

Décision dans l'incertain

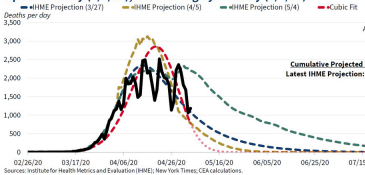
Comment modéliser une épidémie ?

Vendredi 29 Mai 2020

Comment modéliser une épidémie ?

United States Daily COVID-19 Deaths: Actual Data, IHME/UW Model Projections, & Cubic Fit.

Updated today (5/5/20), data through yesterday (5/4/20).



MONEY & MARKETS

White House economic adviser Kevin Hassett created a baffling 'cubic model' showing coronavirus deaths would hit 0 by May 15. It turns out he used a basic Excel function for his controversial projection.

JOSEPH ZEBALLOS-ROIG

MAY 1, 2020, 9:13 AM



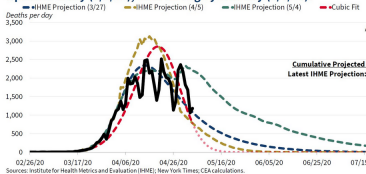
- Beaucoup de modèles sont proposés ...
- ... avec des enjeux énormes

Comment juger la pertinence d'un modèle ?

Comment modéliser une épidémie ?

United States Daily COVID-19 Deaths: Actual Data, IHME/UW Model Projections, & Cubic Fit.

Updated today (5/5/20), data through yesterday (5/4/20).



MONEY & MARKETS

White House economic adviser Kevin Hassett created a baffling 'cubic model' showing coronavirus deaths would hit 0 by May 15. It turns out he used a basic Excel function for his controversial projection.

JOSEPH ZEBALLOS-ROHS
MAY 1, 2020, 9:13 AM

f FACEBOOK t TWITTER u REEDIT in LINKEDIN e EMAIL



- Beaucoup de modèles sont proposés ...
- ... avec des enjeux énormes

Comment juger la pertinence d'un modèle ?

Nous verrons ici comment utiliser un modèle *probabiliste* pour modéliser une épidémie

Le travail présenté ci-après est adapté très largement de celui de Cosma Shalizi dans sa présentation Epidemic Models (16 avril 2020)

D'une modélisation probabiliste au modèle SIR

Le modèle SIR part des présupposés suivants

Supposons que nous ayons une maladie avec 3 états possibles

- S (*Susceptible*) : sain, mais peut être contaminé
- I (*Infectious*) : malade, et peut contaminer d'autres personnes
- R (*Removed*) : guéri, et ne peut plus être contaminé

Jeu d'hypothèses

- **Contagion** : un S peut être contaminé quand il rencontre un I
- **Guérison** au bout d'un certain temps, les I se transforment spontanément en R
- **Mixité** : la *probabilité* qu'un S rencontre un I dépend du nombre total de S , I et R

Proposons un premier modèle *discret*, suivant les hypothèses précédentes

- Nous avons une population totale de n
- Le pas de temps de discrétisation δ est petit
- On note (S_k, I_k, R_k) l'état de l'épidémie au pas de temps $t_k = t_0 + k\delta$
 - Pendant un temps δ , chaque personne S rencontre $p\delta$ personnes, avec une probabilité de contagion de c
 - Nous avons une probabilité $\frac{I_k}{n}$ que chaque personne rencontrée soit infectée
 - la probabilité qu'une personne S soit contaminée durant l'intervalle de temps est dès-lors

$$\begin{aligned}\mathbb{P}(S \rightarrow I) &= 1 - \mathbb{P}(S \rightarrow S) \\ &= 1 - \underbrace{\left(1 - c \frac{I_k}{n}\right)^{p\delta}}_{\text{(aucune rencontre ne mène à une contamination)}} \\ \mathbb{P}(S \rightarrow I) &\approx p\delta \times c \times \frac{I_k}{n} \quad \text{si } n \text{ grand}\end{aligned}$$

