

Examen de Mathématiques et Physique en Interaction Corrigé et commentaires

Durée : 1 heure 30 minutes

Tous les documents sont interdits. Le matériel électronique est interdit, à l'exception d'une montre à affichage digital. Merci d'éteindre et de ranger dans vos sacs vos téléphones et autres objets connectés.

On s'intéresse à la modélisation de la propagation d'une épidémie qui affecte les grenouilles de l'Yvette. Ce problème est constitué de deux parties complémentaires et indépendantes (à l'exception de la question II.5). Il faut traiter les deux parties.

Dans la **Partie I**, on étudie un modèle continu (à l'échelle d'une population), déterministe, qui modélise une flambée épidémique où une proportion non-négligeable de la population est infectée.

Dans la **Partie II**, on étudie un modèle discret (à l'échelle des individus), probabiliste, qui modélise le début possible d'une épidémie.

Les questions ne sont pas classées par ordre de difficulté croissante ; vous pouvez admettre le résultat d'une question pour traiter la suite du sujet.

Partie I

On introduit le modèle SIR de propagation d'une épidémie. Dans ce modèle, les individus sont soit susceptibles (S), soit infectieux (I) soit écartés (R). Les proportions relatives des individus susceptibles, infectés, écartés à un instant $t \in \mathbb{R}$ sont notées respectivement $s(t) \in [0, 1]$, $i(t) \in [0, 1]$, $r(t) \in [0, 1]$, et on a en permanence $s(t) + i(t) + r(t) = 1$.

Dans ce modèle, les individus susceptibles sont infectés à une vitesse qui dépend à la fois des proportions d'individus susceptibles et infectieux (si personne n'est infectieux, ou si personne n'est susceptible, alors l'épidémie ne peut pas progresser). Par ailleurs l'infection dure un certain temps, et les individus infectés deviennent progressivement écartés (par guérison ou par mort). On obtient le système d'équations différentielles suivant :

$$\begin{cases} \frac{ds}{dt} = -R_0 is \\ \frac{di}{dt} = R_0 is - i \\ \frac{dr}{dt} = i, \end{cases} \quad (\text{SIR})$$

où R_0 , le *taux de reproduction*, est une constante. On supposera toujours $R_0 > 1$. La configuration initiale à $t = 0$ est un triplet quelconque (s_0, i_0, r_0) de réels positifs tels que $s_0 + i_0 + r_0 = 1$.

Question I.1: Montrer que les seuls points d'équilibre de cette dynamique sont les configurations où aucun individu n'est infecté : $\{(s, 0, 1-s), s \in [0, 1]\}$.

Corrigé: Par définition, dans cette dynamique à temps continu, un point d'équilibre satisfait

$$\frac{ds}{dt} = \frac{di}{dt} = \frac{dr}{dt} = 0.$$

Or

$$\frac{dr}{dt} = 0 \Leftrightarrow i = 0,$$

donc un point d'équilibre vérifie nécessairement $i = 0$; réciproquement si $i = 0$ on vérifie que $\frac{ds}{dt} = \frac{di}{dt} = 0$. Finalement, les points d'équilibre sont les triplets (s, i, r) de réels positifs vérifiant $i = 0$, c'est-à-dire les points $(s, 0, 1-s)$ où $s \in]0, 1]$.

Commentaire: Un nombre important de copies ont tenté de résoudre explicitement cette équation différentielle, ce qui n'est pas possible sans introduire la fonction spéciale W de Lambert. Parmi les démarches malheureuses, on citera celle consistant à prétendre que $s(t) = s(0) \exp(-R_0 it)$, ce qui n'est pas possible puisque i est constante. Il est vrai que $s(t) = s(0) \exp(-R_0 I(t))$ où I est la primitive de i qui s'annule en 0, mais ça ne servait pas à grand chose, ni dans cette question ni dans la suite du problème.

On cherche à étudier la stabilité de ces points d'équilibre.

