

TIPE – MCOT

Modélisation du développement d'une épidémie

J'ai choisi ce thème car en m'intéressant au problème durant la crise, j'ai été surpris de voir à quel point la propagation d'un virus est liée à des notions mathématiques. En effet, il est fascinant de voir que l'on peut modéliser une situation assez proche de la réalité avec un faible nombre de variables.

De plus, modéliser ce genre d'épidémie s'inscrit dans le thème au programme étant donné que cette crise relève un nouveau défi mondial, celui de mieux connaître l'impact de nos décisions afin d'être mieux préparés pour une prochaine épidémie.

96 mots

Positionnement thématique :

INFORMATIQUE (Informatique pratique), MATHÉMATIQUES (Mathématiques appliquées).

Mots-Clés :

Français

- Épidémie COVID19
- Étude de population
- Modélisation SEIR
- Propagation d'un virus
- Simulation

Anglais

- COVID19 outbreak
- Population study
- SEIR model
- Spread of a virus
- Simulation

Bibliographie commentée :

Le monde est confronté depuis le début de l'année 2020 à l'épidémie de Covid-19. Après la déclaration de l'état d'urgence de santé publique international le 30 janvier par l'OMS alors que le virus était encore contenu en Chine continentale, le nombre de contaminés a malgré l'avertissement dépassé les 5 millions dans le monde en seulement 5 mois ⁽¹⁾. Aucun pays n'était prêt pour une telle épidémie, si bien que certains prenaient des décisions contraires à celles de leurs voisins. En effet quand beaucoup de pays étaient confinés, d'autres ne faisaient rien pour maintenir une bonne économie en attendant l'immunité de groupe.

Ainsi, depuis le début du 21^{ème} siècle nous en sommes à la 5^{ème} grande flambée infectieuse, et malgré celle d'Ebola précédemment où l'on avait déjà été négligent, la mondialisation continue de favoriser la propagation des virus ⁽²⁾. Une fois cette crise terminée, il faudra en tirer des leçons afin de ne pas refaire certaines erreurs et de pouvoir être mieux préparé à ces épidémies. Pour cela, il est possible dès l'apparition d'un virus de modéliser sa propagation numériquement. En effet, à l'aide de certaines informations sur celui-ci que l'on peut obtenir avant qu'il ne devienne trop inquiétant, on pourrait en le comparant aussi à d'anciennes grandes contagions, en déduire rapidement quelles sont les mesures à prendre contre sa propagation afin d'éviter de futures crises d'abord sociétales mais qui ont aussi un gros impact sur l'économie mondiale ⁽³⁾.

Tout d'abord, le modèle SIR (sain / infecté / rétabli) permet en classant la population dans 3 catégories d'établir un modèle simple d'une situation pourtant complexe. Une personne susceptible n'a pas encore contracté la maladie et peut encore être contaminée avec une probabilité σ lors d'un contact avec une personne infectée. Enfin, une fois la période de guérison moyenne passée qu'on note λ , la personne infectée est rétablie (donc ne transmet plus le virus) et a acquis une immunité qui l'empêche d'être de nouveau infectée ⁽⁴⁾. Ainsi, en modifiant les valeurs de σ , λ et la proportion de personnes infectées au début de la simulation, on peut déjà avoir une certaine évolution d'un virus.

Ensuite, il est possible d'intégrer d'autres données au modèle. En effet, il est possible de rajouter une catégorie de personnes exposées qui a été infectée mais qui n'est pas encore contagieuse pendant une période d'incubation ϵ , avant de passer dans la catégorie infectée (modèle SEIR). De plus, on peut rajouter les taux de mortalité de la maladie, de natalité afin de complexifier encore les équations différentielles et ainsi de rendre le modèle plus réaliste ⁽⁵⁾.

