse 1 :

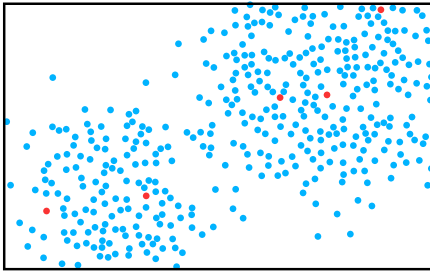
À chaque instant, toute personne infectée est en contact avec le même nombre  $M$  de personnes susceptibles.

Nous visualiserons les êtres humains comme des boules de billard en mouvement. Même simpliste, cette image permet de mieux comprendre cette première hypothèse. La figure suivante représente une telle population relativement homogène où cinq personnes infectées (en rouge) ont été introduites dans la population susceptible (en bleu).



La figure suivante montre une population concentrée autour de deux centres urbains. L'hypothèse 1 est probablement raisonnable dans une région où la densité est à peu près constante et où les habitudes de vie des citoyens sont similaires, mais elle est moins convaincante pour la seconde figure : une personne susceptible ou infectée en milieu rural ne rencontre pas autant de personnes, susceptibles ou infectées, que celles vivant dans les centres urbains.





Et cette hypothèse se justifie plus facilement au début d'une épidémie, lorsque le nombre de personnes infectées est infime par rapport à la taille de la population.

Le modèle le plus simple ajoute une seconde hypothèse :

**Hypothèse 2 :**

Une personne infectée a probabilité  $p_1$  de contaminer une personne susceptible qu'elle rencontre.

Ceci permet d'introduire un taux de transmission par personne infectée :  $\beta = Mp_1$  est le nombre moyen de personnes susceptibles qu'une personne infectée infecte en une journée. Ce qui nous intéresse est l'évolution du nombre d'infectés. Si  $I(n)$  désigne ce nombre au jour  $n$ , alors, au jour  $n + 1$ , il y aura

$$I(n+1) = I(n) + \beta I(n) = (1 + \beta)I(n) \quad (*)$$

personnes infectées, c'est-à-dire qu'à chaque jour, le nombre d'infectés est multiplié par  $(1 + \beta)$ . Après  $n$  jours le nombre d'infectés est

$$I(n) = I(0) \cdot (1 + \beta)^n \quad (**)$$

où  $I(0)$  est le nombre d'infectés au jour  $n=0$ , souvent choisi comme le jour où le premier cas est observé. Ce modèle très simple prédit donc une croissance exponentielle. Cette croissance devrait être visible sur un graphe du logarithme (en base 10) des  $I(n)$  puisque

$$\log_{10} I(n) = \log_{10} I(0) + n \log_{10} (1 + \beta)$$

décrit une droite en fonction de  $n$ . Pour comparer divers pays, nous utiliserons la proportion d'infectés

$$i(n) = I(n)/N,$$

où  $N$  est la population étudiée.

Mais, comment pouvons-nous estimer ces  $i(n)$ ? Ce n'est pas si facile car les statistiques officielles ne donnent que les *cas déclarés*. Nous avons quand même fait un essai ! Supposons que le nombre  $I(n)$  de personnes infectées soit proportionnel au nombre cumulé de cas déclarés  $CD(n)$ . Cette hypothèse n'est pas farfelue. Il n'est pas irréaliste de supposer qu'une fraction de la population infectée passe inaperçue. Si la constante de proportionnalité est  $c$ :

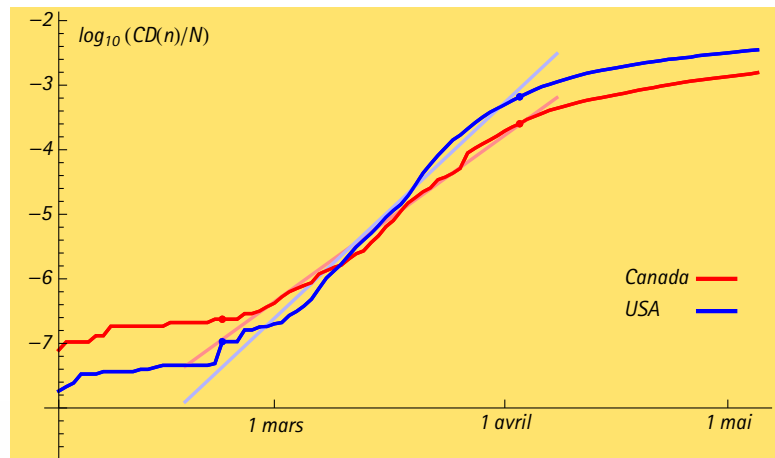
$$CD(n) = c I(n),$$

alors le logarithme de  $CD(n)/N$  est aussi linéaire en  $n$  :

$$\begin{aligned} \log_{10}(CD(n)/N) &= \log_{10}(c I(n)) - \log_{10} N \\ &= (\log_{10}(c) + \log_{10} I(0) - \log_{10} N) \\ &\quad + n \log_{10}(1 + \beta). \end{aligned}$$

Nous avons simplement translaté la courbe vers le haut, sans changer sa pente, et ceci nous permet d'estimer  $\beta$  même si la constante  $c$  demeure inconnue.

