

Simulation épidémiologique

Simulation épidémiologique et modèles à compartiments

Antonin LEFEVRE¹ 

¹Université de Reims

Abstract

Ce projet a pour objectifs de modéliser une **épidémie**, et de simuler sa propagation au sein d'une population d'individus. Le taux d'infection, de guérison et de létalité seront les paramètres de notre modèle, que nous étudierons afin de mieux comprendre leur impact sur la vitesse de propagation du virus. On choisira des populations de quelques centaines à quelques dizaines de milliers d'individus et on supposera qu'un individu immunisé le reste toute sa vie. Cette modélisation sera réalisée avec matplotlib et plotly indépendamment. Ce modèle statistique sera ensuite comparé aux modèles mathématiques **SIR** et **SIDR**, qui appartiennent à la famille des modèles compartimentaux en épidémiologie, qui reposent sur des équations différentielles.

Key Points

- Modèle python
- Modèle SIR et SIDR
- Comparaisons

Correspondence to

Antonin LEFEVRE
antoninlefevre45@icloud.com

Data Availability

Tous les codes sont disponibles sur [GitHub](#).

Keywords Simulation, Épidémiologie, Modèles à compartiments

TABLE DES MATIÈRES

1. Le modèle python	3
1.1. Simulation avec différents taux d'infection	3
1.2. Simulation avec différentes létalités	5
1.3. Simulation avec différents taux d'immunité	6
1.4. Mise en place d'un confinement	7
1.5. Simuler le SARS-cov-2	8
1.6. Simulation avec différents R_0 et cas limites	9
2. Le modèle SIR	10
2.1. Modélisation mathématique SIR	10
2.2. Comparaison SIR / modèle python	12
3. Le modèle SIDR	12
3.1. Modélisation mathématique SIDR	12
3.2. Comparaison SIDR / modèle python	14
4. Conclusion	15

1.. LE MODÈLE PYTHON

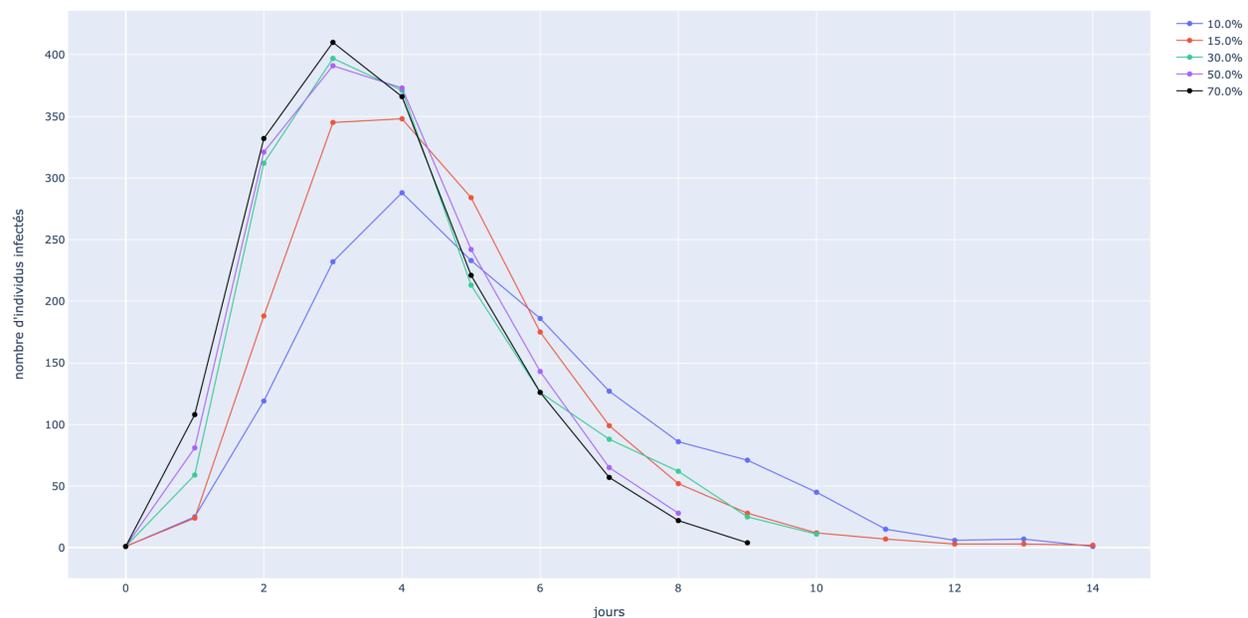
Lors de la simulation, plusieurs paramètres sont en jeu. Il y a tout d'abord le nombre d'individus dans la population, ici on choisira une population entre 1 et 20000 individus pour avoir un temps d'exécution raisonnable. Ce nombre restera constant durant toute la simulation, nous supposons que toute la population est susceptible d'être contaminée, et que la natalité compense la mortalité. La repartition de ces individus sera donnée par le paramètre `variance_population`, et qui variera entre 0 et 2 environ. Plus ce nombre est grand, plus la population sera étalée.

Ensuite, le paramètre `rayon_contamination` correspond à la zone dans laquelle un individu infecté peut transmettre le virus à un individu sain, avec une probabilité nommée `infectiosité`, qui est comprise entre 0 et 1. Chaque individu infecté reste infecté jusqu'à temps qu'il devienne immunisé ou qu'il décède. Après qu'un individu soit infecté, il a une probabilité p de devenir immunisé et donc de ne plus transmettre le virus, et une probabilité d de décéder suite à l'infection. Ces deux paramètres sont compris entre 0 et 1 également. Pour utiliser le paramètre λ du modèle SIR, qui est le taux de retraitement, on fixera $p = 0$ et $d = \lambda$. Nous ne considérerons pas pour l'instant l'apparition d'un vaccin durant la simulation. Seule l'immunité innée des individus est prise en compte. On notera R_0 le taux de reproduction du virus, qui est le nombre moyen d'individus qu'une personne infectieuse infecte tant qu'elle est contagieuse, et qui est égale au rapport de l'infectiosité sur le taux de retraitement (probabilité d'être immunisé ou de décéder).