Cependant, ce système n'est pas la seule façon de modéliser le problème. En effet, il est possible de modéliser la population comme des points en mouvement dans un espace donné, et qui pourraient se transmettre le virus lorsqu'ils rentreraient en contact. Ainsi, on peut complexifier le modèle en changeant le comportement de déplacement des points afin d'avoir une situation la plus proche de la réalité possible : regroupements dans les magasins ou les écoles, non-respect du confinement ou alors restrictions de déplacement ⁽⁶⁾.

Comme le virus est normalement incontrôlable avant que l'immunité de groupe fasse effet (dans notre modèle cela veut dire que suffisamment de personnes sont rétablies et donc immunisées pour que le virus arrête de se propager), l'objectif reste de contrôler sa progression pour que le nombre de personnes contaminées reste sous le seuil critique du nombre places en hôpital pour éviter une surcharge de ceux-ci ⁽⁷⁾. Ainsi, une fois développé, notre modèle mathématique permettrait d'aider à prendre les bonnes décisions en ayant une prédiction fiable de quel sera leur impact pour éviter ce problème.

605 mots

(reste à ajouter des différentes problématiques étudiées par les scientifiques)

Problématique retenue :

Il s'agit dans un premier temps de mettre en évidence quels paramètres c'est-à-dire quelles mesures prises contre la propagation du virus sont les plus efficaces dans le modèle puis comparer ce dernier à la réalité afin de savoir comment réagir en cas de nouvelle épidémie.

45 mots

Objectifs du TIPE :

Pour répondre à la problématique, il va falloir tout d'abord réussir à modéliser le plus fidèlement possible la configuration de la population durant la crise en France de sorte à avoir quelque chose de comparable à la réalité. De plus, il faudra que la modélisation soit paramétrable de façon à pouvoir comparer les différentes simulations correspondant au comportement de la population face au virus. Enfin, j'aimerais étudier la vraisemblance / les limites du modèle par rapport à la vraie évolution de la crise.

83 mots

Références bibliographiques :

- 1) WIKIPÉDIA : [Timeline of the COVID-19 pandemic](#)
- 2) OMS : [How the 4 biggest outbreaks since the start of this century shattered some long-standing myths](#)
- 3) OLGA JONAS (2014) : [Menaces sur la santé mondiale du XXI^e siècle](#), Risque pandémique
- 4) FRANÇOIS RECHENMANN (28/11/2011) : [Modéliser la propagation d'une épidémie](#)
- 5) CORENTIN BAYETTE (30/03/2020) : [Modélisation d'une épidémie](#), Partie 1
- 6) CORENTIN BAYETTE (03/04/2020) : [Modélisation d'une épidémie](#), Partie 2
- 7) WIKIPÉDIA : [Flattening the curve](#)

Théorie / Expériences :

Introduction :

Pour effectuer le modèle d'une propagation d'épidémie, il faut d'abord comprendre son fonctionnement dans "la vraie vie", puis faire des hypothèses simplifiant et essayant de rester le plus fidèle possible au comportement réel de cette diffusion. Ainsi, pour que le virus se propage, il faut qu'une personne infectée rencontre une personne saine c'est-à-dire n'ayant jamais contractée la maladie, puisqu'une fois qu'une personne a été infectée par un virus, elle est immunisée (au moins pour une

certaine durée, cela peut dépendre aussi des virus, mais c'est une hypothèse qu'on va faire pour les simulations). La probabilité que de nouvelles personnes se retrouvent infectées au virus dépend donc de 3 paramètres : le nombre de personnes infectées car plus il y en a plus la population saine va en rencontrer ; le nombre de personnes saines, pour la même raison (en effet s'il n'y a plus assez de personnes saines, plus personne ne pourra être infecté) ; et le coefficient de transmission, cela correspond au nombre de personnes que va contaminer une seule personne infectée : plus ce nombre est élevé, plus on dit qu'une maladie est contagieuse ; c'est le R_0 dont on a pu entendre parler durant l'épidémie de covid-19. De plus si ce nombre est inférieur à 1, alors une personne infectée contaminera en moyenne moins d'une personne, donc la propagation du virus reculera. Ensuite, à chaque instant, une partie de la population infectée va guérir et se retrouver rétablie selon une certaine probabilité.