Question I.2: Montrer que, en plus de $s(t) + i(t) + r(t)$, il y a une autre quantité conservée par l'équation différentielle (SIR):

$$H = R_0(i + s) - \ln(s).$$

On pourra utiliser à profit la formule

$$\frac{dH}{dt} = \frac{dH}{di} \frac{di}{dt} + \frac{dH}{ds} \frac{ds}{dt}.$$

Corrigé: On a

$$\frac{dH}{di} = R_0 \quad \frac{dH}{ds} = R_0 - \frac{1}{s}$$

et donc

$$\begin{aligned} \frac{dH}{dt} &= R_0(R_0is - i) - (R_0 - \frac{1}{s})(R_0is) \\ &= R_0^2is - R_0i - R_0^2is + R_0i \\ &= 0; \end{aligned}$$

c'est la définition d'une quantité conservée.

Commentaire: Cette question était une variante de l'objet du TD1, qui visait à montrer que l'énergie totale était conservée dans un système mécanique sans frottement.

Question I.3: On considère n'importe quelle trajectoire $s(t), i(t), r(t)$ vérifiant l'équation (SIR). Montrer que s est une fonction décroissante, et en déduire qu'elle converge quand $t \rightarrow +\infty$ et quand $t \rightarrow -\infty$. Montrer aussi que r est une fonction croissante, et en déduire qu'elle converge également. Montrer finalement que i converge.

Corrigé: Puisque i, s, R_0 sont trois quantités positives, $\frac{ds}{dt}$ est négatif, et donc s est décroissante sur \mathbb{R} . Puisque $s \in [0, 1]$ par hypothèse de modélisation, elle est bornée, et on en déduit que s converge en $\pm\infty$.

De la même manière $\frac{dr}{dt}$ est positif, donc r est une fonction croissante du temps, et appartient à $[0, 1]$; ainsi r converge en $\pm\infty$.

Finalement, puisque $i = 1 - s - r$, on en déduit que i converge également en $\pm\infty$.

Commentaire: Attention à ne pas prétendre que, puisque s est décroissante sur \mathbb{R} et appartient à $[0, 1]$, alors elle tend vers 0 en $+\infty$!

Question I.4: Étudier la fonction

$$s \mapsto R_0s - \ln(s)$$

sur l'intervalle $]0, 1]$. Montrer en particulier que les courbes $\{H = \text{constante}\}$ intersectent le lieu d'équilibre $\{i = 0\}$, soit en deux points, soit en un seul point.

Corrigé: Notons h la fonction qui à s associe $R_0s - \ln(s)$. Sa dérivée est $h'(s) = R_0 - \frac{1}{s}$, et on peut dresser le tableau de variation de h : elle tend vers $+\infty$ en 0, est décroissante sur $]0, \frac{1}{R_0}]$, vaut $1 + \ln(R_0)$ en $\frac{1}{R_0}$, puis est croissante sur $[\frac{1}{R_0}, 1]$, et vaut R_0 en 1.

On remarque que $h(s) = H(s, 0)$, autrement dit h est la restriction de H à $\{i = 0\}$. Dès lors, pour tout $C \in \mathbb{R}$, on a

$$\{H = C\} \cap \{i = 0\} = \{h = C\},$$

et, vu le tableau de variation de h , on peut conclure: $\{H = C\}$ intersecte $\{s = 0\}$

