

# Comprendre la propagation d'un virus en utilisant les modèles compartimentaux (partie 1)



## Motivation

Avec la récente apparition du Coronavirus, et sa fulgurante expansion en quelques mois, beaucoup de gens ont aujourd'hui été confrontés à la réalité d'une pandémie.

A travers cet article, j'aimerais présenter de manière simple l'aspect mathématique de la propagation d'une épidémie dans une population, et aider les lecteurs à comprendre comment les scientifiques peuvent prédire l'évolution d'une maladie.

## Introduction

Pour modéliser la propagation d'une épidémie (au sens large du terme), les épidémiologistes utilisent ce que nous appelons des **modèles compartimentaux**. Ces modélisations permettent aux scientifiques de répondre à des questions importantes, telles que :

- Quel % de la population sera affecté par l'épidémie ?
- Quelle sera la gravité de cette épidémie ?
- Quand ferons-nous face au « pic » de l'épidémie ?
- Etc...

On dit de ces modèles qu'ils sont compartimentaux car chaque individu de la population étudiée est à tout instant  $t$  de l'épidémie placé dans un *compartiment* ou *état*. Dans cet article, je présenterai le modèle **SIR** à 3 compartiments. La compréhension de ce modèle est fondamentale, puisqu'il peut non seulement s'appliquer à beaucoup de cas du monde réel, mais il sert aussi de pilier fondateur à des modèles plus complexes, comme ceux que l'on pourrait utiliser pour modéliser le Coronavirus.

## Fondements du modèle SIR

Le modèle SIR tient son nom de l'anglais, et est un acronyme des 3 lettres correspondant aux 3 compartiments dans lesquels nous allons placer la population :

### S : Susceptible

Comme son nom l'indique, une personne « *susceptible* » est une personne susceptible de contracter le virus si elle est en contact avec ce dernier. Dans le cas de l'apparition d'un nouveau virus, nous pouvons considérer que toute personne est susceptible d'être contaminée.

### I : Infected

Une personne « *infected* » est une personne qui a contracté le virus, et qui peut le transmettre à une personne *susceptible* si elle entre en contact avec cette dernière.

### R : Recovered

Une personne « *recovered* » est une personne qui a précédemment contracté le virus, mais qui aujourd'hui est une personne saine et ne peut **ni contracter ni transmettre ce virus**. (Cette hypothèse est valable pour la grande majorité des maladies).

**Remarque 1 :** Dans un modèle compartimental (SIR ou autre), toute personne se trouve **obligatoirement** dans **un unique** compartiment à un instant  $t$  donné. Dans notre cas, une personne est toujours soit **S**, soit **I**, soit **R**.

**Remarque 2 :** Dans la suite de l'article, nous noterons  $S(t)$ ,  $I(t)$ , et  $R(t)$  le nombre de personnes *susceptible*, *infected*, *recovered* au temps  $t$ , et nous considérerons que l'unité de temps est le jour.

**Remarque 3 :** Ce modèle est valable dans le cas d'une épidémie où le **taux de mortalité** est faible ou nul. Lorsque ce n'est pas le cas, l'ajout d'un nouveau compartiment **D** (de l'anglais **D**eceased qui signifie décédé) est nécessaire, mais je ne rentrerai pas beaucoup dans les détails dans cet article, et nous considérerons que :

$$S(t) + I(t) + R(t) = N \quad (1)$$

$\forall t$ , où  $N$  est la taille de la population étudiée, et que le **graphe de transition des états** est le suivant :

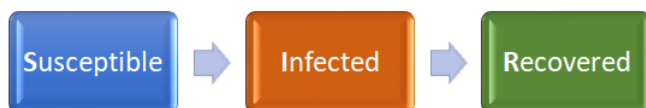


Figure 1: Graphe de transition des états du modèle SIR

## Les équations qui gouvernent le modèle SIR

A ce stade, vous aurez compris que l'objectif de la modélisation est de trouver des formules pour les fonctions  $S(t)$ ,  $I(t)$ ,  $R(t)$ , qui nous permettraient de comprendre et de suivre l'évolution de la maladie. Si nous pouvions tracer et observer ces 3 courbes, alors nous pourrions répondre à toutes les questions soulevées précédemment !

Malheureusement, il n'est pas évident de trouver des formules exactes directement pour ces 3 fonctions. En revanche, il est beaucoup plus simple de trouver les **taux d'accroissement** de ces fonctions, et donc d'exprimer le modèle sous la forme d'un système d'équations différentielles, en fonction de  $\frac{dS}{dt}$ ,  $\frac{dI}{dt}$ , et  $\frac{dR}{dt}$ .