De ce qui précède, nous déduisons que si $n \gg 1$

Les contagions et guérisons sont probabilistes

- Le nombre de contagion C_k ($S \rightarrow I$) suit une *loi binomiale*

$$C_k \sim \text{binom}(S_k, p\delta \frac{I_k}{n})$$

- Le nombre de rémission D_k ($I \rightarrow R$) suit aussi une loi binomiale

$$D_k \sim \text{binom}(I_k, \gamma\delta)$$

avec γ taux de guérison (indépendant de n)

Nous obtenons les équations probabilistes suivantes ...

$$\begin{cases} S_{k+1} = S_k - C_k \\ I_{k+1} = I_k + C_k - D_k \\ R_{k+1} = R_k + D_k \end{cases}$$

avec

- $C_k \sim \text{binom}(S_k, \frac{\beta}{n}\delta I_k)$: nombre de contagions au pas de temps k
- $D_k \sim \text{binom}(I_k, \gamma\delta)$: nombre de rémissions au pas de temps k
- $\beta = pc$ taux de contagion

Nous obtenons les équations probabilistes suivantes ...

$$\begin{cases} S_{k+1} = S_k - C_k \\ I_{k+1} = I_k + C_k - D_k \\ R_{k+1} = R_k + D_k \end{cases}$$

avec

- $C_k \sim \text{binom}(S_k, \frac{\beta}{n}\delta I_k)$: nombre de contagions au pas de temps k
- $D_k \sim \text{binom}(I_k, \gamma\delta)$: nombre de rémissions au pas de temps k
- $\beta = pc$ taux de contagion

*Nous obtenons un modèle Markovien !
Le système suit une dynamique aléatoire*

Nous obtenons les équations probabilistes suivantes ...

$$\begin{cases} S_{k+1} = S_k - C_k \\ I_{k+1} = I_k + C_k - D_k \\ R_{k+1} = R_k + D_k \end{cases}$$

avec

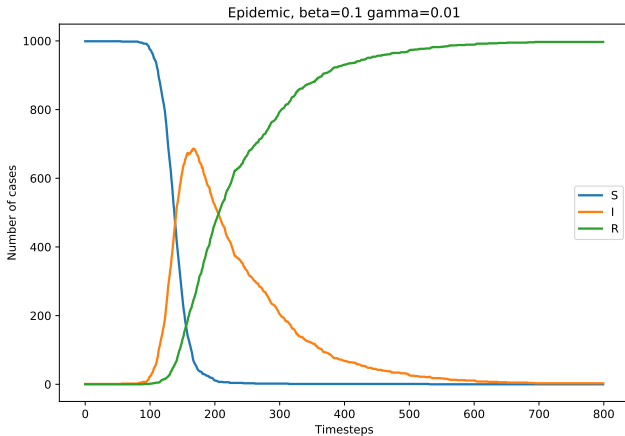
- $C_k \sim \text{binom}(S_k, \frac{\beta}{n}\delta I_k)$: nombre de contagions au pas de temps k
- $D_k \sim \text{binom}(I_k, \gamma\delta)$: nombre de rémissions au pas de temps k
- $\beta = pc$ taux de contagion

Nous obtenons un modèle Markovien !

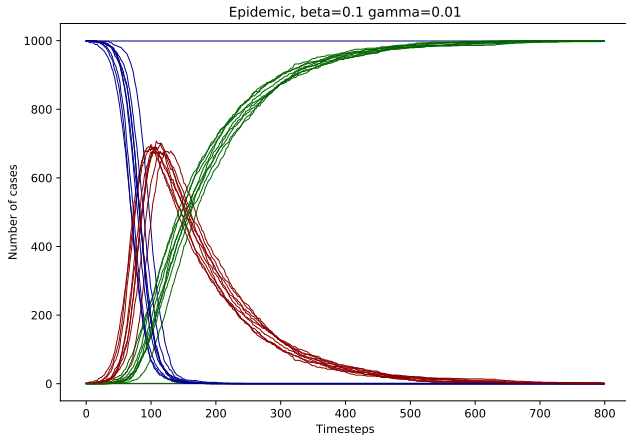
Le système suit une dynamique aléatoire

Notons que le passage entre les trois états se fait uniquement dans le sens $S \rightarrow I \rightarrow R$!

