

***Risque de pandémie en assurance
prévoyance : construction d'un modèle de
propagation dans l'espace***

Guillaume JOUBERT

Slim FOU DHAILI

Promotion 2014

Synthèse

Les dernières années ont vu plusieurs pandémies frapper la population mondiale avec une rapidité et des dégâts considérables, qui ont marqué la conscience collective. On mentionne souvent la grippe espagnole qui a sévi à la fin de la première guerre mondiale (1918-1919) et plus récemment, le virus Ebola qui s'est révélé très meurtrier entre 2013 et 2016 et qui a même montré de nouveaux sursauts en 2017 et 2018.

Plusieurs grandes instances et figures du monde s'accordent à dire que le risque de pandémie est à prendre avec beaucoup de précaution mais qu'il n'est pas suffisamment considéré aujourd'hui. Bill Gates, par exemple, est intervenu en février 2017 lors du sommet sur la sécurité organisé à Munich pour alerter la communauté internationale sur le risque d'une pandémie dans les 10 ou 15 ans à venir, et en appelle à se préparer avec autant de sérieux que ce qui était fait auparavant contre le risque nucléaire.

Le risque de pandémie est en effet grandissant, au vu notamment de l'augmentation de la population mondiale et de ses flux migratoires mais également de la menace du bioterrorisme. Aujourd'hui, la communauté internationale ne semble pas aussi préparée à ce risque qu'aux risques terroristes ou nucléaire par exemple, que ce soit au niveau des installations, des dispositifs de sécurité et d'intervention. Le même constat peut être fait pour les assureurs car les couvertures de réassurance existantes et le modèle Solvabilité 2 sont encore relativement primaires sur ce risque.

Si nous reprenons le parallèle avec le risque terroriste, risque qui s'est considérablement accru ces dernières années (entre 2015 et 2017, on compte en Europe pas moins de 10 attentats ayant entraînés le décès de 5 individus ou plus, pour un total de près de 340 morts). Avant les événements du 11 septembre 2001, les dispositifs de sécurité et d'intervention étaient à des niveaux qui n'ont rien à voir avec ce qu'il se fait aujourd'hui. Ce sont ces attentats du World Trade Center qui ont bouleversé le monde et la façon de surveiller et maîtriser (autant que possible) ce risque. Il en est de même dans le monde assurantiel où ce risque est désormais beaucoup plus suivi et considéré, il suffit par exemple de constater l'évolution des dispositifs de sécurité et du coût des assurances lors de grands événements sportifs tels que les Jeux Olympiques ou les Coupes du Monde de Football. Ainsi, on peut supposer que c'est l'absence de pandémies suffisamment « sévères » qui explique que les gouvernements et assureurs ne considèrent pas encore ce risque avec suffisamment de mesures et d'attention.

Il existe dans la littérature plusieurs modèles pour permettre aux assureurs (entre autres acteurs) de mieux appréhender le risque pandémique. Ces modèles peuvent être classés en trois catégories : les modèles selon une approche par scénario (réapparition de pandémie historique, exemple 1918), les modèles d'estimation statistique (fréquence – sévérité) et enfin les modèles épidémiologiques.

C'est cette dernière approche qui permet de reproduire le plus fidèlement possible les effets d'une épidémie en simulant sa propagation au sein d'une population.

L'objectif de ce mémoire est ainsi de présenter le modèle épidémiologique que nous avons construit et qui permet de représenter le plus fidèlement possible la diffusion d'une pandémie dans l'espace.

Dans un premier temps, notre travail a consisté à coder le modèle épidémiologique le plus répandu dans la littérature : le modèle SEIR dont l'acronyme désigne les 4 états distincts dans lesquels un individu peut se trouver lors d'une épidémie (Susceptible / Exposé / Infectieux / Rétabli).

Nous avons commencé par tester les différents paramètres de notre modèle en appliquant à une population de 10 000 individus un spectre de 100 000 épidémies, celles-ci étant la combinaison de 100 taux de reproduction et 1 000 taux de létalité différents. Le taux de reproduction est une métrique qui caractérise la faculté d'une épidémie à se propager avec plus ou moins de virulence au sein d'une

population, indépendamment de son caractère mortel. Dans ce cas standard, chaque paire d'individus a donc toujours la même probabilité de se transmettre le virus.

Toutefois, ce modèle comporte une limite préjudiciable pour l'exploitation par un assureur des résultats produits : en effet, il ne prend pas en compte la répartition géographique des individus du portefeuille.

Nous avons ainsi souhaité développer un nouveau modèle épidémiologique, qui tient compte de ce paramètre géographique, c'est-à-dire de la localisation des individus au sein d'un territoire donné.

Pour ce faire, nous avons délimité un territoire et y disposer les individus de façon pertinente. Nous avons testé notre modèle sur le périmètre de la France métropolitaine et avons utilisé des données réelles de la démographie française pour répartir les individus afin que notre modèle soit construit sur les meilleures hypothèses. Nous avons testé trois niveaux de mailles pour la disposition des individus puisque nous avons procédé à une répartition par régions administratives, puis par départements et enfin par communes. Nous avons retenu ce dernier niveau de maille de répartition qui est le plus fin et qui nous a permis d'aboutir aux résultats les plus probants.

Dans ce nouveau modèle, la probabilité de transmission du virus entre deux individus n'est plus constante mais décroissante avec la distance géographique qui les sépare. Cette phase nous a permis de calibrer les paramètres de notre modèle géographique afin que, dans des conditions identiques au premier scénario, les résultats soient également similaires en termes de nombre d'infectés, temps d'épidémie, etc. En d'autres termes, nous construisons un modèle plus général, dont le premier scénario en serait un cas particulier, à savoir le plus standard où tous les individus seraient séparés de la même distance (ou plus exactement où tous les individus auraient la même probabilité de transmission).

Au-delà de modéliser le territoire de la France où se situent nos portefeuilles d'assurés, il a été important de représenter également les populations des pays frontaliers (Allemagne, Belgique, Espagne, Italie, etc.) afin de ne pas fonctionner dans un système fermé et tenir compte de la densité réelle de la population autour des départements aux extrémités de la France. C'est un élément déterminant en effet dans la pertinence de modèle car cela permet d'une part de simuler des pandémies qui frapperaient la France en provenance de pays étrangers et d'autre part, pour un individu infecté près de la frontière, d'autoriser la transmission du virus via ces populations limitrophes plutôt que de restreindre la transmission via les français, ce qui n'aurait pas été réaliste. De même que pour la France, notre maille de répartition est d'un niveau fin puisque nous répartissons les individus sur des communes et selon la démographie réelle de ces pays frontaliers.

Enfin, après avoir testé et garanti la robustesse de notre calibration du modèle, nous avons ainsi cherché à illustrer la pertinence de l'intégration de ce critère géographique en appliquant notre modèle à des portefeuilles fictifs d'assureurs (aux répartitions géographiques d'assurés différentes) pour en mesurer les éventuelles différences dans les résultats de sinistralités obtenus.

Notre modèle montre ainsi que la région Ile-de-France est la plus touchée (au sens où le taux de surmortalité induite par une pandémie est le plus élevé). Si ce résultat était prévisible compte tenu de sa densité de population nettement supérieure au reste de la France, notre modèle nous donne des indicateurs quantitatifs (comparaison du taux de surmortalité d'une région à la moyenne nationale par exemple), et nous fournit également des résultats moins évidents : la région Centre-Val-de-Loire par exemple, bien qu'à faible densité de population, a un taux de surmortalité supérieur à la moyenne nationale, c'est sa proximité avec l'Ile-de-France qui explique ce résultat.

Au-delà d'avoir apporté cette dimension supplémentaire aux modèles épidémiologiques déjà existants avec le facteur géographique, une force importante de notre modèle réside dans sa souplesse

d'utilisation et sa capacité à être adapté à d'autres territoires ou à être étendu, par exemple à l'Europe voire au monde entier. De même, il peut justement être utilisé sur d'autres types de pandémies car les taux de reproduction et létalité peuvent aisément être ajustés pour étendre le modèle.

Ainsi, il permet d'être une base robuste pour un assureur désireux de mieux maîtriser son risque pandémique, ce qui est particulièrement utile dans des cas de tarification (que ce soit en assurance directe lors de la réponse à un appel d'offres pour assurer une entreprise par exemple ou que ce soit en réassurance pour un portefeuille plus large) mais également pour construire un modèle interne de solvabilité.

Mots clés : pandémie, risque pandémique, prévoyance, assurance de personnes, modèle interne, tarification, géographie, démographie.

Abstract

There are several models in the literature to enable insurers (among other actors) to better understand the pandemic risk. These models can be separated into three main approaches: the scenario approach (reappearance of a historical pandemic, example 1918), the statistical estimation approach (frequency - severity), the approach via an epidemiological model.

It is this last approach which makes it possible to reproduce as closely as possible the effects of an epidemic by simulating its propagation within a population.

The purpose of this presentation is to develop two epidemiological models of pandemic diffusion.

For the former, our work consisted in coding the most widespread epidemiological model in the literature: the SEIR model whose acronym refers to the 4 distinct states in which an individual may be in an epidemic: Susceptible / Exposed / Infectious / Recovered.

We started by testing our model parameters by running it on a population of 10,000 individuals and a spectrum of 100,000 epidemics, these being the combination of 100 reproduction rates and 1,000 different case fatality rates. Reproduction rate is a metric that characterizes the ability of an epidemic to spread with more or less virulence within a population, regardless of its lethal nature. In this standard case, each pair of individuals therefore always has the same probability of transmitting the virus.

However, this model has a detrimental limit for the exploitation by an insurer of the results produced: it does not take into account the geographical distribution of the individuals in the portfolio.

We wanted to develop a second epidemiological model that takes into account this geographical parameter, that is to say, the location of individuals within a given territory.

To do this, we had to delimit a territory and dispose the individuals in a relevant way. We tested our model on the perimeter of metropolitan France and used real data from French demography to allocate individuals so that our model is built on the best assumptions. We tested three levels of mesh for the disposition of the individuals since we proceeded to a distribution by administrative regions, then by departments and finally by communes. We chose this last level of distribution grid which is the finest and thus showed the most convincing results.

In this second model, the probability of transmission of the virus between two individuals is no longer be constant but decreasing with the geographical distance that separates them. This phase enabled us to calibrate the parameters of our geographical model so that, under the same conditions of the first scenario, the results are also similar in terms of number of infected, epidemic time, etc.

Beyond modeling the territory of France where our portfolios of policyholders are located, it was important to also represent the populations of the bordering countries (Germany, Belgium, Spain, Italy, etc.) in order not to function in a closed system and take into account the real density of the population around the departments in the borders of France. This is a decisive factor in the relevance of our model because it enables on the one hand to simulate pandemics that would hit France from foreign countries and on the other hand, for an infected individual near the border, enable transmission of the virus via these neighboring populations rather than restricting transmission via the individuals based in France, which would have not been realistic. As for France, our distribution grid is of a fine level since we locate individuals into cities and according to the actual demography of these border countries.

Eventually, after having tested and guaranteed the robustness of our model calibration, we have sought to illustrate the relevance of the integration of this geographical criterion by applying our model to fictitious insurers' portfolios to measure the discrepancies in the results.

Our model shows us that the « Ile-de-France » region is the most affected (in the sense that the excess mortality rate induced by a pandemic is the highest). If this result was predictable given its density far above the rest of France, our model gives us quantitative indicators (of which a comparison between the excess mortality rate of a region and the national average rate), and also provides us with less obvious results: the « Centre-Val-de-Loire » region, for instance, although with a low population density, has an excess mortality rate higher than the national average, it is its proximity to the Ile-de-France region which explains this result.

Beyond adding this additional dimension to existing epidemiological models with the geographical factor, an important strength of our model lies in its flexibility of use and its ability to be adapted to other territories or regions, for example to Europe or even to the whole world. Similarly, it can be used for other types of pandemics because reproduction and lethality rates can easily be adjusted to extend the model. It thus provides a robust basis for an insurer wishing to better control its risk. This is particularly useful in cases of underwriting (whether in direct insurance when responding to a call for tender to insure a corporate business for example or in reinsurance for a larger portfolio) but also to build an internal model of solvency.

Key words: pandemic, pandemic risk, foresight, life insurance, internal model, pricing, geography, demography.

Remerciements

Parce que ce mémoire a été le fruit de longs mois de travail mais qu'il n'aurait pas été le même sans l'aide de nombreuses personnes, nous tenons à remercier sincèrement :

- Olivier Della Santina, notre tuteur, pour avoir accepté de nous encadrer, nous avoir fourni des conseils francs et précieux ;
- Nicolas Neuville, Nicolas Richer et Bruno Laporte pour nous avoir apporté les éléments techniques indispensables ;
- Noémie Boccardi et Alice Hunaut pour leur soutien sans faille et leurs encouragements permanents malgré des calendriers professionnels et personnels très chargés ;
- Le personnel enseignant et administratif du CEA pour cette formation extrêmement riche, intense et bénéfique pour nos carrières respectives ;
- Les membres du jury qui ont accepté de recevoir et juger notre travail.

Synthèse	2
Abstract	5
Remerciements	7
1. Introduction	11
2. Notion de pandémie	14
2.1. Définition et caractéristiques	14
2.2. Quelques exemples historiques	15
2.2.1. La peste noire (1349 – 1351)	16
2.2.2. La grippe espagnole / virus H1N1 (1918 – 1919).....	16
2.2.3. La grippe asiatique / virus H2N2 (1957 – 1958).....	17
2.2.4. La grippe de Hong-Kong / virus H3N2 (1968 – 1969).....	18
2.2.5. La grippe A / virus H1N1 (2009 – 2010).....	18
2.2.6. Ebola (2013 – 2016)	18
2.3. Les pandémies : des maladies émergentes.....	19
2.3.1. Classification des différents types de pandémies	19
2.3.2. Chronologie des crises sanitaires récentes.....	20
2.3.3. Emergence d'une pandémie.....	20
2.3.4. Facteurs favorisant la propagation d'une pandémie	22
2.3.4.1. Densité et mouvements de population	22
2.3.4.2. Propagation par les moustiques	23
2.3.4.3. Commerce mondial des aliments	24
2.3.4.4. Bioterrorisme	24
2.4. Impact d'une pandémie	25
2.4.1. Mortalité	25
2.4.2. Impact économique	26
2.5. Contrôle d'une pandémie	27
2.6. Quelle serait la plus grande menace ?.....	28
2.7. Peut-on s'y préparer ?	29
2.7.1. Caractère imprévisible des pandémies	30
2.7.2. Surveillance mondiale sous l'égide de l'OMS	30
2.7.3. Progrès technologiques dans l'identification des nouveaux virus	30
2.7.4. Plans de préparation, exercices de simulation	31
2.8. D'autres pandémies dans un futur proche ?	31
3. Appréhension du risque pandémique dans le monde de l'assurance	33
3.1. Risques au passif pour l'assureur et garanties impactées.....	33
3.1.1. En cas de décès	33
3.1.2. En cas de vie	33
3.1.3. Frais de soin	33

3.1.4.	Arrêt de travail (incapacité / invalidité).....	33
3.1.5.	Garanties Dommages.....	34
3.2.	Risques à l'actif.....	34
3.2.1.	Risque de marché.....	34
3.2.2.	Risque de crédit.....	34
3.2.3.	Risque opérationnel.....	35
3.3.	Mesure du risque dans Solvabilité 2.....	35
3.3.1.	Description rapide de la directive et de ses trois piliers	35
3.3.2.	Formule standard ou modèle interne.....	36
3.3.3.	Focus sur le risque Pandémie	37
3.4.	Réassurance : Le BCAC comme principale couverture existante	38
3.4.1.	Origine de la couverture	38
3.4.2.	Caractéristiques.....	38
3.4.3.	Limites de la couverture BCAC	42
4.	Etat de l'art.....	44
4.1.	Modèles actuariels (approche de masse)	44
4.2.	Modèles épidémiologiques (approche individuelle).....	44
4.2.1.	Introduction et grands principes	44
4.2.2.	Les différents états	45
4.2.3.	Facteurs de développement.....	45
4.2.4.	Les différents modèles compartimentaux déterministes	46
4.2.4.1.	Le modèle SI.....	46
4.2.4.2.	Le modèle SIR.....	46
4.2.4.3.	Le modèle SEIR.....	46
5.	Développement de notre modèle.....	47
5.1.	Mode opératoire.....	47
5.1.1.	Construction d'un réseau.....	47
5.1.2.	Utilisation du modèle SEIR stochastique.....	49
5.1.2.1.	Temps d'incubation et d'infectiosité.....	51
5.1.2.2.	Transmission d'un virus entre deux individus.....	54
5.1.3.	Le taux de reproduction de base R_0	55
5.1.4.	Le taux de létalité.....	56
5.1.5.	Dimensionnement de ces taux	58
5.1.5.1.	Taux de reproduction R_0	58
5.1.5.2.	Taux de létalité	60
5.1.6.	Fonctionnement du package « <i>epinet</i> » dans R.....	61
5.1.7.	Modèle stochastique.....	61
5.2.	Synthèse du mode opératoire	62

5.3.	Scénario 1 : cas simple, sans facteur géographique.....	63
5.4.	Mesure de l'influence du R0 de l'épidémie sur le pourcentage d'individus infectés	66
5.5.	Mesure du taux de surmortalité	67
5.6.	Mesure de la sinistralité décès de l'assureur	68
6.	Modélisation de la propagation spatiale de l'épidémie	70
6.1.	Disposition des individus sur le territoire	70
6.1.1.	Répartition par région ou département.....	70
6.1.1.1.	Modélisation de la France	70
6.1.1.2.	Modélisation des pays frontaliers	74
6.1.2.	Répartition par commune	76
6.1.2.1.	Modélisation de la densité de population en France	78
6.1.2.2.	Modélisation de la densité de population des pays frontaliers.....	78
6.1.2.3.	Résultat : densité de population ainsi obtenue.....	79
6.2.	Prise en compte de la distance	80
6.2.1.	Rappel du fonctionnement de base.....	80
6.2.2.	Comment s'assurer du correct dimensionnement des p_{ij} ?	81
6.3.	Scénario 2 : avec facteur géographique / assurés répartis de la même façon que la population totale.....	84
6.4.	Scénario 3 : avec facteur géographique / assurés selon des répartitions particulières	89
6.5.	Mesure de la sinistralité décès de l'assureur	90
6.6.	Construction d'un module de « data visualisation »	95
7.	Perspectives	97
7.1.	Avantages et limites de notre modèle	97
7.1.1.	Avantages.....	97
7.1.2.	Limites	97
7.2.	Mesure du risque	99
7.3.	Tarification (assurance directe et réassurance)	99
7.4.	Autres perspectives	99
8.	Conclusion	101
9.	Bibliographie	103
10.	Glossaire	104

1. Introduction

Le terme « pandémie » vient du grec ancien $\pi\acute{\alpha}\nu$ / $p\acute{\alpha}\nu$ « tous », et $\delta\eta\mu\omicron\varsigma$ / $d\acute{\epsilon}\mu\omicron\varsigma$ « peuple » et désigne une épidémie qui se répand sur une large zone géographique et touche une partie importante de la population. C'est une définition relativement large et dont les frontières sont floues, elle est d'ailleurs sans cesse revue et discutée par de nombreux acteurs, dont l'OMS¹, un des principaux experts dans ce domaine. Dans le langage courant, tout le monde s'accorde à dire qu'il s'agit en tout cas d'une maladie à grande échelle et qui a des conséquences importantes.

En effet, plusieurs exemples historiques, dont certains extrêmement récents voire actuels, nous prouvent que la menace est bien réelle et importante. Nous pouvons notamment citer les deux exemples peut-être les plus connus : la grippe espagnole d'une part, qui a sévi essentiellement en 1918 et 1919 et Ebola d'autre part, qui a provoqué de lourds dégâts entre 2013 et 2016 avec quelques réapparitions en 2017 et 2018. Ces cas de pandémie montrent les conséquences dramatiques sur les plans économique, social et sanitaire que cela pouvait engendrer.

Au travers de nos différentes lectures, recherches et expériences professionnelles, il apparaît que ce risque, bien qu'il soit envisagé et mesuré, n'est pas le plus surveillé et n'est pas le plus mentionné et considéré par les assureurs, comparativement aux autres types de risques auxquels ils sont soumis.

A titre d'exemple, nous avons obtenu le CRP (« Comprehensive Risk Profile ») 2017/18 d'un assureur français de taille moyenne, qui répertorie les 15 risques considérés comme les plus importants pour cet acteur, classés selon leur estimation de fréquence d'occurrence et leur estimation d'impact en cas de réalisation (sur le résultat, la solvabilité, le risque d'image, etc.). Le risque pandémique figure bien dans cette liste du CRP mais son positionnement est intéressant :

- La fréquence d'occurrence est estimée comme étant très faible (tous les 10-30 ans), ce qui le place dans les fréquences les plus faibles, au même niveau par exemple que le risque de réputation ;
- Son impact est en revanche estimé comme important, ce qui le place au même niveau que les risques réglementaires actuels (IDD, PRIIPS, contrats responsables, clauses de désignation) mais en-dessous des risques de marché (crise sur les marchés actions, crise de crédit en Europe)

Cette hiérarchie relève d'une évaluation de la part du management et peut nous interpeller ou en tout cas susciter des questions, notamment concernant son impact qui est estimé être du même ordre de grandeur que les réformes réglementaires qui animent le marché de l'assurance aujourd'hui. Nous pouvons alors imaginer que, tant que les différents managements de la profession n'ont pas réellement expérimenté et fait face à de très sévères pandémies (plus sévères que celles qui ont déjà existé) le risque pandémique restera surveillé et appréhendé comme il l'est aujourd'hui, sans que l'on observe une croissance de son exigence en couverture.

Nous pouvons d'ailleurs faire le parallèle avec le risque terroriste qui a véritablement émergé au moment des très graves attentats qu'ont été ceux du World Trade Center à New-York le 11 septembre 2001. C'est à partir de cet événement tragique que le risque terroriste a été véritablement appréhendé, de façon beaucoup plus significative qu'auparavant, que ce soit par les gouvernements de façon générale (accords internationaux, traques militaires, mesures de sécurité dans les aéroports, etc.) et par les assureurs d'un point de vue plus spécifique (notamment les réassureurs).

1

<http://archive.wikiwix.com/cache/?url=http%3A%2F%2Fwww.arretsurimages.net%2Fcontenu.php%3Fd%3D2264>

Pourtant, les impacts d'une pandémie sont à plusieurs niveaux puisqu'elle peut engendrer une augmentation des décès, des frais de santé (consultations médicales, dépenses pharmaceutiques) et des arrêts de travail (incapacité à se rendre sur son lieu de travail et/ou de travailler). Une pandémie peut même affecter les garanties dommages à travers la perte d'exploitation agricole (champs d'exploitation détruits par le virus ou la bactérie).

Pour les assureurs, ces risques sont autant de risques biométriques auxquels ils pourraient avoir à faire face, ce qui représente un coût certain d'indemnisation. Mais il ne faut pas omettre les risques opérationnels auxquels ils seraient eux-mêmes soumis via les arrêts de travail de leurs propres salariés ainsi que les risques collatéraux se produisant sur les marchés financiers (risque de crédit, de liquidité, écartement des spreads, etc.).

Mais comment mesurer et tenter de prévenir les prochaines pandémies ? Quelles sont les plus grandes menaces ? Peut-on faire en sorte de mieux prévoir leur occurrence et/ou leur propagation ? Comment est-ce que ce risque va se développer le bioterrorisme ? Ces questions trouvent difficilement leurs réponses aujourd'hui mais plusieurs signes sont encourageants et ont tendance à montrer que les choses évoluent favorablement. Nous pouvons citer en outre :

- les différents programmes de surveillance mondiale, sous l'égide de l'OMS et d'autres organisations mondiales ;
- les progrès technologiques qui enrichissent notre connaissance, permettent de continuer à recenser toujours plus de virus/bactéries et d'identifier toujours plus rapidement leurs caractéristiques ;
- les plans de préparation et exercices de simulation mis en place à travers le monde, ce qui aboutit à des réactions plus rapides de la part des organisations et permet d'apporter des réponses le plus tôt possible après la survenance d'une pandémie, afin d'endiguer au maximum sa propagation.

Aujourd'hui, les assureurs n'ont pas une multitude d'options pour se couvrir spécifiquement contre le risque pandémique. Il existe une solution de marché des assureurs collectifs en prévoyance en France, initiée en 2015 par le BCAC mais qui ne couvre que le risque Décès. Autrement, la principale alternative choisie par les assureurs est la souscription individuelle à une couverture de réassurance en stop loss.