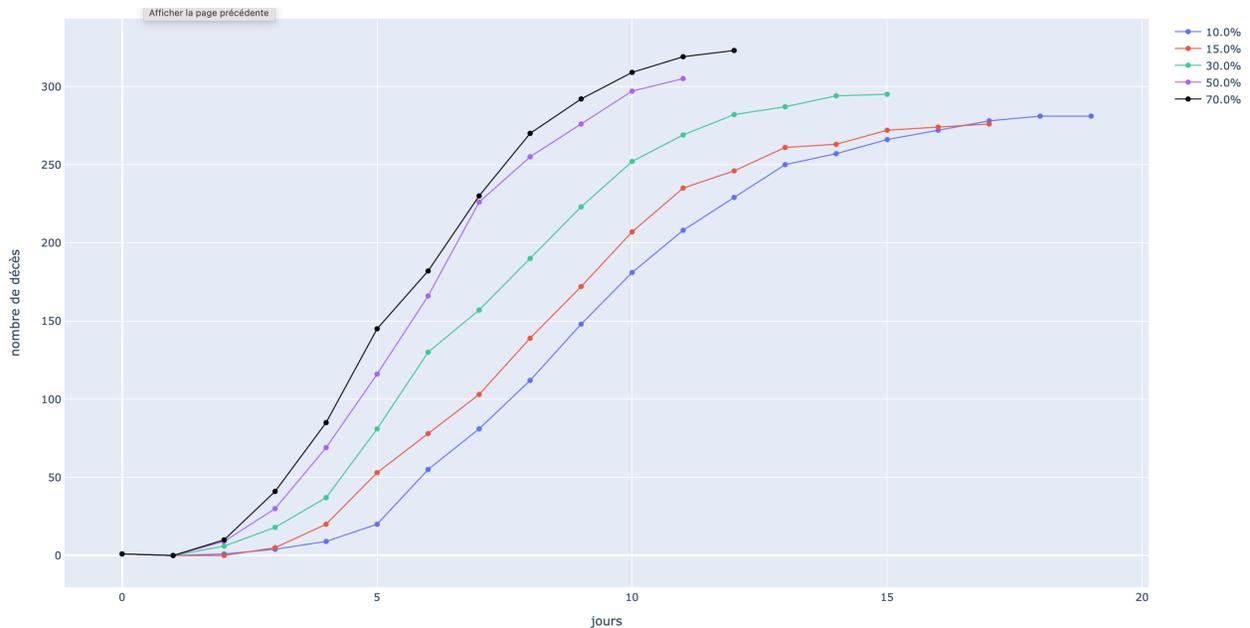
Les simulations de cette section sont réalisées avec le modèle python, et affichées avec la librairie `plotly`

1.1.. Simulation avec différents taux d'infection

Faisons varier le taux d'infection entre 10%, 15%, 30%, 50% et 70% et observons la courbe du nombre de personnes infectés.



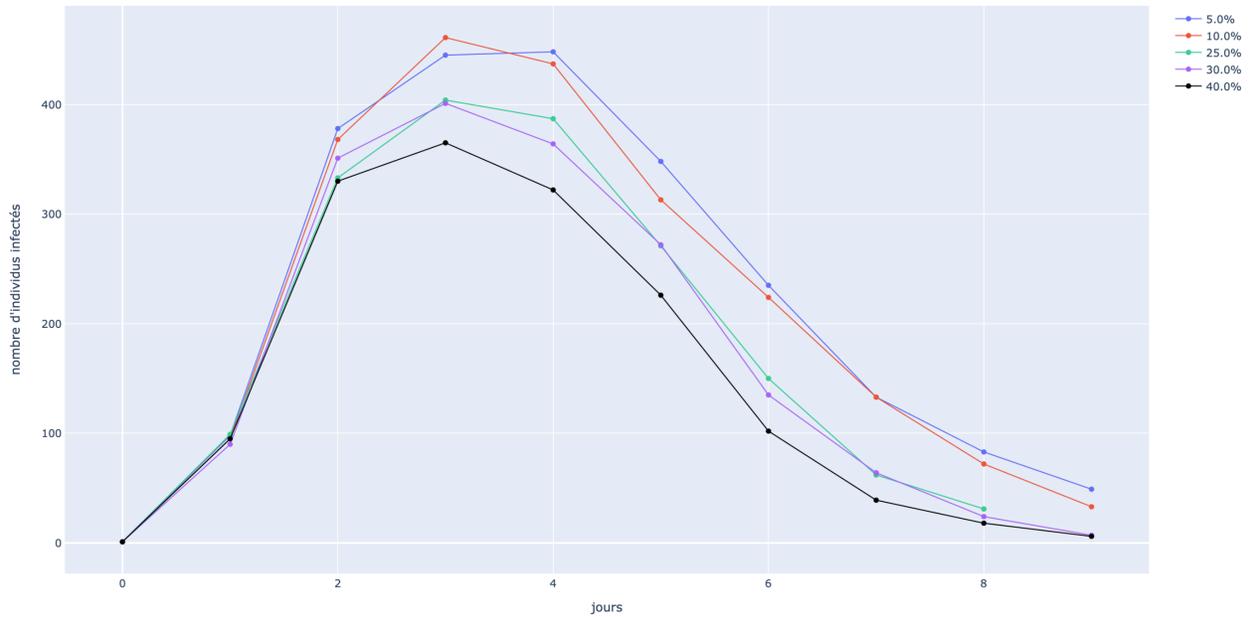
On remarque naturellement que plus l'infectiosité est grande, plus le nombre maximum de personnes infectées est grand. Mais également que le nombre d'infectés chute plus rapidement avec une infectiosité élevée. Intéressons nous maintenant au nombre de décès en fonction de ces différentes valeurs d'infectiosité.



Les courbes sont quasiment linéaires sur une grande partie, jusqu'à atteindre un certain seuil, qui est très proche malgré des infectiosités très différentes. La différence se fait dans la vitesse à laquelle le seuil est atteint, plus l'infectiosité est grande plus le seuil est atteint rapidement, il est atteint en une dizaine de jours pour une infectiosité de l'ordre de 70% et en une quinzaine pour une infectiosité de l'ordre de 10%.