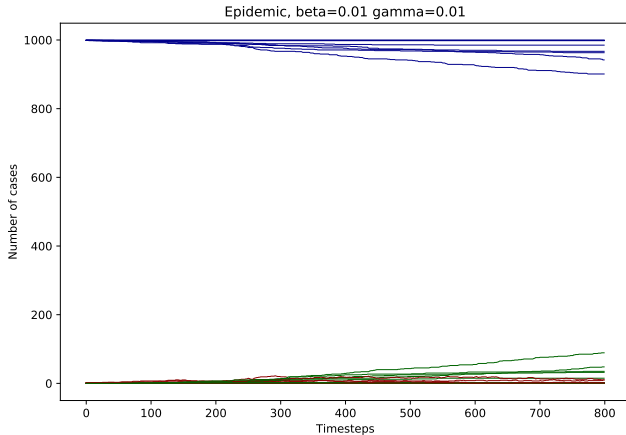
Si nous faisons tourner une simulation, nous avons ...



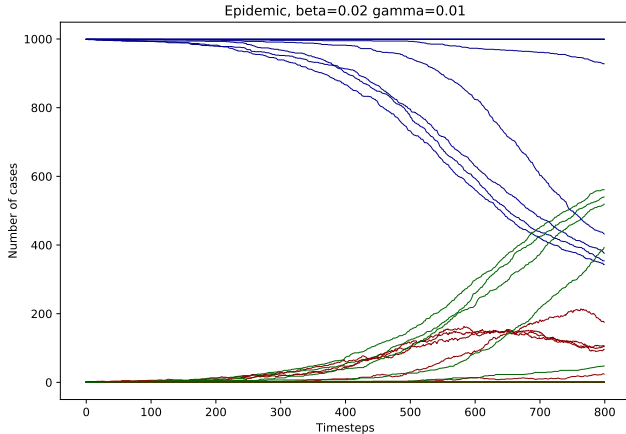
Mais n'oublions pas que nous avons une dynamique aléatoire !
En faisant tourner 100 simulations, nous avons



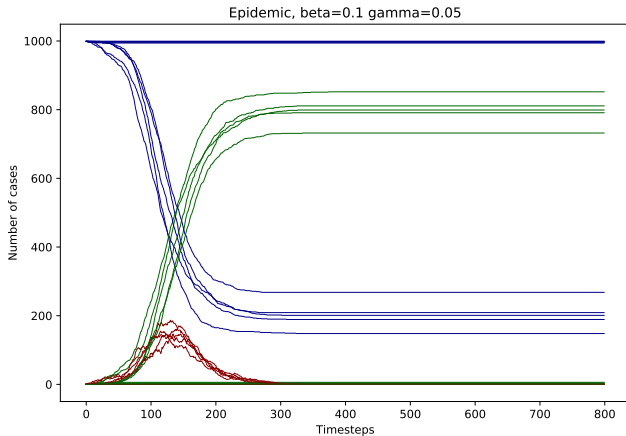
Si la contagion devient plus difficile



Niveau de contagion intermédiaire



Probabilité de guérison plus élevée



Que se passe-t-il si le pas de temps très petit ($\delta \rightarrow 0$) ?

- On a :

$$\begin{aligned}\frac{S_{k+1} - S_k}{\delta} &\sim -\frac{1}{\delta} \text{binom}(S_k, \beta \delta I_k / n) \\ \frac{I_{k+1} - I_k}{\delta} &\sim \frac{1}{\delta} \text{binom}(S_k, \beta \delta I_k / n) - \frac{1}{\delta} \text{binom}(I_k, \gamma \delta)\end{aligned}$$

- En passant à l'espérance conditionnelle
(si $X \sim \text{binom}(n, p)$, $\mathbb{E}(X) = np$)

$$\begin{aligned}\mathbb{E}\left[\frac{S_{k+1} - S_k}{\delta} \mid S_k = s_k, I_k = i_k\right] &= -\frac{\beta}{n} s_k i_k \\ \mathbb{E}\left[\frac{I_{k+1} - I_k}{\delta} \mid S_k = s_k, I_k = i_k\right] &= \frac{\beta}{n} s_k i_k - \gamma i_k\end{aligned}$$

Qu'en est-il des fluctuations autour de la moyenne ?