Nous nous proposons donc dans la suite d'établir les formules de ces taux d'accroissement pour décrire notre modèle.

**Cas de  $\frac{dS}{dt}$  :**

D'après nos hypothèses et la définition du modèle SIR, nous savons que  $S(t+1) \leq S(t)$  (i.e demain, il y aura soit autant de personnes susceptibles d'être contaminées, soit moins puisque de nouvelles personnes l'ont été) :  $S(t)$  est une fonction décroissante, sa dérivée  $\frac{dS}{dt}$  est donc négative.

Un taux d'accroissement élevé de  $S(t)$  résulte d'un grand nombre d'interactions entre les **S** et les **I** : si nous avons beaucoup de personnes **Susceptibles** et beaucoup de personnes **Infected** au temps  $t$ , alors la probabilité d'avoir beaucoup d'infections (et donc beaucoup moins de **S** à  $t+1$ ) est grande. Nous pouvons donc affirmer que  $\frac{dS}{dt}$  est proportionnelle à  $S \cdot I$

À cela, nous ajouterons un coefficient de proportionnalité  $\beta$ , que l'on appelle le **taux d'infection** (*infection rate* en anglais) ou **force de l'infection**. Pour simplifier, nous considérerons que ce paramètre dépend de deux facteurs :

1. le nombre de personnes moyen qu'une personne rencontre dans la journée (que l'on peut réduire grâce au confinement).
2. la **probabilité de contagion** : probabilité qu'une personne **I** transmette le virus à une personne **S** lors d'un contact (que l'on peut réduire avec le port d'un masque ou de gants par exemple).

Nous poserons donc :

$$\frac{dS}{dt} = -\beta \cdot S \cdot I \quad (2)$$

avec  $\beta > 0$

**Cas de  $\frac{dI}{dt}$  :**

Nous nous intéressons maintenant à la variation du nombre de personnes *infected* d'un jour  $t$  au lendemain  $t+1$ .

Cette variation est simplement égale au nombre de personnes *susceptibles* qui ont été infectées ( $\frac{dS}{dt}$  !), moins le nombre de personnes infectées qui ont vaincu le virus et sont maintenant guéries.

Si une personne reste malade (*infected*) en moyenne  $\lambda$  jours, alors nous pouvons dire qu'en moyenne, une proportion  $\gamma = \frac{1}{\lambda}$  des personnes infectées guérit chaque jour (i.e quitte le compartiment **I** pour devenir des **R**).

Nous noterons donc :

$$\frac{dI}{dt} = \beta \cdot S \cdot I - \gamma \cdot I \quad (3)$$

où  $\gamma$  représente la proportion de personnes infectées qui guérissent chaque jour, aussi appelé le *taux de guérison* (ou *recovery rate* en anglais).

Cas de  $\frac{dR}{dt}$  :

Par définition,  $\frac{dR}{dt}$  est égale au nombre de personnes infectées qui guérissent chaque jour. Il s'en suit que :

$$\frac{dR}{dt} = \gamma \cdot I \quad (4)$$

**Remarque 1 :** Notons que  $\frac{dR}{dt}$  est une fonction positive, ce qui est en accord avec les hypothèses du modèle SIR puisqu'une personne *recovered* ne peut pas redevenir *infected*, donc  $R(t)$  est croissante.

**Remarque 2 :** Nous pouvons facilement vérifier que :

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$$

ce qui est en accord avec l'équation (1).

Pour conclure, le système d'équations (différentielles) qui gouverne le modèle SIR est le suivant :

$$\begin{cases} S(t) + I(t) + R(t) = N \\ \frac{dS}{dt} = -\beta \cdot S \cdot I \\ \frac{dI}{dt} = \beta \cdot S \cdot I - \gamma \cdot I \\ \frac{dR}{dt} = \gamma \cdot I \end{cases} \quad (5)$$

où :

- $S(t)$  : Nombre de personnes *susceptibles* au temps  $t$
- $I(t)$  : Nombre de personnes *infected* au temps  $t$
- $R(t)$  : Nombre de personnes *recovered* au temps  $t$
- $N$  : Taille de la population étudiée
- $\beta$  : Taux d'infection
- $\lambda$  : Durée moyenne d'une infection (en jours)
- $\gamma$  : Taux de guérison ( $= \frac{1}{\lambda}$ )