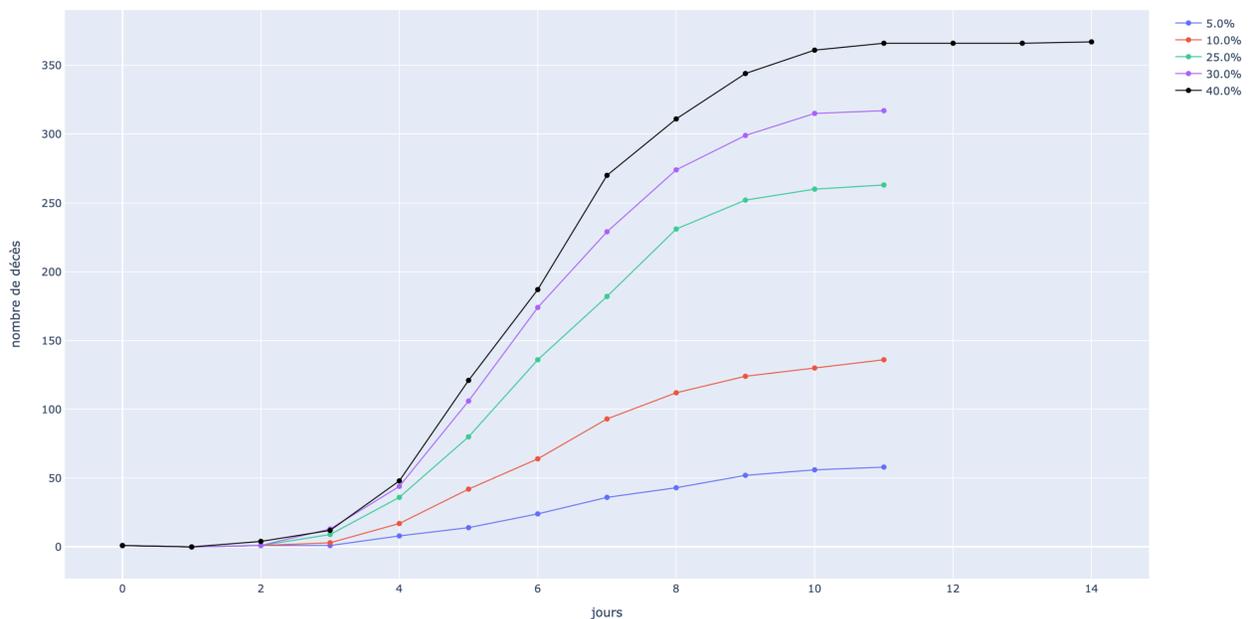
1.2.. Simulation avec différentes létalités

Faisons à présent varier la létalité entre 5%, 10%, 25%, 30% et 40%. Et observons l'évolution des courbes du nombre d'individus infectés.



Ici, une létalité élevée aplatis la courbe des infectés, en hauteur et largeur. En effet, si le virus ne tue pas beaucoup, il y a alors plus de personnes susceptibles de le transmettre, et sur une plus grande période. Au contraire d'une létalité élevée, qui élimine rapidement les individus infectés, avant qu'ils puissent transmettre le virus.

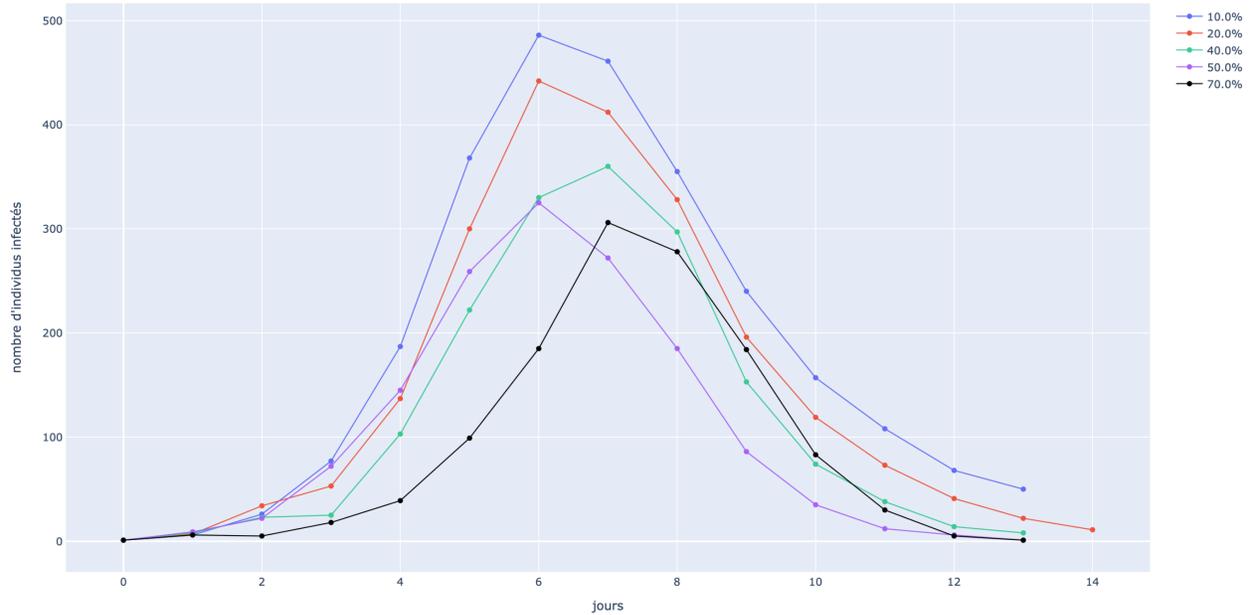
Cependant même si le nombre d'infectés est plus faible avec une grande létalité, le nombre de décès est nettement plus élevé. Comme le montre le graphique ci-dessous.