- Si on regarde la variance (si $X \sim \text{binom}(n, p)$, $\mathbb{V}(X) = np(1 - p)$)

$$\mathbb{V}\left[\frac{S_{k+1} - S_k}{\delta} \mid S_k = s_k, I_k = i_k\right] = \frac{\beta}{\delta n} s_k i_k \left(1 - \frac{\beta}{n} \delta i_k\right)$$

- Sous un jeu d'hypothèses précis (par ex. $n\delta \rightarrow \infty$) nous pouvons obtenir $\mathbb{V}\left[\frac{S_{k+1} - S_k}{\delta} \mid S_k = s_k, I_k = i_k\right] \rightarrow 0$
- Le même raisonnement tient pour $\mathbb{V}\left[\frac{I_{k+1} - I_k}{\delta}\right]$
- Dans ce cas, nous pouvons négliger les variations autour de l'espérance dans notre modèle

Qu'en est-il des fluctuations autour de la moyenne ?

- Si on regarde la variance (si $X \sim \text{binom}(n, p)$, $\mathbb{V}(X) = np(1 - p)$)

$$\mathbb{V}\left[\frac{S_{k+1} - S_k}{\delta} \mid S_k = s_k, I_k = i_k\right] = \frac{\beta}{\delta n} s_k i_k \left(1 - \frac{\beta}{n} \delta i_k\right)$$

- Sous un jeu d'hypothèses précis (par ex. $n\delta \rightarrow \infty$) nous pouvons obtenir $\mathbb{V}\left[\frac{S_{k+1} - S_k}{\delta} \mid S_k = s_k, I_k = i_k\right] \rightarrow 0$
- Le même raisonnement tient pour $\mathbb{V}\left[\frac{I_{k+1} - I_k}{\delta}\right]$
- Dans ce cas, nous pouvons négliger les variations autour de l'espérance dans notre modèle

Ce passage n'est pas trop rigoureux mais nécessite des notions qui seront vues durant le cours de processus stochastique de 2e année

Lorsque $n \rightarrow \infty$ et $\delta \rightarrow 0$ on obtient le modèle SIR continu :

Modèle SIR

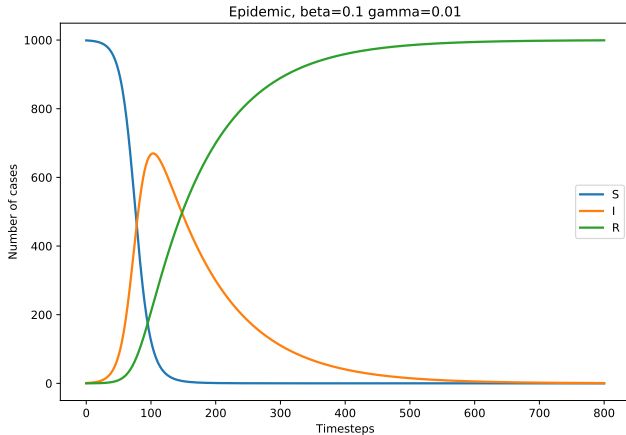
Le modèle SIR correspond aux équations différentielles suivantes :

$$\begin{cases} \dot{S} = -\frac{\beta}{n}SI \\ \dot{I} = \frac{\beta}{n}IS - \gamma I \\ \dot{R} = \gamma I \end{cases}$$

Ces équations sont déterministes et non-linéaires !

Simulation avec les paramètres originaux

On résout le système d'équation différentielle numériquement



Qu'est-ce que le nombre de reproduction basique R_0 ?