Ce sont donc ces données pour la COVID-19 que nous avons portées sur le graphique ci-dessous.



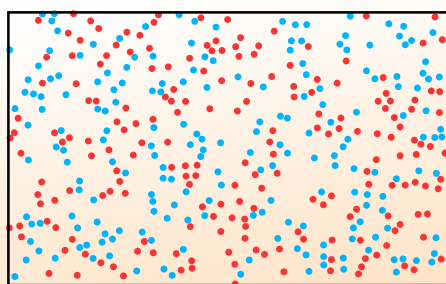
Pour la période allant du 22 février au 4 avril, la progression semble linéaire. Cette progression se traduit en un  $\beta$  d'environ un tiers de personne par jour pour les États-Unis et d'un quart pour le Canada. Ces graphes permettent également de constater que la progression exponentielle ne s'étend pas au-delà de la fin mars, vraisemblablement à cause des mesures de confinement et d'hygiène imposées par les gouvernements.

Mais est-il justifié d'utiliser les cas déclarés pour estimer la proportion  $i(n)$  d'infectés ?

C'est certainement périlleux : beaucoup de personnes infectées sont asymptomatiques et, pour la COVID-19, la plupart des tests étaient réservés aux personnes ayant voyagé ou ayant des symptômes prononcés. Nos estimés des  $i(n)$  sont loin d'être parfaits ! Et il y a plus grave : sur la période du mois de mars, certaines personnes infectées auront guéri, ce que le modèle le plus simple semble ignorer ! Il est temps de passer à un modèle plus fin.

## SIR, un modèle plus fin

Pour le second modèle, nous conservons l'hypothèse 2, mais nous raffinons l'hypothèse 1. En effet il est plus naturel de supposer un nombre moyen constant  $M$  de rencontres quotidiennes d'un individu infectieux et, si le virus se propage à une partie importante de la population, alors parmi ces  $M$  personnes, certaines seront déjà infectées ou immunisées : ces contacts généreront alors moins de nouvelles infections.



Par exemple, la figure ci-dessus montre une situation où la moitié de la population est infectée. Clairement, le nombre de personnes susceptibles autour d'une personne infectée a diminué par rapport à la figure de la première page.

### Hypothèse 1' :

Deux personnes quelconques ont probabilité  $p_2$  de se rencontrer une journée donnée.

Donc, il y a probabilité  $p = p_1 p_2$  que deux personnes, l'une infectée, l'autre susceptible se rencontrent une journée donnée et que ce contact résulte en une infection. Soit maintenant  $S(n)$  le nombre de personnes susceptibles au  $n$ -ième jour. Comme précédemment  $I(n)$  désigne le nombre de personnes infectées. La probabilité qu'une personne susceptible en particulier se fasse infecter entre le  $n$ -ième jour et le  $(n+1)$ -ième jour est alors  $p \cdot I(n)$ . Ainsi,  $p \cdot S(n) I(n)$  personnes quittent le groupe des susceptibles vers le groupe des infectés et, au jour  $n+1$ ,

$$S(n+1) = S(n) - p \cdot S(n) I(n).$$

Ce terme  $p \cdot S(n) I(n)$  qui dépend des deux populations en présence apparaît dans la description de plusieurs phénomènes en chimie, physique et écologie : il porte le nom de *loi d'action de masse*.

Le premier modèle simple supposait que les personnes infectées demeuraient indéfiniment contagieuses. Ceci n'est pas le cas et requiert une autre hypothèse :

### Hypothèse 3 :

Chaque jour, une personne infectée arrête d'être contagieuse avec une probabilité  $q$ .

Remarquons que l'hypothèse 3 capture les deux conclusions possibles d'une infection : la guérison ou le décès. Le symbole  $R(n)$  sera le nombre de ces personnes ainsi *retirées* des populations susceptibles et infectées. Pour les maladies où le taux de mortalité est très faible, cette catégorie devient celle des personnes *rétablies*. Il est aussi possible de distinguer les personnes qui guérissent avec probabilité  $q_g$  de celles qui décèdent avec probabilité  $q_d$ . Dans ce cas, il faudra que  $q = q_g + q_d$ .