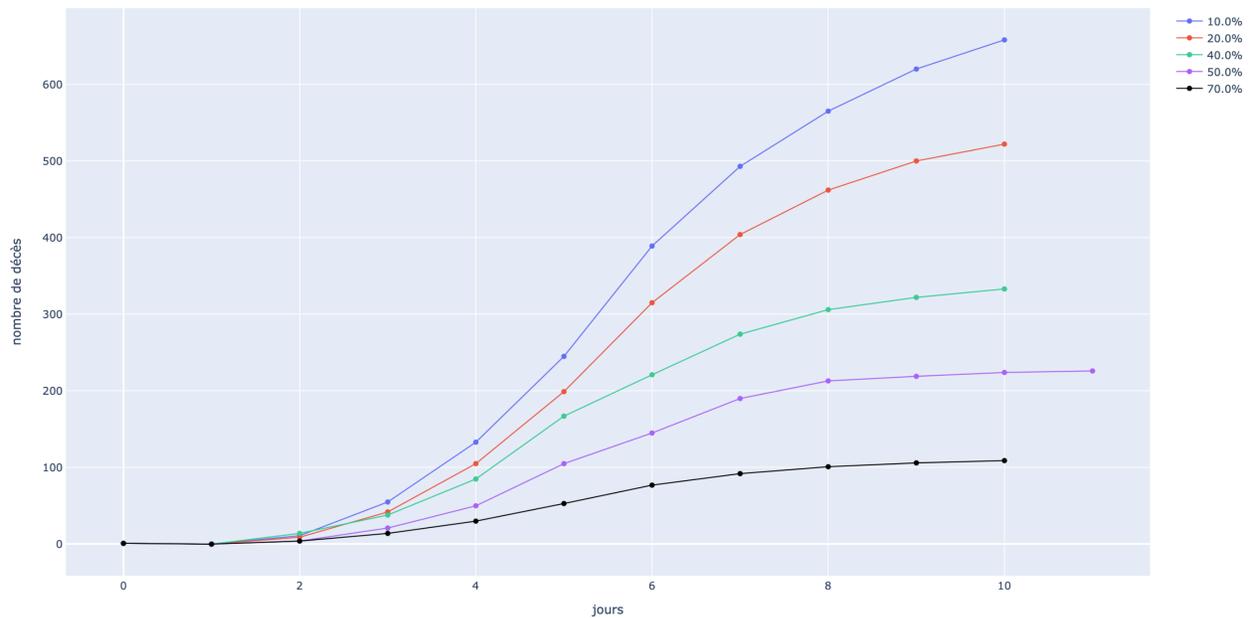
Le nombre de décès avec une létalité de 40% est 7 fois plus grand qu'avec une létalité de 5%. Ce qui représente des différences énormes sur une population de plusieurs millions d'individus.

1.3.. Simulation avec différents taux d'immunité

Enfin, faisons varier le taux d'immunité entre 10%, 20%, 40%, 50% et 70%; et observons les courbes du nombre d'individus infectés.



Sans grande surprise, si le taux d'immunité est très élevé, le maximum de personnes infectés et drastiquement réduit par rapport à un taux d'immunité plus faible, mais il est également atteint plus lentement. La propagation est freinée. On observe cette même tendance si l'on compare maintenant les courbes des décès.



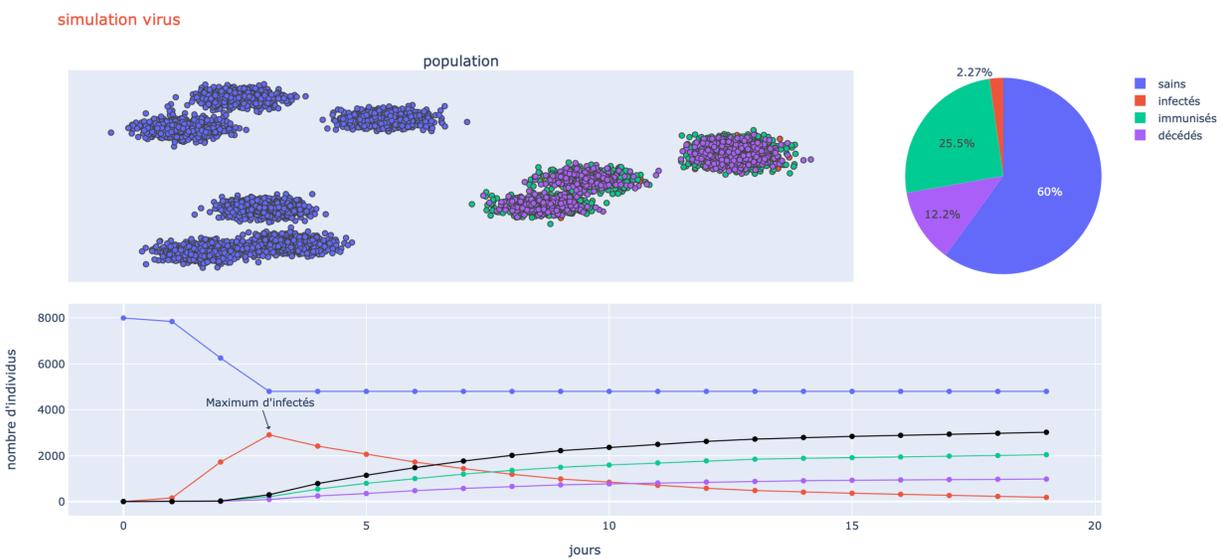
Le nombre de décès est immensément plus grand lorsque l'immunité est quasi inexistante. Avec un taux d'immunité égale à 10%, on a 6 fois plus de décès qu'avec un taux à 70%.

1.4.. *Mise en place d'un confinement*

1.4.1.. *Confinement dès le premier cas détecté:*

On peut simuler un confinement de la population, en diminuant la variance de celle-ci et en augmentant le nombre de cluster, c'est à dire le nombre de groupes distincts physiquement, d'individus dans la population.