Premier Modèle :

Dans notre premier modèle, à chaque instant une personne contaminée devrait rencontrer un certain nombre M d'autres individus dont S/n sont saines, et à chaque contact, le virus a une probabilité P de se transmettre. D'abord, M dépend de plusieurs choses dont la densité de population, la distance parcourue par les personnes pendant cet instant ou encore le rayon à partir duquel on considère que deux personnes sont entrées en contact (ce sera important pour la suite). Ensuite, S est le nombre d'individus sains et n celui total dans l'étude. On peut donc établir en multipliant par le nombre d'infectés et avec $\sigma = MP$ qu'on appelle le taux d'incidence :

$$\text{Nombre de nouveaux cas} = \text{Nombre d'infectés} * \frac{S}{n} * M * P = \frac{I * S}{n} * MP = \sigma \frac{I * S}{n}$$

De plus, à chaque instant, une certaine proportion λ des infectés guérit, d'où :

$$\text{Nombre de nouveaux rétablis} = \text{Nombre d'infectés} * \lambda = \lambda * I$$

Enfin, en retranchant ou en ajoutant aux bonnes catégories (par exemple on retire le nombre de nouveaux cas du nombre de sains pour l'ajouter à celui des infectés), on obtient un premier système d'équation différentielles correspondant au modèle que l'on vient d'établir :

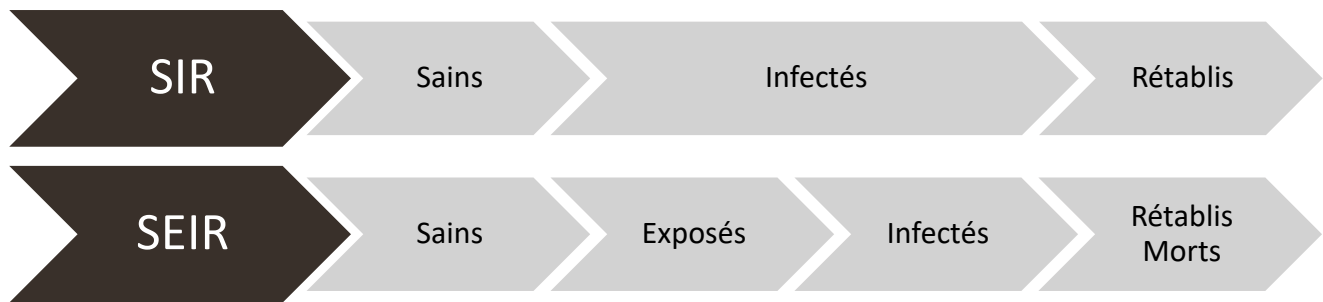
$$\begin{cases} \frac{dS(t)}{dt} = -\sigma * \frac{S(t) * I(t)}{n} \\ \frac{dI(t)}{dt} = \sigma * \frac{S(t) * I(t)}{n} - \lambda * I(t) \\ \frac{dR(t)}{dt} = \lambda * I(t) \end{cases}$$

Remarques :

- Je n'ai pas utilisé les lettres grecques utilisées conventionnellement pour un modèle SIR ou SEIR, car comme je rajoute des catégories d'une façon similaire après, j'ai préféré changer ces lettres pour mieux m'y retrouver lors de la programmation.
- Ce type de système est très compliqué à résoudre « à la main » (peut seulement se "simplifier" en équations différentielles non linéaires), c'est pourquoi on le résoudra via la méthode d'Euler en Python
- Ici, le R_0 évoqué plus haut dépend de deux paramètres, σ et λ . En effet, σ correspond au produit du taux de contact avec la probabilité de contagion et λ la durée de guérison en moyenne d'une personne infectée, donc on a $R_0 = \lambda * \sigma$