Nombre de reproduction basique R_0

On définit R_0 comme le nombre moyen de nouvelles infections si on ajoute un agent infectieux dans une population non atteinte



Nous avons trois régimes :

- $R_0 < 1$: Sous-critique
- $R_0 > 1$: Super-critique
- $R_0 = 1$: Critique

On suppose qu'initialement : $S(0) \approx n$. Alors

$$\begin{aligned} \dot{I} &= \frac{\beta}{n}SI - \gamma I \\ &= \left(\beta \frac{S}{n} - \gamma\right)I \\ &\approx (\beta - \gamma)I \end{aligned}$$

d'où une croissance exponentielle de malades au début de l'épidémie

$$I(t) \approx I_0 e^{(\beta - \gamma)t}$$

La croissance devient *exponentielle* si $\beta > \gamma$

On en déduit

$$R_0 = \frac{\beta}{\gamma}$$

Le modèle SIR comporte seulement trois états ! Des variantes existent :

- Etat d'incubation E des gens qui ont été exposés à la maladie mais ne sont pas encore dans l'état I (modèle SEIR)
- Etat asymptomatique A des gens ne présentant pas de symptôme mais contaminant
- Possibilité éventuelle de retomber malade (passage $R \rightarrow I$)

Le modèle dépend donc intrinséquement des caractéristiques du virus ...

On n'est jamais trop prudent

Un modèle est un *outil*, et ne doit jamais être pris au pied de la lettre !

Une extension usuelle du modèle SIR est de prendre en compte les échanges entre plusieurs régions $i = 1, \dots, N$

$$\dot{I}_i = \frac{\beta}{n} S_i I_i - \gamma I_i - \mu_{ij} I_i + \mu_{ji} I_j$$

où μ_{ij} est la proportion de personnes voyageant de la région i à j

Une extension usuelle du modèle SIR est de prendre en compte les échanges entre plusieurs régions $i = 1, \dots, N$

$$\dot{I}_i = \frac{\beta}{n} S_i I_i - \gamma I_i - \mu_{ij} I_i + \mu_{ji} I_j$$

où μ_{ij} est la proportion de personnes voyageant de la région i à j

Pendant l'épidémie, μ_{ij} a très souvent été estimé avec les données téléphoniques

Un modèle plus détaillé



Source : Wikipedia

Nous avons vu précédemment le modèle SIR, qui reposait sur un certain nombre d'hypothèses

Analysons maintenant comment pouvons-nous adapter ce modèle sur un graphe

- Certaines personnes sont plus sociables que d'autres ...
- Intéressons nous maintenant au *graphe des relations* (si i (Igor) connaît j (Jeanne), alors $a_{ij} = 1$, sinon 0)

Définitions (rappel)

- Le *degré* d'un noeud i est égal à

$$\deg(i) = \sum_{j=1}^n a_{ij}$$

- Pour tout $k \in \mathbb{N}$, on note $p(k)$ la proportion des noeuds ayant un degré égal à k

Comment la topologie du réseau influe-t-elle sur la diffusion de l'épidémie ?

- Notons τ le taux de transmission de la maladie
- Supposons que le noeud j soit infecté par un noeud adjacent i
- R_0 est le nombre moyen de nouvelles infections causées par j , soit

$$R_0 := \tau \times (\text{nombre d'amis de Jeanne} - 1)$$

(on enlève le noeud adjacent i qui a transmis la maladie à Jeanne)

Sur le graphe, notons le nombre moyen d'amis comme

$$\mathbb{E}[K] := \sum_{k=1}^{\infty} kp(k)$$

Attention ! Sur le graphe

nombre moyen d'amis \neq nombre moyen d'amis de mes amis

et Jeanne est une amie d'Igor ...

Attention ! Sur le graphe

nombre moyen d'amis \neq nombre moyen d'amis de mes amis

et Jeanne est une amie d'Igor ...

