

Projet Quantification d'Incertitudes

Propagation du coronavirus MERS-CoV en Corée du Sud

2019-2020

1 Description du problème

Une épidémie d'infection au coronavirus MERS a touché de nombreux pays entre 2012 et 2015, avec un pic remarquable en Corée du Sud, qui a tracé avec précision l'évolution de la propagation de ce virus, cf [Chang]. Les informations sur ce virus et cette épidémie sont tirées de [FranceTVInfo].

1.1 Qu'est-ce que le coronavirus MERS ? Quels sont les symptômes de l'infection ?

Les coronavirus sont une sous-famille de virus infectant les voies digestives et respiratoires supérieures de nombreuses espèces (oiseaux, mammifères). La surface de la coque de ces virus présente des prolongements suggérant une "couronne" - d'où le nom de coronavirus, abrégé CoV. Six souches de coronavirus sont susceptibles d'atteindre l'homme, parmi lesquels le tristement célèbre SRAS-CoV, responsable du syndrome respiratoire aigu sévère.

Comme SRAS-Cov, le coronavirus MERS provoque une infection des poumons, et les personnes touchées souffrent de fièvre, de toux et de difficultés respiratoires. A la différence du SRAS, il génère également une défaillance rénale. Il n'existe actuellement aucun vaccin ou traitement pour cette maladie. Ce virus est-il aussi dangereux que celui du SRAS ?

Selon le recensement de l'Organisation mondiale de la santé, 36% des personnes atteintes par le coronavirus MERS décèdent.

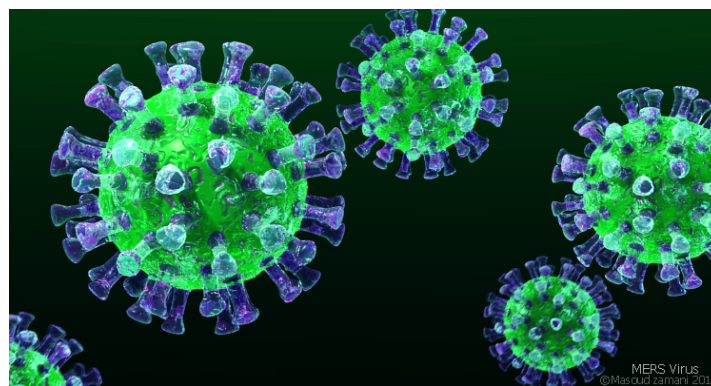


Figure 1: Middle East respiratory syndrome virus (MERS-CoV) 3-D image

Ce virus apparaît ainsi plus mortel que celui causant le SRAS (qui avait fait près de 800 morts dans le monde en 2003). Mais il semble en revanche bien moins contagieux. Selon l'OMS, le virus ne passerait pas aisément d'une personne à une autre (une proximité prolongée avec un patient semble nécessaire).

La maladie semble se transmettre par des voies similaires à celle de la grippe (absorption de postillons, mise à la bouche de mains ayant été en contact avec le virus).

1.2 Quelle est l'origine de l'épidémie en Corée du Sud ?

Le premier cas humain d'infection au virus a été constaté post-mortem en 2012 chez un Saoudien. MERS signifie "Middle-East Respiratory Syndrome", soit syndrome respiratoire du Moyen-Orient. En France, on devrait donc plutôt parler de coronavirus SRMO.

Plus de 20 pays ont été touchés par MERS-CoV depuis 2012. La plupart des cas recensés l'ont été en Arabie saoudite (plus de 400 décès recensés) et dans la péninsule arabique. Au printemps 2013, deux cas ont été confirmés en France, chez un voyageur de retour de Dubaï, puis chez son voisin de chambre à l'hôpital.

L'origine de l'actuelle épidémie en Corée du Sud est imputée à un homme de 68 ans, diagnostiqué deux semaines après son retour d'un séjour au Moyen-Orient au cours duquel il a passé deux jours en Arabie saoudite, premier foyer de la maladie.

1.3 Modèle de propagation d'épidémie: le modèle SIR

Différents modèles de propagation d'épidémie ont été proposés dans la littérature, cf [Kramer]. Parmi ceux-ci, le modèle SIR propose de distinguer trois groupes au sein d'une population:

- Le groupe S des personnes *Susceptible*, c'est-à dire pouvant être contaminées;
- Le groupe I des personnes *Infectious*, c'est-à dire étant contaminées et contagieuses;
- Le groupe R des personnes *Removed*, c'est-à dire étant soit mortes victimes de la maladie, soit guéries.