- zéro fois si $C < 1 + \ln(R_0)$,

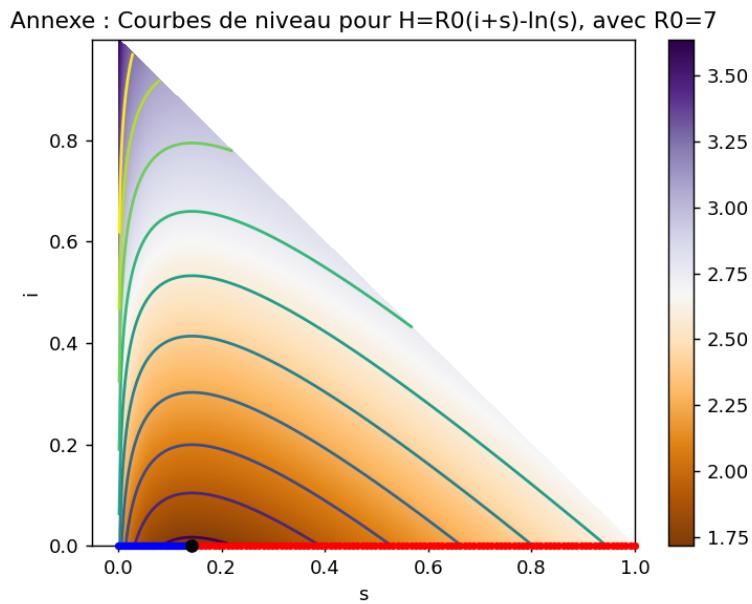
- une fois si $C = 1 + \ln(R_0)$,
- deux fois si $1 + \ln(R_0) < C \leq R_0$,
- et à nouveau une fois si $C > R_0$.

La fonction H étant croissante par rapport à i , on sait que $H(i, s) \leq h(s) \leq 1 + \ln(R_0)$ pour tout i, s , et donc que $\{H = C\}$ est vide si $C < 1 + \ln(R_0)$; finalement la première condition n'arrive jamais, et les courbes $\{H = C\}$ intersectent $\{i = 0\}$ soit en un point soit en deux points.

Commentaire: Curieusement, cette question, relativement indépendante du reste du sujet et accessible sans avoir appris son cours, a été très mal réussie.

Question I.5: Dans l'annexe (à rendre avec la copie) on a dessiné les courbes de niveau de la fonction H dans le plan (s, i) , le long desquelles le système évolue. Rajouter des flèches pour préciser dans quel sens évolue la dynamique, et préciser la valeur du point marqué. Représenter les points d'équilibre instables et les points d'équilibre stables (n'oubliez pas de légendier).

Corrigé:



Les flèches vont vers la gauche, sur chacune des courbes : s est une fonction décroissante du temps ! Il suit que, pour toute configuration initiale proche d'un des points d'équilibre bleu (ou le point noir en $\frac{1}{R_0}$), la trajectoire va tendre, en temps positif, vers un point d'équilibre proche de ce point d'équilibre. A l'inverse, les points d'équilibre représentés en rouge sont instables : pour toute configuration initiale proche, mais pas sur l'axe $\{i = 0\}$, on va avoir une vraie flambée épidémique (en suivant l'une des courbes de niveau de H de la gauche vers la droite) pour finir en l'un des points bleus.

Commentaire: cette question cristallisait le degré de compréhension du modèle.

Description complète du modèle: Le modèle *SIR* est le modèle le plus simple pour représenter la propagation d'une épidémie. Dans ce modèle, une immunité de masse se forme en $s = \frac{1}{R_0}$: tant que s est trop grand (s'il y a trop d'individus susceptibles), l'épidémie va progresser, et il va y avoir de plus en plus d'individus infectés (d'après l'allure des courbes $\{H = C\}$). Une fois que $s < \frac{1}{R_0}$, il n'y a plus assez d'individus susceptibles, et i commence à décroître.