Au-delà des solutions individuelles de réassurance, le référentiel Solvabilité 2 n'impose pas non plus aux assureurs une forte exigence en fonds propres puisque sa formule standard (qui repose sur un modèle de Swiss Re (2007)) est assez simpliste et présente quelques limites, notamment le fait de séparer les scénarii de stress sur les garanties Décès et Santé.

En termes de modélisation, il existe aujourd'hui plusieurs modèles épidémiologiques, déterministes et stochastiques, dont le modèle SEIR (Susceptible, Exposé, Infectieux, Rétabli) est un des plus répandus tout en étant relativement simple d'utilisation. Ce modèle distingue la population en 4 états selon lesquels peut se trouver un individu au cours d'une épidémie. Il simule ensuite la propagation d'une épidémie en fonction des contacts (voies de transmission du virus) qui peuvent exister au sein de ce réseau d'individus.

Les modèles SEIR que nous avons vus être appliqués dans nos recherches et notre bibliographie fonctionnent sur un modèle théorique où tous les individus sont dans un seul et même ensemble (comme s'ils étaient dans une boîte noire) et chaque paire d'individus a la même probabilité de se transmettre le virus que n'importe quelle autre paire. Ils ne tiennent pas compte de critères propres à des individus ou à des paires d'individus, qui existent pourtant dans la réalité et sont déterminants.

Nous pensons en particulier au paramètre de la distance géographique entre personnes, que ce soit au sein d'un même département, région, pays ou même de façon plus générale dans le monde avec

la séparation des personnes. En effet, il est évident qu'un virus se transmet beaucoup plus facilement entre deux individus si ceux-ci sont voisins plutôt que si l'un vit à l'opposé du globe par rapport à l'autre.

Ainsi, ce paramètre de distance géographique nous a paru essentiel et très important à prendre en compte dans la modélisation d'une épidémie, et cela présente un intérêt tout particulier pour un assureur car cela lui permet de mesurer le risque qui lui est propre et donc de mieux le maîtriser, en particulier lorsqu'il a un portefeuille d'assurés réparti de façon atypique par rapport au milieu (en général le pays) dans lequel il exerce.

Nous pouvons citer deux exemples dans lesquels cet intérêt peut se manifester de façon concrète :

- un assureur plutôt régional et dont le portefeuille est concentré dans une certaine partie de la France, souhaite se réassurer à titre individuel : il a alors tout intérêt à savoir si sa zone de couverture est plutôt surexposée ou sous-exposée au risque pandémique par rapport à la moyenne nationale (ce qui lui permet donc de savoir s'il est gagnant ou perdant en souscrivant à la couverture BCAC, tarifée selon la mortalité nationale) ;
- un assureur national avec un portefeuille réparti plutôt de façon uniforme sur le territoire et qui répond à un appel d'offre grands comptes pour assurer / reprendre un portefeuille d'assurés (particuliers et/ou entreprises) d'une certaine région de la France : le modèle pandémique avec paramètre géographique lui permettra de savoir si c'est un portefeuille qui se situe dans des régions plus ou moins exposées au risque pandémique et donc évaluer la future sinistralité.

Ce mémoire a donc pour ambition de présenter la solution que nous avons construite en partant du modèle SEIR et en y intégrant le paramètre géographique pour la propagation d'une pandémie, de sorte à mesurer en France (périmètre de notre étude) les différences d'exposition par région au risque pandémique.

Pour cela, nous commencerons par dresser l'état des lieux des enjeux qui gravitent autour de la notion de pandémie en présentant plusieurs exemples historiques, les facteurs qui favorisent ou non une pandémie, les grandeurs qui la caractérisent et évaluer quelles pourraient être les prochaines menaces dans le futur. Puis, nous présenterons l'état de l'art en termes de modèles actuariels / épidémiologiques, nous rappellerons les risques auxquels doivent faire face les assureurs en cas de pandémie, comment Solvabilité 2 propose de se couvrir et les solutions de réassurance qui existent aujourd'hui. Enfin nous expliquerons le fonctionnement du modèle SEIR et comment nous avons procédé pour intégrer le paramètre géographique et dimensionner les différents paramètres afin de refléter au maximum la réalité géographique de la France en termes de répartition des individus sur le territoire. Notre modèle sera éprouvé sur plusieurs scénarii et plusieurs répartitions afin d'assurer sa robustesse et présenter des résultats directement utilisables pour un assureur.

2. Notion de pandémie

2.1. Définition et caractéristiques

Commençons par définir précisément le terme de « pandémie » afin de comprendre ce qu'il recoupe ainsi que la différence avec les termes « épidémie » et « endémie ». En effet, si le terme d'épidémie est plutôt bien connu et cerné par le grand public, l'endémie et la pandémie sont deux autres notions et l'on a souvent du mal à expliquer clairement ce qui les différencie.

Une **épidémie** désigne le développement et la propagation rapide d'une maladie à un endroit et un instant donné. Dans le langage courant, on parle généralement d'épidémie lorsque la maladie est une maladie infectieuse, c'est-à-dire une maladie transmise par un agent infectieux ou un micro-organisme (bactérie, virus, champignon, parasite, etc.) comme la grippe, mais cela peut concerner d'autres types de maladie comme l'obésité, le suicide ou l'addiction aux smartphones par exemple.

On parle ensuite d'**endémie** lorsqu'une maladie persiste dans une zone géographique limitée, soit parce qu'elle y reste en permanence, soit parce qu'elle y revient à des époques semblables. Sa présence est donc connue, signalée mais cela ne signifie pas qu'elle soit en progression ou qu'elle se répande, contrairement à l'épidémie. Ainsi, dans certains pays où une endémie est présente, une vaccination peut être nécessaire pour les visiter. On peut citer le paludisme dans certaines zones tropicales, la fièvre jaune en Afrique / Amérique du Sud ou encore l'hépatite A en Thaïlande.

Enfin, on parle de **pandémie** lorsqu'une maladie touche une partie très importante de la population, ce qui est très souvent synonyme de propagation internationale. Citons quelques-unes des plus grandes pandémies de maladies infectieuses vécues par l'humanité : la peste noire, le choléra, la grippe, la tuberculose, la variole ou encore plus récemment le SIDA. Les exemples historiques les plus cités sont détaillés un peu plus loin dans ce mémoire.

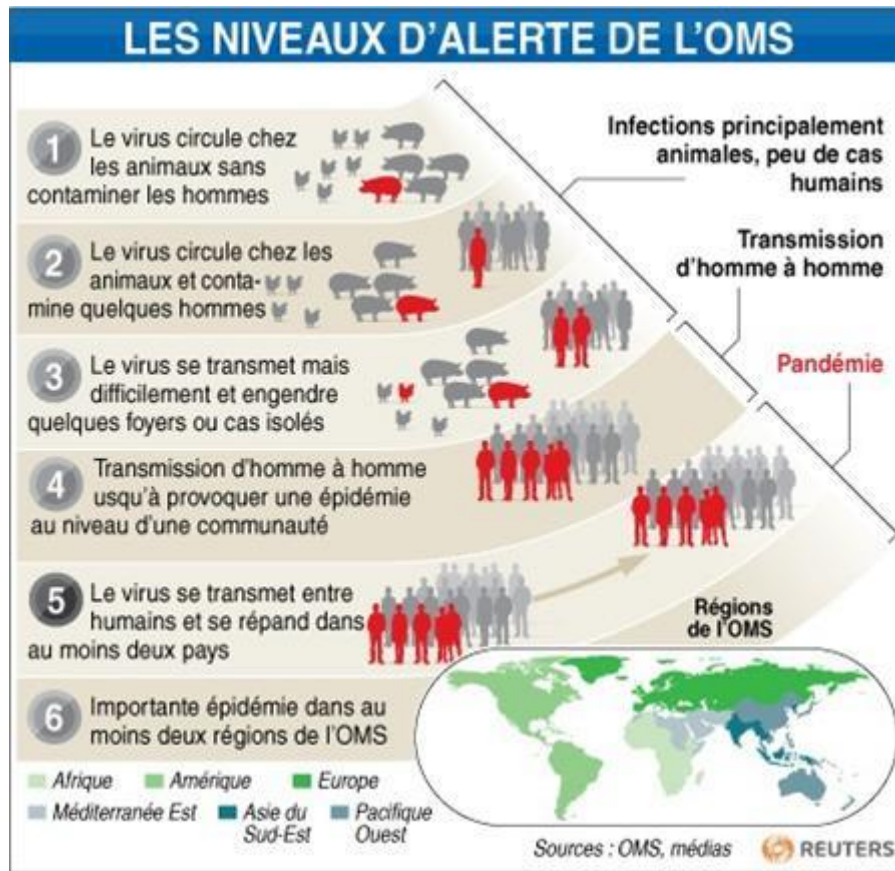
Des questions légitimes se posent alors : quand parle-t-on de pandémie et plus d'épidémie ? Existe-t-il des seuils et si oui, quels sont-ils ? Est-ce qu'il y a une instance chargée de qualifier officiellement une maladie en pandémie ?

De façon intuitive, c'est bien sûr l'OMS qui, en collaboration avec ses Etats Membres, s'occupe de plusieurs activités telle que la coordination de l'action internationale, la qualification de la phase dans laquelle se trouve la pandémie en cours, l'assistance aux populations, la sélection de la souche du vaccin antipandémique et sa production, ainsi que la synthèse de toutes les données sur la pandémie.

L'OMS propose d'ailleurs sa propre définition d'une pandémie : « les pandémies sont des événements rares mais itératifs, invariablement associés à une grande morbidité, à une mortalité importante, et à une perturbation considérable de la vie sociale et économique. La vulnérabilité de la population, conjuguée à la très grande contagiosité des virus grippaux, a pour conséquence une propagation rapide dans le monde entier, en général en moins d'un an. »

Pour mesurer la sévérité d'une pandémie l'OMS a défini une échelle à six niveaux d'alerte :

➤ **Les niveaux d'alerte de l'OMS :**



Pour citer quelques exemples :

- pour l'épidémie de grippe aviaire H5N1 en France en 2005 le niveau d'alerte était de 3 : la transmission d'homme à homme n'avait pas lieu, et le nombre d'infection était limité.
- pour la grippe A(H1N1) de 2009, le niveau d'alerte s'élevait à 6 à l'échelle mondiale, soit le niveau maximal.

2.2. Quelques exemples historiques

Dans l'historique de notre civilisation, plusieurs pandémies sont survenues. Même s'il est difficile pour les organisations d'avoir un inventaire exhaustif et précis (notamment pour cause de manque de données fiables à certaines périodes), les exemples ci-dessous sont ceux qui sont le plus souvent cités et que nous avons croisé le plus souvent dans nos recherches et lectures, car elles sont à la fois récentes et ont significativement marqué notre population et nos modes de vie.

2.2.1. La peste noire (1349 – 1351)

La peste noire a frappé l'Europe au milieu du 14^{ème} siècle, cette pandémie de peste est partie de Chine amenée par les Mongols en Crimée qui faisaient le siège du port Génois de Caffa. Quand les hostilités se sont terminées, les bateaux génois sont repartis et ont emmenés avec eux les hommes, les rats et les puces.

Cette peste noire est due à une bactérie : le bacille qui vit chez des rats qui peut être transmis à l'homme via les puces des rats.

A partir de cette contamination à Caffa, la bactérie de la peste a voyagé avec ces rats et ces puces dans les navires génois et est arrivé progressivement dans tous les ports de la méditerranée. La propagation de la maladie s'est faite en 3-4 ans vers le nord de l'Europe. Cette progression s'est accompagnée d'un nombre de décès considérable. S'il n'existe pas de statistiques très fiables sur les données disponibles de l'époque, on estime toutefois qu'un tiers de la population des villes touchées en Europe est morte pendant le passage de la peste noire.

2.2.2. La grippe espagnole / virus H1N1 (1918 – 1919)

Un des exemples les plus cités par les spécialistes pour illustrer le caractère meurtrier d'une pandémie est la grippe de 1918, célèbre pour être la première grande pandémie de l'ère moderne et pour sa virulence très forte. Elle est aussi appelée « grippe espagnole », non pas en raison de son origine mais parce que l'Espagne fut le premier pays à être entendu par l'opinion publique sur les ravages et la portée internationale de cette pandémie, tandis que d'autres pays avaient déjà été touchés (notamment la France) mais faisaient davantage écho aux faits de guerre (l'Espagne était neutre).

Bien que l'origine ne soit pas complètement certaine, le virus père serait vraisemblablement d'origine aviaire et apparu en Chine pour la première fois, avec des caractéristiques assez communes en particulier une faible mortalité : la période d'incubation est de 2 à 3 jours et est suivie par 3 à 5 jours de symptômes, en particulier de la fièvre et un affaiblissement important des défenses immunitaires.

En revanche, sa contagiosité plus forte que les épidémies habituelles et en période de guerre a provoqué sa propagation très rapide aux Etats-Unis puis en Europe et dans le reste du globe, en raison des mouvements de troupes entre continents. Elle s'est manifestée par 3 vagues de mortalité, la 1^{ère} ayant lieu au printemps 2018 avec le virus père puis les 2 autres de septembre à décembre 2018 et de février à mai 2019 avec le virus de souche H1N1, mutation du virus père et plus létal que ce dernier.

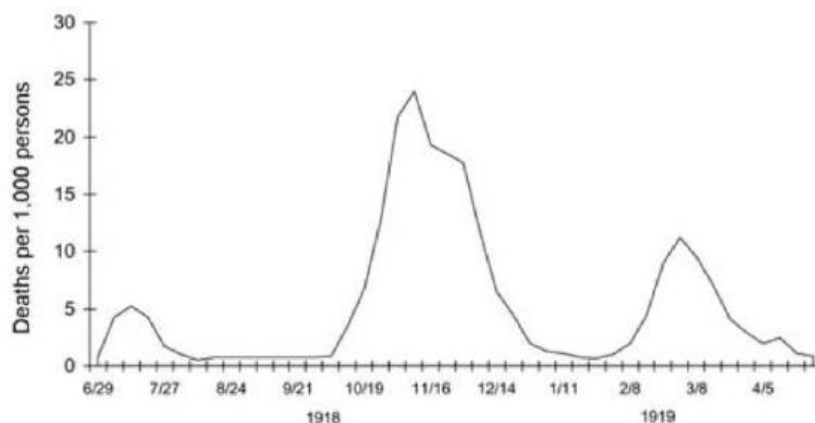
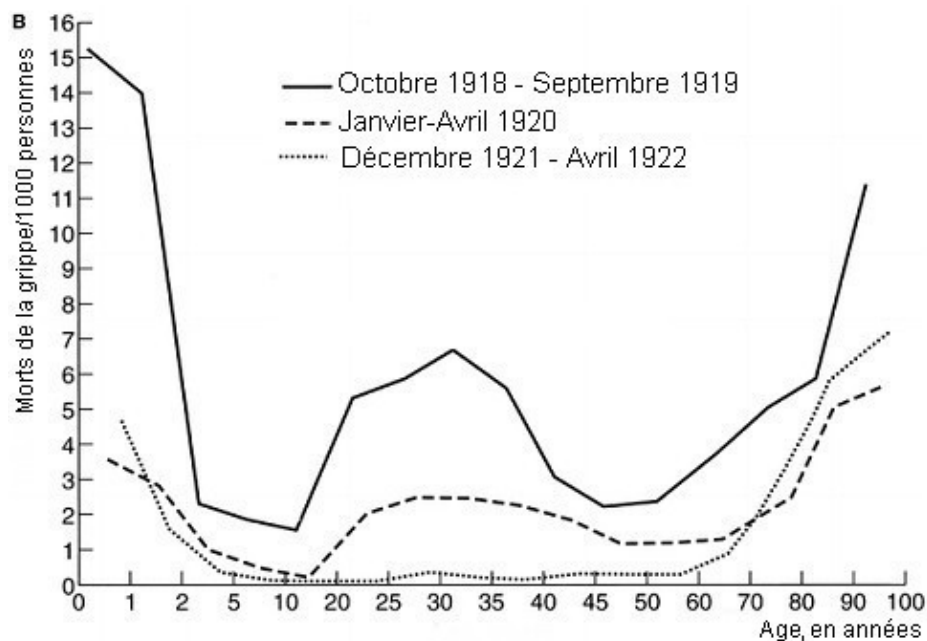


Figure 1. Three pandemic waves: weekly combined influenza and pneumonia mortality, United Kingdom, 1918–1919 (21).

En termes de mortalité, le bilan reste aujourd'hui incertain en raison de la période de guerre et faute d'outils et de méthodes statistiques suffisamment fiables à l'époque. Il semble néanmoins se centrer entre 20 et 50 millions de morts, même si certaines sources estiment qu'il pourrait être plutôt proche des 100 millions². Quoiqu'il en soit, avec une population mondiale proche des 2 milliards à l'époque, on estime un taux de morbidité autour de 50%.

Enfin, notons le caractère atypique de la répartition des décès par âge de la grippe. La courbe de mortalité d'une grippe standard est en « U », traduisant le fait que les jeunes enfants et personnes âgées sont généralement les personnes les plus touchées car plus fragiles. Celle de la grippe espagnole est en « W » avec un 3^{ème} pic de mortalité sur la tranche des jeunes adultes (autour de 30 ans), qui s'explique par la proximité de ces tranches de population dans les ateliers, casernes et tranchées et leur forte présence dans les flux migratoires de l'époque, pour des raisons professionnelles ou de guerre.



La forte contagiosité du virus et la période de guerre (en particulier la proximité et la mobilité des populations) sont les 2 principaux éléments qui, par corrélation, ont permis d'aboutir au bilan dévastateur de la grippe espagnole que l'on connaît aujourd'hui. Ces facteurs seront déterminants dans l'étude de notre modèle de propagation.

2.2.3. La grippe asiatique / virus H2N2 (1957 – 1958)

La grippe asiatique a été transportée par le virus H2N2, issu d'une mutation d'un virus humain chez le canard. Les premiers cas ont été identifiés en Chine en février 1957 et se sont ensuite propagés jusqu'aux Etats-Unis en juin de la même année. Son origine en partie humaine a permis d'atténuer la virulence de ses effets sur la population et le bilan est estimé entre 1 et 2 millions de morts selon l'OMS, certaines sources allant jusqu'à 4 millions.

² <https://www.ncbi.nlm.nih.gov/pubmed/11875246?dopt=Abstract>

2.2.4. La grippe de Hong-Kong / virus H3N2 (1968 – 1969)

Le virus H3N2 est une évolution du virus H2N2 (on parle de « virus réassorti ») et cause une nouvelle pandémie de grippe, 10 ans après la grippe asiatique. Comme la précédente, elle démarre de Chine en juillet 1968 et s'étend rapidement en Asie du Sud-Est, en Inde, en Australie puis aux Etats-Unis et en Europe. Le bilan est chiffré autour du million de morts dans le monde.

2.2.5. La grippe A / virus H1N1 (2009 – 2010)

La grippe A a été officiellement reconnue par l'OMS le 11 juin 2009 comme la première pandémie du XXI^e siècle. Elle serait vraisemblablement apparue en Amérique du Nord, soit dans des élevages industriels de porcs au Mexique soit via l'import de porcs d'Asie (la source est discutée, comme souvent pour les pandémies).

Si l'on observe le bilan global, son taux de mortalité a été assez faible et même inférieur à une grippe saisonnière. En revanche, le taux d'atteinte a été très élevé et supérieur à celui des gripes observées jusqu'à présent et par ailleurs, les taux de mortalité ont été très disparates selon les pays et catégories de population. Fait surprenant qui a d'ailleurs inquiété l'OMS, les populations aisées et d'ordinaire moins fragiles ont été cette fois atteintes à des niveaux importants.

La fin de la pandémie est décrétée par l'OMS en août 2010, après des polémiques en janvier 2010 sur le potentiel lobbying de l'industrie pharmaceutique pour surestimer la menace de la pandémie³.

2.2.6. Ebola (2013 – 2016)

Bien qu'il ne s'agisse pas d'une pandémie car sa portée n'a pas été complètement internationale, le virus Ebola est la dernière épidémie en date à avoir eu une ampleur et une sévérité très importantes, faisant notamment surgir la crainte d'une nouvelle pandémie.

Le virus a été découvert en 1976 en Afrique Centrale lorsqu'il s'est manifesté simultanément à Nzara et à Yambuku, cette dernière ville étant située près de la rivière Ebola, qui a donné son nom au virus. Un jeune médecin belge de 27 ans, Peter Piot, avait alors réussi à contenir cette première poussée du virus, qui a fait près de 300 morts⁴.

En décembre 2013, le virus se manifeste de nouveau mais va s'avérer beaucoup plus meurtrier par la suite. Il apparaît dans le Sud-Est de la Guinée, près de la frontière avec le Libéria et la Sierra Leone. Pendant les premiers mois de 2014, le virus frappe la région Sud de la Guinée puis arrive dans les 2 pays limitrophes en avril et mai, le nombre de décès s'accélère considérablement. Le premier cas hors d'Afrique est confirmé le 30 septembre 2014⁵. La gestion du risque pandémique se fait à l'échelle mondiale sous l'égide de l'OMS, qui met fin à l'état d'urgence de santé publique de portée internationale le 29 mars 2016⁶.

³ http://www.lemonde.fr/epidemie-grippe-a/article/2010/01/27/l-oms-somme-de-s-expliquer-sur-la-grippe-a_1297307_1225408.html#ens_id=1185166

⁴ http://www.lemonde.fr/planete/article/2014/08/10/1976-a-la-decouverte-du-virus-ebola_4469215_3244.html

⁵ <http://www.lefigaro.fr/flash-actu/2014/09/30/97001-20140930FILWWW00458-ebola-premier-cas-d-infection-diagnostique-aux-etats-unis.php>

⁶ <http://www.who.int/mediacentre/news/statements/2016/end-of-ebola-pheic/fr/>

Le dernier rapport de situation⁷ fait état de 28 616 cas confirmés, probables et suspects ont été notifiés en Guinée, au Libéria et en Sierra Leone, dont 11 310 décès soit un taux de létalité de 40%. Dans de très petites proportions, le virus aura atteint d'autres pays africains comme le Mali, le Nigéria et le Sénégal mais également d'autres continents avec des cas aux Etats-Unis, au Royaume-Uni, en Espagne et en Italie.

2.3. Les pandémies : des maladies émergentes

2.3.1. Classification des différents types de pandémies

Les pandémies peuvent être classées selon une catégorisation réalisée par l'Institute of Medicine aux Etats Unis en 1992⁸ :

1) Infections passées de l'animal à l'homme et devenues épidémiques en population humaine

Exemples : SIDA, SRAS, EBOLA

Ces infections sont le plus souvent des virus. Ces virus se trouvent au départ chez les animaux et certains de ces virus vont franchir ce qu'on appelle la barrière d'espèce : ils vont alors s'adapter à l'homme et éventuellement pouvoir se transmettre d'homme à homme, et c'est à ce moment-là le potentiel pandémique existe.

Les exemples les plus récents sont le SIDA au début du 20^{ème} siècle (virus initialement présent chez le singe et transmis à l'homme), le SRAS (virus présent chez la chauve-souris « porteur sain » et transmis à l'homme via un autre animal dit « hôte intermédiaire » : la civette) qui a été le modèle de pandémie au début du 21^{ème} siècle puis Ebola (virus également présent chez la chauve-souris et transmis à l'homme via le singe).

2) Infections dues à des pathogènes ayant muté et contre lesquelles il n'existe pas de protection immunitaire (exemple : grippe)

Exemple : Grippe

Ces infections peuvent échapper aux protections immunitaires existantes puisque ayant mutées elles ne sont plus susceptibles d'être neutralisées par des anticorps ou d'autres mécanismes de réponse immunitaire.

L'exemple le plus typique est celui de la grippe : le virus de la grippe circule aussi bien dans la faune aviaire que chez les porcins et éventuellement chez l'homme. Quand un nouveau virus de la grippe apparaît, les personnes n'ont pas d'immunité contre ce nouveau virus et à ce moment-là une pandémie grippale peut démarrer. A chaque fois qu'un nouveau virus de la grippe est généré par ces mécanismes de réassortiments, un tiers de la population mondiale est touché : c'était le cas pour les trois grandes gripes du XX^{ème} siècle (grippe espagnole, grippe asiatique et grippe de Hong-Kong) et également pour la grippe de 2009 (deux milliards de personnes touchées).

⁷ http://apps.who.int/iris/bitstream/10665/208876/1/ebolasitre_2June2016_fre.pdf?ua=1

⁸ « Emerging infections : Microbial threats to health in the United States », Institute of Medicine, 1992

3) Infections ayant récemment gagné de nouvelles zones géographiques

Exemples : fièvre du Nil Occidental, Chikungunya, Zika

La fièvre du Nil occidental (West Nile), arrivée en 1999 aux Etats-Unis, était un virus jusqu'alors présent exclusivement dans les pays méditerranéens, ce virus est arrivé à New-York été 1999 : le virus a touché tout d'abord les oiseaux, puis a été transmis à l'homme via des insectes dits « vecteurs », cette épidémie a balayé en quatre ans les Etats-Unis d'Est en Ouest faisant plusieurs dizaines de milliers de victimes.