Dans ce modèle, au  $n$ -ième jour, les personnes infectées voient leur nombre augmenter de celles qui viennent d'être infectées et diminuer de celles qui viennent de guérir ou de mourir :

$$I(n+1) = I(n) + p \cdot S(n) I(n) - q I(n).$$

Finalement le nombre de personnes retirées de la population s'accroît justement par ce même terme  $qI(n)$  :

$$R(n+1) = R(n) + qI(n).$$

Remarquons que la population totale (incluant les décès parmi les retirés) donnée par la somme de  $S(n)$ ,  $I(n)$  et  $R(n)$  est constante et égale à :

$$\begin{aligned} S(n+1) + I(n+1) + R(n+1) \\ = S(n) + I(n) + R(n) = N. \end{aligned}$$

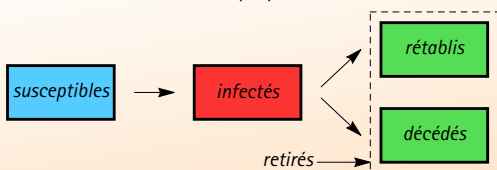
Il est aussi possible de diviser les trois équations par  $N$ . Alors

$$s(n) = \frac{S(n)}{N}, \quad i(n) = \frac{I(n)}{N}, \quad r(n) = \frac{R(n)}{N}$$

seront les fractions de la population qui sont susceptibles, infectées et retirées ; elles prennent leurs valeurs dans l'intervalle  $[0, 1]$  et leur somme est égale à 1. Introduisons les nouveaux paramètres  $\beta = pN$  et  $\gamma = q = q_g + q_d$ . Le modèle SIR<sup>1</sup> est décrit par les trois équations récursives

$$\begin{aligned} s(n+1) &= s(n) - \beta s(n) \cdot i(n), \\ i(n+1) &= i(n) + \beta s(n) \cdot i(n) - \gamma i(n), \\ r(n+1) &= r(n) + \gamma i(n). \end{aligned}$$

Le nom SIR vient du fait que la population est partagée entre les trois compartiments des individus *susceptibles*, *infectés* et *retirés* (ou *rétablis*); le diagramme suivant capture le flot entre ces trois populations.



Étant donné des conditions initiales  $s(0)$ ,  $i(0)$  et  $r(0)$ , il existe une unique solution  $s(n)$ ,  $i(n)$  et  $r(n)$  de ces équations récursives. Pour une modélisation à partir du début de l'épidémie, on prend des conditions initiales  $r(0) = 0$  et  $i(0)$  très petit, parce que seule une petite fraction de la population est infectée. Parce qu'elles simplifient grandement la complexité du phénomène, les hypothèses 1', 2 et 3 permettent de capturer un phénomène *probabiliste* par un modèle complètement *déterministe*.

1. aussi appelé modèle SIR discret, par opposition au modèle continu avec des équations différentielles.

## Que prédit le modèle SIR ?

L'analyse du premier modèle était simple : le nombre (ou la proportion) d'infectés y croît de façon exponentielle et le seul paramètre  $\beta$  mesure le taux de croissance de cette exponentielle. Le modèle SIR, avec ses deux paramètres  $\beta$  et  $\gamma$ , est beaucoup plus riche. Explorons-le.

Puisque  $s(n+1) - s(n) < 0$  si  $s \neq 0$  et  $i \neq 0$ , la proportion  $s(n)$  de susceptibles décroît. De façon semblable,  $r(n+1) - r(n) > 0$  si  $i \neq 0$ , et la proportion  $r(n)$  de retirés croît. Regardons maintenant la variation journalière  $i(n+1) - i(n)$  des infectés. L'équation qui régit  $i(n)$  peut se réécrire

$$\begin{aligned} i(n+1) - i(n) &= \beta i(n) \cdot \left( s(n) - \frac{\gamma}{\beta} \right) \\ &= \beta i(n) \cdot (s(n) - R_0^{-1}), \end{aligned}$$