Deuxième Modèle :

Ce deuxième modèle va reprendre la base du premier mais en essayant de corriger certaines lacunes pour se montrer plus proche de la réalité en ne changeant pas grand-chose. En effet, on va ajouter deux nouvelles catégories fonctionnant exactement comme les précédentes : la première correspondant à une catégorie intermédiaire de personnes exposées, c'est-à-dire qu'elles ont « attrapé le virus », mais que celui-ci ne s'est pas encore développé pour se transmettre à d'autres personnes ou encore provoquer certains symptômes chez l'hôte. La seconde est une catégorie parallèle aux rétablis, celle des personnes décédées à cause du virus ; à chaque instant, une proportion μ des infectés meurt de l'infection.



$$\begin{cases} \frac{dS(t)}{dt} = -\sigma * \frac{S(t) * I(t)}{n} \\ \frac{dE(t)}{dt} = \sigma * \frac{S(t) * I(t)}{n} - \varepsilon * E(t) \\ \frac{dI(t)}{dt} = \varepsilon * E(t) - \lambda * I(t) - \mu * I(t) \\ \frac{dR(t)}{dt} = \lambda * I(t) \\ \frac{dM(t)}{dt} = \mu * I(t) \end{cases}$$

Ainsi, on peut établir le deuxième modèle nommé SEIR, plus précis que le modèle SIR

Remarques :

- En épidémiologie, le modèle SEIR prend en compte l'évolution de la démographie, c'est-à-dire les taux de natalité et de mortalité naturelle de la population. Dans notre modèle, on fait l'approximation que la durée de l'épidémie est suffisamment courte pour pouvoir négliger ces deux taux.

Modèle Complet :

Désormais, on sait complexifier notre modèle en ajoutant des catégories afin de s'approcher d'une situation réelle. Ainsi, dans le cas du covid-19 et dans la modélisation qu'on fait :

$$\begin{cases} \frac{dS(t)}{dt} = -\sigma * S(t) * \frac{I(t) + C(t) + A(t)}{n} \\ \frac{dE(t)}{dt} = \sigma * S(t) * \frac{I(t) + C(t) + A(t)}{n} - \varepsilon * E(t) \\ \frac{dC(t)}{dt} = \varepsilon * E(t) - \gamma * C(t) \\ \frac{dI(t)}{dt} = (1 - \alpha)\gamma * C(t) - (\lambda + \chi + \mu) * I(t) \\ \frac{dA(t)}{dt} = \alpha\gamma * C(t) - \lambda * A(t) \\ \frac{dQ(t)}{dt} = \chi * I(t) - (\lambda + \mu) * Q(t) \\ \frac{dR(t)}{dt} = \lambda * (I(t) + A(t) + Q(t)) \\ \frac{dM(t)}{dt} = \mu * (I(t) + Q(t)) \end{cases}$$

- Il existe une période où la personne contaminée sera contagieuse sans avoir les symptômes : c'est la catégorie C qui se place entre les exposés et les infectés.
- Il existe des personnes asymptomatiques qui comme les contagieux n'ont pas de symptômes mais peuvent transmettre le virus : c'est une catégorie parallèle à celle des infectés avec le même taux de guérison ; les personnes vont dans l'une ou dans l'autre selon le paramètre α .
- A chaque instant, les personnes infectées ont une probabilité χ d'être identifiées comme telles et d'être mises en quarantaine afin de limiter la propagation du virus : elles n'interagissent plus avec le reste de la population