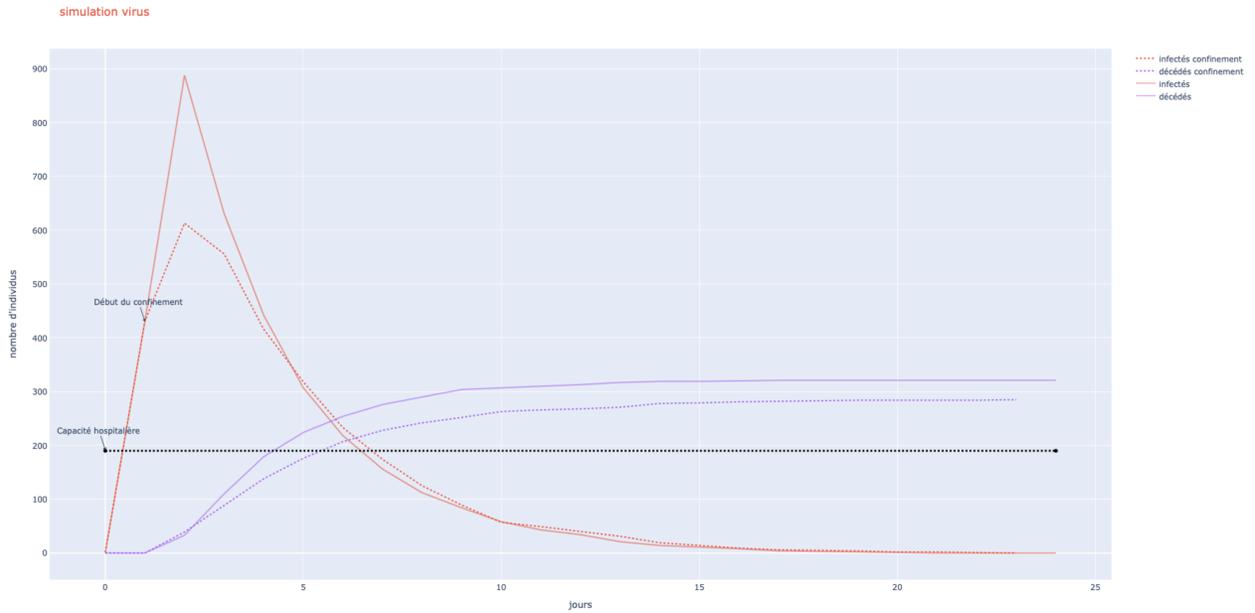
Par exemple avec une variance de 0.6 et 10 centers (modifiable dans la fonction make_blobs), on observe que l'épidémie est globalement sous contrôle, le pic d'infectés est très bas. On compte à la fin, 12% de décès sur l'ensemble de la population.



1.4.2.. *Confinement après dépassement de la capacité hospitalière:*

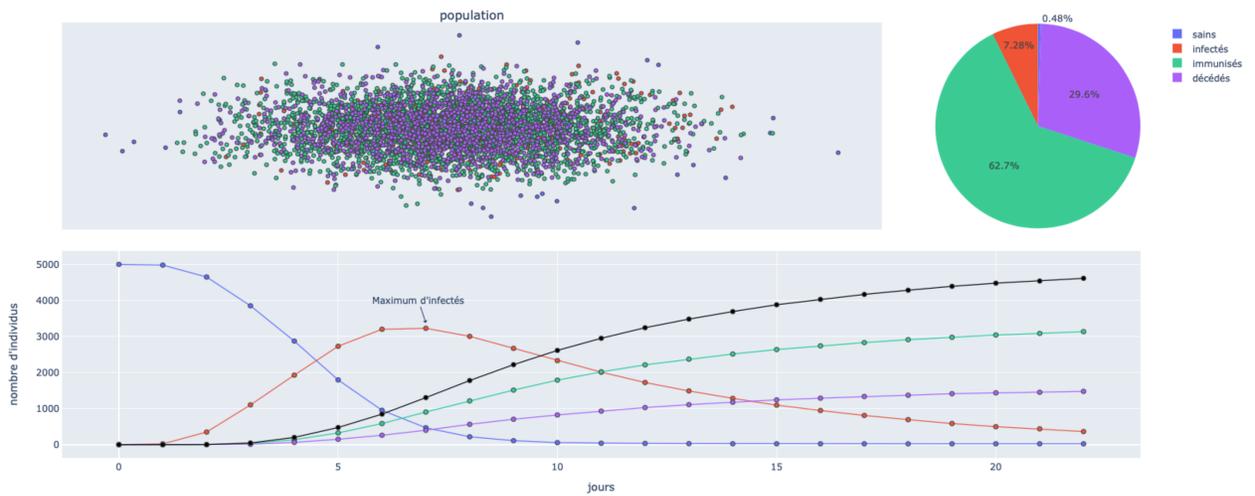
Une fois le seuil dépassé, l'infectiosité est divisé par 8 et le rayon de contamination est divisé par 4.

Ici, la courbe pleine représente l'épidémie sans l'application du confinement, et la courbe en pointillé avec application du confinement à partir du jour ou le nombre de personnes infectés est supérieur à la capacité hospitalière. On remarque que le nombre de personnes infectés diminue quelques jours après le début du confinement, et que le nombre de décès journaliers atteint un seuil inférieur au cas ou il n'y a pas de confinement.



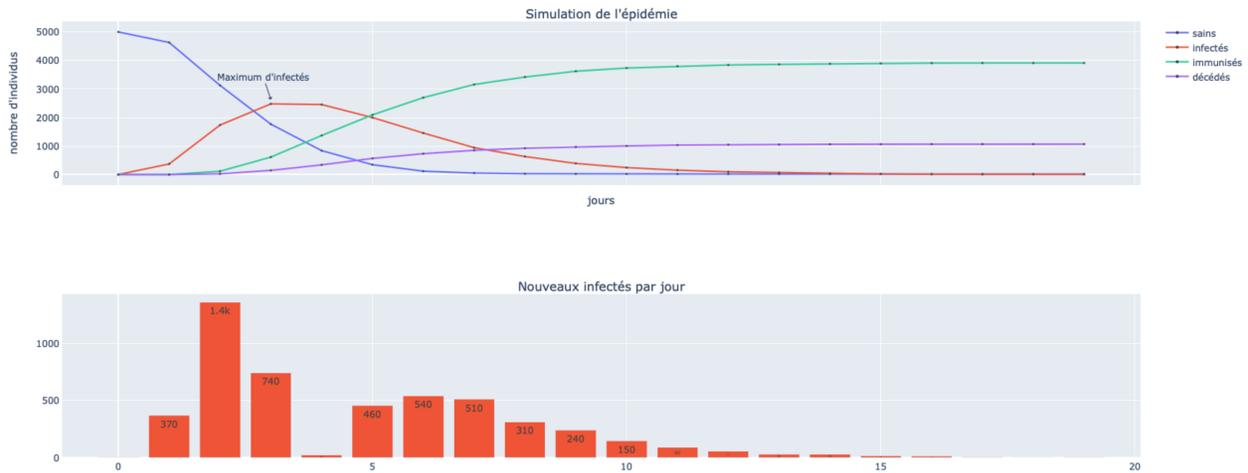
1.5.. Simuler le SARS-cov-2

Cette simulation est à prendre avec beaucoup de précautions, car elle ne reflète pas la réalité. Nous prendrons ici comme paramètres, un taux d'infection de 17%, un taux d'immunité de 10% et une létalité de 0.5%, qui sont des estimations données par les médias en été 2020.