Le Chikungunya existait au départ dans l'est de l'Afrique et en Inde, en 2005 il s'est retrouvé à la Réunion, un tiers de la population réunionnaise a été touché avec ces formes invalidantes de douleurs rhumatismales.

Le virus Zika était au départ présent dans les zones rurales de l'Afrique Centrale, ce virus s'est ensuite propagé via des moustiques « vecteurs » (*Aedes Aegypti* notamment : moustique connu pour avoir « colonisé » les zones urbaines et pour avoir véhiculé – avant Zika – le virus de la dengue le virus et celui de la fièvre jaune).

2.3.2. Chronologie des crises sanitaires récentes

Sur un les 40-50 dernières années, on dénombre les crises sanitaires suivantes :

- 1976 : Ebola (Afrique Centrale)
- 1981 : SIDA
- 1996 : Maladie de la « vache folle »
- 1997 : Grippe aviaire
- 1999 : Fièvre du Nil Occidental (Etats-Unis)
- 2003 : SRAS
- 2005 : Chikungunya (La Réunion)
- 2009 : Grippe pandémique A(H1N1)
- 2011 : Escherichia coli 0104 :H4
- 2012 : MERS
- 2014 : Ebola (Afrique de l'Ouest)
- 2015 : Zika

Nous avons donc connu une crise grave tous les 5 ans, causée le plus souvent par un virus.

Comment-est ce que ces virus se transmettent à l'homme, c'est que nous verrons dans le paragraphe suivant.

2.3.3. Emergence d'une pandémie

La transmission du virus à l'homme suit le schéma suivant :

Au départ les virus vivent dans ce qu'on appelle des « **réservoirs animaux** » ou des « **porteurs sains** » : ce sont des chauves-souris, des oiseaux migrateurs, des rats, des animaux qui ne vivent pas nécessairement au contact immédiat de l'homme et qui se sont parfaitement adaptés à la présence de ces virus, c'est-à-dire que ces animaux ne vont pas présenter de symptômes quand ils sont infectés.

Ces virus sont ensuite transmis à l'homme le plus souvent par un autre animal qu'on appelle « **hôte intermédiaire** » ou « **amplificateur** ». Cet animal vit au contact de l'homme et n'est pas habitué à ce virus, dès lors qu'il est contaminé par le porteur sain cet animal présente des symptômes et développe une maladie. Comme cet animal vit au contact de l'homme il y a alors un risque de transmission

Quelques exemples :

- Pour la maladie d'Ebola : les réservoirs animaux est la chauve-souris, les hôtes intermédiaires sont les grands primates (en Afrique centrale avant chaque épidémie d'Ebola on retrouve une épidémie chez les grands primates), les grands primates touchés par Ebola sont malades et deviennent pour les chasseurs de primates des proies plus faciles. Les chasseurs de primates se contaminent alors en découpant la viande de primate et ensuite ces chasseurs amènent la maladie dans leur village.
- Pour la grippe aviaire : les réservoirs animaux sont les oiseaux migrateurs et les hôtes intermédiaires sont les poules ou les porcins, ce virus de la grippe aviaire va alors se retrouver au contact de l'homme via les poules et éventuellement infecter l'homme.
- Pour le SRAS : les réservoirs animaux sont encore la chauve-souris, les hôtes intermédiaires sont les civettes : petits carnivores qui étaient consommés dans les restaurants de Canton et qui ont infecté par voie respiratoire l'homme (personnel et clients du restaurant).

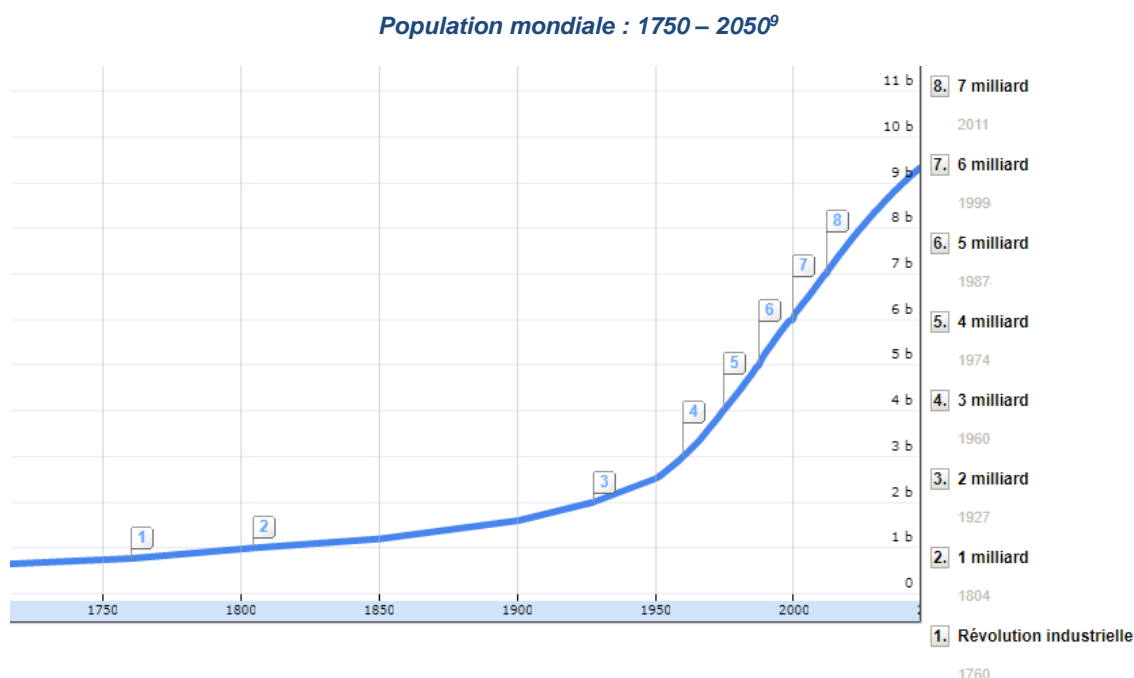
Une fois que le virus a réussi le passage de l'animal à l'homme (franchissement de la barrière d'espèce), il n'est pas nécessairement capable de se transmettre d'homme à homme : tout dépend de la capacité du virus à s'adapter pour pouvoir se propager chez l'homme.

Par exemple, pour la grippe aviaire cette transmission d'homme à homme n'est jamais arrivée : on dénombre plusieurs centaines de cas de transmissions de la poule à l'homme mais aucun cas de de transmission interhumaine.

Lorsque l'homme est contaminé et que la transmission interhumaine est possible, l'ampleur de la propagation de la maladie va dépendre d'un certain nombre de facteurs.

2.3.4. Facteurs favorisant la propagation d'une pandémie

2.3.4.1. Densité et mouvements de population



L'augmentation de la population humaine joue un rôle majeur dans la diffusion des virus à l'origine des pandémies. La population mondiale a été multipliée par quatre au XX^e siècle : nous sommes passés d'un milliard et demi à six milliards et nous sommes déjà à on est déjà 7 milliards aujourd'hui.

La constitution de grandes mégapoles qui pour certaines maladies joue également un rôle important.

Enfin le développement du trafic aérien contribue lui aussi beaucoup à cette diffusion : un virus qui infecterait une personne peut se retrouver à l'autre bout de la planète en quelques jours seulement, c'est ce qu'illustre bien l'exemple du SRAS :

En 2003 le SRAS fait son apparition à Canton mais l'épidémie locale est alors bien maîtrisée, puis le 21 février 2003 un des médecins qui traitait ses patients à Canton, se rend à l'hôtel Métropole de Hong-Kong, et en l'espace de 48 heures infecte dix personnes, ces dix personnes dans la semaine qui suit vont se retrouver dans six pays différents et c'est le démarrage de la grande pandémie mondiale.

⁹ <http://www.worldometers.info/fr/population-mondiale/>

Tournant de l'épidémie de SRAS : février 2003¹⁰



En une semaine ce virus a touché la planète entière. Pour la peste noire il avait fallu six mois pour que les bateaux génois qui partaient de Caffa en Mer Noire arrivent finalement à Marseille...

D'autres vecteurs de transmission que le trafic aérien peuvent permettre à un virus de se propager sur l'ensemble de la planète.

2.3.4.2. Propagation par les moustiques

Les moustiques peuvent aussi servir de vecteurs de transmission des virus, on les appelle « moustiques vecteurs » et ce phénomène est souvent appelé « colonisation vectorielle ».

La fièvre jaune est un exemple de maladie transmise par les moustiques : la maladie qui touchait initialement l'Afrique de l'Ouest a été introduite aux Etats-Unis au XVII^{ème} siècle par le commerce des esclaves : les navires étaient infestés de moustiques (*Aedes Aegypti*) qui portaient cette maladie.

Plus récemment nous avons l'exemple du virus Zika transmis là aussi par les moustiques *Aedes Aegypti* et *Aedes Albopictus*.

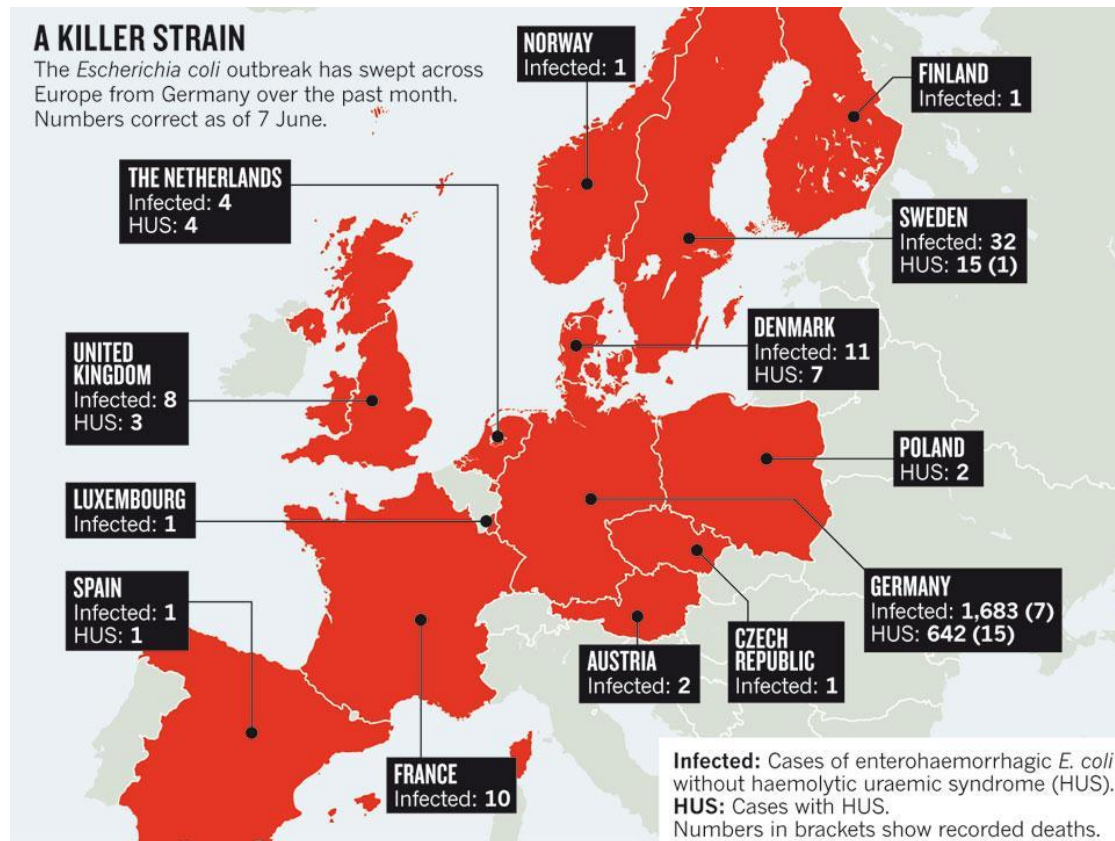
¹⁰ Severe Acute Respiratory Syndrome, Christian et al, CID 2004.

2.3.4.3. Commerce mondial des aliments

Le commerce mondial des aliments est un autre facteur de dissémination de maladies.

En 2011 le développement de la bactérie *Escherichia coli* en est une illustration :

Propagation de l'Escherichia coli : de l'Allemagne à l'ensemble de l'Europe¹¹



En 2011 des syndromes d'insuffisance rénale apparaissent chez une centaine de patients en Allemagne, l'aliment porteur de la bactérie était le fenugrec égyptien : graines égyptiennes importées en Europe et à l'origine d'une contamination qui touché plusieurs centaines de personnes.

2.3.4.4. Bioterrorisme

Enfin un dernier facteur de propagation peut être la volonté de l'homme d'utiliser ces bactéries ou virus à des fins bioterroristes.

On peut alors citer l'exemple de l'accident du centre de recherche militaire soviétique de Sverdlovsk en 1979. Dans ce centre de recherche militaire, des chercheurs soviétiques essayaient de produire des armes bactériologiques à base d'anthrax.

Un accident s'est produit dans cette usine Sverdlovsk, et a causé la fuite de ces spores d'anthrax. Cette fuite avec le vent a touché plusieurs villages voisins causant une centaine de morts.

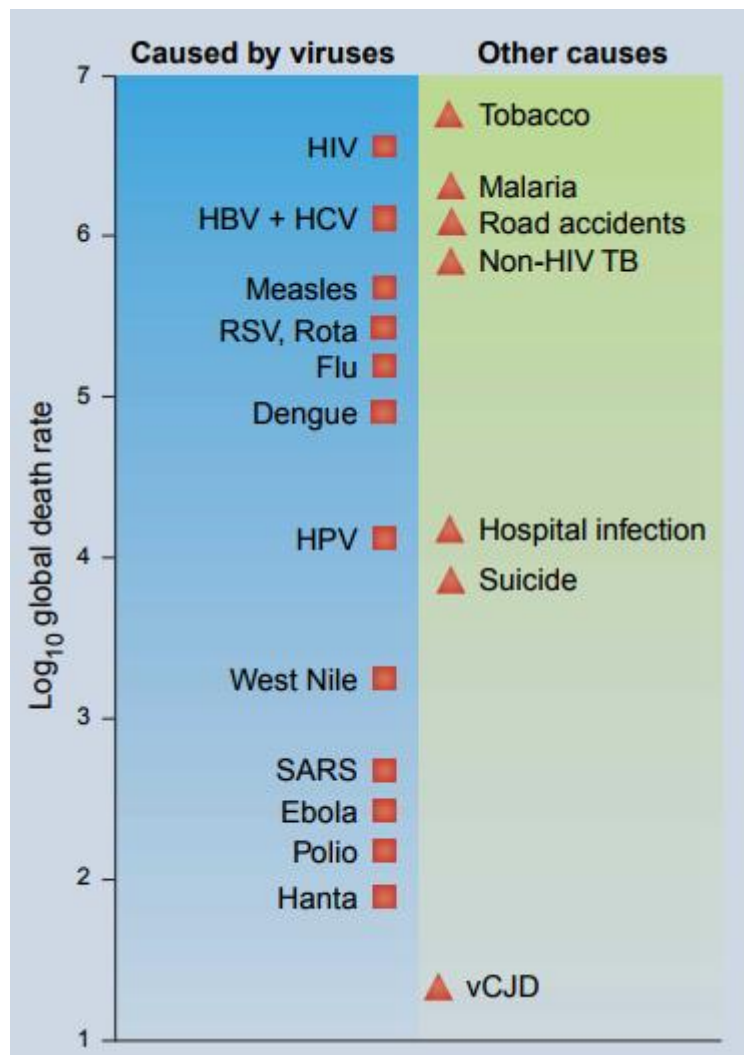
¹¹ Nature, Juin 2011 (<https://www.nature.com/news/2011/110607/full/474137a.html>).

2.4. Impact d'une pandémie

2.4.1. Mortalité

Un critère assez naturel pour évaluer l'impact d'une pandémie est la mortalité, c'est-à-dire le nombre de victimes de ces virus.

Nombre de décès dans le monde en 2003 classés par cause (Virus à gauche, autres causes à droite)¹²



On constate sur ce graphique, que les virus responsables des pandémies les plus « médiatiques » (SRAS et Ebola) ont finalement un impact en termes de mortalité limité (moins d'un millier de victimes dans le monde en 2003, sensiblement moins de morts que le tabac ou les accidents de la route par exemple).

Pour autant, ces virus par leur capacité à voyager et par l'absence de maîtrise qu'on a de leur transmission sont préoccupants.

¹² Social and environmental risk factors in the emergence of infectious diseases, Weiss, Nature Medicine, 2004.

Le SRAS et l'Ebola sont des virus contagieux après le début des symptômes, c'est-à-dire que l'on peut identifier les personnes infectées (parce qu'elles présentent les manifestations cliniques de ces maladies), les isoler et ainsi bloquer la transmission.

D'autres virus présents sur ce graphique sont plus dangereux parce qu'ils sont contagieux avant le début des symptômes : il y a alors un risque de propagation alors que la maladie ne s'est pas encore déclarée, c'est là que nous trouvons les virus les plus virulents : SIDA et Hépatite C (plus d'un million de victimes).

2.4.2. Impact économique

L'impact d'une pandémie est également économique.

Si l'on se concentre sur l'épidémie de SRAS de 2003 qui a démarré à Hong-Kong :

Taux de croissance du PIB de Hong-Kong sur la période 2000-2010



On constate sur ce graphique au moment de l'épidémie du SRAS en 2003 une chute du PIB Hong-Kongais.

L'impact économique se fait principalement sur les secteurs des transports et du tourisme mais les autres secteurs sont également touchés avec une hausse de l'absentéisme.

On peut citer également une étude de la banque mondiale réalisée en 2006 visant à évaluer les conséquences économiques d'une pandémie grippale¹³ :

Table 2 Possible economic impacts of flu pandemic

	<u>Mild</u>	<u>Moderate</u>	<u>Severe</u>
<i>(% change in GDP, first-year)</i>			
World	-0.7	-2.0	-4.8
High-income	-0.7	-2.0	-4.7
Developing	-0.6	-2.1	-5.3
East Asia	-0.8	-3.5	-8.7
Europe and Central Asia	-2.1	-4.8	-9.9
Middle-East & North Africa	-0.7	-2.8	-7.0
South Asia	-0.6	-2.1	-4.9
Deaths (millions)	1.4	14.2	71.1

Source: World Bank calculations based on McKibbin & Sidorenko (2006).

Dans son scénario sévère, la banque mondiale estime qu'une pandémie grippale pourrait causer 70 millions de décès dans le monde et une baisse de 5% du PIB mondial.

2.5. Contrôle d'une pandémie

Plusieurs méthodes existent pour tenter de contrôler le développement d'une pandémie, c'est-à-dire d'en limiter la propagation.

Isolement et quarantaine

Les mesures d'isolement et de quarantaine s'avèrent efficaces si elles sont déployées au tout début d'une épidémie, c'est-à-dire lorsque le nombre de personnes infectées est encore raisonnable.

On peut citer à cet effet l'exemple du Coronavirus du MERS (Middle East Respiratory Syndrom), virus qui provoque des infections respiratoires qui sont transmises par les chameaux dans les pays du Golfe : en France il y a eu deux cas en 2013, un patient revenu de Dubaï avait infecté une deuxième personne sur le territoire français dans un hôpital du nord de la France. Ces deux personnes avaient eu des contacts dans les jours précédents leur hospitalisation avec 162 personnes¹⁴. Pour chacune de ces personnes exposées il fallait déterminer le niveau de proximité avec le patient, et en déduire les mesures à appliquer (isolement, quarantaine, et a minima surveillance quotidienne par l'Institut de veille sanitaire pour s'assurer qu'elles ne développaient pas les symptômes).

Ces méthodes ont fonctionné puisqu'il n'y a pas eu de cas supplémentaires en France, mais elles nécessitent la mobilisation du personnel médical, ce dernier pouvant être très rapidement débordé dans une pandémie qui serait majeure.

¹³ Evaluating the Economic Consequences of Avian Influenza, Burns et al, 2006.

¹⁴<https://france3-regions.francetvinfo.fr/hauts-de-france/2013/05/12/d-ou-vient-le-coronavirus-162-personnes-concernees-par-l-enquete-etiologicalue-250011.html>

Restriction des déplacements

Les déplacements peuvent être limités pour atténuer la propagation d'une pandémie. On peut citer l'exemple de la Chine en 2003 pendant l'épidémie de SRAS : le gouvernement chinois avait mis en place des mesures de restrictions imposées au déplacement des personnes.

Ces mesures contribuent au coût économique d'une pandémie, nous avons l'exemple de la compagnie aérienne Cathay Pacific basée à Hong-Kong, qui a enregistré des déficits record à cette période et a failli déposer le bilan¹⁵.

2.6. Quelle serait la plus grande menace ?

Les spécialistes s'accordent à dire que le virus le plus dévastateur serait un virus **transmis par voie respiratoire** : parce qu'il serait alors très difficile de contrôler la contagion (les virus transmissibles par les fluides biologiques – sang, sperme - sont plus faciles à contrôler).

Ce virus serait par ailleurs **hautement contagieux**, il aurait un nombre de reproduction élevé (nommé R_0 , il s'agit du nombre de cas secondaires pour une personne infectée dans une population totalement susceptibles).

Concrètement ce nombre de reproduction R_0 définit le facteur de contagion du virus :

- Si ce nombre est égal à 1 : chaque malade ne transmet la maladie qu'à une seule autre personne, on reste alors sur un niveau de contagion « plat ».
- Si ce nombre est supérieur à 1 : chaque malade transmet la maladie à plus d'une personne, on alors une augmentation du nombre de cas et on se retrouve en période épidémique.
- Si ce nombre est inférieur à 1 : c'est la période où l'épidémie est en décroissance

Pour citer quelques exemples, ce nombre de reproduction de base est de l'ordre de 2 à 3 pour le SRAS et pour les gripes.

Un autre paramètre important est **le taux de létalité** : c'est la proportion de décès par cas infecté.

Pour citer quelques exemples, pour la grippe espagnole on estime le taux de létalité entre 1% et 1 pour mille, pour le virus de la grippe A(H1N1) de 2009 heureusement il était plutôt de l'ordre de un pour dix mille et c'est ce niveau assez faible qui explique le peu de victimes causées par cette épidémie.

Un autre facteur très important est **le temps de génération**, c'est le temps qu'il y a entre chaque « chaîne de transmission », c'est-à-dire c'est le nombre de jours qui s'écoulent entre le moment où une personne est infectée et celui où elle devient contagieuse.

Plus ce temps est faible (exemple pour la grippe où il est inférieur à 3 jours) plus la propagation de l'épidémie va être sévère.

Un temps plus élevé (exemple pour la variole où il est de l'ordre de 14 jours) laisse le temps de mettre en place des mesures de contrôle.

¹⁵ <https://www.wsj.com/articles/SB106014765860149100>

Enfin dernier facteur : **une maladie contagieuse avant les symptômes** est redoutable, en effet il est alors impossible de mettre en place des mesures de contrôle puisqu'on ignore qui est contagieux dans la population.

La question qui se pose alors est : est-ce qu'une maladie réunit tous ces facteurs ? Malheureusement oui et il s'agit de la grippe.

La grippe est l'agent pathogène qui fait le plus peur parce que transmis par voie respiratoire avec des R0 qui se situent entre 2 et 3, un taux de létalité qui peut être très élevé, un temps de génération de l'ordre de trois jours et est contagieux avant les symptômes.

Si on prend l'exemple de la pandémie grippale A(H1N1) de 2009. Tous les critères étaient réunis à l'exception du taux de létalité : ce taux estimé inférieur à 1 pour dix mille, explique que sur 2 milliards de personnes infectées (un tiers de la population mondiale) on dénombre au final moins de 200 000 morts.

2.7. Peut-on s'y préparer ?

2.7.1. Anticiper les pandémies futures

Parmi les projets en cours visant à anticiper les pandémies futures on peut citer le Global Virome Project¹⁶.

Ce projet américain lancé en 2016 par USAID (agence des Etats-Unis pour le développement international) a pour objectif d'identifier en l'espace de dix ans, 99% des virus animaux représentant une menace pour la santé humaine. Le coût de ce projet est estimé à 3,5 milliards de dollars¹⁷.

Concrètement l'objectif du projet est de séquencer tous les virus animaux afin d'en obtenir le catalogue complet et de disposer pour les transmissions futures de ces virus vers l'homme de test diagnostique et de vaccin.