## Analyse qualitative des équations du modèle SIR

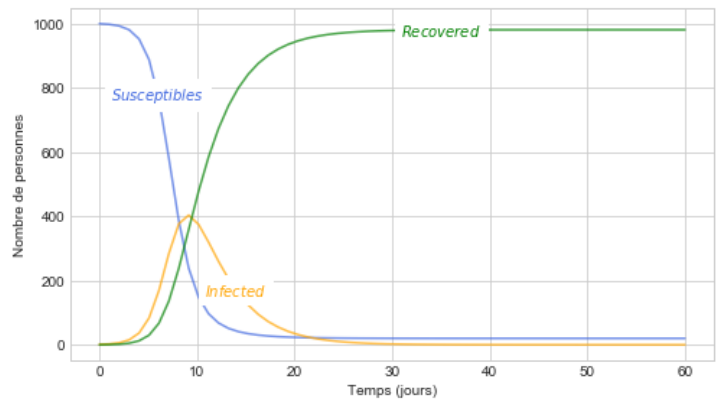
Le système d'équations (5) obtenu dans la partie précédente n'est pas évident à résoudre analytiquement à cause de la non-linéarité introduite par le produit de  $S$  et  $I$ .

Dans cet article, l'objectif n'est pas de résoudre ce système mais plutôt de l'analyser qualitativement et de comprendre comment se propage une épidémie.

Nous nous proposons donc d'analyser des scénarios d'épidémie pour des conditions initiales différentes. Nous observerons la propagation d'une épidémie sur 2 mois, et sur une population de taille  $N = 1000$ .

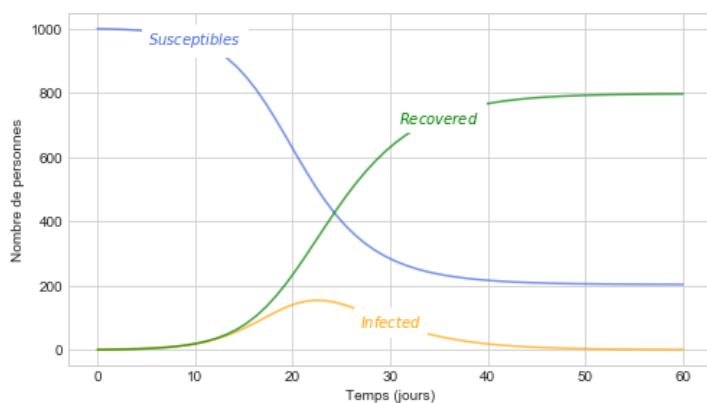
**Note :** Dans la suite, nous noterons  $S(0)$ ,  $I(0)$  et  $R(0)$  respectivement le nombre de personnes *susceptibles*, *infected* et *recovered* au temps  $t_0$  (état initial du système).

**Scénario 1 :**  $S(0) = 999$   $I(0) = 1$   $R(0) = 0$   $\beta = 1.3$   $\gamma = 0.3$



Ce premier scénario est le scénario classique d'une épidémie dévastatrice. À  $t_0$ , le virus vient d'émerger. Très peu de personnes sont contaminées, 1 sur les 1000 ici. La valeur de  $\beta$  est forte par rapport à  $\gamma$  : **soit parce que le virus est très contagieux, soit parce que les contacts entre les personnes sont nombreux**. Quelle que soit la raison, le nombre de personnes infectées monte en flèche, pour atteindre **le pic de l'épidémie** en moins de 10 jours. À un moment donné, le nombre de personnes *susceptibles* devient faible, donc les contacts entre les  $S$  et les  $I$  diminuent, le terme  $\gamma \cdot I$  devient plus grand que  $\beta \cdot S \cdot I$  (équation (3)), ce qui implique une décroissance de la courbe jaune des personnes infectées. En 20 jours, quasiment toute la population a été touchée par le virus et est soit *infected*, soit *recovered*.

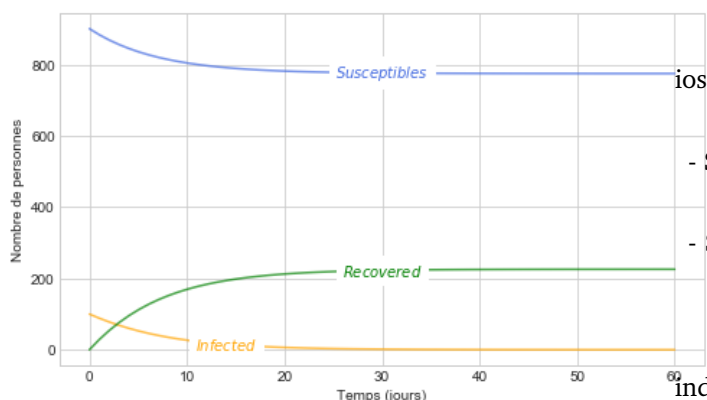
**Scénario 2 :**  $S(0) = 999$   $I(0) = 1$   $R(0) = 0$   $\beta = 0.6$   $\gamma = 0.3$