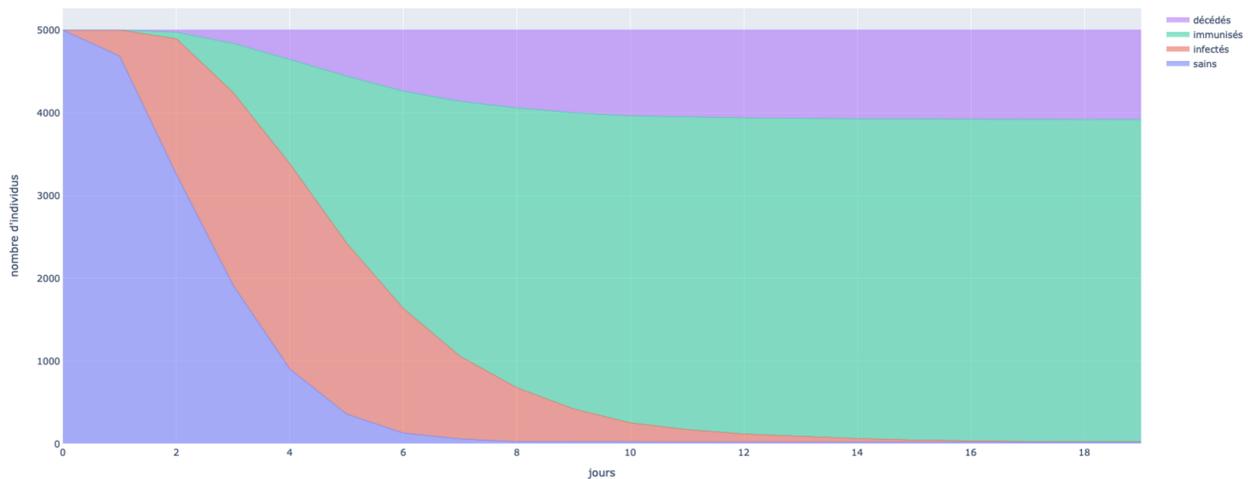


Avec cette simulation on observe que le nombre d'infectés augmente rapidement dès le début pour arriver à son maximum au bout de 7 jours, puis diminue très lentement pendant plus de 20 jours. Au final on compte quasiment 30% de décès, et plus de la moitié deviennent immunisés.

En s'intéressant maintenant aux variations du nombre de nouveaux infectés chaque jour, on retrouve bien le fait que l'augmentation du nombre infectés est plus rapide que la diminution de ce nombre :

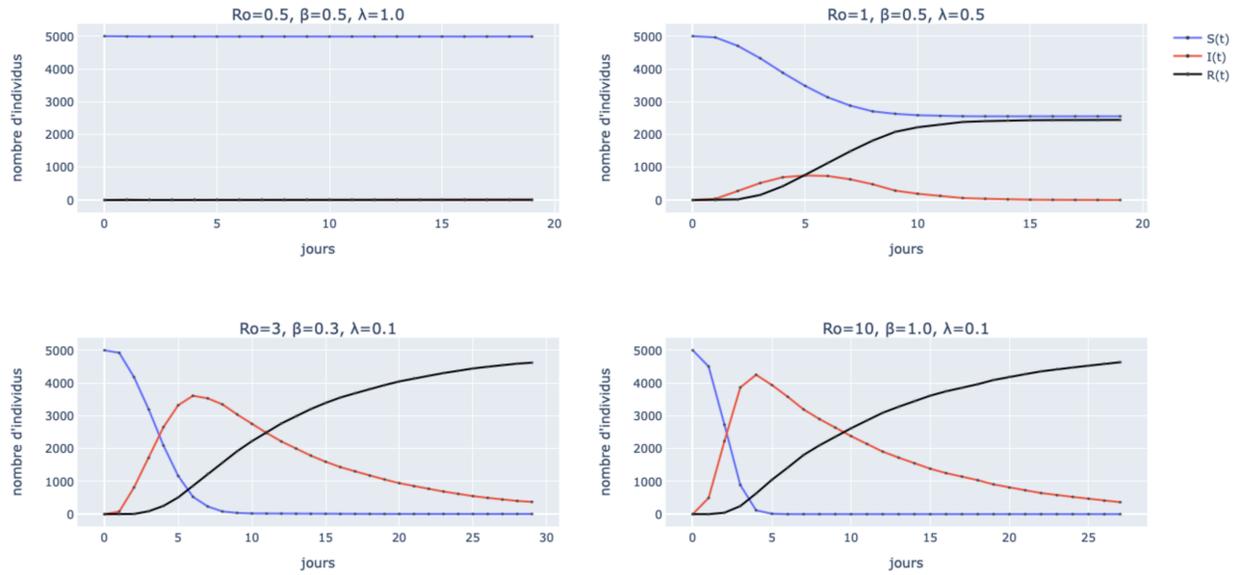


Si on veut maintenant visualiser la taille des catégories de la population :



1.6.. Simulation avec différents R_0 et cas limites

On fait varier ici le taux de reproduction de base R_0 du virus, dont l'expression est $R_0 = \frac{\beta}{\lambda}$, ou ici β est l'infectiosité du virus, et λ la probabilité pour qu'un individu ne puisse plus transmettre le virus (pour se faire, on fixe $p = 0$ et $d = \lambda$). On remarque que la situation est très différente pour un R_0 inférieur et supérieur à 1. Pour un $R_0 < 1$, peu d'individus sont infectés, et le virus ne propage pas. À la frontière $R_0 = 1$, le nombre de personnes retirées atteint quasiment le nombre de personnes saines, mais sans l'atteindre. On parle, pour un $R_0 \leq 1$ d'équilibre. Il faut attendre un $R_0 > 1$ pour que la courbe des individus retirés passe au dessus de celle des individus sains, et provoque ainsi la propagation du virus, qui est d'autant plus importante et rapide que la valeur du R_0 est grande. Simulation réalisée avec le modèle python.