où nous avons introduit le paramètre

$$R_0 = \beta/\gamma.$$

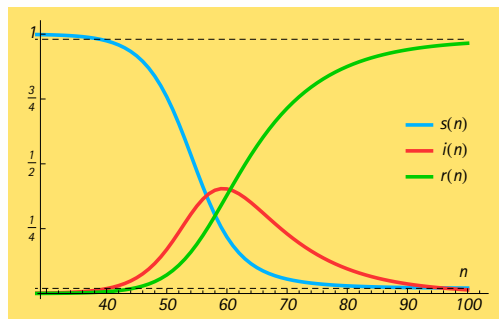
Ce nombre crucial est appelé le *taux de reproduction de base*. C'est lui qui indique si la maladie va s'étendre et devenir une épidémie ou, au contraire, va s'éteindre. En effet, si  $R_0 < 1$ , alors  $1/R_0 > 1$ , et comme  $s(n) \in [0, 1]$ , alors  $i(n+1) - i(n) < 0$ . La proportion d'infectés décroît et la maladie ne se propage pas. Par contre, si  $R_0 > 1$  et donc  $1/R_0 < 1$ , alors, lorsque  $s(n)$  est assez grand, ce qui est le cas quand seulement quelques individus sont infectés, la proportion d'infectés augmente et une épidémie s'enclenche.

Mais jusqu'où l'épidémie sévira-t-elle? Au fur et à mesure que les proportions d'infectés  $i(n)$  et de personnes rétablies et décédées  $r(n)$  croissent,  $s(n)$  décroît. Dès que  $s(n)$  passe en deçà de  $1/R_0$ , alors  $i(n+1) - i(n)$  devient négatif et le nombre d'infectés diminue jusqu'à s'éteindre. L'épidémie disparaît avant que toutes les personnes susceptibles aient contracté la maladie. C'est ce qu'on appelle l'*immunité de groupe*.

Voici une première simulation avec  $\gamma \approx 1/10$  et  $R_0 \approx 3,8$ , des valeurs à l'intérieur des fenêtres acceptées pour la COVID-19 au

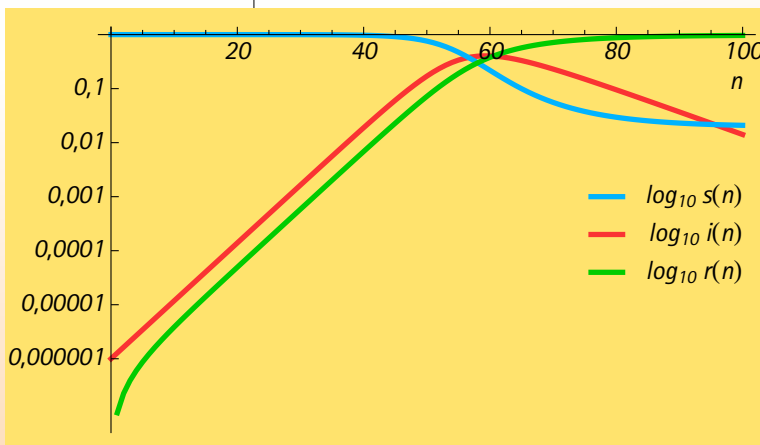


moment d'écrire ce texte (fin juin 2020)<sup>2</sup>. La proportion initiale des infectés a été posée à 0,0001% d'infectés, ce qui correspondrait au Québec à environ 8 personnes infectées. Les données  $s(0)$ ,  $s(1)$ ,  $s(2)$ ,  $s(3)$ , ..., ont été calculées, puis jointes par une courbe, de même pour les  $i(n)$  et  $r(n)$ . Sur le graphique ci-dessous, les premiers jours de l'épidémie ont été retirés puisque les proportions d'infectés et de retirés sont pratiquement nulles. Le pic de l'épidémie est atteint au 59<sup>e</sup> jour



avec plus de 40% d'infectés. L'immunité de groupe  $\gamma$  est visible : l'épidémie meurt avant que toute la population ne soit infectée, en en épargnant ainsi environ 2%.

Le graphique ci-dessous utilise une échelle logarithmique pour les trois populations. Le logarithme des  $i(n)$  y croît linéairement pendant les quelques 70 premiers jours ; une bonne approximation pendant cette période est donnée par  $i(n) \approx i(0)(1 + \beta - \gamma)^n$ , et le premier modèle simple du début aurait pu reproduire ce comportement. Il ne peut cependant pas expliquer la décroissance du nombre d'infectés !