Supposons dans ce scénario que nous soyons face au même virus que le précédent. Les conditions initiales sont identiques, hormis une valeur de  $\beta$  plus faible. Nous émettrons donc l'hypothèse (certes irréaliste...) qu'à  $t_0$ , toute la population était au courant de l'émergence d'un virus, et que les personnes ont volontairement diminué les contacts humains. Nous pouvons remarquer deux changements principaux par rapport au scénario précédent :

1. Lorsque le système atteint un état stable et que l'épidémie prend fin, environ 200 personnes (soit 20% de la population) n'ont jamais été contaminées par le virus.
2. Le nombre de personnes infectées  $I(t)$  ne croît pas aussi rapidement, et son pic est bien plus bas : **la courbe s'est aplatie**. Ce terme est largement utilisé aujourd'hui, et lorsque les scientifiques parlent **d'aplatir la courbe**, ils espèrent en fait **influer sur le paramètre  $\beta$**  (réduire le taux d'infection). En effet, comme le montre ce scénario, **aplatir la courbe des personnes infectées revient indirectement à minimiser la proportion de la population qui sera touchée à la fin de l'épidémie**. Cette relation est mise en évidence à travers les équations du modèle SIR.

**Scénario 3 :**  $S(0) = 900$   $I(0) = 100$   $R(0) = 0$   $\beta = 0.2$   
 $\gamma = 0.3$



Pour ce 3ème scénario, nous avons fait 2 principaux changements :

1. Pour la première fois, nous avons injecté une valeur de  $\beta$  plus faible que  $\gamma$ .
2. Au début de l'épidémie, nous avons 100 personnes infectées au lieu d'une.

Intuitivement, nous pourrions penser qu'avec 100 personnes infectées au lieu d'une à  $t_0$ , la propagation de l'épidémie se ferait rapidement. Cette intuition est bonne ! Lorsque le nombre de personnes contagieuses est important à  $t_0$ , l'épidémie a de grandes chances de se propager très vite. Seulement cette fois-ci, **la valeur de  $\beta$  est inférieure à celle de  $\gamma$** . En d'autres termes, **les individus guérissent plus vite que l'épidémie ne se propage !** Et pour cause, la courbe jaune  $I(t)$  n'a à aucun moment été croissante, et il n'y a pas eu d'épidémie.

Ce scénario indique donc que **même avec un grand nombre de personnes infectées à l'état initial, le virus va se «bloquer» et ne pourra pas se propager si  $\beta < \gamma$** , n'entraînant donc pas d'épidémie.

Vous aurez maintenant compris à ce niveau de l'article que ce ne sont pas tant les valeurs de  $\beta$  et  $\gamma$  seules qui importent, mais **le rapport** de ces deux valeurs, que l'on appelle le *taux de reproduction de base*, plus connu sous le nom de  $R_0$ .

## Le taux de reproduction de base $R_0$ et l'immunité collective

Le taux de reproduction de base, ou  $R_0$  est défini comme **le rapport entre le taux d'infection et le taux de guérison** :

$$R_0 = \frac{\beta}{\gamma} \quad (6)$$

En d'autres termes, il s'agit du **nombre moyen d'individus qu'infecte une personne dans une population où tout le monde est susceptible**.

Comme nous l'avons compris de l'analyse des scénarios précédents :

- Si  $R_0 > 1$ , le virus se propage plus vite que les gens ne guérissent, et il y aura donc une épidémie.
- Si  $R_0 < 1$ , le virus se propage moins vite que les gens ne guérissent, et la contagion ne se fera pas assez rapidement pour entraîner une épidémie.

Vous comprenez maintenant l'importance de cet indicateur, qui nous permet de mesurer en un clin d'oeil la puissance de la propagation du virus, et dont l'interprétation est relativement simple.

Néanmoins, son utilisation ne s'arrête pas là. Pour des  $R_0 > 1$ , cet indicateur nous permet aussi de calculer **le seuil d'immunité collective**, ou **immunité grégaire** (*herd immunity* en anglais) sous certaines conditions.

L'immunité collective est une technique visant à bloquer la propagation du virus si un certain pourcentage de la population est immunisé (soit car il a été contaminé par le passé et est aujourd'hui *recovered*, soit par le biais d'un vaccin). Le seuil d'immunité collective, noté  $p_c$ , représente le pourcentage de la population qui doit être immunisé au virus pour bloquer l'évolution de la maladie.