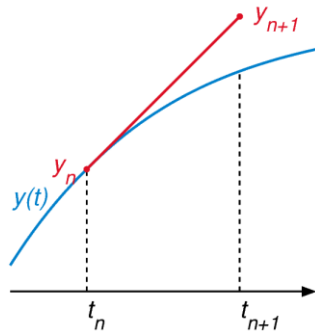
Enfin, on pourra décider d'appliquer un confinement à partir d'un certain niveau de personnes infectées (seuil début confinement noté SDC) et qui s'arrêtera en dessous d'un seuil bas (seuil fin confinement noté SFC). On pourra aussi décider de la proportion (PPC) de la population qui respectera ce confinement qui s'appliquera équitablement sur toutes les catégories de population. Ainsi, seulement $PPC * I$ infectés et seulement $PPC * S$ sains pourront interagir entre eux, donc lors d'un confinement on a une équation différentielle de d'habitude :

$$\text{Nombre de nouveaux cas} = \sigma * \frac{\text{Sains} * PPC * \text{Infectés} * PPC}{n} = (\sigma * PPC^2) \frac{S * I}{n}$$

De cette manière, on peut établir que le taux d'incidence lors du confinement égal à $\sigma * PPC^2$, où σ est celui sans confinement

Ce modèle final n'étant toujours pas résoluble à la main (surtout si on ajoute le confinement conditionnel), on va donc le résoudre numériquement sur Python avec la méthode d'Euler.

Méthode d'Euler :



D'après la formule de Taylor-Young à l'ordre 1, pour une fonction dérivable f , on a quand Δt tend vers 0 : $f(t + \Delta t) = f(t) + \Delta t * f'(t)$. Ainsi, lorsqu'on a une relation entre f et sa dérivée, c'est-à-dire une équation différentielle du premier ordre telle que $f'(t) + a(t) * f(t) = b(t)$, on peut utiliser la formule précédente pour résoudre le problème : c'est la méthode d'Euler qu'on pourra appliquer en résolution numérique sur Python pour avoir un résultat très proche de celui souhaité si l'on prend Δt faible devant la durée de l'étude.

Conclusion sur la théorie :

Avec ces systèmes d'équations, il est aisé d'ajouter plus de catégories pour s'approcher encore plus d'une situation réelle ; plus l'on fera des hypothèses se rapprochant de la réalité, plus on aura des résultats vraisemblables. Cependant, si l'on prend l'exemple de la France, à l'instar des deux confinements de 2020, les taux sont constamment modifiés par le contexte actuel : le premier confinement fut déclaré aux alentours de 1000 nouveaux cas alors que le deuxième fut annoncé à 35000 nouveaux cas. Ainsi, à cause de nombreux paramètres extérieurs à l'épidémie (notamment sociaux et économiques), il est quasi impossible même en ajoutant de plus en plus d'hypothèses de prévoir l'évolution d'une épidémie aussi complexe au-delà d'un futur proche à l'aide de ce modèle.

Modélisation informatique :

J'ai réalisé deux modèles différents de la propagation d'une épidémie. Dans ces deux modèles, il est défini plusieurs états que l'on peut avoir par rapport à l'épidémie :

- Sain → personne saine pouvant être contaminée
- Exposé → personne qui vient d'être exposée au virus, pas encore contagieuse et qui n'a pas encore de symptômes
- Contagieux → personne contagieuse mais qui n'a pas encore de symptômes
- Infecté → personne infectée (contagieuse et symptômes)
- Asymptomatique → personne infectée mais sans symptômes
- Rétabli → personne rétablie, désormais immunisée
- Mort → personne décédée à cause de la maladie
- Isolé → personne infectée identifiée et que l'on a pu isoler
- Confiné → personne confinée quelque soit son état (modèle physique uniquement)

Le premier modèle est mathématique : il correspond à la résolution des équations différentielles vues dans la partie théorique grâce à la méthode d'Euler. La simulation prend en paramètres tous les taux vus dans la partie théorique, en plus des seuils haut et bas du confinement et sa proportion de la population le respectant associée.