Le paradoxe de l'amitié

Mes amis ont en moyenne plus d'amis que moi

- Jeanne *n'est pas* un noeud aléatoire, mais un noeud atteint par une *arête aléatoire*
- En moyenne, les noeuds avec une connectivité élevée ont une probabilité plus grande d'être atteint par une arête prise de manière aléatoire
- Dans l'ensemble, chaque noeud j
 - apparaît comme ami d'ami $\deg(j)$ fois
 - et contribue pour un nombre d'amis égal à $\deg(j)$ à la moyenne du nombre d'amis d'ami

On a alors, si on note m_2 le nombre moyen d'amis d'ami

$$m_2 = \frac{\sum_{v \in \mathcal{V}} \deg^2(v)}{\sum_{v \in \mathcal{V}} \deg(v)}$$

Or nous avons $\sum_{v \in \mathcal{V}} \deg(v) = \sum_{k=1}^{\infty} kp(k)$. D'où

$$m_2 = \frac{\sum_{k=1}^{\infty} k^2 p(k)}{\sum_{k=1}^{\infty} kp(k)}$$

Par Cauchy-Schwarz, on peut montrer que

$$\left(\sum_{k=1}^{\infty} kp(k) \right)^2 \leq \underbrace{\sum_{k=1}^{\infty} p(k)}_{=1} \times \sum_{k=1}^{\infty} k^2 p(k)$$

On prouve alors le paradoxe de l'amitié :

$$\sum_{k=1}^{\infty} kp(k) \leq m_2$$

On a alors

$$\begin{aligned} R_0 &= \tau \mathbb{E}[\text{Nombre d'amis de Jeanne} - 1] \\ &= \tau \frac{\sum_{k=1}^{\infty} (k-1)kp(k)}{\sum_{k=1}^{\infty} kp(k)} \\ &= \tau \frac{\mathbb{E}[K^2 - K]}{\mathbb{E}[K]} \end{aligned}$$

On a alors

$$\begin{aligned} R_0 &= \tau \mathbb{E}[\text{Nombre d'amis de Jeanne} - 1] \\ &= \tau \frac{\sum_{k=1}^{\infty} (k-1)kp(k)}{\sum_{k=1}^{\infty} kp(k)} \\ &= \tau \frac{\mathbb{E}[K^2 - K]}{\mathbb{E}[K]} \end{aligned}$$

Or nous avons $\mathbb{V}[K] = \mathbb{E}[K^2] - \mathbb{E}[K]^2$. D'où

$$R_0 = \tau \left(\frac{\mathbb{V}[K] + \mathbb{E}[K]^2 - \mathbb{E}[K]}{\mathbb{E}[K]} \right)$$

On souhaite avoir $R_0 < 1$, soit

$$\tau \left(\frac{\mathbb{V}[K]}{\mathbb{E}[K]} + \mathbb{E}[K] - 1 \right) < 1$$

Pour cela, on peut :

- Diminuer la transmission τ (*sortir masqué*)
- Diminuer le degré moyen $\mathbb{E}[K]$ (*rester chez soi*)
- Diminuer la variance $\mathbb{V}[K]$ (*enlever les noeuds très connectés*)

Comment simuler l'épidémie sur un réseau ?

- *Graphes aléatoires*

- En pratique, il est très difficile d'avoir accès à des graphes réalistes (hormis pour Facebook et quelques opérateurs téléphoniques)
- On se contente souvent de *graphes aléatoires* pour les simulations !
Par exemple, en utilisant des graphes d'Erdős-Rényi

- *Modèles agents*

- Un modèle agent est une simulation avec n agents où chacun des agents suit et interagit avec les autres agents suivant un comportement programmé
- Des comportements systémiques peuvent parfois émerger des interactions individuelles

- La présentation de Cosma Shalizi dont cette conférence est largement inspirée :
<http://www.stat.cmu.edu/~cshalizi/dm/20/lectures/special/epidemics.html>
- Sur la diffusion épidémique sur un graphe, voir l'interview de Claire Mathieu :
<https://lejournel.cnrs.fr/articles/des-graphes-pour-planifier-le-deconfinement>
- Sur les graphes aléatoires, voir les travaux du CERMICS !
- Sur les modèles agents, voir la présentation de Guy Theraulaz :
<https://www.college-de-france.fr/site/colloque-2018/symposium-2018-10-19-11h50.htm>