L'évolution de ces trois populations est décrit par un système d'équations différentielles non-linéaires du premier ordre:

$$\begin{aligned}\frac{dS}{dt}(t) &= -\beta \frac{S(t)I(t)}{N(t)} \\ \frac{dI}{dt}(t) &= \beta \frac{S(t)I(t)}{N(t)} - \nu I(t) \\ \frac{dR}{dt}(t) &= \nu I(t)\end{aligned}$$

Ce modèle correspond à une transition des personnes du groupe S au groupe I proportionnellement au produit des effectifs des groupes S et I , modélisant ainsi une propagation par promiscuité entre les personnes de ces deux groupes. L'effectif du groupe R correspond à un prélèvement d'une portion de la population du groupe I . On remarque que la population totale $N(t) = S(t) + I(t) + R(t)$ est constante, sa dérivée temporelle étant nulle.

Des versions plus sophistiquées de ce modèle existent, prenant en compte la dynamique d'évolution de la population (naissance, mortalité par causes externes), mais la version présentée est bien adaptée à la description d'une épidémie de courte durée.

La solution de ce modèle est déterminée par la connaissance des paramètres β et ν ainsi que par la population initiale $N(0)$ et l'effectif initial $I(0)$ du groupe I . Pour R , on prend comme condition initiale $R(0) = 0$. L'unité de temps est le jour et la durée de simulation du modèle est limitée à $t \in [0, T]$ avec $T = 300$.

1.4 Objectifs de l'étude

Lors de l'apparition d'une épidémie, de nombreuses questions sont posées à l'autorité publique. Nous retiendrons les trois questions suivantes:

- L'épidémie va-t-elle se propager à toute la population en l'absence de mesures spécifiques?
- Quelles mesures peuvent être prises pour atténuer ou enrayer l'épidémie?
- Ces mesures étant prises, quand se produira le pic de l'épidémie et quelle sera son ampleur?

L'estimation du *taux de reproduction de base* R_0 de l'épidémie est d'une importance capitale, puisqu'il correspond au nombre moyen de personnes pouvant être infectées par une personne déjà infectée. De la connaissance de R_0 résultent les réponses aux trois questions retenues. Une valeur $R_0 < 1$ indique que l'épidémie va s'éteindre d'elle-même alors qu'une valeur $R_0 > 1$ indique que l'épidémie va s'étendre à toute la population S . Dans le cadre du modèle SIR, on a $R_0 = \frac{\beta}{\nu}$. La calibration du modèle SIR conduit à adopter un modèle probabiliste pour le vecteur (β, ν) , permettant de calculer la probabilité de l'événement $R_0 > 1$.

Par ailleurs, R_0 admet une expression alternative $R_0 = \rho c d = \frac{\rho c}{\nu}$ où ρ est la probabilité de transmission de la maladie, c le nombre de contacts par unité de temps et $d = 1/\nu$ la durée de la période contagieuse. Cette expression permet de quantifier l'effet de politiques publiques agissant sur ces trois paramètres:

- Diminution de ρ via des mesures d'hygiène (lavage des mains, masques de protection);
- Diminution de c par des incitations à diminuer la vie sociale (fermeture de lieux publics, isolement des malades, restrictions de déplacements);
- Diminution de d par l'utilisation d'antiviraux ou de vaccins *impossible dans le cas de MERS-CoV puisqu'aucun traitement préventif ni curatif n'était disponible au moment de l'épidémie.*

2 Travail à réaliser

2.1 Modélisation probabiliste

La population de Corée du Sud compte environ $N(0) = 51\,000\,000$ individus au début de l'épidémie. On prend en compte une incertitude de l'ordre de 0.1% sur ce nombre, ce qui conduit à modéliser $N(0)$ par une loi uniforme $\mathcal{U}(50\,950\,000, 51\,050\,000)$.