Le second est un modèle "physique" de la situation, c'est-à-dire qu'il simule le déplacement de personnes représentées par des points dans un espace fermé à deux dimensions. Ces points placés aléatoirement ont une direction déterminée à l'initialisation du modèle, puis se déplacent selon celle-ci à une certaine vitesse pendant toute la durée de l'étude ; s'ils rencontrent un des bords de l'espace, ils "rebondissent" sur celui-ci (c'est-à-dire que sa direction est inversée par rapport à la normale de la bordure rencontrée).

Avant de lancer une simulation, on contamine un certain nombre de patients 0 de l'épidémie, puis lors des déplacements, si une personne contagieuse (soit appartenant au groupe infectés, asymptomatiques ou rétablis) entre en contact d'une personne saine (dans un certain rayon autour de celle-ci), cette dernière aura une probabilité σ d'être exposée. Ensuite, après des périodes ε et γ , les personnes deviennent respectivement contagieuses puis asymptomatique avec une probabilité α , sinon infectée. Enfin, ces deux derniers groupes finissent par se rétablir après une durée λ bien que les infectés (seul groupe avec des symptômes) ont une probabilité μ de mourir à chaque instant.

De plus, le programme permet de passer en argument les seuils haut et bas du confinement (mettre > 1 en seuil haut pour ne pas confiner) et la proportion de la population infectée qui sera identifiée puis isolée pour éviter trop de contamination. Pour la première mesure, on arrêtera simplement le déplacement de la population respectueuse du confinement (et on empêchera la contamination par l'une de ces personnes), tandis que pour la deuxième, on sortira les personnes infectées de l'étude le temps qu'ils se rétablissent.

Pour visualiser les résultats des simulations, on peut faire une animation de la propagation du virus pour le premier modèle, avec l'affichage du nombre de cas en parallèle (avec la multiplication des catégories et le nombre de points importants qu'il faut bouger à chaque image, celle-ci n'est plus intéressante : pour vraiment voir la propagation, il faut enregistrer chaque image à part puis tout mettre bout à bout dans un logiciel de montage vidéo). Pour les deux modèles, on peut aussi afficher les courbes d'évolutions des cas en fonction du temps pendant la simulation sur un graphique.

Enfin, ces modèles ont de grosses différences sur la complexité de leur code mais aussi sur leurs avantages et leurs défauts. En effet le premier permet une grande modularité : on pourrait modifier la façon dont ils bougent pour modéliser des comportements très complexes comme le rassemblement dans un magasin malgré le confinement, ou le manque de respect des gestes barrières dans certains lieux de regroupement, contrairement au second qui ne s'occupe pas des positions des personnes. Ainsi, on peut augmenter la précision du modèle "physique" en modifiant les déplacements, alors que le deuxième est limité de ce côté. Cependant, ce dernier est beaucoup plus rapide à coder (une fois que l'on acquies la partie théorique), mais aussi bien plus rapide à simuler pour l'ordinateur car de complexité linéaire en la durée de la simulation (et ne dépend pas de la taille de la population) : les opérations seront toujours les mêmes dans la fonction propagation quelle que soit la taille des différents groupes, et celle-ci seront répétées autant de fois que la simulation dure ; ce n'est pas le cas pour le second : il est long à modéliser car il faut réfléchir à beaucoup de comportements que va avoir la population et toutes les opérations et tous les tests effectués pour déplacer et faire changer d'état chaque personnes une par une augmente largement le temps de calcul qui devient très long dès 1000 individus.

DOT :

- 1) Étude théorique de la propagation d'une épidémie et compréhension des modèles SIR et SEIR.
- 2) Programmation de la résolution du système différentiel par la méthode d'Euler.
- 3) Programmation du modèle discret.
- 4) Détermination des liens entre les paramètres des deux modèles.
- 5) Complétion des deux modèles avec de nouvelles catégories et paramètres
- 6) Étude empirique de l'influence des paramètres sur l'évolution d'une épidémie et comparaison des deux modèles.