Certaines pandémies passées mettent en lumière les limites d'une telle démarche : Ebola était un virus bien connu, cela n'a pas empêché sa propagation quand il est arrivé en Afrique de l'Ouest, de même pour Zika lorsqu'il est arrivé en Polynésie Française et au Brésil : la construction d'un « catalogue des virus » aura une utilité mais n'empêchera pas la propagation de pandémies futures qui ont souvent un caractère imprévisible.

¹⁶ <http://www.globalviromeproject.org/>

¹⁷ The World Is Not Ready for the Next Pandemic, Time Magazine, Mai 2017.

2.7.1. Caractère imprévisible des pandémies

Si l'on prend l'exemple des pandémies récentes, on peut mettre en lumière le caractère hautement imprévisible de ces phénomènes, en effet ces pandémies n'ont absolument pas été prédites.

Les pandémies sont toujours arrivées là où on ne les attendait pas :

- le SRAS en 2003 : à Hong-Kong, transmis par des civettes dans les restaurants,
- le Coronavirus du MERS en 2012 : transmis par les chameaux des pays du Golfe
- Ebola en 2014-2015 : virus connu en Afrique centrale mais qui est arrivé en Afrique de l'ouest à la frontière Guinée, Sierra Leone, Liberia sans que l'on sache comment
- La Grippe A(H1N1) en 2012 : les gripes historiquement venaient plutôt d'Asie, celle-ci est arrivée du Mexique
- Zika en 2013-2014 : propagé en Polynésie française alors que le virus était connu en Afrique centrale.

Il semble donc compliqué de prédire les pandémies, la solution pourrait alors plutôt prendre la forme d'une surveillance et la réactivité.

2.7.2. Surveillance mondiale sous l'égide de l'OMS

La surveillance est organisée à l'échelle mondiale par l'OMS avec le Global Outbreak Alert and Response Network (GOARN). Ce réseau mondial d'alerte d'action en cas d'épidémie a pour objectif de contribuer à la sécurité sanitaire mondiale en luttant contre la propagation internationale des épidémies, en veillant à ce que les états touchés bénéficient rapidement d'une assistance technique appropriée et en contribuant à la préparation aux épidémies.

D'autres moyens de surveillances moins formels existent également, les chercheurs obtiennent beaucoup d'informations sur les débuts de pandémie à partir des recherches de mots clés sur internet. Un logiciel dédié a notamment été développé pour l'OMS (GPHIN pour Global Public Health Intelligence Network), ce logiciel traque des mots clés comme « infections respiratoires », « encéphalite » etc. et lorsqu'il retrouve dans une même région du monde sur les moteurs de recherche et les réseaux sociaux un nombre important de recherche sur ces sujets crée une alerte qui remonte à l'OMS. La Commission Européenne a développé un outil similaire : Medisys.

2.7.3. Progrès technologiques dans l'identification des nouveaux virus

L'identification des nouveaux virus est un sujet déterminant dans le contrôle du développement d'une pandémie, en effet ce n'est qu'une fois que le nouveau virus responsable d'une pandémie est identifié (c'est-à-dire que sa séquence génétique est connue) que les tests diagnostiques (tests qui permettent de déterminer les personnes infectés) peuvent être produits et qu'un vaccin peut éventuellement être conçu.

Pour identifier les nouveaux virus, les chercheurs utilisent des méthodes de séquençage qui leur permettent d'amplifier tout matériel génomique à l'intérieur d'un échantillon. Les séquences inconnues des nouveaux virus sont ainsi identifiées et peuvent être comparées aux séquences des virus connus pour déterminer de quel virus connu ce nouveau virus est le plus proche.

Les progrès technologiques (puissance de calcul des machines démultipliée, progrès de la bio-informatique) ont permis de réduire le temps nécessaire pour identifier les nouveaux virus (plusieurs mois auparavant, une semaine aujourd'hui), ce gain de temps est précieux car il permet de démarrer plus rapidement les mesures de prévention et les premiers traitements des personnes infectées.

2.7.4. Plans de préparation, exercices de simulation

Pour terminer nous pouvons citer un exemple d'exercice de simulation à grande échelle qui a été réalisé en 2002 aux Etats-Unis. Ce plan de préparation qui s'appelait « Dark Winter » réunissait d'anciens membres du gouvernement et des journalistes. Pendant 48 heures ils mimaient les différentes phases d'une épidémie de variole.

Les conclusions de cet exercice de simulation mettaient en exergue le manque de coordination entre les différentes chaînes du gouvernement, l'incapacité à augmenter les chaînes de production de vaccins, une communication désastreuse avec les médias menant à des émeutes et à une intervention de l'armée.

En matière de préparation aux scénarios de crise pandémique, on peut également citer l'exemple du Royaume-Uni. Les anglais ont créé un groupe interministériel du nom de « Cobra » qui « survit » aux changements de gouvernement et de premier ministre, c'est ainsi le même groupe qui gère toutes les crises depuis une quinzaine d'années, permettant l'installation d'une relation de confiance entre ses membres et une capitalisation de l'expérience de ce groupe au fur et à mesure des crises traversées.

2.8. D'autres pandémies dans un futur proche ?

Le nombre de victimes d'Ebola a finalement été significativement inférieur aux 4 pandémies grippales du XX^{ème} et XXI^{ème} siècle, notamment grâce à des moyens médicaux (humains et matériels) plus pointus mais également grâce à une meilleure communication et organisation internationale, qui permet d'apporter une réponse aux épidémies plus adaptée et plus rapide qu'auparavant.

Au-delà du bilan chiffré, Ebola a permis une prise de conscience collective et à grande échelle que le risque biologique (pandémie ou bioterrorisme) était un danger réel qui pouvait avoir de graves conséquences dans un délai très court, en particulier avec la population grandissante et l'augmentation permanente des flux migratoires.

Ces phénomènes ont pour effet d'accroître la densité donc la proximité des individus sur le globe et c'est un facteur essentiel qui, comme nous avons pu le voir à travers la grippe espagnole pendant la guerre, permet la propagation rapide d'un virus, indépendamment de sa contagiosité intrinsèque.

Le 18 février 2017, Bill Gates (fondateur de Microsoft et désormais mécène) est intervenu devant la Conférence de Sécurité de Munich¹⁸ en expliquant qu'il « est assez probable que le monde va vivre une telle épidémie dans les 10 à 15 ans à venir », prenant justement pour exemples la grippe espagnole de 1918 et l'épidémie Ebola de 2014. Il a ainsi appelé les Etats à investir dans la recherche pour développer des technologies capables de créer des vaccins en quelques mois et se préparer à une pandémie avec le même sérieux qu'ils se préparaient à une attaque nucléaire durant la Guerre froide.

¹⁸https://www.challenges.fr/monde/pour-bill-gates-le-monde-doit-se-preparer-a-une-pandemie-globale_455214

Selon ses propos, « qu'ils apparaissent dans la nature ou dans les mains d'un terroriste, les épidémiologistes disent qu'un pathogène transmis dans l'air et se propageant rapidement peut tuer 30 millions de personnes en moins d'un an. [...] Le coût global de la préparation à une pandémie est estimé à 3,4 milliards de dollars par an. La perte annuelle qu'une pandémie provoquerait pourrait atteindre 570 milliards ».

Déjà en début d'année 2015¹⁹ mais avec moins de médiatisation, ce même message était délivré par Peter Daszak, président de l'organisation américaine EcoHealth Alliance qui dirige des programmes de recherche sur la santé internationale, dans une interview à Matt McCarthy²⁰. L'investissement nécessaire à une meilleure prévention des pandémies est très important (Daszak estime que 6,3 milliards de dollars sont nécessaires pour découvrir tous les virus infectant les mammifères) mais représente toujours une fraction du coût de traitement a posteriori quand celles-ci frappent.

Dans la prédiction des pandémies, qu'il compare à la prédiction des séismes, Peter Daszak met en avant la corrélation entre la transmission des maladies et les mouvements commerciaux mondiaux (eux-mêmes corrélés à l'exposition des humains à ces maladies). C'est ce qu'on appelle l'épidémiologie économique, qui prône une modélisation de la survenance des épidémies en fonction du commerce. Selon l'EcoHealth Alliance, les 3 facteurs favorables à la propagation d'un virus chez les humains sont la densité de la population humaine, la diversité de la faune et les changements dans l'utilisation des sols.

Dans son article, McCarthy mentionne le virus de Nipah²¹ et le virus de la vallée du Rift²² comme étant les 2 virus susceptibles de causer la prochaine pandémie. Le virus de Nipah est transmis par les chauves-souris, il est extrêmement létal mais peu contagieux, il n'existe pas encore de vaccin et le traitement symptomatique intensif reste la principale méthode de prise en charge de cette infection chez l'homme. Le virus de la vallée du Rift est quant à lui transmis par les moustiques et possède des symptômes similaires à Ebola. Bien qu'il soit moins létal que Nipah ou Ebola, c'est sa transmission par les moustiques qui inquiète car cela signifie qu'elle peut passer par les avions.

En conclusion, au-delà de connaître le nom de la prochaine pandémie, nous retenons dans le cadre de ce mémoire les facteurs de propagation énumérés ci-dessus. Parmi ces facteurs, nous allons nous concentrer en particulier sur celui de la densité humaine, qui est celui dont nous pouvons mesurer l'impact le plus facilement avec nos outils et modèles actuariels. Par ailleurs, ce facteur de densité humaine a un impact direct sur la pollution et les évolutions technologiques qui eux-mêmes peuvent impacter les 2 autres facteurs que sont la diversité de la faune et l'utilisation des sols. Ces facteurs biologiques sont difficilement mesurables d'un point de vue actuariel et nous choisissons de ne pas y concentrer nos efforts dans notre modèle.

¹⁹http://www.slate.com/articles/health_and_science/medical_examiner/2015/01/preparing_for_pandemics_what_diseases_will_be_the_next_ebolas.html

²⁰<https://www.ecohealthalliance.org/personnel/dr-peter-daszak>

²¹<http://www.who.int/csr/disease/nipah/fr/>

²²<http://www.who.int/mediacentre/factsheets/fs207/fr/>

3. Appréhension du risque pandémique dans le monde de l'assurance

3.1. Risques au passif pour l'assureur et garanties impactées

3.1.1. En cas de décès

La garantie décès permet le versement d'un capital ou d'une rente à un bénéficiaire en cas de décès de l'assuré avant le terme du contrat. Il peut s'agir d'un capital ou de rentes conjoint et/ou éducation pour subvenir de façon prolongée aux besoins du conjoint endeuillé et/ou du financement de l'éducation des enfants.

La durée de garantie peut contractuellement être délimitée à une période d'une ou plusieurs années (temporaire décès), mais peut également s'étendre jusqu'au décès de l'assuré (vie entière).

Lors d'une pandémie, certaines personnes pourront être fragilisées et décéder des suites de la maladie, ce qui va contraindre l'assureur à verser les capitaux et rentes (sous réserve que le décès en cas de pandémie soit également couvert par le contrat). Sauf à ce que l'assureur ait modélisé l'occurrence d'une pandémie dans son tarif, les décès interviennent plus tôt que prévu par le modèle qui avait « prévu » des décès selon une certaine table de mortalité, certains décès n'ayant même pas été « prévus ».

3.1.2. En cas de vie

A l'inverse, les décès liés à la pandémie peuvent impacter de manière plus favorable (l'assureur lorsque celui-ci porte des garanties en cas de vie. C'est le cas des contrats qui prévoient le versement d'un capital à une échéance donnée si l'assuré est toujours vivant ou des contrats qui garantissent le versement d'une rente (à l'assuré ou à un bénéficiaire désigné) tant que l'assuré/le bénéficiaire est en vie (rentes viagères).

3.1.3. Frais de soin

Les contrats de Santé traditionnels peuvent prévoir le remboursement total ou partiel des frais médicaux occasionnés à la suite d'une maladie ou d'un accident, que ce soit des frais de consultation médicale chez un médecin, des frais pharmaceutiques, des frais d'hospitalisation médicale/chirurgicale, du forfait journalier hospitalier, ou encore des frais d'analyse et de laboratoire. Cette garantie est naturellement fortement impactée en cas de pandémie.

3.1.4. Arrêt de travail (incapacité / invalidité)

En cas d'un arrêt de travail causé par un accident ou une maladie, les contrats de Prévoyance incluent souvent une garantie qui prévoit le versement d'un complément de revenu par rapport aux indemnités de la sécurité sociale pendant une période déterminée. On distingue la garantie incapacité temporaire de travail et la garantie invalidité.

En cas d'incapacité temporaire partielle ou totale d'exercer une activité professionnelle suite à une maladie ou à un accident, une indemnité journalière forfaitaire est versée en compensation. Le début du versement des indemnités est variable, il peut être au premier jour de l'arrêt de travail ou intervenir après une période de franchise (appelée « délai de carence »). De même, le montant de l'indemnité peut dépendre du salaire ou du manque à gagner de l'assuré. En cas de pandémie, certains assurés malades ne peuvent se rendre à leur travail et bénéficient donc de cette garantie.

En cas d'invalidité, c'est-à-dire en cas de réduction partielle ou totale de certaines aptitudes consécutivement à une maladie ou à un accident, l'assuré couvert par cette garantie reçoit un capital

ou une rente d'invalidité. Le déclenchement de cette garantie en cas de pandémie n'est pas évident à mesurer puisque cela suppose que la pandémie se traduit par une maladie contagieuse et invalidante, c'est-à-dire un état dans lequel les malades restent invalides pendant une certaine durée tout en restant vivants, ce qui n'est a priori pas fréquent.

3.1.5. Garanties Dommages

Ce n'est pas faire une hypothèse forte que de dire que les garanties dommages n'interviennent que très peu en cas de pandémie. En effet, l'impact éventuel d'un tel événement serait la perturbation des flux migratoires et de circulation sur les axes routiers.

La partie de la population qui se sent le plus à l'abri aura tendance à ce cloîtrer chez elle et limiter ses déplacements et donc les risques d'accident.

A l'inverse, la partie de la population qui se sent davantage menacée par la pandémie aura tendance à fuir sa région d'habitat et se diriger vers un autre endroit. Ce phénomène provoquerait donc une augmentation de la circulation routière et donc des risques d'accident.

Nous voyons ainsi que ces phénomènes s'opposent et l'impact sur les garanties en cas d'accident est très difficile à évaluer, pour peu qu'il soit significatif, ce qui n'est pas évident. Il en est de même pour les autres garanties Dommages les plus fréquentes que sont l'incendie, le vol, le bris de glace, etc.

L'impact étant jugé non significatif et n'étant pas le centre d'intérêt de notre mémoire, nous choisissons de ne pas nous concentrer sur les garanties Dommages.

3.2. Risques à l'actif

3.2.1. Risque de marché

En cas de pandémie, les risques à l'actif ne sont pas ceux auxquels on pense en premier, à l'inverse des risques au passif qui sont plus intuitifs. Pourtant, il est très probable qu'une pandémie aurait un impact important sur les marchés financiers en raison de son impact sur le comportement des investisseurs et sur celui des consommateurs (mouvements éventuels de panique, perturbation des flux migratoires, etc.).

Les investisseurs pourraient ainsi afficher une aversion au risque, les valeurs refuge et le secteur pharmaceutique et médical surperformeraient, tandis que les indices des pays les plus touchés chuteraient probablement de manière significative. Si l'impact dépendrait bien sûr de la gravité de la pandémie, le risque systémique d'une réaction en chaîne ne peut être exclu, avec une augmentation soudaine des retraits entraînant une crise de liquidité, des arbitrages entre les unités de compte et les valeurs en euros, et enfin des rachats massifs en épargne et en assurance des emprunteurs en cas de crise immobilière.

3.2.2. Risque de crédit

Par ailleurs, on peut aussi imaginer qu'une pandémie aurait un impact sur la solvabilité de certaines entreprises, ce qui pourrait engendrer des faillites et un risque de crédit accru pour les assureurs. En particulier, le risque de pandémie est très surveillé par les réassureurs puisque ce seraient a priori les plus à même de supporter une majeure partie du coût. Un tel risque pourrait alors, si ces acteurs n'y étaient pas suffisamment préparés, menacer leur solvabilité.

Ainsi, leurs entreprises cédantes verraient leur risque de crédit augmenter, potentiellement de façon significative.

A titre d'exemple, cet élément de risque est pris en compte dans l'évaluation ORSA de SwissLife France puisque un des scénarii pour l'entité SwissLife Prévoyance et Santé (« SLPS ») est un scénario

de pandémie et, outre les chocs au passif (taux de mortalité, frais de soin, volumes de production et rachats, frais généraux, politique tarifaire), il est également modélisé un choc à l'actif sur les spreads de crédit : AAA +37bp ; A +75bp ; BBB +100bp ; +200bp sinon.

3.2.3. Risque opérationnel

Bien évidemment, (peut-être malheureusement pour nous), les employés du secteur de l'assurance n'ont pas de raison d'être davantage épargnés par la pandémie que des employés d'autres secteurs, toutes choses égales par ailleurs.

Ainsi en cas de pandémie, un assureur pourrait avoir une partie de ses salariés malades et/ou en arrêt de travail voire décédés, éventuellement à des postes clés. De fait, il s'agit là d'un risque opérationnel pour l'assureur.

3.3. Mesure du risque dans Solvabilité 2

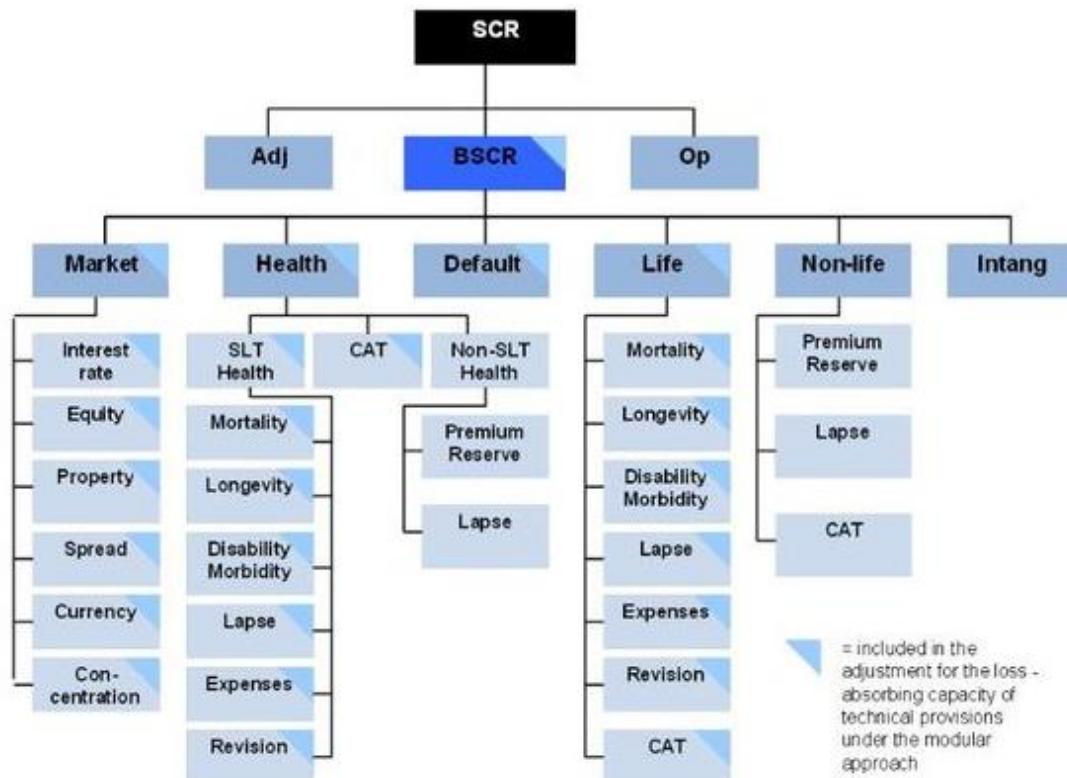
3.3.1. Description rapide de la directive et de ses trois piliers

De façon synthétique, rappelons simplement que Solvabilité 2 est une réforme réglementaire européenne qui s'applique au secteur de l'assurance depuis le 1^{er} janvier 2016. Celle-ci a pour objectif de mieux adapter les exigences en fonds propres des assureurs et réassureurs aux risques encourus par ces compagnies. Certains risques n'étaient en effet pas captés (ou pas intégralement) dans l'ancienne directive Solvabilité 1, qui était relativement simpliste du fait que l'exigence en fonds propres était calculée par un pourcentage des primes, sinistres ou réserves.

Dans la lignée des normes Bâle 2 dans le secteur de la Banque, Solvabilité 2 a étoffé le cadre prudentiel pour intégrer davantage les risques propres à chaque compagnie tels que les risques opérationnels, de réassurance, d'exposition aux marchés financiers, etc.

La directive repose sur trois piliers :

- **Le pilier 1** définit les normes quantitatives pour le calcul des provisions techniques et des fonds propres. Pour cela, deux seuils réglementaires sont définis :
 - MCR (*Minimum Capital Requirement soit minimum de capital requis*) : niveau minimum de fonds propres réglementaire en-dessous duquel l'intervention de l'autorité de contrôle (ACPR dans le cas de la France) sera automatique. En principe, son non-respect constitue le seuil déclencheur du retrait d'agrément ;
 - SCR (*Solvency Capital Requirement soit capital de solvabilité requis*) : montant de fonds propres estimé comme nécessaire pour absorber le choc provoqué par une sinistralité exceptionnelle (calibrée sur une fréquence d'occurrence à 200 ans). Son calcul est basé sur l'exposition aux risques liés à l'activité des organismes d'assurance, c'est-à-dire principalement le risque de souscription, le risque de crédit, le risque opérationnel, le risque de liquidité et le risque de marché. Il est décomposé en sous-modules de risques :



- **Le pilier 2** complète le premier au sens où il régit les normes qualitatives. Il définit d'une part les règles de gouvernance et de gestion interne des risques pour les compagnies et implémente d'autre part un système d'évaluation propre des risques d'une compagnie, appelé ORSA (*Own Risk and Solvency Assessment*) ;
- **Enfin, le pilier 3** définit l'enveloppe des deux premiers, à savoir qu'il régit la communication d'informations au public et aux autorités de contrôle. Il vise à harmoniser au niveau européen les informations publiées par les organismes d'assurance ainsi que celles remises aux superviseurs

3.3.2. Formule standard ou modèle interne

Pour le calcul des exigences quantitatives (MCR et SCR notamment), la directive prévoit les formules de calcul à appliquer, qui constituent ce qu'on appelle le « modèle standard » et qui sont les mêmes pour tous les acteurs européens. Cela permet d'harmoniser les exigences entre pays et de faciliter les comparaisons.

Certains acteurs sont soumis à Solvabilité 2 et sont de petite taille, au sens qu'ils ont relativement peu de moyens humains et financiers à mettre en face de l'exigence de calcul imposée par cette réglementation qui a complexifié le cadre. Pour ceux-ci (plusieurs benchmarks sur le marché qu'il s'agit souvent de mutuelles ou institutions de prévoyance à taille réduite car très régionales par exemple), la formule standard a l'avantage indéniable de fournir un ensemble de formules déjà prêtes, qu'il reste à appliquer à ses propres grandeurs.

D'autres acteurs en revanche (souvent les plus importants du marché) ne souhaitent pas se voir appliquer des exigences en fonds propres qui sont standard pour le marché européen (qui ont donc été construites en quelque sorte sur la base d'observations moyennes) mais préfèrent que celles-ci reflètent davantage leurs propres caractéristiques.

C'est pour cela que la directive laisse le choix aux acteurs d'appliquer ce qu'on appelle un « modèle interne », c'est-à-dire que les acteurs peuvent adapter la formule standard à leur propre profil de

risque et ainsi calibrer les exigences en fonds propres de chaque module, ce qui nécessite un temps et des moyens (humains, financiers, informatiques) importants. Pour éviter toute dérive au sein du marché, il est évident qu'un modèle interne doit faire l'approbation du régulateur national avant de pouvoir être appliqué.

Il existe une troisième notion de modèle, dite « modèle interne partiel » qui est une solution hybride entre les deux précédentes. Dans ce cas, l'adaptation des formules de calcul se fait sur une partie seulement des modules de calcul qui composent le SCR.

3.3.3. Focus sur le risque Pandémie

On peut voir dans la « pieuvre » des modules de risque qui composent le SCR, aucun n'est spécifiquement identifié au risque de pandémie. En effet, ce risque est capté et couvert de façon indirecte dans les deux sous-modules que sont le SCR Vie CAT et le SCR Santé CAT Pandémie, définis comme suit.