Lors des premiers jours de l'évolution de l'épidémie, le recueil d'information sur les effectifs des groupes I et R a permis d'estimer les valeurs de β et de ν par ajustement du modèle aux observations. L'incertitude d'estimation conduit à adopter les hypothèses suivantes:

- β_{ini} suit une loi Beta $\mathcal{B}(5.5, 5.5, 0.215, 0.248)$
- ν suit une loi Gamma de paramètres $\mathcal{G}(7.3, 975, 0.021)$

Par ailleurs, le nombre initial de personnes infectées $I(0)$ est mal connu, on suppose qu'il est distribué selon une loi Beta de paramètres $\mathcal{B}(2, 5, 1, 3)$.

1. Créer les lois marginales de β_{ini} , ν , $I(0)$, $N(0)$. On utilisera pour cela les objets **Beta**, **Gamma**, **Uniform**.
2. On introduit la variable aléatoire $R_0 = \frac{\beta_{ini}}{\nu}$. Créer la loi de R_0 en utilisant l'arithmétique des distributions. Tracer sa densité de probabilité. Commenter sur la possibilité d'une extinction spontanée de l'épidémie. Calculer la probabilité pour que $R_0 < 1$.

On décide de prendre en compte les mesures de contrôle de l'épidémie prises par le gouvernement Coréen. Ces mesures d'hygiène et de contrôle des activités sociales conduisent à changer la loi de β en β_{new} , à partir d'une date t_{new} où ces mesures prennent leur plein effet. Cette date est mal connue, modélisée par une distribution uniforme $\mathcal{U}(24, 29)$. Le paramètre β_{new} suit alors une distribution Gamma de paramètres $\mathcal{G}(2.8, 2870, 0.0183)$.

3. Créer la nouvelle distribution de R_0 . L'épidémie va-t'elle s'éteindre?

Le modèle complet est donc fonction des sources d'incertitudes suivantes, supposées indépendantes:

- La population totale $N(0)$;
- Le nombre initial de personnes infectées $I(0)$;
- Le paramètre β_{ini} à $t = 0$;
- Le paramètre β_{30} à $t = t_{new}$;
- Le paramètre t_{new} ;
- Le paramètre ν .

4. Créer le vecteur $X = (N(0), I(0), \beta_{ini}, \beta_{new}, t_{new}, \nu)$ des sources d'incertitudes

2.2 Création de la variable d'intérêt

On s'intéresse désormais à l'intensité de l'épidémie, c'est-à dire au nombre maximum $Y = f(X)$ de personnes infectées durant la période $[0, T]$.

1. A l'aide de la fonction Python `computeMax` fournie, créer la fonction f telle que $Y = f(X)$. On utilisera pour cela l'objet `PythonFunction`. Penser à renseigner le paramètre `n_cpus=-1` pour effectuer les évaluations en parallèle sur l'ensemble des cœurs disponibles.
2. Créer le vecteur aléatoire Y .
3. Produire un échantillon de taille $n = 10000$ de Y , estimer sa moyenne, son écart-type et tracer une estimation par noyaux de sa densité. Pour cela on utilisera l'objet `KernelSmoothing`. On archivera cet échantillon au format CSV via la méthode `exportToCSVFile` de la classe `Sample`.

2.3 Calcul de probabilité d'événements rares

La gestion de crise nécessite de dimensionner les services de secours pour accueillir les personnes infectées le temps que dure l'épidémie. On estime qu'à partir de $T = 300$ un traitement aura été mis au point, faisant entrer l'épidémie dans une autre phase non modélisée ici.

La capacité d'accueil des hôpitaux pour les malades devant être mis à l'isolement complet est de 1000 places. On redoute donc l'événement $E = \{Y > 1000\}$.

1. Utiliser l'échantillon de Y pour estimer $\mathbb{P}(E)$.
2. Mettre en œuvre la méthode FORM pour calculer une approximation de la probabilité de l'événement E (classe `ThresholdEvent`). L'algorithme FORM utilisera l'algorithme d'optimisation `Cobyla`, les dérivées de f étant difficiles à calculer précisément, et sera initialisé par le point moyen de X . On indiquera le nombre d'appels à l'évaluation du modèle. Commenter.
3. La différence entre les approximations FORM et SORM incite à valider le calcul par une méthode de simulation. Quel est l'ordre de grandeur du nombre d'évaluations du modèle à réaliser pour estimer cette probabilité avec un coefficient de variation de 0.1 en utilisant une méthode de Monte Carlo simple?
4. Le budget de simulation est limité à 100000 évaluations de f . En utilisant les classes `ProbabilitySimulationAlgorithm` et `MonteCarloExperiment`, confirmer l'ordre de grandeur de la probabilité obtenue par la méthode FORM. On utilisera les méthodes `setBlockSize()` et `setMaximumOuterSampling()` de la classe `ProbabilitySimulationAlgorithm` pour répartir l'effort de simulation en limitant le nombre total d'itérations (*outer sampling*) à 1000, avec des blocs de taille 100. Quel est le coefficient de variation de l'estimateur? Conclure sur le risque de ne pas pouvoir accueillir les personnes infectées.