2.. LE MODÈLE SIR

2.1.. Modélisation mathématique SIR

Le modèle SIR est un modèle mathématique déterministe à compartiments permettant de simuler de façon simplifiée l'évolution d'une population face à une épidémie. Ce modèle fait partie d'une famille de modèles compartimentaux en épidémiologie, qui ont commencé à être mis en pratique avec le SIDA dans les années 1980, ils sont à l'heure actuelle de plus en plus présents lors des décisions politiques concernant la gestion de problèmes sanitaires. Dans ce modèle en particulier, la population est divisée en 3 catégories : sains (S), infectés (I) et removed (R) (removed, c'est à dire dans l'incapacité de transmettre le virus). Ce modèle étudie la taille de ces 3 catégories au cours du temps, tel que $S(t)$, $I(t)$ et $R(t)$ soient le nombre d'individus dans chaque catégorie à l'instant t, et ou le nombre d'individu par compartiment est gouverné par un ensemble d'équations différentielles. On peut donc remarquer que, soit N le nombre d'individus dans la population, on a :

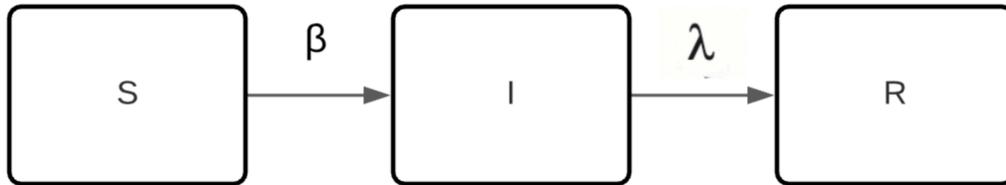
$$\forall t, \quad N = S(t) + I(t) + R(t) \tag{1}$$

On introduit maintenant nos paramètres :

- β : le taux d'infection du virus, qui correspond à la probabilité d'être infecté après avoir été en contact avec un individu infecté
- λ : le taux de retraitement, qui correspond à la probabilité de ne plus pouvoir transmettre le virus, soit après être immunisé soit après le décès

On peut alors définir R_0 le taux de reproduction de base du virus, qui vaut $\frac{\beta}{\lambda}$ et qui représente le nombre moyen de personnes infectées par un seul individu.

Le modèle **SIR** peut donc être représenté par le diagramme suivant :



Puis on introduit nos équations différentielles qui régissent le modèle :

$$\left\{ \begin{array}{l} \frac{dS(t)}{dt} = -\frac{\beta S(t)I(t)}{N} \\ \frac{dI(t)}{dt} = \frac{\beta S(t)I(t)}{N} - \lambda I(t) \\ \frac{dR(t)}{dt} = \lambda I(t) \end{array} \right.$$

On remarque que :

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \tag{2}$$

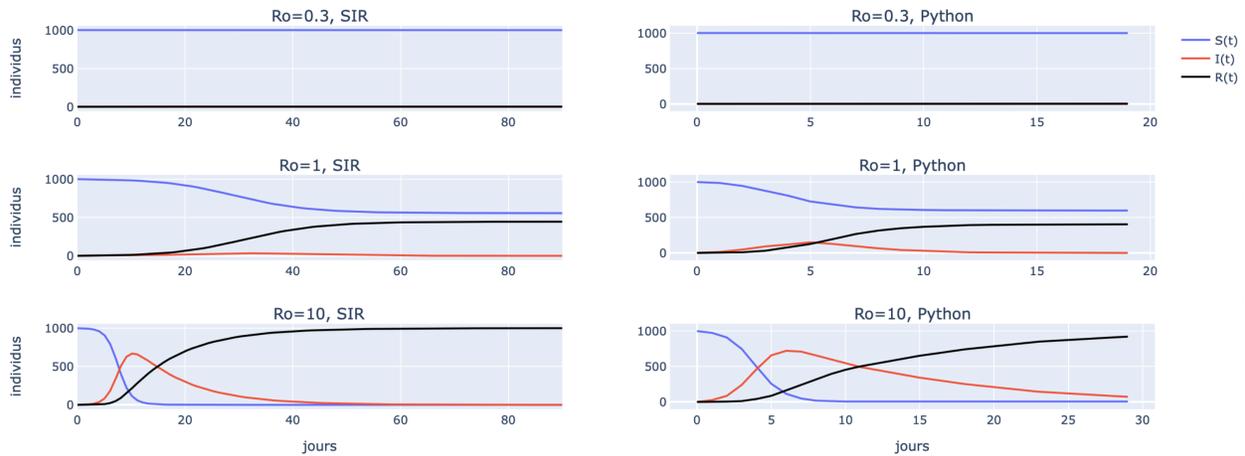
Ce qui confirme le fait que $\forall t, S(t) + I(t) + R(t) = \text{constante} = N$ car la dérivée de la somme est nulle.

La première équation différentielle correspond au nombre de personnes saines, dont le signe est négatif, en effet la fonction S est décroissante pour tout t , car le nombre de personnes saines ne peut que diminuer (conformément au modèle dans lequel on se place, où les personnes infectés ne peuvent redevenir saines). Ensuite, la deuxième équation a le signe de la différence entre le nombre de personnes saines $\beta \frac{S(t)I(t)}{N}$ et retirés $\lambda I(t)$, donc le nombre de personnes infectés diminue sur une période dt quand le nombre de personnes saines est plus grande que le nombre de personnes retirées. La dernière équation, elle, correspond au nombre de personnes retirées qui est simplement le produit du nombre d'individus infectés à l'instant t par le taux de retraitement.

Le vecteur unitaire $[S_0, I_0, R_0]$ correspond aux nombres d'individus sains, infectés et retirés à $t = 0$. Dans ce projet, aucun individu n'est immunisé à $t = 0$ et la valeur de I_0 est 1.

2.2.. Comparaison SIR / modèle python

Comparons maintenant le modèle mathématique SIR et le modèle statistique créé avec python en prenant différentes valeurs de R_0 . Tous les autres paramètres sont identiques.



Les courbes sont assez similaires, peu importe le R_0 . La petite différence au niveau des abscisses est due au fait que les deux modèles n'ont pas la même représentation temporelle. Le modèle créé en python est donc conforme au modèle mathématique.