➤ **SCR Vie Cat**

$$SCR_{Vie\ CAT} = 1,5\text{‰} \times [Capital\ sous\ risque]$$

où le capital sous risque correspond à la somme assurée que l'on aurait à verser si tous les assurés décédaient immédiatement. Cette formule représente donc un risque d'une surmortalité de 1,5 ‰ pour une pandémie bicentenaire. Ce choc a été calibré d'après le modèle épidémiologique de Swiss Re (2007), selon lequel la surmortalité engendrée par une pandémie bicentenaire comprise entre 1,0 ‰ et 1,5 ‰ dans les pays développés et pour une population dont la répartition par âge reflète celle d'un portefeuille assuré.

Par exemple, pour un assureur qui a un q_x de 0,15%, cela représenterait donc un doublement de ses prestations décès sur une année.

➤ **SCR Santé Cat Pandémie**

$$SCR_{Santé\ CAT\ Pandémie} = 0,075\text{‰} \times E + 0,4 \times F$$

où E est l'exposition de l'assureur aux garanties invalidité et F l'exposition aux garanties frais d'hospitalisation et frais de soin, en considérant qu'en cas de pandémie, 1% des personnes seraient hospitalisées et que 20 % iraient consulter un médecin pour un traitement médical (et que donc les 79% restants n'engageraient aucune démarche particulière).

La principale limite de la formule standard concernant la mesure du risque de pandémie réside justement dans cette séparation des impacts avec d'un côté le scénario de stress sur les garanties Vie et de l'autre un scénario différent de stress sur les garanties Santé. Ainsi, les interactions entre ces garanties (compensations et mutualisations) ne sont pas captées lorsque les calculs sont faits de façon distincte.

Pour davantage de détails sur la mesure du risque de pandémie dans Solvabilité 2, le lecteur pourra se référer aux études de Speisser, R. (2013) et Mandhouj, K. (2011), qui le décrivent de façon plus détaillée et proposent la construction d'un modèle interne pour palier à ces limites.

Pour conclure, on peut donc dire que la pandémie représente un risque majeur puisqu'il est considéré dans le référentiel Solvabilité 2. Néanmoins, la basicité de la formule nous laisse entrevoir des voies d'amélioration pour mieux appréhender ce risque. C'est ce que va permettre le modèle que nous avons construit et que nous présentons dans la suite de ce mémoire.

3.4. Réassurance : Le BCAC comme principale couverture existante

3.4.1. Origine de la couverture

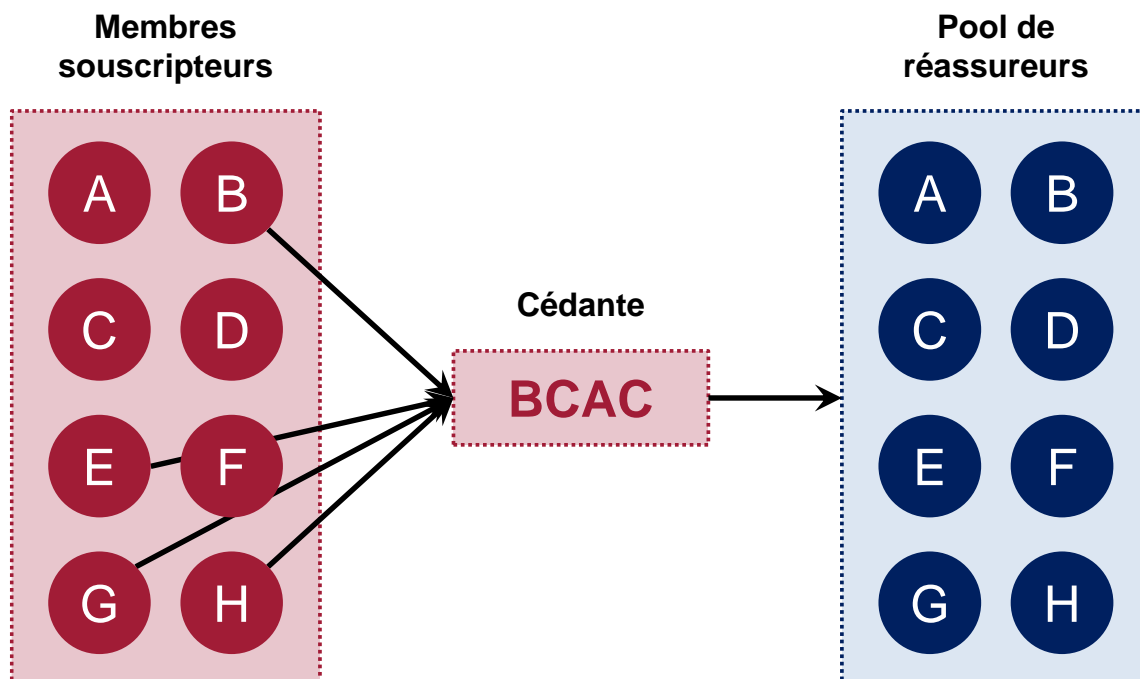
Actuellement, le risque de pandémie est pris en compte de façon très limitée par les assureurs et réassureurs, c'est une garantie beaucoup moins fréquente et qui ne fait pas partie des principaux enjeux stratégiques du marché international de l'assurance, compte tenu du caractère ponctuel des événements pandémiques (c'est-à-dire qu'il est de survenance rare ou avec des impacts limités pour les assureurs).

En 2014, lorsque le virus Ebola se répand à grande échelle en Afrique de l'Ouest et devient extrêmement meurtrier (11 000 décès pour 28 000 cas soit un taux de létalité autour de 40%), la profession accélère la réflexion pour obtenir une couverture de marché et le BCAC construit avec Guy Carpenter (courtier en réassurance) une couverture du risque de pandémie à l'échelle nationale, appliquée pour la première fois sur l'exercice 2015.

Didier Weckner (alors président du comité directeur du BCAC et directeur général délégué d'AXA France Vie) se déclare alors « [...] fier d'avoir pu organiser, à travers le BCAC, avec d'autres grands assureurs le lancement d'une protection de place contre les catastrophes majeures. Cette capacité à innover démontre la vigueur, le réalisme et la réactivité de notre profession et du BCAC ».

3.4.2. Caractéristiques

Cette couverture de réassurance est matérialisée par un traité non-proportionnel, avec un pool international de réassureurs (dont l'un est apériteur) et dont la souscription par les assureurs membres du BCAC (cédante) est facultative. Pour l'exercice 2018-2019, 4 acteurs y ont souscrit (contre 7 en 2015, 8 en 2016 et 5 en 2017).



La capacité totale s'élève à 99m€ pour l'exercice 2018-2019, répartie de la façon suivante entre les différents membres souscripteurs du BCAC :

Membre	2015 ²³	2016	2017	2018-19
A	5,9	10	-	-
B	4,5	5	5	5
C	4	3,7	-	-
D	0,3	0,7	-	-
E	21,3	20	20	40
F	-	50	50	-
G	11,6	34,8	50	50
H	2,7	3,8	3,8	4
Capacité totale	50,3	128	128,8	99

Les engagements se décomposent ainsi au sein du pool de réassureurs :

Réassureur	Rating (solidité financière) au 31/08/2017	2015	2016	2017	2018-19
A	AA- / Stable (S&P)	<u>16,50</u>	<u>32,00</u>	<u>32,80</u>	-
B	AA (S&P)	10,00	25,00	22,00	-
C	A+ (S&P)	2,00	15,75	16,50	5,00
D	A+ / Stable (S&P)	2,00	11,00	11,00	-
E	A- / Strong (S&P)	6,50	10,50	10,25	5,00
F	A- / Stable (A.M. Best)	2,00	9,25	9,50	-
G	A / Stable (S&P)	2,00	9,25	9,50	5,00
H	AA- / Stable (S&P)	7,00	9,25	5,25	4,00
I	A+ / Stable (S&P)	-	-	4,00	3,00
J	A- / positive (S&P - credit)	-	2,00	2,00	-
K	AA- / Stable (S&P)	2,00	2,00	2,00	2,00
L	A- / stable (S&P - credit)	-	-	2,00	-
M	A- / stable (S&P)	-	2,00	2,00	-
N	Non disponible	-	-	-	<u>75,00</u>
Capacité totale (EUR m)		50,00	128,00	128,80	99,00

Les montants soulignés dans le tableau indiquent le réassureur qui est apériteur du traité.

²³ Il convient de noter que la capacité totale signée pour l'exercice 2015 est de 50m€ (et non de 50,3m) mais pour des raisons de confidentialité de la part du BCAC, nous n'avons pas connaissance de l'acteur pour qui la capacité a été réduite de 0,3m par rapport à la version dont nous disposons.

Le traité a pour objectif de couvrir une dérive significative de la mortalité et porte sur les catégories suivantes (au sens de l'article A344-2 du code des assurances) :

- 03 : Contrats individuels d'assurance temporaire décès (y compris groupes ouverts)
- 06 : Contrats collectifs d'assurance en cas de décès

Jusqu'en 2017, la prime de réassurance était payée par le BCAC (cédante) et était différente pour chaque réassureur en fonction de : son engagement pris dans le pool, son caractère d'apériteur et sa notation financière. Les tarifs ci-dessous s'entendent frais d'intermédiation compris.

Exercice 2015 :

Engagement du réassureur (E)	Coefficient appliqué à E
E < 10 M€	1,54%
10 M€ ≤ E	1,64%
Bonus pour l'apériteur	+0,16%

Exercice 2016 :

Engagement du réassureur (E)	Coefficient appliqué à E
E < 10 M€	1,55%
10 M€ ≤ E < 25 M€	1,61%
25 M€ ≤ E	1,73%
Bonus pour l'apériteur	+0,23%

Exercice 2017 :

Engagement du réassureur (E)	Coefficient appliqué à E
E < 10 M€	1,53%
11 M€ ≤ E < 22 M€	1,58%
22 M€ ≤ E	1,63%
Bonus pour l'apériteur	+0,20%
Bonus si notation S&P (ou équivalent A.M. Best) long terme au 1 ^{er} janvier 2017 ≥ AA-	+0,05%

Jusqu'à l'exercice 2017, la couverture du BCAC fonctionnait ainsi :

- Un Stop loss de fréquence : i.e. couvre la surmortalité au-delà de la mortalité moyenne
- Un achat groupé de capacité : chaque membre décide de la capacité qu'il achète, mais la couverture ne prévoit pas de mutualisation des capacités ni de la sinistralité entre les membres
- Un Stop Loss ayant une priorité et une portée identiques pour l'ensemble des membres :
 - Surmortalité de 0,10% (par rapport à la mortalité N-1 du portefeuille propre au souscripteur)
 - Portée de 0,05% (toujours par rapport à la mortalité N-1 du souscripteur)

Illustration du fonctionnement de cette couverture sur un exemple :

Paramètre	Valeur du paramètre
Portefeuille assuré	300 000 assurés
Nombre de décès en année N-1	360
Priorité	0,10% au-delà de la sinistralité N-1 soit $300 + 360 = 660$ décès
Portée	$0,05\% \times$ le nombre d'assurés soit 150 décès couverts
Capacité	15M€
Prestation payée par décès au-delà de la portée	15M€/150 soit 100K€ par décès

Exemple de pandémie :

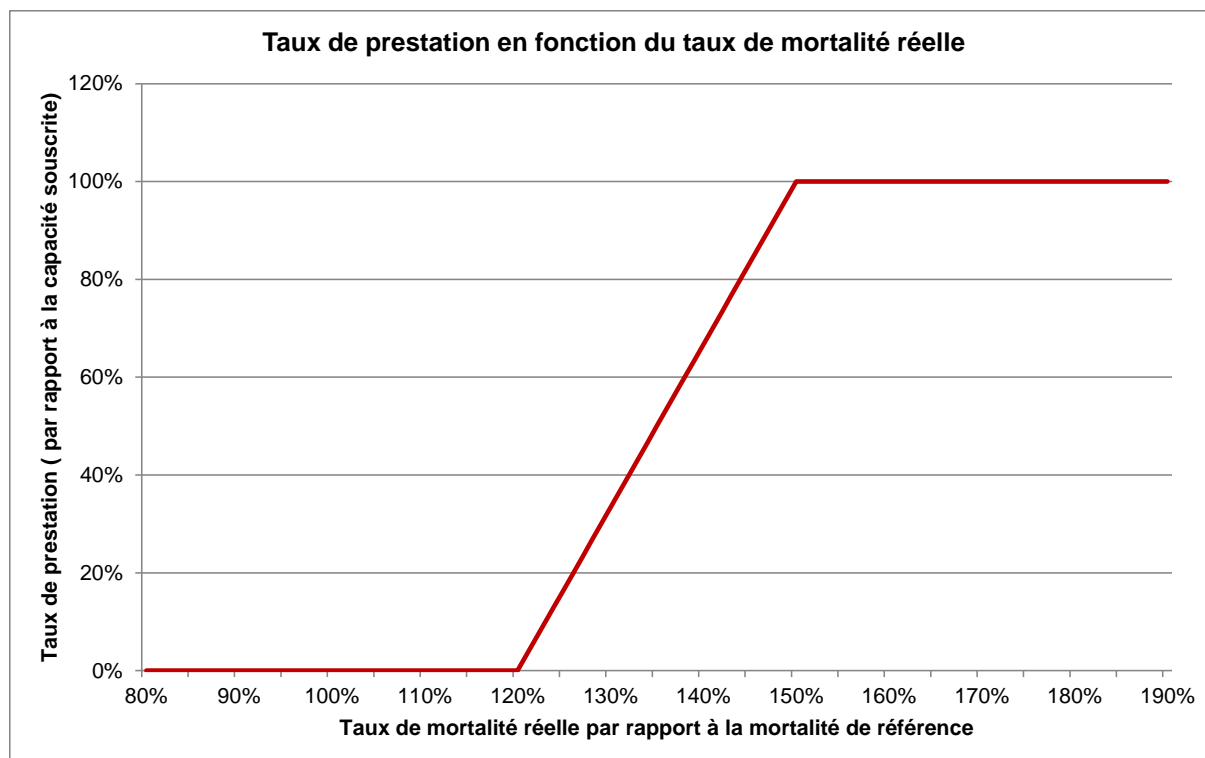
- 700 décès en année N, soit 40 décès au-delà de la priorité
- La prestation payée par les réassureurs est de $40 \times 100\text{k€}$

Pour l'exercice 2018, la couverture a évolué pour devenir un traité sous forme paramétrique. Auparavant, la dérive de mortalité se mesurait par rapport à la mortalité N-1 de chaque acteur, désormais la référence utilisée est celle de l'indice INSEE du nombre de décès annuel sur la population 20 – 65 ans en France²⁴.

Ce nombre est d'environ 100.000 par an, la couverture est déclenchée si ce nombre dépasse 20% (priorité) de la sinistralité annuelle 2017, et s'étend jusqu'à 50% de celle-ci (portée de 30%). Il s'agit donc d'un traité stop-loss de 30% en excess de 20%. Ce système présente l'avantage d'être extrêmement simple à gérer puisque ne dépendant que de données statistiques publiques.

Concrètement, la prestation remboursée à un assureur souscripteur est une fonction très classique du taux de mortalité réelle de la population française (selon le périmètre défini à l'instant) :

²⁴ <https://www.insee.fr/fr/statistiques/3124970?sommaire=3053204&q=T74>



Le traité est souscrit pour une durée de 2 ans, mais est résiliable au bout d'un an par le BCAC seulement (pas par les réassureurs) s'il n'y a pas de sinistre. La durée de 2 ans permet d'être assuré d'avoir une protection minimum d'un an devant soi au cas où une pandémie se déclarerait en fin de première année. A l'inverse s'il n'y a pas de pandémie la 1ère année le BCAC résilie et renouvelle pour une nouvelle période de 2 ans. Ce principe de résiliation/reconduction permet, en l'absence de déclenchement de la couverture (donc absence de pandémie ou en tout cas de forte dérive de mortalité), d'actualiser l'année de référence pour la mortalité.

Par exemple, pour le traité actuel 2018-2019 (mortalité de référence = 2017) :

- S'il n'y a pas de pandémie en 2018, la couverture est résiliée et reconduite sur 2019-2020 en prenant l'année 2018 comme nouvelle référence qui est donc « non polluée » par une mortalité atypique ;
- A l'inverse, si 2018 est touchée par une pandémie, la couverture est maintenue sur 2019, avec toujours l'année 2017 en référence (puisque la mortalité 2018 ne peut plus servir de référence)

En termes de tarif, les assureurs souscripteurs payent une prime de réassurance égale à 1,52% de leur capacité souscrite, directement au réassureur apériteur.

3.4.3. Limites de la couverture BCAC

Cette couverture a aujourd'hui le mérite d'exister et d'amorcer la protection contre ce risque de façon concrète mais est encore considérée par beaucoup d'assureurs comme trop peu intéressante en raison de la franchise importante et la faiblesse de la portée (cf. exemple ci-dessus où on voit que le nombre de décès couverts est très faible pour un assureur pris individuellement). On note d'ailleurs que certains membres du BCAC en sont sortis et que certains assureurs ont souscrit à des couvertures en réassurance à titre individuel afin d'avoir une couverture plus adaptée à leur portefeuille et à leurs besoins. Une autre limite réside dans le fait d'avoir des capacités ouvertes.

Enfin, une dernière limite de cette couverture que nous pouvons mentionner, et non des moindres, est qu'elle ne couvre que le risque Décès. En effet, le libellé « Pandémie » n'apparaît pas dans le traité de

réassurance et les garanties Santé ne sont pas couvertes alors même que les frais d'hospitalisation sont évidemment augmentés de façon significative par une pandémie. De même, les garanties incapacité-invalidité ne sont pas couvertes par ce traité et il est évident qu'une pandémie les affecte de façon significative.

Il serait intéressant et naturel d'inclure une réassurance des frais de santé et des garanties incapacité-invalidité ainsi qu'adapter la couverture Décès de façon adaptée en trouvant une alternative à l'indemnisation forfaitaire qui existe aujourd'hui dans la couverture du BCAC.

4. Etat de l'art

4.1. Modèles actuariels (approche de masse)

Les modèles actuariels sont des modèles de projection basés sur la mortalité historique. Dans ces modèles, la courbe de mortalité par âge est lissée ce qui permet de déterminer une tendance. On calcule ensuite la surmortalité empirique (additive ou multiplicative) par rapport à la courbe lissée et on ajuste une loi de probabilité pour estimer différents quantiles de la distribution de la surmortalité aux niveaux de confiance souhaités.

4.2. Modèles épidémiologiques (approche individuelle)

4.2.1. Introduction et grands principes

Les premiers modèles épidémiologiques ont été élaborés par William Ogilvy Kermack et Anderson Gray McKendrick durant les premières décennies du XXe siècle. Ces modèles sont appelés des modèles compartimentaux car leur principe est de créer des compartiments de population selon leur état (cf. ci-dessous) : un compartiment S qui regroupe les individus Susceptibles de contracter la maladie et un compartiment I regroupant les individus Infectés (ou Infectieux, selon le modèle).

Le modèle nécessite ensuite des règles de passage d'un compartiment à l'autre, celles-ci déterminent le nom de plusieurs modèles simples.

Le plus simple, le modèle SI, est constitué uniquement de 2 compartiments avec une règle qui définit (généralement sous forme d'équation à plusieurs paramètres) la proportion $p(S \rightarrow I)$ d'individus susceptibles qui devient infectée. S'il peut exister symétriquement une proportion $p(I \rightarrow S)$ d'individus infectés qui peut être guéris, le modèle est alors SIS.

Il existe ainsi de nombreux modèles, certains plus compliqués que d'autres en fonction des règles de passages et du nombre de variables qui entrent en jeu. Dans la suite de ce mémoire, nous avons choisi de nous concentrer sur le modèle le plus répandu, le modèle SEIR, auquel nous avons ajouté une dimension : le paramètre géographique de distance entre les individus.

4.2.2. Les différents états

La théorie des modèles épidémiologiques présente huit états, catégorisés dans huit compartiments donc :

Compartiment	Description des individus
S	Susceptibles d'être infectés : non immunisés contre la maladie et pouvant être contaminés par d'autres individus.
E	Exposés (ou Latents) : contaminés par la maladie mais qui ne peuvent pas encore transmettre eux-mêmes la maladie. Ils présentent les symptômes de la maladie mais sont encore dans la période de latence avant de devenir eux-mêmes infectieux (compartiment I).
I	Infectieux : infectés et contagieux. Il existe des modèles qui divisent ce compartiment en sous-compartiment pour distinguer différents de types d'infections, degrés de contagiosité ou de concentrations du virus par exemple.
R	Rétablis : guéris et immunisés contre la réinfection. Il s'agit d'un des trois états (avec D et Q) qui peuvent suivre l'état infectieux
D	Décédés
Q	Quarantaine : toujours infectieux mais isolé du reste de la population par une mise en quarantaine. Certains modèles peuvent trouver intéressant d'introduire ce compartiment et de le distinguer du D, pour identifier les éventuels individus qui pourraient être guéris et revenir dans le compartiment S ou R par exemple.
M	Immunisés dès la naissance : individus ne pouvant ni être contaminés ni être contagieux.
C	Porteurs (carriers) : individus portant la maladie et pouvant la transmettre, sans toutefois en présenter les symptômes.

4.2.3. Facteurs de développement

Les principaux facteurs intervenant dans la contamination des individus et le développement de la pandémie sont :

- L'âge des individus : généralement, les individus les plus jeunes et les plus âgés sont ceux les plus touchés par une maladie ;
- Leur vulnérabilité face à la maladie : outre leur âge, on vise par là leur état de santé générale et leur capacité intrinsèque à être facilement contaminé ou non, ainsi qu'à la durée pendant laquelle ils porteront le virus ;
- L'emplacement géographique : c'est également un facteur important puisqu'il représente une certaine qualité des infrastructures sanitaires de l'habitat et la facilité d'accès aux soins. Ce sont des éléments qui peuvent sensiblement atténuer l'impact d'une pandémie, de même que le niveau de préparation du pays et la rapidité avec laquelle il pourra y réagir ;
- Le phénomène d'anti-sélection : généralement, une population assurée a un taux de mortalité légèrement inférieur à celui de la population globale dans la mesure où ce sont des individus ayant davantage et plus rapidement accès à des soins médicaux. Ce phénomène est à prendre en compte dans l'évaluation de notre modèle.

4.2.4. Les différents modèles compartimentaux déterministes

4.2.4.1. Le modèle SI

Ce modèle à deux états (Susceptible, Infecté) a été développé par Hamer en 1906.

4.2.4.2. Le modèle SIR

Ce modèle à trois états (Susceptible, Infecté, Retiré) a été développé par Kermak et McKendrick en 1927.

4.2.4.3. Le modèle SEIR

Par l'introduction d'un quatrième état (E pour Exposé), le modèle SEIR est celui qui correspond le mieux à la modélisation d'une épidémie de grippe. L'état E permet en effet de modéliser la période d'incubation pendant laquelle l'individu est infecté mais pas encore contagieux.

5. Développement de notre modèle

5.1. Mode opératoire

5.1.1. Construction d'un réseau

La propagation d'une épidémie et son impact sur la population totale dépendent bien sûr du nombre de personnes « susceptibles » en $t=0$ mais également de la façon dont les personnes sont « connectées ». En effet, comme nous l'avons vu au travers des exemples historiques, toutes choses égales par ailleurs, une épidémie se propage d'autant plus vite que les connexions entre individus sont nombreuses, ce qui est facilement compréhensible. On entend par connexion le lien entre deux personnes qui permet la transmission du virus : la transmission sanguine ou par voie aérienne en sont les exemples les plus intuitifs, mais on peut penser à d'autres exemples de connexions tels que les rapports sexuels (pour les maladies sexuellement transmissibles) ou les connexions informatiques (pour les virus informatiques).

Il y a donc un réel enjeu dans le point de départ qu'est la construction de notre réseau. Pour se rapprocher davantage de la réalité, nous ne considérons pas les contaminations de façon aléatoire entre les individus (c'est-à-dire tout individu infecté peut contaminer n'importe quel autre individu) mais nous allons modéliser des liens dans la population qui, pour deux individus donnés, seront existants ou non (modèle binaire). Ainsi, pour deux individus pris aléatoirement dans la population et outre les paramètres propres au virus, celui-ci ne pourra se transmettre que s'il existe un lien. Si le lien n'existe pas, la transmission est impossible.