Le modèle SIR ne donne qu'une description simplifiée de l'évolution d'une épidémie. On décide de modéliser l'incertitude liée à ce modèle sous la forme d'un facteur multiplicatif r distribué selon une loi Beta de paramètres $\mathcal{B}(2, 4, 0.9, 1.2)$

1. Créer le vecteur \tilde{X} correspondant à cette modélisation augmentée.
2. Créer le modèle augmenté $\tilde{f} = rf$ en créant une version modifiée de `computeMax`, nommée `computeMaxFuzzy`.
3. Créer l'événement \tilde{E} correspondant à situation redoutée et incluant r .
4. Quantifier la probabilité de \tilde{E} . Commentaire par rapport à la modélisation précédente?

2.4 Analyse de sensibilité

1. Calculer les indices de Sobol du premier ordre et les indices de Sobol totaux du premier ordre pour les différentes sources d'incertitude à l'aide d'une méthode d'échantillonnage utilisant l'algorithme de Saltelli avec une suite à discrétance faible de Sobol. On utilisera une taille d'échantillon de 4096 pour la génération des matrices élémentaires, cf `ResourceMap.SetAsString('SobolIndicesExperiment-SamplingMethod', 'QMC')`, `SobolIndicesExperiment`, `SaltelliSensitivityAlgorithm`. On tracera ces indices à l'aide de la méthode `DrawSobolIndices` de `SaltelliSensitivityAlgorithm`. Commenter.

2. Estimer l'espérance de \tilde{X} conditionnée à l'événement espéré approché \hat{E} (On générera un échantillon de \tilde{X} de grande taille puis on ne gardera que les réalisations conduisant à \hat{E} , jusqu'à ce que ce nombre de réalisations vaille 100 ou jusqu'à ce que 100000 évaluations de \tilde{f} aient été réalisées).

En déduire des indices de sensibilité de la probabilité par rapport aux variables $N(0)$, $I(0)$, β_{ini} , β_{new} , t_{new} , ν , r . Comparer à la hiérarchisation des variables obtenue à partir des indices de Sobol.

Comment interprétez-vous ces indices? Pourquoi le paramètre β_{new} est-il si peu influent? Pourquoi ν est-il si peu influent? Pourquoi $N(0)$ et $I(0)$ jouent-ils des rôles si différents?

Est-ce que la maîtrise de Y suffit à contrôler les effets de l'épidémie?

3 Evaluation

Le projet est à réaliser en binôme. Chaque binôme devra rendre un unique notebook python (.ipynb) incluant les réponses aux questions et le code python associé. Le fichier sera nommé *NomEtudiant1_NomEtudiant2_projetQI.ipynb* et devra être déposé sur madoc avant le 11/03/2020 à 16h par un seul des membres du binôme.

References

- [Chang] Hyuk-Jun Chang, *Estimation of basic reproduction number of the Middle East respiratory syndrome coronavirus (MERS-CoV) during the outbreak in South Korea 2015*, BioMed Eng OnLine (2017) 16:79
- [FranceTVInfo] https://www.francetvinfo.fr/sante/maladie/epidemie-de-coronavirus-mers-en-coree-du-sud-quatre-choses-a-savoir_937815.html
- [Kramer] Alexander Krämer, Mirjam Kretzschmar, Klaus Krickeberg, *Modern Infectious Disease Epidemiology-Concepts, Methods, Mathematical Models, and Public Health*, Statistics for Biology and Health, Springer (2010)
- [CDC] *MERS Clinical Features* <https://www.cdc.gov/coronavirus/mers/clinical-features.html>