3.. LE MODÈLE SIDR

3.1.. Modélisation mathématique SIDR

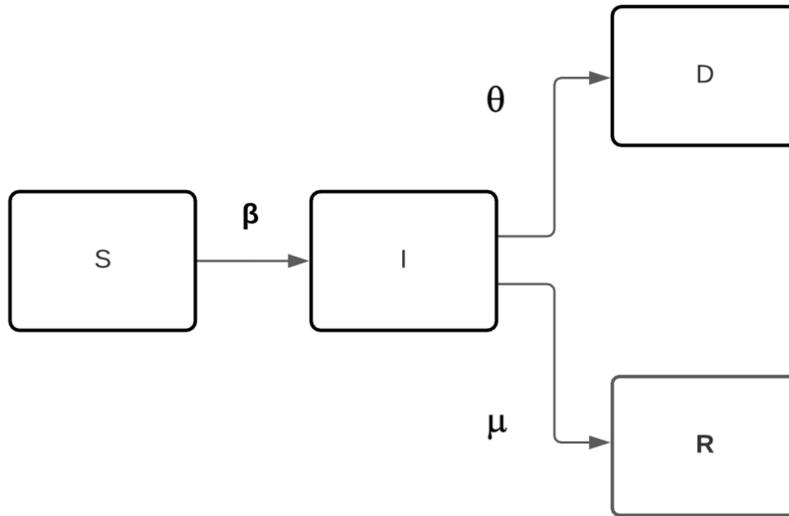
Ce modèle, tout comme le modèle SIR, appartient à la famille des modèles compartimentaux en épidémiologie. Cependant, il est un peu plus précis, car il découpe la population en 4 compartiments. On a alors $\forall t, S(t)$ le nombre d'individus sains, $I(t)$ le nombre d'individus infectés, $D(t)$ le nombre d'individus décédés, et $R(t)$ le nombre de personnes rétablies. Comme précédemment, soit N la taille de la population :

$$\forall t, N = S(t) + I(t) + D(t) + R(t) \tag{3}$$

Les paramètres de ce modèle sont :

- β : le taux d'infection
- θ : le taux de mortalité
- μ : le taux de guérison

On représente alors ce modèle comme ceci :



Les équations qui régissent ce modèle sont :

$$\left\{ \begin{array}{l} \frac{dS(t)}{dt} = -\frac{\beta S(t)I(t)}{N} \\ \frac{dI(t)}{dt} = \frac{\beta S(t)I(t)}{N} - (\mu + \theta)I(t) \\ \frac{dD(t)}{dt} = \theta I(t) \\ \frac{dR(t)}{dt} = \mu I(t) \end{array} \right.$$

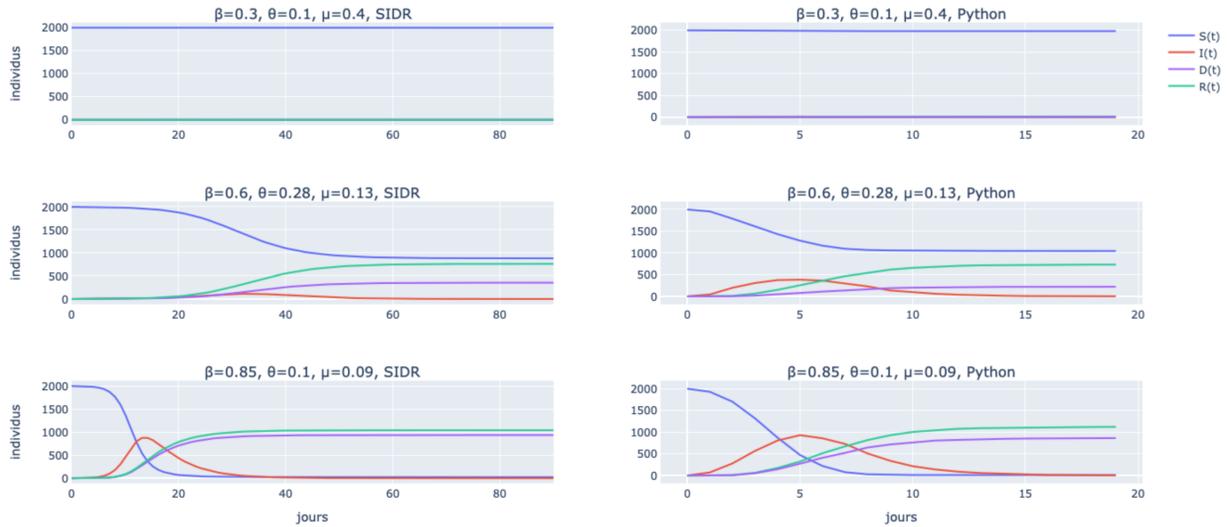
De même, nous avons la somme des équations différentielles nulle, qui justifie que N est constant :

$$\begin{aligned} \frac{dS(t)}{dt} + \frac{dI(t)}{dt} + \frac{dD(t)}{dt} + \frac{dR(t)}{dt} &= -\frac{\beta S(t)I(t)}{N} + \frac{\beta S(t)I(t)}{N} - (\mu + \theta)I(t) + \theta I(t) + \mu I(t) \\ &= -(\mu + \theta)I(t) + \theta I(t) + \mu I(t) \\ &= 0 \end{aligned}$$

L'expression du nombre de personnes saines est le même que le modèle SIR, l'expression du nombre d'individus infectés est la différence entre le nombre d'individus sains et la somme des individus décédés et guéris. Pour le nombre de personnes décédés et guéris, on fait simplement le produit du coefficient correspondant et de la taille de la population infectée.

3.2.. Comparaison SIDR / modèle python

On compare à présent le modèle SIDR avec notre modèle python, pour différents paramètres de l'épidémie, les autres étant identiques.



Les courbes des individus sains, décédés et immunisés sont assez similaires sur ces 3 simulations, cependant la courbe des individus infectés est un peu plus "volatile" dans notre modèle python.

4.. CONCLUSION

L'étude épidémiologique reste au final un domaine très théorique. Même si les résultats et observations obtenues semblent assez proches de la réalité et des modèles mathématiques, ces modélisations ne prennent pas en compte certains autres facteurs, tels que la vaccination, la mise en quarantaine des individus dès qu'ils sont infectés, ou même le déplacement des individus dans la population. Néanmoins, cela nous aide à comprendre comment un virus se comporte dans une population, quel est l'influence de certains paramètres sur sa propagation et comment l'évolution de l'épidémie peut être grossièrement appréhendée.