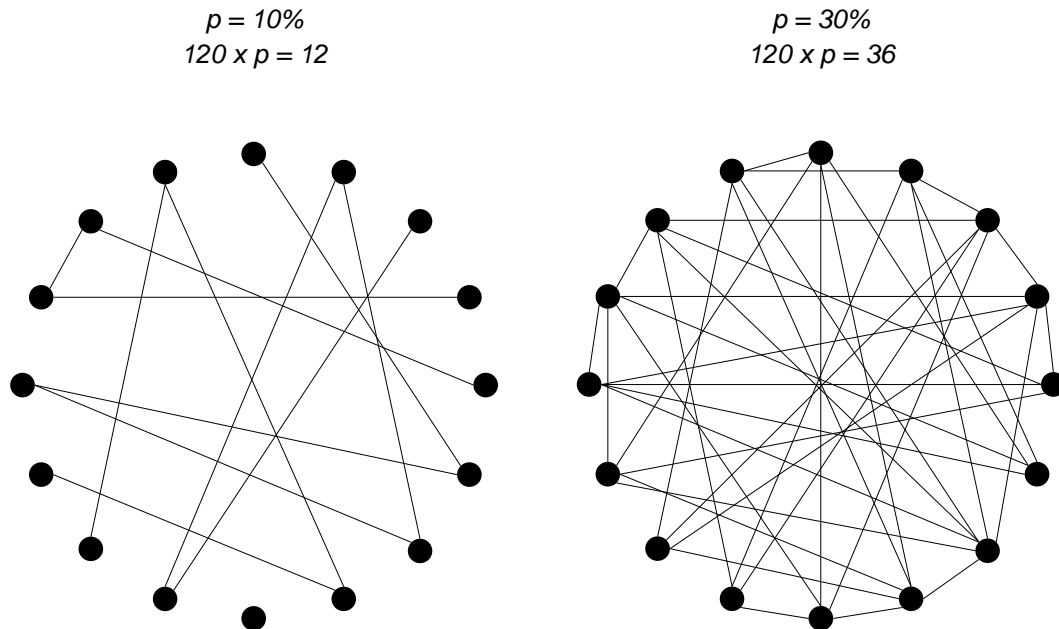
L'objectif est de se rapprocher d'une structure sociale réelle avec la possibilité d'intégrer les différents paramètres qui peuvent conditionner l'existence d'un lien social entre 2 individus donnés : appartenance familiale, distance géographique (résidentielle et professionnelle), appartenance à des groupes sociaux ou religieux (associations, clubs de sports, etc.).

Pour ce faire, nous utilisons le modèle de (Groendyke et *al.*, 2010) qui passe par la génération d'un graphe dit d'Erdős-Rényi. Considérons une population de taille constante composée de N individus représentée par un graphe G à $S = \{1, \dots, N\}$ sommets. Pour tout $(i,j) \in S^2$ avec $i \neq j$, l'arête $\{i,j\}$ est dans G avec la probabilité p_{ij} , indépendamment de toutes les autres paires de sommets. S'il y a une arête, on estime que le virus pourra être véhiculé entre les deux individus i et j , autrement on estime que la contamination est impossible.

Dans le cas particulier où $p_{ij} = p$ pour tout $(i,j) \in S^2$ avec $i \neq j$, autrement dit la probabilité de présence d'une arête entre 2 sommets est la même pour tous les sommets, cela revient à prendre successivement toutes les paires du graphe et effectuer un tirage de Bernoulli avec cette probabilité p (les tirages de Bernoulli étant indépendants entre eux).

➤ **Exemple de construction de réseau avec $N = 16$**

Le nombre maximum de relations est de $[N \times (N-1) / 2] = 120$. On réalise donc un tirage de Bernoulli sur chacune de ces 120 liaisons possibles avec une probabilité p . L'espérance du nombre d'arêtes construites est ainsi de $(120 \times p)$. Ci-dessous deux représentations du réseau obtenu après un tirage aléatoire avec des probabilités $p=10\%$ et $p=30\%$



Le cas général de construction d'un réseau est décrit par la formule suivante :

$$(E1) \quad \ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \sum_m \eta_m X_{\{i,j\},m}$$

Qui peut se réécrire de façon équivalente sous la forme :

$$p_{ij} = \frac{\exp(\sum_m \eta_m X_{\{i,j\},m})}{1 + \exp(\sum_m \eta_m X_{\{i,j\},m})}$$

où :

- les η_m sont des paramètres réels
- les $X_{\{i,j\},m}$ sont les termes (i,j) de m matrices dyadiques

Concrètement, cela signifie que pour construire notre réseau, la probabilité p_{ij} d'établir la liaison entre i et j va dépendre de m jeux de variables $(X_{\{i,j\},m})$ caractéristiques du couple (i,j) , pondérés par les η_m .

Le cas le plus simple consiste à avoir $m = 1$ et $\forall(i,j) X_{\{i,j\},1} = 1$. Le cas particulier évoqué ci-dessus correspond à cette condition combinée à $\eta_1 = p$.

Concernant les autres termes, au-delà des pondérations η_m , cette formule laisse une grande liberté de choix sur les variables intervenant pour la probabilité p_{ij} . On peut par exemple imaginer :

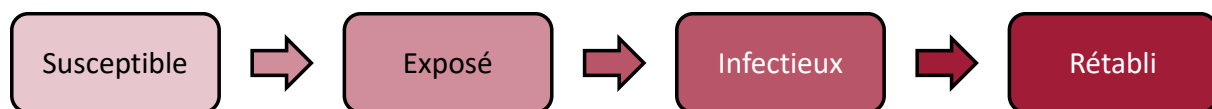
- $X_{\{i,j\},m} = d_{ij} - \hat{d}$ avec d_{ij} la distance (norme à définir) et \hat{d} la moyenne des d_{ij}
- $X_{\{i,j\},m} = \begin{cases} 1 & \text{si } i \text{ et } j \text{ partagent un critère commun (ex: habitent dans le même département)} \\ 0 & \text{sinon} \end{cases}$

5.1.2. Utilisation du modèle SEIR stochastique

Pour les raisons préalablement évoquées, nous choisissons d'utiliser le modèle SEIR, qui est le plus répandu pour modéliser des épidémies grâce à l'introduction de l'état E qui s'apparente à la période d'incubation d'un virus.

Pour rappel, les 4 différents états sont donc :

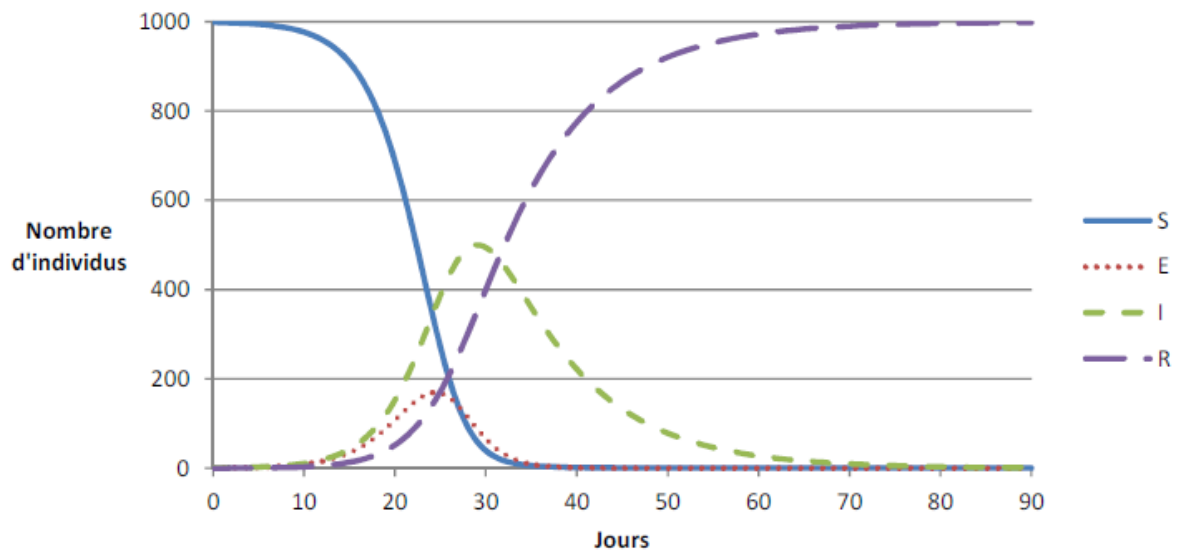
- S = Susceptibles d'être infectés : non immunisés contre la maladie et pouvant être contaminés par d'autres individus.
- E = Exposés (ou Latents) : contaminés par la maladie mais qui ne peuvent pas encore transmettre eux-mêmes la maladie. Ils présentent les symptômes de la maladie mais sont encore dans la période de latence avant de devenir eux-mêmes infectieux (compartiment I).
- I = Infectieux : infectés et contagieux
- R = Rétablis : guéris et immunisés contre la réinfection



Il est important de noter que les individus R sont des individus qui « sortent » du modèle, c'est-à-dire qu'ils ne peuvent plus revenir dans un des états précédents. Ainsi, ce modèle ne distingue pas les individus décédés des individus restés en vie mais guéris du virus. Ces 2 classes sont confondues. Par la suite, nous verrons que notre modélisation des décès se fait par l'application d'un taux de mortalité (sur base d'observations réelles et historiques) à cette population R.

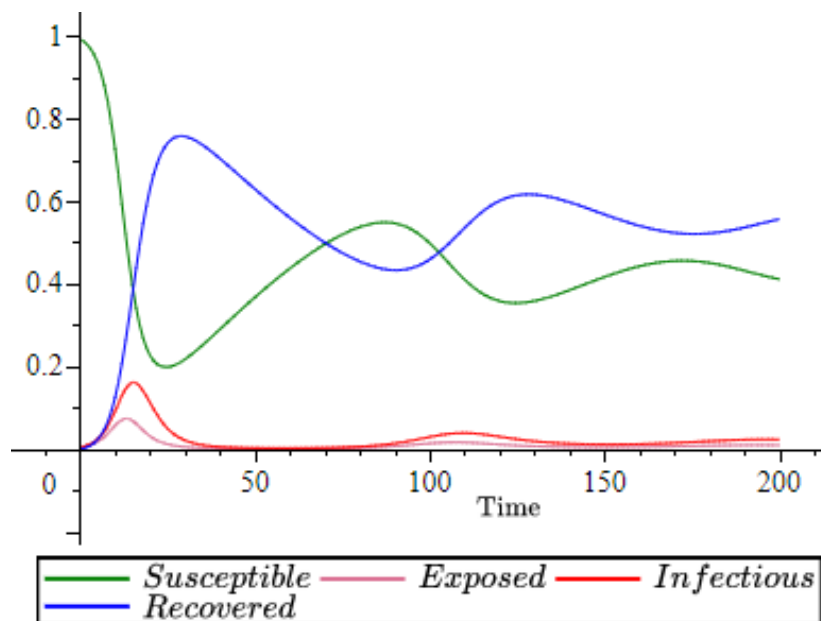
Il existe un modèle SEIR déterministe qui est régi par un système d'équations différentielles dont les paramètres (probabilités de passage d'un état à un autre) sont des constantes. Le principe du modèle SEIR déterministe est d'introduire un individu infecté dans une population d'individus susceptibles. Les équations différentielles et leurs paramètres régissent ensuite l'évolution dans le temps de chacune des populations.

Ci-dessous un exemple simple²⁵ :



Dans cet exemple, toute la population est infectée et termine dans l'état rétabli (vivant ou décédé), ce qui n'est pas réaliste. Le modèle déterministe peut être affiné sur les paramètres ou sur les hypothèses, par exemple : est-ce que des naissances sont modélisées ou est-ce que l'hypothèse est d'avoir un groupe fermé ?

L'exemple ci-dessous²⁶ est celui d'un modèle déterministe où la population globale n'est pas fermée puisqu'on y autorise les naissances :



²⁵ Speisser, R. (2013)

²⁶ Günter Edenharter. *The SEIR model with births and deaths*. Récupéré le 25 février 2017 de <https://maple.cloud/#doc=6407056173039616>

Si le modèle déterministe a l'avantage d'être relativement simple à modéliser et à représenter, il a l'inconvénient majeur de ne représenter qu'un scénario précis, qui est totalement conditionné par les paramètres des équations différentielles. Il ne tient pas compte des aléas de transmission, taux d'infection, rapidités de rétablissements, etc. De même, la position de départ de l'individu infecté et les éloignements entre tous les individus de la population ne sont pas des facteurs influents, or il est évident que ce sont des éléments très importants dans la propagation d'une pandémie.

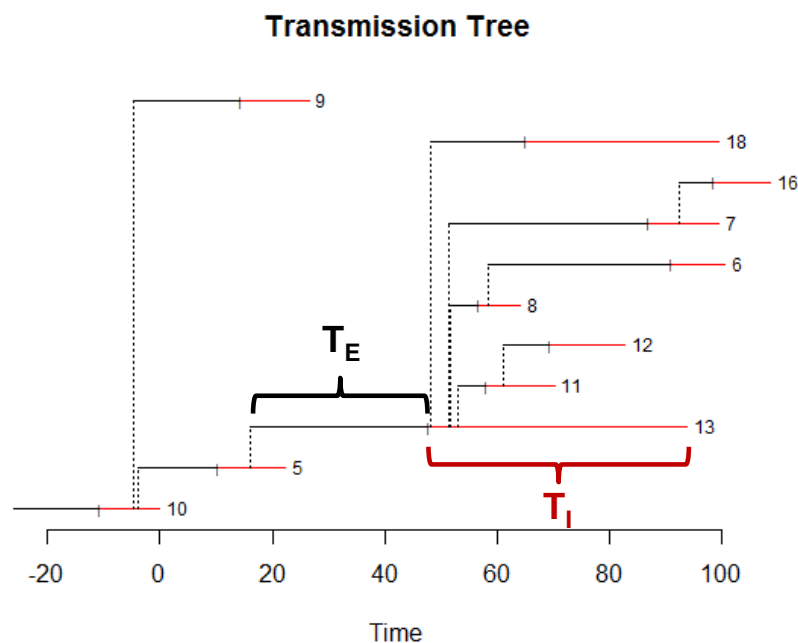
Nous choisissons donc plutôt de construire un modèle SEIR stochastique en associant des distributions de probabilité aux variables et paramètres du modèle. Cela permet de se rapprocher davantage de la réalité en introduisant de l'aléa dans les périodes d'incubation, d'infection et taux de transmission.

5.1.2.1. Temps d'incubation et d'infectiosité

Dans le modèle SEIR stochastique, on définit les variables aléatoires suivantes :

- T_E correspond à la période de latence (ou d'incubation), i.e. la période pendant laquelle un individu reste dans l'état E
- T_I , de façon similaire, correspond à la période pendant laquelle un individu reste dans l'état I, appelée période d'infectiosité (ou de contagion)

Nous utilisons le logiciel R, dans lequel il existe plusieurs packages de base permettant de travailler sur les modèles SEIR. En particulier, le package « *epinet* »²⁷ permet de représenter graphiquement de façon simple une épidémie :



L'exemple ci-dessus est construit avec une population de départ de $N = 20$ individus et une probabilité $p = 10\%$ pour la construction du réseau.

²⁷ Package public et disponible sur <https://cran.r-project.org/web/packages/epinet/index.html>

Lecture du graphique :

- L'individu 10 est le patient zéro, i.e. le premier individu contaminé.

Dans le package, ce patient zéro est déterminé de façon aléatoire (méthode non précisée) parmi tous les individus susceptibles. La fonction `set.seed()` dans R permet néanmoins (quand cela s'avère nécessaire) de « mémoriser » une réalisation d'un processus stochastique, ce qui permet de reproduire le même scénario quand on fait plusieurs tirages d'un processus aléatoire. Dans notre exemple du modèle SEIR, cette fonction permet donc de retrouver à chaque tirage le même patient zéro et la même « trajectoire » de l'épidémie.

- La ligne noire décrit sa période en état E : il est contaminé mais ne peut pas transmettre. Il passe ensuite en état I, décrit par la ligne rouge. C'est pendant cette période qu'il peut contaminer les autres individus, les transmissions sont représentées par les lignes verticales et pointillées.
- Ainsi, on voit que pendant sa période infectieuse (I), l'individu 10 transmet le virus aux individus 5 et 9

In fine, cette épidémie aura touché 11 individus, les 9 autres sont restés dans l'état susceptible (S) et n'ont jamais été contaminés. Le fait qu'un individu de la population de départ ne soit pas contaminé peut s'expliquer de trois façons :

- 1) Initialement, lors de la construction du réseau, tous les tirages de Bernoulli (sur chacune des paires qu'il forme avec les autres individus) ont « échoué » et ont abouti à ce qu'il n'y ait jamais de liaison créée. Dans ce cas, peu importe la pandémie (sévérité, point de départ, etc.), l'absence de liaison avec le reste de la population le prémunit contre toute contamination. En faisant le parallèle avec le monde réel, cela pourrait représenter une personne maintenue isolée dans un bunker ou sur une autre planète par exemple ;
- 2) L'individu est en contact avec au moins un autre qui devient infectieux (I) à un moment donné de la pandémie mais ce dernier se rétablit (R) avant même de transmettre le virus ;
- 3) L'individu est en contact avec d'autres individus mais aucune de ses liaisons directes n'a été contaminée, pour une des deux premières raisons.

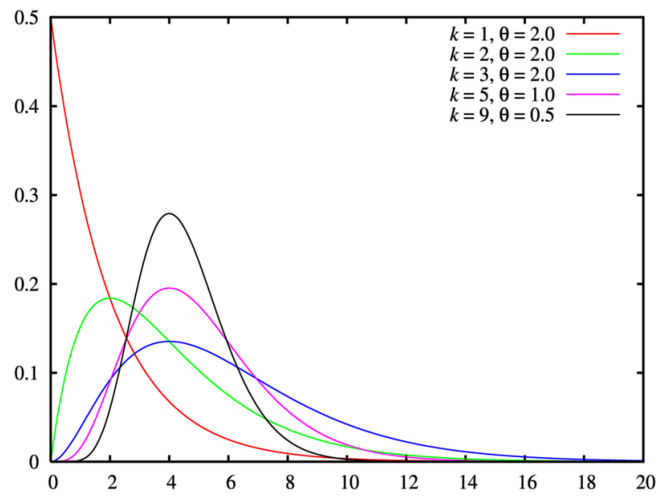
En lisant le deuxième motif de non-contamination potentielle, la question naturelle qui s'en suit est donc de savoir : comment sont régies les périodes T_E et T_I , ainsi que la transmission entre deux individus ?

Nous nous appuyons sur les travaux de Groendyke et *al.* (2010) ainsi que ceux de Ray et Marzouk (2008) selon lesquels ces 2 variables suivent des lois Gamma de paramètres respectifs (k_E, θ_E) et (k_I, θ_I) . Ces lois Gamma sont en effet les plus utilisées dès lors qu'il s'agit de modéliser des phénomènes temporels, ce qui est exactement notre objet.

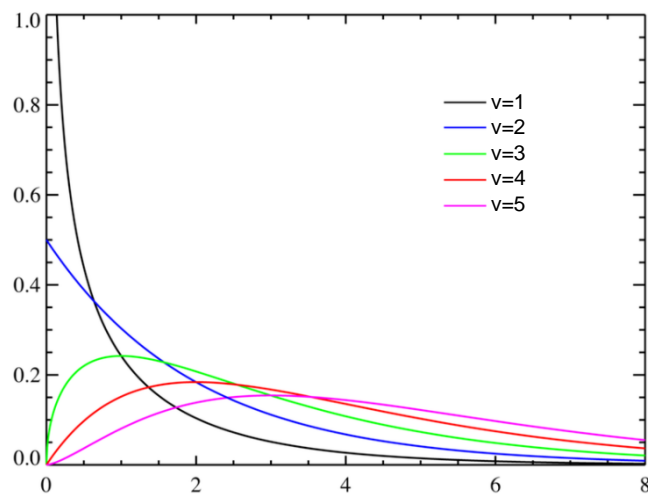
Les lois Gamma incluent entre autres la loi du χ^2 et la loi exponentielle, c'est-à-dire que ces dernières se retrouvent pour des valeurs particulières dans les paramètres de loi Gamma. Nous aurions pu utiliser une de ces deux lois, comme l'ont fait Britton et *al.* (2002) qui avaient choisi une loi exponentielle pour modéliser les périodes T_E et T_I .

Ci-dessous la comparaison des lois de densité des trois lois mentionnées :

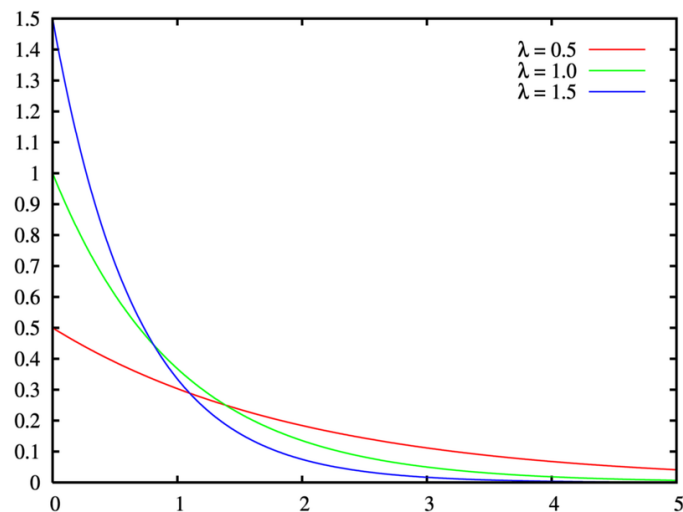
➤ **Loi Gamma**



➤ **Loi du χ^2** (= loi Gamma de paramètres $k = v/2$ et $\theta = 2$)



➤ **Loi Exponentielle** (= loi Gamma de paramètres $k = 1$ et $\theta = 1/\lambda$)



Nous choisissons de conserver la loi Gamma, qui offre davantage de flexibilité dans la modélisation d'une épidémie, justement parce qu'elle dispose de deux paramètres (contre un seul pour la loi du χ^2 et la loi exponentielle) qui permettent de modifier la forme et l'échelle de sa densité de probabilité. Ce nombre de paramètres reste suffisamment restreint pour éviter d'avoir un modèle trop compliqué.

Par souci de simplification, nous choisissons également de retenir la même loi pour modéliser les états E et I :

$$\begin{cases} k_E = k_I = k \\ \theta_E = \theta_I = \theta \end{cases}$$

Nous faisons donc l'hypothèse que la période d'incubation T_E et la période d'infectiosité T_I sont semblables, ce qui est une hypothèse simplificatrice mais qui reste suffisamment réaliste. Ce choix et le dimensionnement de ces grandeurs est détaillé par la suite, nous verrons que c'est une approximation réaliste et que nous retenons une valeur moyenne de 5,5 jours pour ces deux périodes, ce qui est en ligne avec les valeurs réelles observées aujourd'hui. Notre modèle et le package « epinet » de R (voir ci-après) permettent sans difficulté de les modéliser de façon différente.

Rappelons que pour une loi $T \sim \text{Gamma}(k, \theta)$:

$$\begin{cases} k, \theta > 0 \\ E[T] = k\theta \\ V[T] = k\theta^2 \end{cases}$$

Ces 2 variables k et θ vont donc déterminer le temps d'incubation et le temps d'infectiosité et nous allons les dimensionner par la suite. Ce sont des variables propres à chaque individu, indépendamment du milieu dans lequel ils évoluent et des interactions possibles avec les autres individus.

Or on sait qu'une question fondamentale dans la taille et l'impact d'une épidémie réside dans sa propagation, c'est-à-dire la capacité de transmission du virus d'un individu à un autre, élément que nous n'avons pas encore analysé à ce stade.

C'est ici qu'intervient le réseau et les différents contacts que nous avons construits préalablement. Ce temps de transmission est une autre variable à déterminer, c'est l'objet du paragraphe suivant.

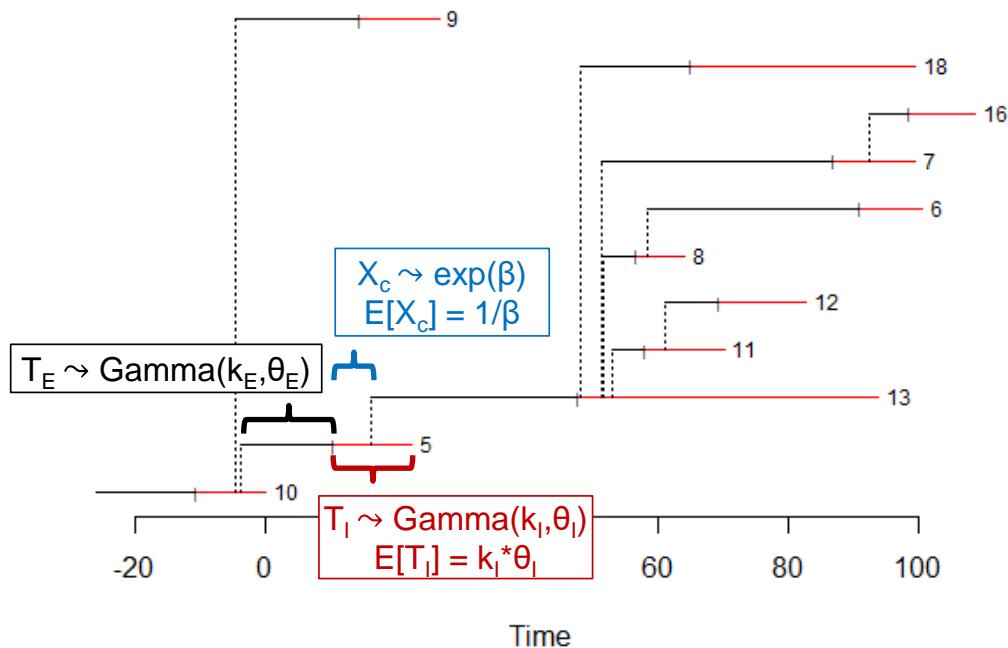
5.1.2.2. Transmission d'un virus entre deux individus

Nous avons vu que dans un premier temps, nous pouvons séparer les paires d'individus en deux ensembles : celles pour lesquelles il existe un contact (possibilité de transmission du virus) et celles pour lesquelles le contact n'a pas été établi (aucune transmission possible).