Dans toute épidémie, il y a un pourcentage d'individus immunisés, soit pour des raisons génétiques (par exemple, chez l'humain, certaines mutations des gènes codant pour les protéines CCR5, CD4, ou le système HLA, assurent un degré d'immunité face à la variole, l'hépatite B, la peste bubonique, et/ou le SIDA), soit par mémoire immunitaire, après une précédente infection ou une vaccination ; au début d'une vague potentielle d'épidémie, on n'est pas forcément à $s = 1, i = 0, r = 0$. Pour lutter efficacement contre une épidémie, dans ce modèle, il faut réduire s (et la vaccination assure le meilleur rapport bénéfice/risque pour le faire), et réduire R_0 , c'est-à-dire limiter la faculté de l'épidémie à se transmettre une fois qu'elle a infecté un individu. Ce sont les mesures prophylactiques, encore appelées gestes barrière : d'une part les mesures d'ordre général telles que tests, traçage, isolement, quarantaine, et d'autre part des mesures moins contraignantes mais adaptées au mode de transmission d'une maladie donnée comme le port d'une protection jetable (masque, gants, préservatif, ...), aération, désinfection, traitement des eaux usées, et ainsi de suite.

Les limitations du modèle *SIR* sont multiples: d'abord, pour certaines maladies comme le Covid-19, l'immunité acquise après l'infection ne dure qu'un certain temps, alors qu'ici on a supposé que les personnes qui avaient été infectées une fois étaient écarté du modèle "pour toujours". En général, à moins que la flambée épidémique soit extrêmement rapide, il faut également prendre en compte le cycle de vie naturel des populations : naissance de nouveaux individus peu ou pas immunisés, et mort d'individus pour d'autres raisons. On peut corriger le modèle *SIR* pour prendre ces effets en compte, notamment en rajoutant des termes dans l'équation différentielle.

D'autres difficultés sont plus inhérentes à une modélisation par une équation différentielle : le modèle *SIR* suppose une population homogène, or en général différentes catégories d'individus ont une immunité préalable différente, ou un R_0 effectif différent. Par ailleurs, ce modèle suppose une grande quantité d'individus dans chaque groupe (susceptible, infecté, écarté), et est mal adapté à l'étude du début ou de la fin d'une épidémie (ou très peu de personnes sont infectées) ainsi qu'à l'étude de la propagation de maladies dans des petites populations (par exemple, pour les espèces en voie de disparition).

Dans les cas d'une faible population au total ou d'un faible nombre d'individus infectés, un modèle probabiliste s'impose pour mieux étudier la propagation, et c'est l'objet de la partie II.

Partie II

Si $i = 0$ et pour certaines valeurs de s , l'épidémie ne peut pas démarrer, d'après les résultats de la partie I. On se propose de conforter ces résultats à l'aide d'un modèle à l'échelle des individus.

Dans les premières phases d'une tentative de flambée épidémique, seul un nombre restreint d'individus (négligeable par rapport à la taille de la population) est infecté. Dans ce régime, les individus susceptibles forment un "réservoir" infini.

Une modélisation raisonnable pour la propagation épidémique est alors le modèle probabiliste à temps discret suivant.

Au temps initial, une seule grenouille est malade. À chaque temps $n \in \mathbb{N}$, chaque individu malade au temps $n-1$ a guéri, mais entre temps il a infecté un nombre aléatoire d'autres grenouilles. On se donne des probabilités

$$p_0, p_1, p_2, \dots$$

dont la somme fait $p_0 + p_1 + p_2 + \dots = 1$. Alors, indépendamment les uns des autres, on suppose que chaque individu malade a contaminé k autres grenouilles avec probabilité p_k . Si à un moment donné plus aucune grenouille n'est infectée, alors l'épidémie s'est éteinte.

Le nombre moyen de grenouilles infectées par individu malade et par unité de temps est noté

$$R = 1 \times p_1 + 2 \times p_2 + 3 \times p_3 + \dots$$

Pour tout $n \in \mathbb{N}$, on note $q_k(n)$ la probabilité que k individus soit infectés au temps n . On a donc $q_1(0) = 1$ et $q_k(0) = 0$ pour tout $k \neq 1$. Le nombre moyen de grenouilles infectées au temps n est noté

$$I_n = 1 \times q_1(n) + 2 \times q_2(n) + 3 \times q_3(n) + \dots$$

Question II.1 Puisqu'en moyenne, une grenouille infectée au temps n résulte en R grenouilles infectées au temps $n+1$, quelle est l'équation de récurrence vérifiée par I_n ? (Un raisonnement heuristique suffira, points bonus pour une démonstration rigoureuse).