Dans une population suffisamment **homogène** (c'est-à-dire une population où l'on considère que chaque individu peut entrer en contact avec tous les autres), le seuil d'immunité collective est défini comme :

$$p_c = 1 - \frac{1}{R_0} \quad (7)$$

Pour avoir une vague idée de cas réels, nous pouvons observer sur le tableau suivant certaines valeurs de  $R_0$  et  $p_c$  pour des maladies contagieuses :

Maladie	$R_0$	$p_c$
Diphtérie	6-7	85%
Rougeole	12-18	83-94%
Oreillons	4-7	75-86%
Coqueluche	12-17	92-94%
Polio	5-7	% 80-86
Rubéole	5-7	83-85%
Variole	6-7	83-85

Tableau 1 : Valeurs de  $R_0$  et  $p_c$  pour des maladies communes

## Vers des modèles plus complexes...

Le modèle SIR présente un certain nombre de limitations, qui peuvent être plus ou moins contraignantes selon l'épidémie étudiée. De mon point de vue, pour une épidémie telle que celle déclenchée par le Coronavirus, il ne capture pas suffisamment la réalité du terrain. Les deux principales améliorations que nous pourrions apporter à ce modèle, et que je détaillerai dans un prochain article, sont les suivantes :

<sup>1</sup>www.laquotidienne.ma

<sup>2</sup>www.lepoint.fr

## 1 - L'ajout de nouveaux compartiments

Le taux de mortalité du Coronavirus varie beaucoup selon les pays. Dans des pays comme le Maroc, ce taux se situerait autour de 5%<sup>1</sup>, tandis que pour des pays comme la Belgique, nous atteindrions les 15%<sup>2</sup>. Ces taux de mortalité me paraissent suffisamment haut pour justifier l'ajout d'un nouveau compartiment **D**.

Il se pourrait aussi que pour une certaine raison, l'immunité d'un individu face au Corona diminue et qu'il redevienne susceptible de contracter le virus au bout d'un certain temps. Peut-être faudrait-il ajouter une transition entre les états **R** et **S** ? Cette question est encore aujourd'hui sans réponse mais mérite d'être considérée.

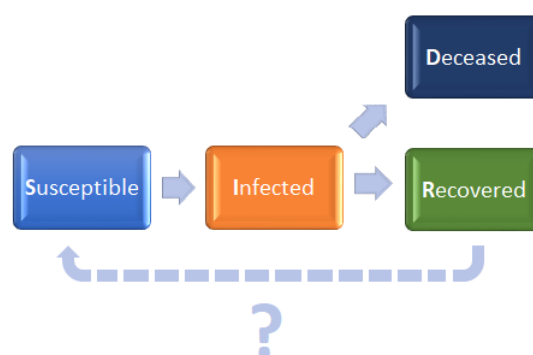


Figure 2: Graphe de transitions des états d'un modèle SIRD, possible pour modéliser le Coronavirus.

Quel que soit le scénario choisi, la méthodologie reste la même : d'abord établir le système d'équations différentielles, puis estimer les paramètres à l'aide de données réelles.

**Remarque** : Si nous observons une épidémie sur des périodes **très** longues, nous pouvons aussi ajouter au modèle des **dynamiques vitales**, en y incluant les naissances et les décès naturels.

## 2 - Passer d'un paramétrage statique à un paramétrage dépendant du temps

Dans le modèle que nous avons décrit, les paramètres à estimer et en particulier  $\beta$  (et donc  $R_0$ , puisque celui-ci dépend directement de  $\beta$ ) sont figés et ne dépendent pas du temps. **Cette hypothèse n'est absolument pas réaliste sur de longues périodes !** En effet, pour un modèle plus précis, il serait judicieux de faire de  $\beta$  (et  $R_0$ ) des variables dépendantes du temps  $\beta(t)$  (et  $R_0(t)$ ). En particulier, nous aimerions modéliser des événements tels que **les confinements, l'adoption de gestes barrière, le port de masques et de gants, les vaccins...**

## Conclusion

Dans cet article, j'espère être parvenu à vous initier aux modèles compartimentaux à travers le modèle de base, le modèle SIR.

Après avoir compris ses fondements, nous avons dérivé les équations qui gouvernent ce modèle. Cette étape est fondamentale, et ne change pas drastiquement d'un modèle compartimental à l'autre.

L'étude de 3 scénarios a permis de comprendre comment la variation du taux d'infection  $\beta$  et des conditions initiales pouvait influencer sur la propagation d'une épidémie. Dans un prochain article, je montrerai comment établir les équations d'un modèle plus complexe, et comment ajuster certains paramètres comme le  $R_0$  pour modéliser des événements tels que le confinement ou la découverte d'un vaccin.