Pour chaque paire possédant un contact c , dès lors qu'un des deux individus est infectieux (état I), on modélise par une loi exponentielle de paramètre β le temps X_c de transmission du virus, c'est-à-dire le temps entre le début de l'état I du premier des deux individus et le moment où le deuxième passe de l'état S à E (i.e. contracte le virus). La loi exponentielle est très répandue pour les modélisations de durée de vie d'un phénomène sans mémoire ou sans vieillissement, ce qui semble parfaitement adapté ici pour un virus (il paraît en effet légitime de supposer que la transmission d'un virus s'effectue indépendamment de son « parcours » précédent).

Complétons le graphique vu précédemment en l'enrichissant de ces notions :

Transmission Tree



Ainsi, pour chaque contact c , on va comparer les observations des variables aléatoires X_c et T_i . Le virus est transmis si et seulement si $X_c \leq T_i$. De fait, sur toute la population de départ, la propagation du virus est conditionnée dans un premier temps par l'existence ou non d'un contact social et dans un deuxième temps par la rapidité de transmission (X_c comparativement à T_i).

Rappelons que pour une loi $X_c \sim \exp(\beta)$:

$$\begin{cases} \beta > 0 \\ E[X_c] = \frac{1}{\beta} \\ V[X_c] = \frac{1}{\beta^2} \end{cases}$$

Pour la suite, lorsque nous dimensionnerons les paramètres de notre modèle, nous pourrons donc retenir que la comparaison entre les valeurs $k_i \theta_i$ (période d'infectiosité) et $\frac{1}{\beta}$ (temps de transmission moyen) aura son importance dans la propagation et donc l'ampleur de l'épidémie.

5.1.3. Le taux de reproduction de base R_0

Comme évoqué ci-avant, la propagation d'une épidémie (au sens rapidité et facilité de transmission du virus) est un élément déterminant dans son impact sur une population donnée. Pour cela, un paramètre est défini pour mesurer cette information. Il est appelé « taux de reproduction de base », noté R_0 et représente le nombre moyen (espérance mathématique) d'individus qu'un infectieux pourra contaminer. Cette valeur est considérée dans une population où il n'y a qu'un seul individu infectieux au sein d'une population de susceptibles et mesurée à l'instant où débute cette période d'infectiosité pour le premier individu. Si la population de départ (outre l'individu infectieux) n'est pas intégralement composée de susceptibles, on parle d'un taux de reproduction net R .

Ainsi, on comprend intuitivement que la position de R_0 par rapport à 1 est révélatrice :

- Si $R_0 < 1$, un individu infectieux contaminera en moyenne moins d'une personne, le virus est donc voué à s'éteindre et ne se propage pas dans la population
- Si $R_0 \geq 1$, le virus se propage dans la population, transformant ainsi une simple maladie au départ en une épidémie

Dans le cas particulier où le réseau a été construit de façon élémentaire avec un tirage de Bernoulli et une probabilité p sur une population initiale de N personnes, et avec les mêmes notations que précédemment, cette quantité peut s'exprimer de la manière suivante (Britton et O'Neill (2002)) :

$$R_0 = N \times p \times P(X_c < T_I)$$

En conservant les lois utilisées précédemment pour $X_c \sim \text{Exp}(\beta)$ et $T_I \sim \text{Gamma}(k_I, \theta_I)$, on peut montrer que :

$$R_0 = N \times p \times \left(1 - \left(\frac{1}{1 + \beta\theta_I} \right)^{k_I} \right)$$

Dans la suite de notre modèle, cette variable sera un des axes principaux de nos analyses, de même que le taux de létalité (voir ci-après). Nous choisissons de modéliser des épidémies dont le R_0 varie entre 0 et 3, le dimensionnement de cette valeur est documenté dans le paragraphe suivant car il est intéressant d'analyser le taux de reproduction d'une épidémie au regard de son taux de létalité.

5.1.4. Le taux de létalité

Comme déjà présenté, le modèle SEIR ne distingue pas dans la population des rétablis (R) les individus vivants des décédés. Or, pour évaluer l'impact pour l'assureur et se rapprocher davantage de la réalité, nous avons besoin de la donnée du nombre de décès relatif à l'épidémie. Pour ce faire, nous appliquons des taux de létalité à la population qui a été infectée.

Le taux de létalité mesure le nombre de décès au nombre de personnes infectées, il caractérise ainsi la dangerosité (au sens qu'il est plus ou moins mortel) d'une épidémie ou d'un virus de façon plus générale. C'est un paramètre bien distinct du taux de reproduction R_0 , qui mesure la capacité d'un virus à se propager plus ou moins rapidement : un virus peut se propager de façon très vive mais pour autant ne pas être mortel du tout, et inversement.

Nous avons souhaité construire un tableau comparatif des taux de reproduction et de létalité pour plusieurs épidémies historiques afin de nous donner des ordres de grandeur et dimensionner les paramètres de notre modèle. Cependant, au travers de nos recherches et de notre bibliographie, nous nous sommes rapidement aperçus qu'il était très difficile d'obtenir des valeurs fiables et sur lesquelles il y avait des consensus. En effet, pour une même épidémie donnée, les taux de létalité indiqués peuvent varier très fortement selon la source d'information.

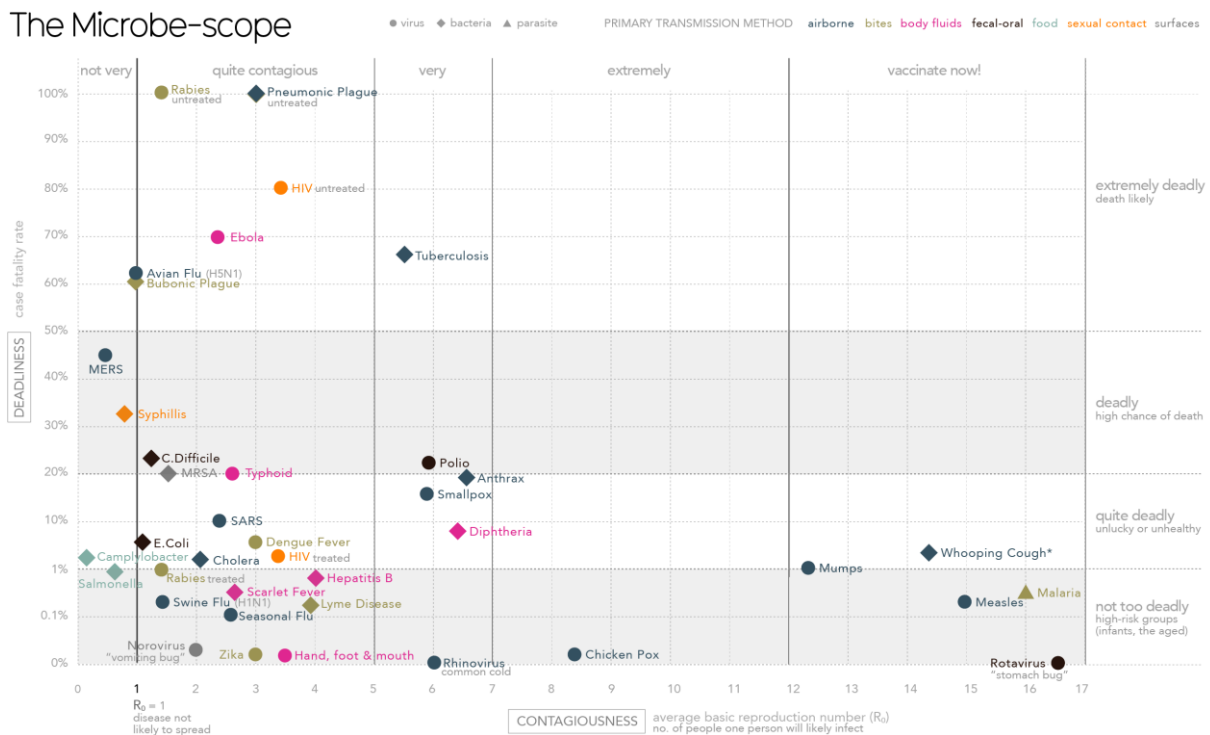
A titre d'exemple, voici un comparatif réalisé par l'APREF dans sa note sur la pandémie (2016), qui montre la variabilité des taux indiqués :

Article	Auteur	Taux d'attaque	Taux de létalité
Pandemic Influenza Risk Management - WHO Interrim Guidance	WHO	1918	2-3%
		1957	<0,2%
		1968	<0,2%
Potential Impact of Pandemic Influenza on the US Life Insurance Industry	Jim Toole, FSA, MAAA	1918	2,17%
		1957	0,23%
		1968	-
Managing Influenza Pandemic Risk	RMS	1918	2,4%
		1957	1,2%
		1968	0,5%

A noter que le taux d'attaque présenté ici rapporte le nombre de personnes infectées à la fin de la pandémie à la population totale.

Cette incertitude sur les différents taux s'explique principalement par l'absence de réel outil de mesure (enquêtes, relevés sanitaires) et quand ces outils pouvaient exister, la fiabilité des données n'est pas certaine, compte tenu du contexte de guerre et/ou de la non exhaustivité des mesures.

Toutefois, nous pouvons nous reposer sur ce graphique réalisé par David McCandless (sur la base de données de l'OMS) qui répertorie les taux de reproduction R_0 (en abscisses) et taux de létalité (en ordonnées) de plusieurs virus, bactéries et parasites :



David McCandless v1.15 / Jan 2016
InformationisBeautiful.net

sources: Centers for Disease Control, World Health Org, CIDRAP, studies
fatality rate for healthy adult in developed nation, * = infants

data: bit.ly/KIB_Microbescope
part of KnowledgeisBeautiful

Pour plus de lisibilité, le lecteur pourra se référer au site Internet²⁸, sur lequel sont également disponibles les données en format Excel.

5.1.5. Dimensionnement de ces taux

Sur la base du graphique précédent et de nos différentes recherches, nous choisissons de choisir de faire varier le taux de reproduction entre 0 et 3 et le taux de létalité dans un intervalle entre 0% et environ 6% :

$$\text{Taux de reproduction } R_0 \in [0 - 3]$$

$$\text{Taux de létalité} \in [0\% - 6\%]$$

Ces intervalles ne sont pas à considérer comme bornés à droite au sens mathématique du terme, il s'agit des intervalles sur lesquels nous souhaitons concentrer l'essentiel de nos valeurs, il n'est pas exclu que les taux prennent ponctuellement des valeurs supérieures à 3 et 6% respectivement puisqu'en pratique, elles seront modélisées par des variables aléatoires sur R_+ .

Ce choix nous paraît être cohérent et réaliste pour être représentatif des gripes (à la fois des gripes faibles et des gripes plus importantes) puisque ce sont essentiellement les gripes que nous modélisons. Nous ne tenons pas compte en effet des événements extrêmes (en taux de létalité ou de reproduction) car ils ne sont pas le reflet de gripes (peste, rage, VIH, rougeole, malaria).

5.1.5.1. Taux de reproduction R_0

Commençons par rappeler que :

- le taux de reproduction n'est que la conséquence de trois paramètres, regroupés dans une seule variable :

$$R_0 = N \times p \times \left(1 - \left(\frac{1}{1 + \beta \theta_I} \right)^{k_I} \right)$$

- $1/\beta$ représente le temps moyen de transmission du virus entre deux individus
- $k_I \times \theta_I$ représente le temps moyen de l'état infecté (et également de l'état d'incubation puisque nous avons choisi $k_E = k_I = k$ et $\theta_E = \theta_I = \theta$).

Nous cherchons donc à trouver trois lois pour modéliser β , k et θ qui aboutiront à une distribution qui nous satisfait pour paramètre résultat R_0 , l'objectif étant que celui-ci soit dans un intervalle $[0 ; 3]$, plutôt centré sur la gauche de cet intervalle pour que notre échantillon d'épidémies soit le plus réaliste possible. En effet, les valeurs entre 0 et 1 représentent les gripes bénignes et doivent être une certaine majorité de notre échantillon tandis que les valeurs entre 2 et 3 représentent les gripes plus virulentes et doivent être sur la queue de distribution.

²⁸ <https://informationisbeautiful.net/visualizations/the-microscope-infectious-diseases-in-context/>

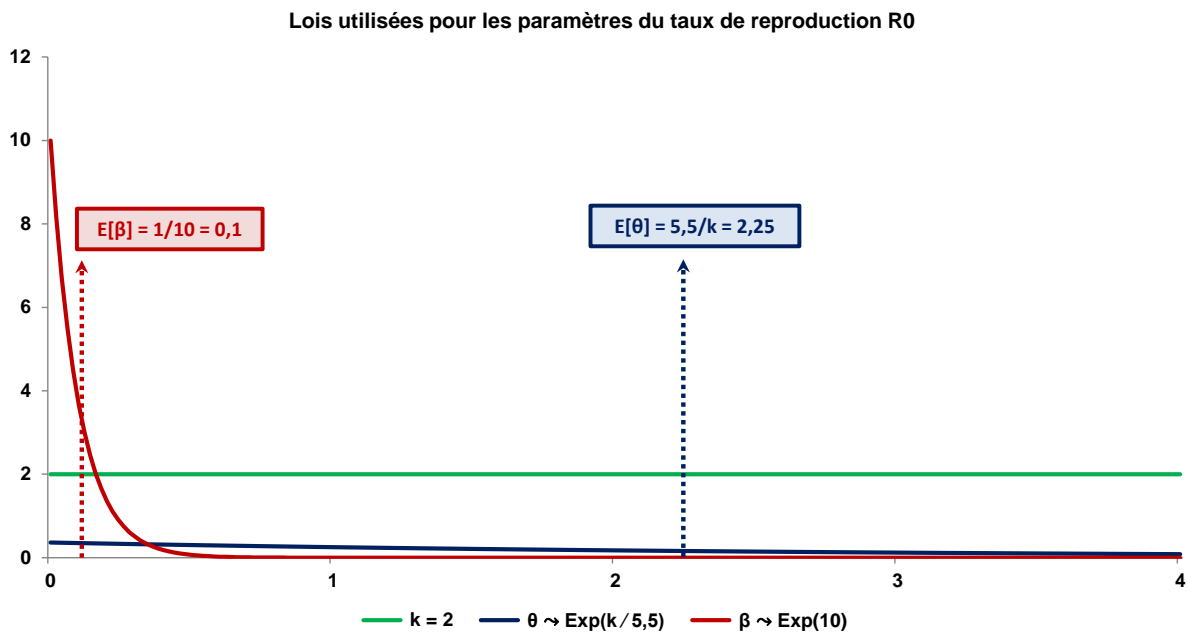
Nous imposons une deuxième contrainte sur le couple (k, θ) puisque nous visons un temps moyen pour T_E et T_I (donné par $k * \theta$ puisqu'on rappelle que ces temps suivent une loi Gamma) entre 5 et 6 jours. C'est en effet une bonne moyenne entre le temps d'incubation d'une grippe (environ 3 jours) et le temps d'infectiosité (1 à 2 semaines), d'après nos différentes recherches sur les données épidémiologiques de grippe. La contrainte sur l'espérance est donc :

$$E[k \times \theta] = 5,5$$

Nous avons donc essayé et combiné de multiples lois pour chacun de ces trois paramètres pour aboutir à une distribution qui nous apparaît réaliste pour le modèle. Notre choix s'est arrêté sur :

- $k = 2$: nous figeons un des deux paramètres soumis à la contrainte ci-dessus
- $\theta \sim \text{Exp}(k/5,5) = \text{Exp}(2/5,5)$: on a bien $E[k \times \theta] = k \times E[\theta] = k \times 5,5/k$
- $\beta \sim \text{Exp}(10)$: en espérance, β prend une valeur moyenne de 0,1 et donc le temps moyen de transmission est de $1/\beta$ soit 10 jours

Nous pouvons synthétiser le choix de ces paramètres ainsi :

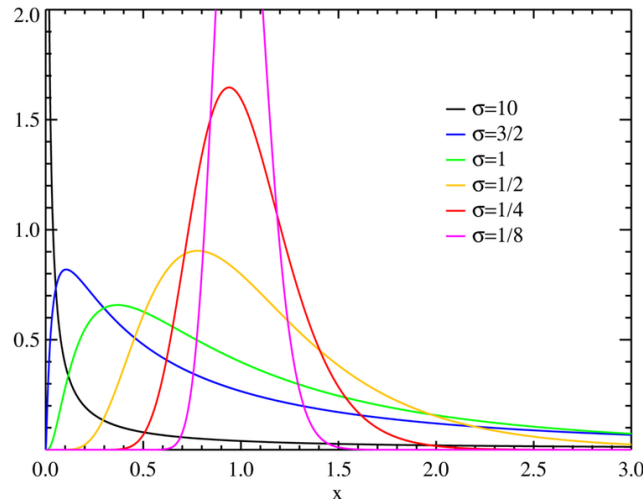


Nous avons vu précédemment que dans le modèle SEIR stochastique, la capacité de propagation d'une épidémie repose sur le sens de l'inégalité entre $k\theta$ et $1/\beta$. Étant donné que nous avons modélisé ces paramètres par des variables aléatoires, il aurait pu être intéressant de calculer et comparer en espérance $E[k\theta]$ et $E[1/\beta]$. Or comme β suit une loi exponentielle, l'espérance $E[1/\beta]$ ne converge pas.

5.1.5.2. Taux de létalité

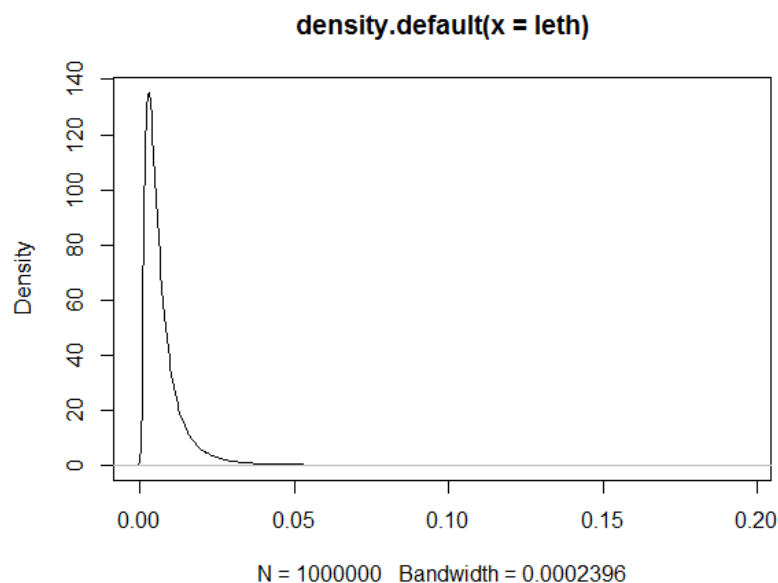
Concernant la loi de distribution du taux de létalité, nous optons pour la loi log-normale. Pour rappel, voici à quoi ressemble la densité de probabilité d'une loi log-normale de paramètres $\mu = 0$ et pour différentes valeurs de σ .

➤ Loi log-normale



Ce choix de la loi log-normale s'explique principalement par le fait qu'elle dispose d'une queue de distribution relativement élevée, ce qui permet de considérer quelques valeurs extrêmes avec des probabilités non négligeables. C'est d'ailleurs cette loi qu'utilise Swiss Re²⁹ pour modéliser le taux de surmortalité, calibrée en pondérant les données de plusieurs épidémies du 20^{ème} siècle aux Etats-Unis.

En termes de paramétrage, nous choisissons de centrer notre taux de létalité sur 0,5% et de calibrer le 2^{ème} paramètre de la loi log-normale pour que la queue de distribution ne soit pas trop épaisse après la valeur de 6%. Ceci nous amène à utiliser la loi log-normale de paramètres $\mu = \log(0,5\%)$ et $\sigma = 0,8$, dont voici la densité de probabilité :



²⁹ Se reporter à "Pandemic influenza : A 21st century model for mortality shocks"

5.1.6. Fonctionnement du package « *epinet* » dans R

Un outil important de notre modèle réside dans le package « *epinet* » du logiciel R, sur lequel que nous avons travaillé. Ce package a été développé par Groendyke et *al.* (2010), tel qu'ils le précisent dans leur œuvre « Bayesian inference for contact networks given epidemic data » (§ 4.6).

Ce package permet de simuler une pandémie selon le modèle SEIR stochastique et en procédant comme nous venons de le décrire précédemment dans le mode opératoire :

- Les entrées à fournir au package sont le nombre N d'individus puis les paramètres η_m et $X_{(i,j),m}$ qui, pour rappel déterminent les probabilités de construction d'un contact entre deux individus i et j
- Le package construit ainsi un réseau selon le tirage de Bernoulli
- Une fois le réseau construit, le package simule une pandémie de façon stochastique grâce à la fonction *SEIR.simulator* qui prend en entrées
 - Le nombre N d'individus et le réseau construit préalablement
 - Les valeurs des paramètres $\beta, k_E, \theta_E, k_I$ et θ_I de la loi exponentielle et des deux lois Gamma

En sortie du package, plusieurs informations sur la pandémie sont disponibles de façon tabulaire ou graphique, que nous verrons sur un exemple simple ci-après.

Pour des explications plus détaillées sur les fonctions du package « *epinet* », le lecteur pourra se référer aux annexes de ce mémoire via les explications inscrites dans notre code R, ou directement à la documentation R³⁰.

5.1.7. Modèle stochastique

La pandémie qui est en sortie de ce package n'est qu'une réalisation du modèle SEIR stochastique. En effet, elle est le fruit d'un processus qui a rencontrés plusieurs aléas :

- Initialement lors de la construction du réseau via les tirages de Bernoulli (paramètres p_{ij})
- Puis lors des durées de latence et d'infectiosité (paramètres $\beta, k_E, \theta_E, k_I$ et θ_I) ainsi que des temps de transmission (paramètre β)

Comme pour tout processus stochastique, il est donc indispensable d'en obtenir plusieurs réalisations (des « runs ») afin de prendre en compte plusieurs tirages aléatoires et ainsi d'obtenir la trajectoire du processus. Nous pouvons ensuite étudier son espérance, sa variance, etc. Dans notre modèle, pour chaque pandémie, c'est-à-dire la donnée d'un réseau et d'un quintuplet $(\beta, k_E, \theta_E, k_I, \theta_I)$, nous effectuons 50 runs.

Nous avons réalisé de nombreux cas de pandémies et testé plusieurs nombres de runs, allant de 10 à 1 000. Le nombre de 50 runs s'avère être un bon compromis puisqu'il est suffisamment petit pour ne pas avoir des temps de calcul trop conséquents. En effet, avec ce nombre de runs, notre modèle demande déjà un temps assez important : près de 10 heures de calcul pour simuler 100 épidémies x 50 runs sur une population de 10 000 individus.

Nous aurions pu utiliser la fonction *set.seed()*, vue précédemment, pour garder en mémoire chacun des runs, mais cela ne nous est pas nécessaire car pour chaque pandémie, nous conservons uniquement la valeur moyenne sur les 50 runs du nombre d'infectés et la durée de la pandémie. Nous

³⁰ <https://www.rdocumentation.org/packages/epinet/versions/2.1.7/topics/SEIR.simulator>

pourrions (notre code le permet aisément) conserver davantage de données sur chacun des runs et observer par exemple des densités ou fonctions de répartition de la pandémie, mais ce n'est pas l'objet de notre mémoire. Nous choisissons de caractériser une pandémie à son nombre d'infectés et sa durée, et nous avons une maille assez fine pour analyser les résultats puisque nous mémorisons pour chaque individu et pour chaque épidémie simulée le nombre moyen de fois où il a été infecté (sur les 50 runs).