Corrigé On trouve immédiatement $I_{n+1} = RI_n$, puisque chaque grenouille en infecte, en moyenne, R autres.

Le moyen le plus confortable de donner une démonstration rigoureuse est d'utiliser le langage des probabilités conditionnelles. Soit $k \in \mathbb{N}$ et supposons qu'au temps n il y ait k grenouilles infectées. Cet événement, qui arrive avec probabilité $q_k(n)$, est noté $A_k(n)$. Alors, conditionnellement à $A_k(n)$, chacune des k grenouilles malades infectant R grenouilles en moyenne, on aura, au temps $n+1$, en moyenne, kR grenouilles malades.

Finalement, par la formule des probabilités totales,

$$I_{n+1} = \sum_k kR\mathbb{P}[A_k(n)] = R \sum_k kq_k(n) = RI_n.$$

Commentaire Certaines personnes ont mal interprété le sujet et trouvé $I_{n+1} = (R-1)I_n$ (en pensant que, dans le calcul de R , on prenait en compte les grenouilles déjà malades).

Question II.2 On suppose $R < 1$. Montrer que $I_n \rightarrow 0$ lorsque $n \rightarrow +\infty$.

Corrigé La suite I_n est une suite géométrique de raison R , et on a supposé $R < 1$.

Question II.3 Montrer l'inégalité de Markov

$$q_0(n) \geq 1 - I_n.$$

Déduire des questions précédentes que, si $R < 1$, alors quand $n \rightarrow +\infty$, la probabilité que l'épidémie se soit éteinte tend vers 1.

Corrigé On a par définition

$$q_0(n) + q_1(n) + q_2(n) + q_3(n) + \dots = 1,$$

et donc, puisque tous les $q_k(n)$ sont positifs ou nuls,

$$q_0(n) + \underbrace{q_1(n) + 2q_2(n) + 3q_3(n) + \dots}_{=I_n} \geq 1.$$

Finalement $q_0(n) \geq 1 - I_n$. On peut maintenant appliquer le théorème du sandwich : on a $1 - I_n \rightarrow 1$ lorsque $n \rightarrow +\infty$, et par ailleurs $q_0(n) \geq 1$ puisque c'est une probabilité, donc $q_0(n) \rightarrow 1$.

Question II.4 On suppose que $R > 1$. Comment qualifier la vitesse de propagation (en moyenne) de l'épidémie ?

Corrigé On a toujours $I_n = I_0 R^n$, donc on a, en moyenne, une propagation exponentielle !

Question II.5 Comparer les résultats de la question précédente avec le comportement asymptotique de la proportion d'infectés dans le modèle (SIR) quand $t \rightarrow -\infty$ (Indice : justifier le remplacement de s par une constante). Quel est le lien entre R et R_0 ?

Corrigé Le début de l'épidémie correspond à t proche de $-\infty$ dans le modèle SIR. Puisque s converge vers s_- en $-\infty$, on peut le remplacer par une constante dans l'équation différentielle portant sur i , et on trouve

$$i(t) = i_0 \exp((R_0 s_- - 1)t).$$

Si t_0 est le temps typique pendant lequel un individu est contagieux, alors en un laps de temps t_0 , i a été multiplié par