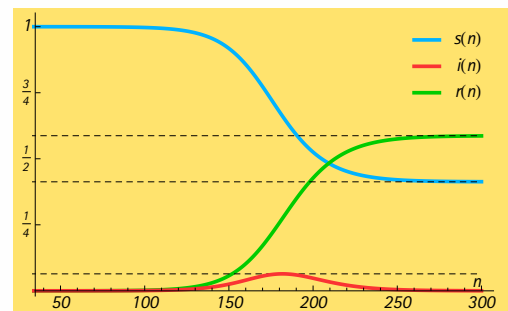


2. Un fichier Excel dans lequel vous pouvez modifier les paramètres  $R_0$  et  $\gamma$  ainsi que les conditions initiales  $i(0)$  et  $r(0)$  se trouve à <http://accromath.uqam.ca/accro/wp-content/uploads/2020/08/sir.xlsx>

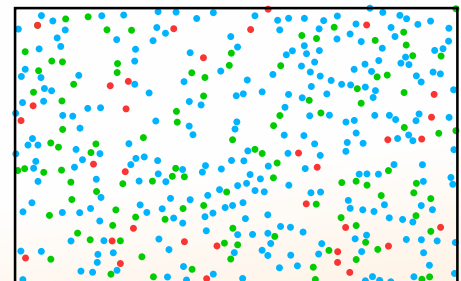
*Mais ce n'est pas ce que le Québec a observé...*

*Quand la santé publique intervient, les paramètres  $\beta$  et  $\gamma$  changent !*

Voici une seconde simulation : les paramètres sont maintenant proches de ceux de la grippe saisonnière, soit  $\gamma \approx 1/7$  et  $R_0 \approx 1,5$ . L'épidémie prend beaucoup plus de temps à s'installer, jamais plus de 6,3% de la population n'est malade en même temps et 42% de la population est épargnée par l'immunité de groupe.



La figure ci-dessous, tracée au pic de l'épidémie, indique le petit nombre des personnes infectées.



*Le taux de reproduction de base  $R_0$  est lié au nombre moyen d'individus qu'au tout début de l'épidémie, une personne infectée infectera au cours de sa période contagieuse.*

Si chaque personne infectée infecte plus d'une personne, alors le nombre de personnes qui vont contracter la maladie croîtra exponentiellement. Nous avons vu plus tôt que les



nombres de cas déclarés au Canada et aux États-Unis avaient en effet cru exponentiellement pendant une bonne partie du mois de mars. Le graphique ci-dessous présente des estimés des proportions  $\hat{i}(n)$  pour le Canada, la Corée du Sud, l'Italie et les États-Unis. Pour obtenir les estimés de  $\hat{i}(n)$  que nous noterons  $\hat{i}(n)$ , nous n'avons conservé que les nouveaux cas déclarés  $(CD(n) - CD(n-1))$  journaliers durant les dix jours précédents. Chacun a été pondéré par la probabilité que ces personnes infectées soient toujours contagieuses. Ainsi  $\hat{i}(n)$  est proportionnel à

$$\frac{1}{N} \sum_{i=0}^9 (1-q)^i (CD(n-i) - CD(n-i-1))$$

où nous avons posé  $q = 1/10$ . (Ceci correspond à un temps de contagion moyen de 10 jours.) Chacune de ces courbes doit être comparée à celle du graphe de  $i(n)$  (en rouge sur les simulations du modèle SIR de la page précédente). La Corée a rapidement maîtrisé l'épidémie et l'Italie l'a suivie. Pour chacun de ces deux pays, les mesures de confinement ont diminué le paramètre  $\beta$  et ramené  $R_0$  à une valeur  $R_0'$  en deçà de 1. Le profil de la courbe des États-Unis indique un  $R_0'$  un tout petit peu en deçà de 1, et celle du Canada un  $R_0'$  se rapprochant de 1.

Mais attention au déconfinement ! Il risque de ramener le  $R_0'$  à un  $R_0'' > 1$ , alors que la proportion de la population qui a été atteinte est bien en deçà de l'immunité de groupe. Donc, il est absolument essentiel que les mesures de distanciation physique soient suffisantes pour maintenir le nouveau  $R_0''$  bien en deçà de 3,8.

### Le travail de la Santé publique

Nous avons vu que le modèle SIR est simpliste. Malgré cela, il est riche d'enseignement : il capture la croissance exponentielle au début de l'épidémie, le pic de l'épidémie et l'aplatissement de la courbe. De plus, il introduit des outils simples mais essentiels pour comprendre la propagation de l'épidémie : le nombre  $R_0$  et le concept d'immunité de groupe.

Le travail des agences de santé publique en est un d'équipe. Aux premières loges se trouvent les équipes médicales qui essaient de comprendre le nouveau virus. En parallèle, les équipes de modélisation raffinent les modèles théoriques, comme le modèle SIR au vu de la compréhension du virus et de ses mécanismes de propagation, tandis que les statisticiens et actuaires aident à l'estimation des paramètres des modèles théoriques à partir des données disponibles. Les sciences mathématiques jouent un rôle important dans ce travail d'équipe.