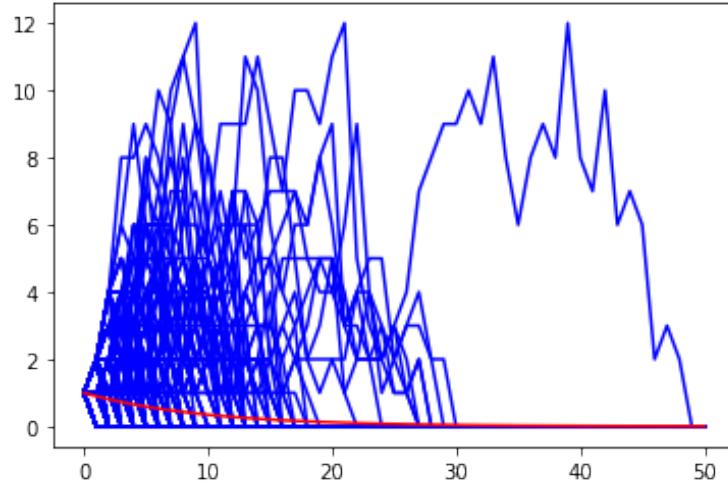
$$R = \exp((R_0 s_- - 1)t_0).$$

On retrouve le lien entre la potentialité d'une flambée épidémique et $R_0 s_-$: on a $s_- > \frac{1}{R_0}$ si et seulement si $R > 1$.

Commentaire Attention à l'homogénéité ! R est une quantité adimensionnée mais R_0 est homogène à l'inverse d'un temps, donc on ne peut pas avoir $R = R_0$, $R = R_0 s$, ou autre chose du même goût.

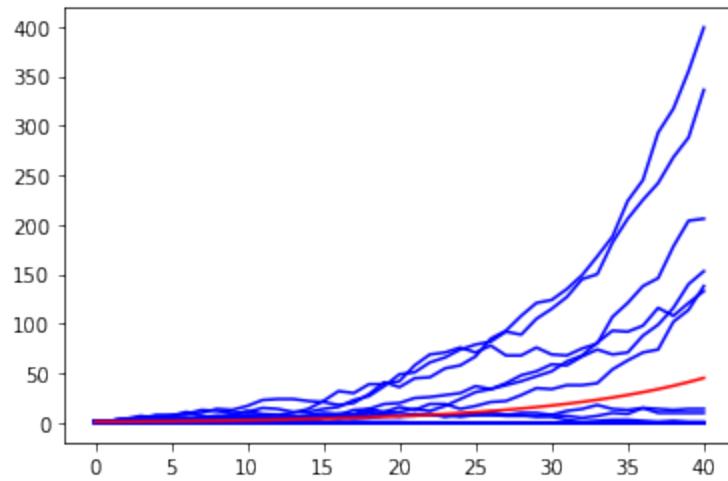
Description complète du modèle Ce modèle étant, par essence, probabiliste, même si $R < 1$ on ne peut pas garantir qu’au bout d’un temps fixé l’épidémie se sera éteinte: il est toujours possible, même si cela arrive avec une probabilité de plus en plus faible, que l’épidémie survive.

S’il y a initialement un seul individu malade, alors les résultats de ce modèle sont très variables : le nombre d’infectés moyen décrit assez mal les différentes trajectoires. On a simulé, par exemple, 50 fois, ce modèle, pour $p_0 = 0.4, p_1 = 0.3, p_2 = 0.3$, correspondant à $R = 0.9$. L’épidémie a tendance à s’eteindre, mais les trajectoires du nombre d’individus infectés en fonction du temps (en bleu) ne suivent pas l’allure de I_n (en rouge).



On voit que, pour chacune de ces 50 simulations, l’épidémie s’est éteinte, mais ça a pu prendre un certain temps pour quelques unes de ces simulations.

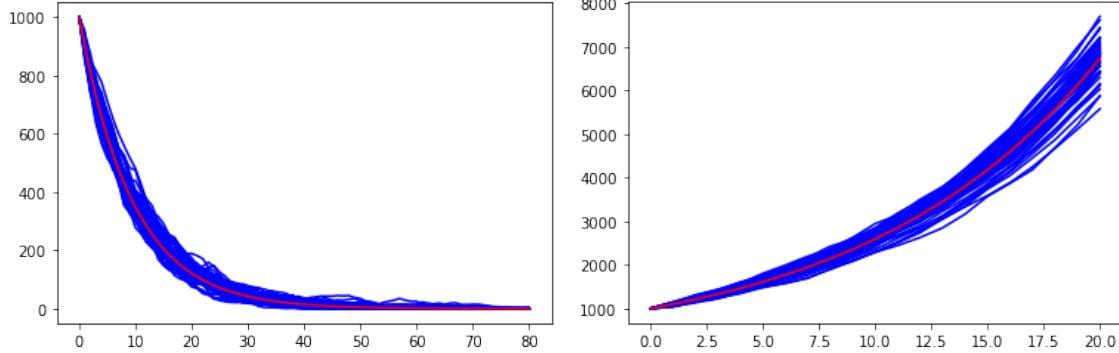
L’une des subtilités de ce modèle est que, si $R > 1$, alors l’épidémie se propage en moyenne, mais il est quand même possible que l’épidémie s’arrête. Par exemple, avec $p_0 = 0.3, p_1 = 0.3, p_2 = 0.4$, on a $R = 1.1 > 1$, mais si on commence avec un seul individu malade, on a déjà une probabilité 0.3 qu’elle n’infecte personne et que l’épidémie s’arrête immédiatement ! On a représenté ici 50 tirages pour ce modèle :



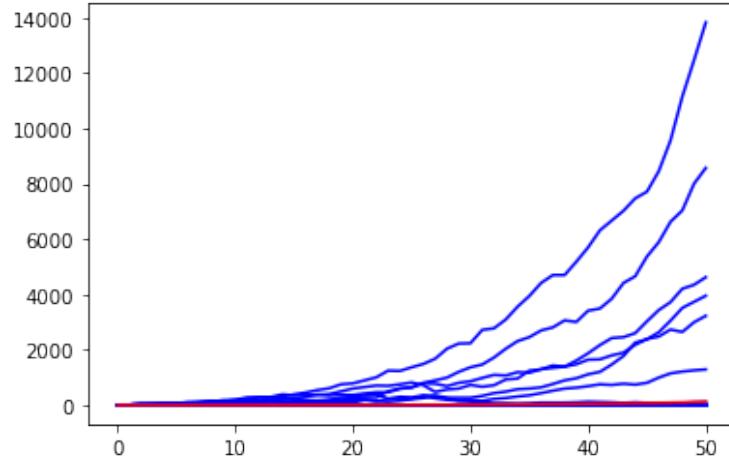
On peut compter “à la main” que, parmi ces 50 essais, on a eu une flambée épidémique dans seulement 8 cas. On constate aussi que, au sein des flambées épidémiques “réussies”, la croissance

est beaucoup plus rapide que la moyenne (courbe rouge). C'est normal, car seuls ces essais comptent pour un nombre d'infectés non nul!

Si on commence avec plus d'individus infectés, alors les effets de la loi des grands nombres se font sentir, et les différentes simulations sont à peu près proches de la courbe moyenne. On a simulé 50 tirages, en partant de 1000 individus infectés, pour les deux lois de probabilité plus haut:



Ce modèle est intéressant pour étudier finement les propriétés de la manière dont l'épidémie se propage. On peut comparer, par exemple, les résultats de $p_0 = 0.3, p_1 = 0.3, p_2 = 0.4$, avec une hypothèse de type "supercluster": une proportion faible d'individus va en infecter beaucoup d'autres, par exemple avec $p_0 = 0.9$ et $p_{11} = 0.1$. Dans les deux cas on a $R = 1.1$, mais dans le modèle supercluster, les flambées épidémiques sont plus violentes, ici avec 200 simulations:



On a une chance plus faible d'une flambée épidémique, mais puisque la moyenne est la même, s'il y a flambée, le nombre de personnes infectées grandit très vite.

On constate donc qu'il est non seulement important de réduire R , c'est-à-dire de réduire R_0 et s en moyenne, mais aussi d'éliminer les possibilités de supercluster.