Dans un deuxième temps, nous multiplions le nombre de pandémies en faisant varier leur quintuplet caractéristique dans des fourchettes de valeurs permettant de simuler des pandémies proches de la réalité (cf. ci-dessus avec le dimensionnement de R_0), grâce aux observations des pandémies historiques.

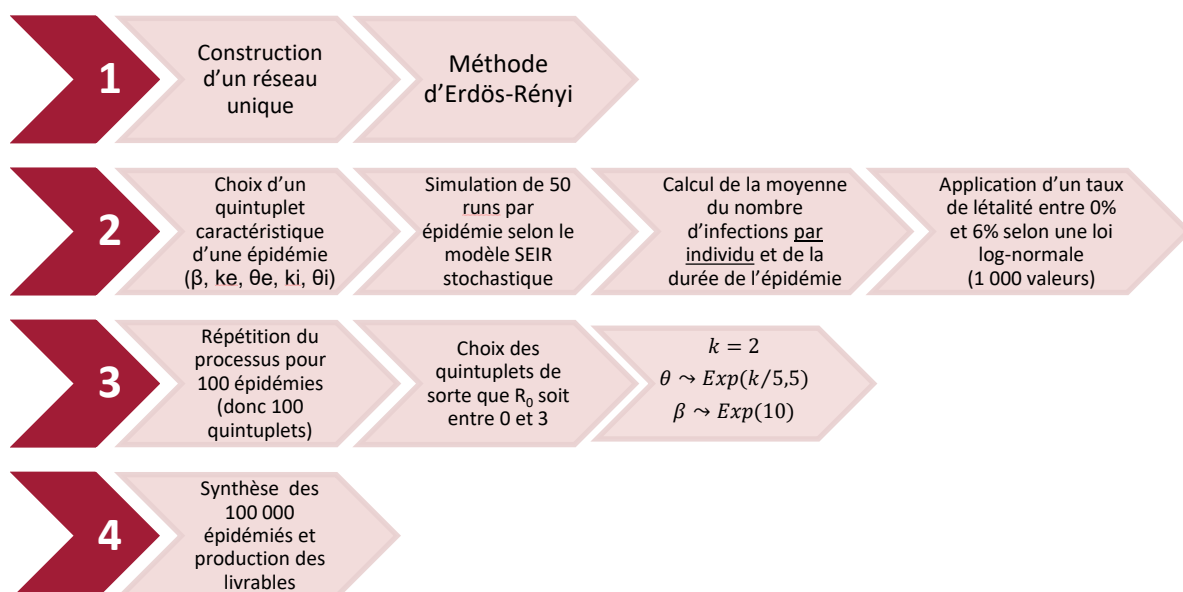
In fine, nous gardons en sortie de code un tableau comme suit (données illustratives) :

Nb moyens d'infections (sur 50 runs)	Epidémie 1	Epidémie 2	...	Epidémie 100
Individu 1	34	15	...	2
Individu 2	46	23	...	5
Individu 3	38	12	...	11
Individu 4	29	27	...	6
...
Individu N-1	44	30	...	9
Individu N	39	17	...	13

Dans cet exemple, l'individu 1 a été contaminé 34 fois sur les 50 simulations de l'épidémie 1.

5.2. Synthèse du mode opératoire

Voici comment nous pouvons résumer notre mode opératoire et la production de nos 100 000 épidémies, qui sont la combinaison de 100 valeurs de R_0 et 1 000 taux de létalité :



Dans notre mode opératoire, on peut voir que le réseau construit au départ est le même qui sert de sous-jacent à toutes nos épidémies (et tous nos runs). C'est une hypothèse importante puisqu'on a vu que les différents contacts étaient déterminants dans la propagation d'une épidémie. Ainsi, tous les résultats et conclusions en sortie de notre modèle sont à mettre au regard de notre réseau unique.

Nous nous sommes interrogés quant à savoir s'il était plus pertinent de conserver ce réseau unique pour toutes les épidémies ou si nous devions simuler aléatoirement un réseau pour chaque épidémie (voire chaque run).

Nous avons opté pour la conservation d'un réseau commun à toutes les épidémies car :

- Cela permet d'avoir des épidémies et runs qui fonctionnent sur une même maille et donc d'obtenir des résultats réellement comparables ;
- Au cours de nos différentes simulations, nous avons vu que les résultats de notre modèle étaient très stables à chaque construction d'un nouveau réseau unique, ce qui conforte notre choix.

L'objectif est que ce réseau soit suffisamment représentatif de la France métropolitaine et d'avoir des résultats réalistes, ce qui est bien le cas avec notre méthode. Cela est vérifié par le fait que nous avons un nombre suffisamment grand d'individus dans la population de départ.

5.3. Scénario 1 : cas simple, sans facteur géographique

Commençons par éprouver notre modèle sur un cas standard, auquel nous allons nous comparer par la suite. A ce stade, le paramètre géographique n'est pas pris en compte dans la construction du réseau, il s'agit simplement d'établir un scénario de référence afin de nous assurer que le calibrage de nos différents paramètres est réaliste et cohérent.

Dans ce scénario donc, les N individus n'ont pas d'emplacement attribué et pour la construction du réseau, nous nous plaçons dans le cas particulier déjà vu précédemment où chaque paire d'individus a la même probabilité d'être en contact selon un tirage de Bernoulli. Nous avons vu que cela revient à ne retenir qu'un paramètre η_1 dans la formule (E1) et $p_{ij} = p \quad \forall (i, j)$.

Il s'agit maintenant de quantifier cette grandeur p , qui représente la probabilité pour une paire (i, j) , tirée aléatoirement parmi les N individus, d'être en contact. Pour cela nous repartons de la formule qui détermine R_0 :

$$R_0 = N \times p \times \left(1 - \left(\frac{1}{1 + \beta\theta_1} \right)^{k_1} \right)$$

Rappelons que nous souhaitons modéliser des épidémies dont le R_0 varie entre 0 et 3. Ainsi, nous choisissons de prendre :

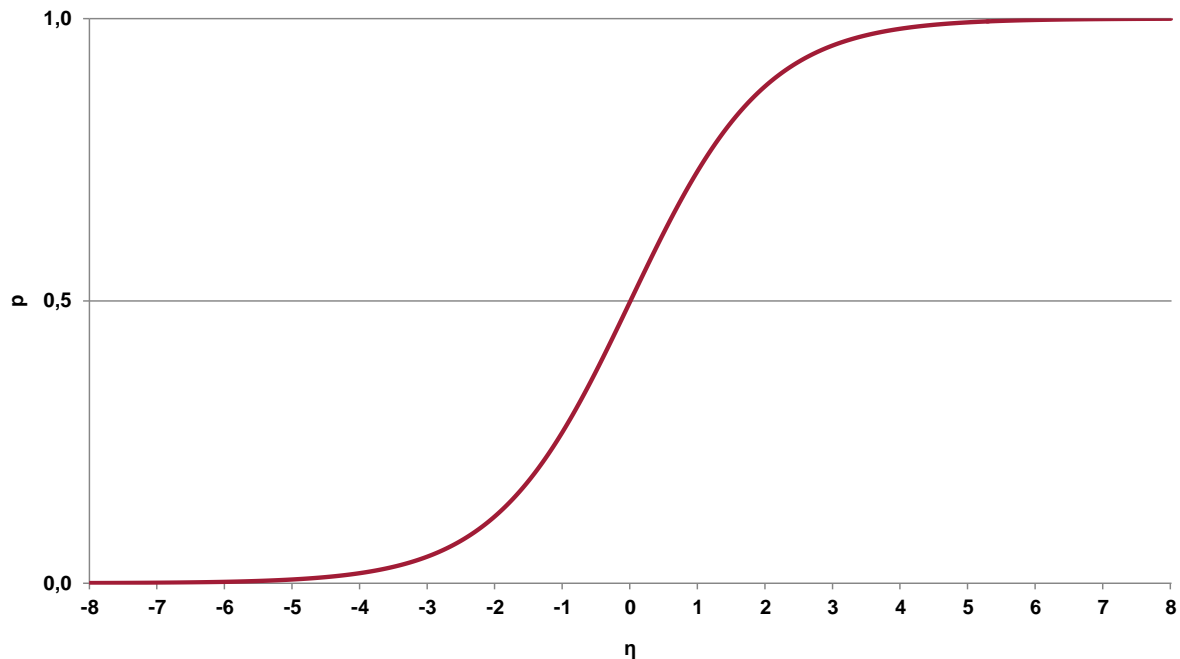
$$p = 3/N$$

Ainsi, le terme précédent la parenthèse est fixé égal à 3, la parenthèse a nécessairement ses valeurs dans $[0 ; 1]$ et ce sont les lois que nous avons choisi précédemment qui permettent d'obtenir le type de distribution souhaitée pour R_0 .

Dans le package *epinet*, la valeur qui sert d'entrée est η , donnée par la formule

$$p = \frac{\exp(\eta)}{1 + \exp(\eta)} \text{ soit } \eta = \ln\left(\frac{p}{1-p}\right)$$

Afin de mieux se représenter les ordres de grandeur, voici la forme de cette fonction :



Dans la mesure où nous avons $p = \frac{3}{N}$, et que N est relativement grand, nous aurons systématiquement des valeurs négatives pour η . Par exemple :

N	p	η
1 000	0,30%	-5,81
2 000	0,15%	-6,50
5 000	0,06%	-7,42
10 000	0,03%	-8,11

Une valeur positive ou nulle pour η impliquerait d'avoir une probabilité de lien $p > 50\%$, soit $N \leq 6$, ce qui est loin d'être réaliste et pertinent pour notre modèle. Nous pouvons donc garder en mémoire les ordres de grandeur suivant : p prend des valeurs plutôt faibles et η prend des valeurs « assez négatives ».

Au-delà de vérifier qu'il est cohérent d'avoir une probabilité décroissante avec la population totale, il est difficile de confronter ce paramètre $p = \frac{3}{N}$ et lui donner une signification concrète. Il faudrait pour cela des données statistiques et des études sur les « contacts » entre individus en France, en ayant défini au préalable la notion de « contact » : lien social, familial, amical, etc. ?

Si le dimensionnement de ce paramètre de façon isolée n'est pas le cœur de notre mémoire, nous avons tout de même tenu à retenir un paramétrage que nous estimons pertinent et suffisamment réaliste dans la mesure où cela amène à un R_0 entre 0 et 3, et nous avons vu que ces valeurs étaient cohérentes avec les gripes que nous souhaitons simuler.

Voici la synthèse des paramètres du scénario 1 :

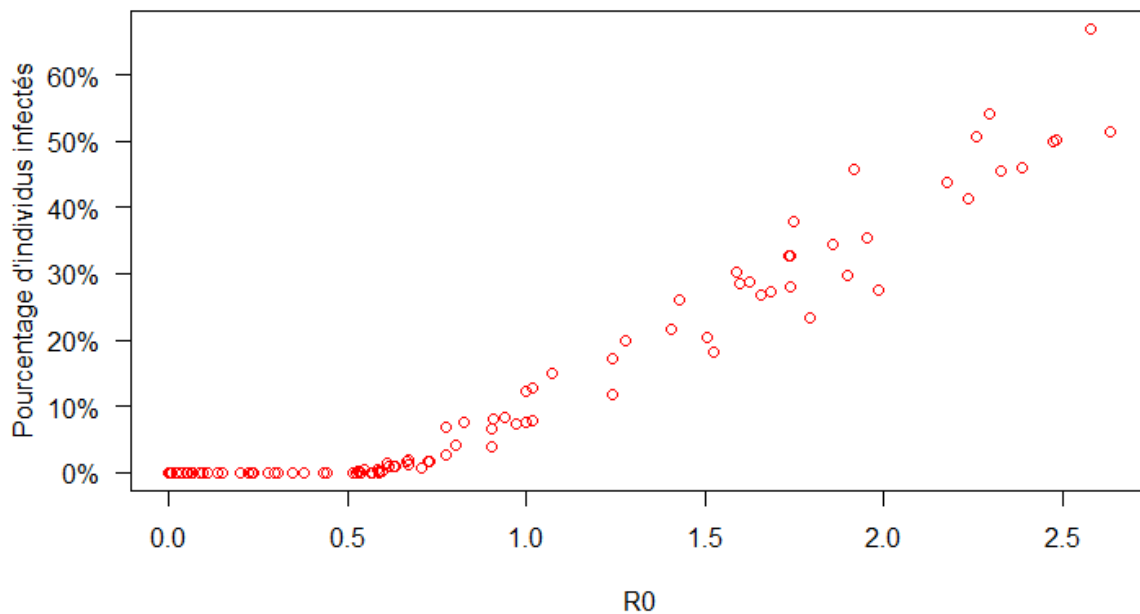
Scénario 1	Valeurs
Nombre d'individus N	10 000
Disposition de la population globale	Aucune disposition particulière
Répartition de la population assurée	Aucune répartition particulière
Paramètre géographique	Non pris en compte
Nombre d'épidémies	100
Nombre de runs par épidémie	50
Paramètres des épidémies (chaque épidémie est modélisée selon le SEIR stochastique avec un quintuplet spécifique, déterminé comme suit) :	
β	Aléatoire selon loi exponentielle $Exp(10)$
$k_e = k_I = k$	Constant égal à 2
$\theta_e = \theta_I = \theta$	Aléatoire selon loi exponentielle $Exp(k/5,5) = Exp(2/5,5)$
Taux de létalité	1 000 valeurs aléatoires selon loi log-normale de paramètres $\mu = \log(0,5\%)$ et $\sigma = 0,8$

Les résultats de ce scénario standard sont les suivants :

5.4. Mesure de l'influence du R_0 de l'épidémie sur le pourcentage d'individus infectés

Nous pouvons tout d'abord mesurer la corrélation entre le R_0 des épidémies que nous modélisons et le pourcentage d'individus infectés au sein de la population initiale :

Impact du R_0 de l'épidémie sur le pourcentage d'individus infectés



Ce graphique confirme l'interprétation *a priori* du paramètre du paramètre R_0 : plus ce dernier est important, plus le nombre d'individus infectés par l'épidémie est important.

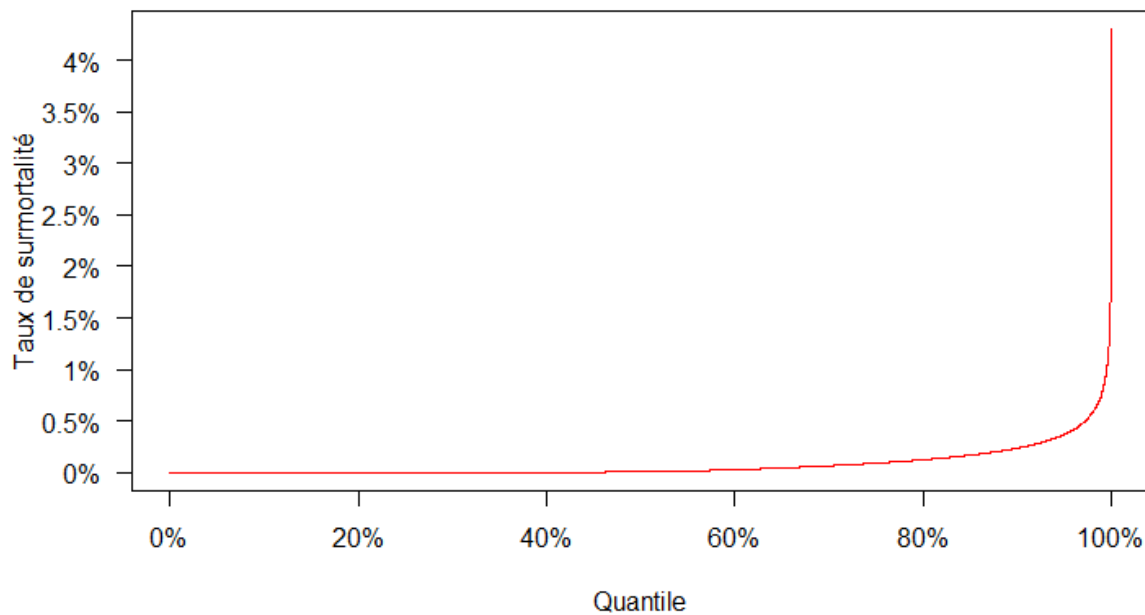
5.5. Mesure du taux de surmortalité

Pour mémoire, pour une épidémie, on définit :

- le taux d'attaque, qui rapporte le nombre de personnes infectées à la population totale,
- le taux de létalité, qui rapporte le nombre de décès au nombre de personnes infectées,
- le taux de surmortalité, qui rapporte le nombre total de décès dus à la pandémie.

Ci-dessous, nous traçons la fonction de répartition de ce taux de surmortalité :

Quantile de répartition de surmortalité



Le risque bicentenaire (à 99,5%) est un taux de surmortalité de 1,01% : la probabilité qu'une épidémie entraîne le décès de plus d'1,01% de la population est de 0,5%.

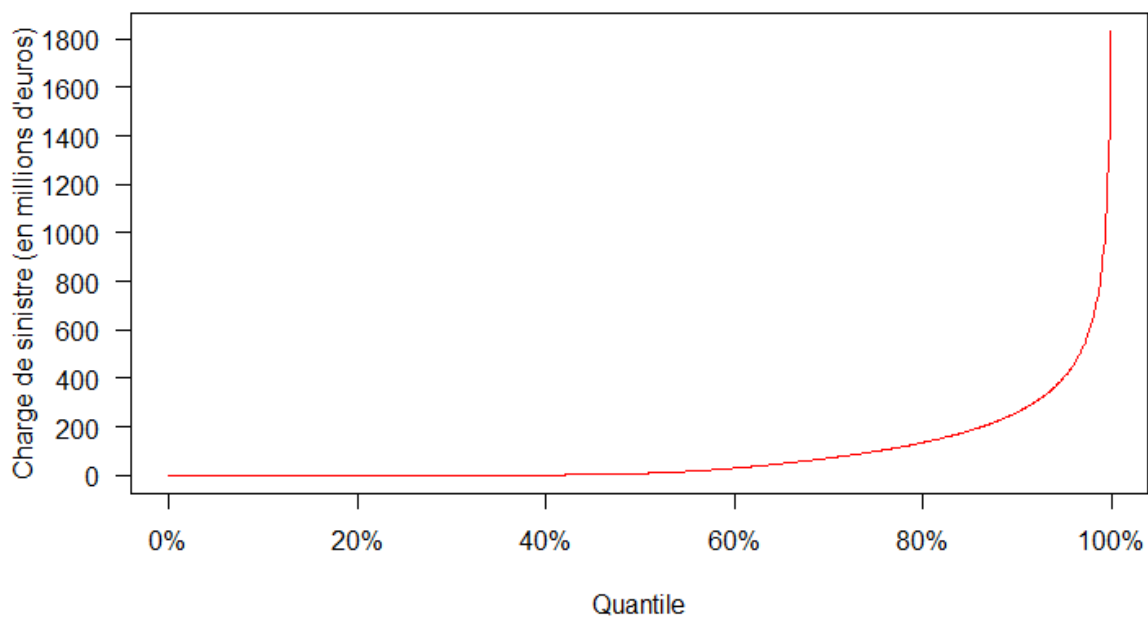
5.6. Mesure de la sinistralité décès de l'assureur

Nous calculons maintenant la charge de sinistre décès induite par ces épidémies sur le portefeuille d'un assureur de 1 millions d'assurés.

Pour ce faire, nous retenons comme coût d'un sinistre individuel un montant de 110 000 euros (moyenne calculé sur le portefeuille Malakoff Médéric).

N'ayant pu simuler nos résultats que sur une population de 10 000 individus (pour des raisons de capacités de calcul), nous prenons l'hypothèse que nos résultats observés (mesure du taux de surmortalité) « se prolongent » sur une population plus grande.

Quantile de répartition de la charge de sinistre décès



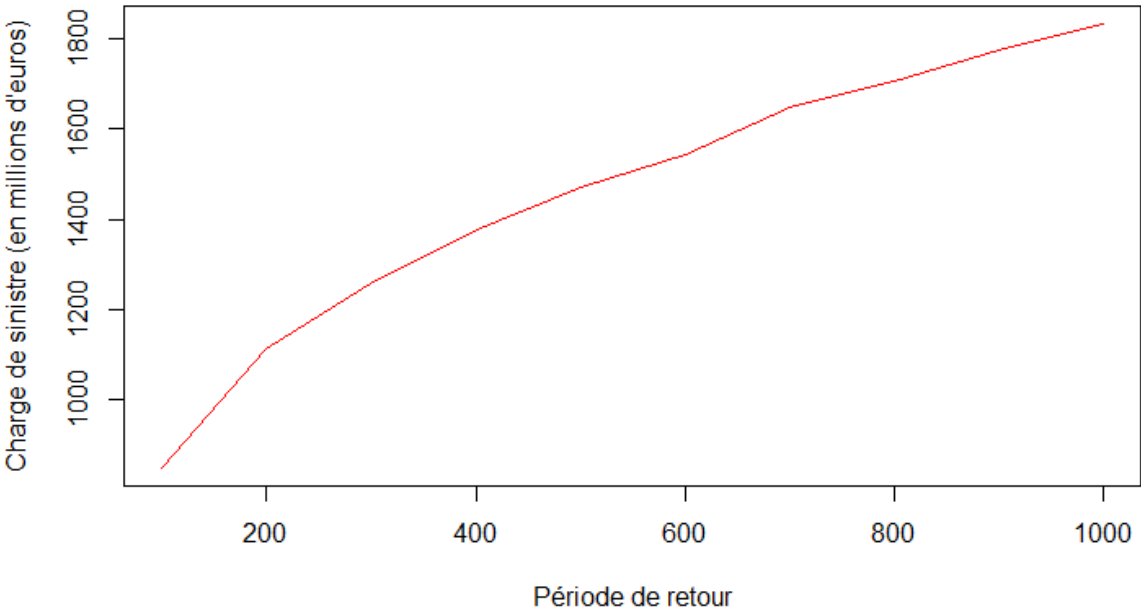
Le risque bicentenaire (à 99,5%) est une charge de sinistre de 1,1Md€ : c'est-à-dire la probabilité qu'une épidémie entraîne pour notre assureur une charge de sinistre décès de plus d'1,1Md€ est de 0,5%.

Les acteurs du marché de l'assurance et de la réassurance utilisent aussi la notion de « période de retour » pour représenter les fonctions de répartition.

Une période de retour de 200 ans, correspond à un risque se produisant une fois tous les 200 ans, soit un risque à 99,5%.

Nous pouvons ainsi représenter nos résultats comme suit :

Quantile de répartition de la charge de sinistre décès



6. Modélisation de la propagation spatiale de l'épidémie

Au-delà d'appliquer uniquement le modèle SEIR à une population complètement théorique, nous avons donc souhaité prendre en compte le paramètre géographique afin de tenir compte du facteur d'éloignement entre individus et de densité de la population globale (et celle de la population d'assurés qui peut en être différente), afin de modéliser une pandémie plus réaliste.

En effet, dans le scénario où une épidémie surgit en France (nous prenons cet exemple car les assurés de notre portefeuille se situent quasi-intégralement en France), on comprendra aisément que son impact ne sera pas le même selon que les individus sont fortement concentrés sur une région ou répartis de façon relativement uniforme sur le territoire. C'est le cas pour l'impact global d'une épidémie mais c'est également important pour le point de vue d'un assureur en particulier : le coût ne sera pas le même selon que sa population assurée est uniformément répartie ou si elle est fortement concentrée dans une région particulière.

Nous avons tenu à intégrer cette dimension dans notre modèle car nous l'estimons importante et pouvant influencer de façon assez significative sur l'impact d'une pandémie (au niveau global et pour un assureur). N'oublions pas en effet que les flux migratoires et la densité des individus sont des facteurs importants dans la propagation d'une pandémie

En réalité et pour être tout à fait complet, le paramètre géographique doit faire intervenir deux sous-paramètres que sont l'adresse de résidence et l'adresse professionnelle.

Ainsi, les salariés d'une entreprise que nous couvrons par un contrat collectif en Santé/Prévoyance seront plus ou moins exposés au virus et se le transmettront plus ou moins rapidement selon :

- qu'ils résident dans une ville de taille importante ou non (facteur d'exposition au nombre de personnes)
- qu'ils ont un trajet plus ou moins long entre leur domicile et leur lieu de travail
- que ces salariés habitent dans un rayon géographique limité ou étendu

Dans notre modèle, nous n'intégrons pas autant de finesse et de complexité, nous nous contentons de modéliser chaque individu par un point géographique. C'est ensuite la distance est les individus qui sert de pondération dans la construction du réseau.

Cet ajout de dimension géographique permet de compléter les différents modèles existants avec un paramètre supplémentaire qui est largement déterminant dans l'impact d'une épidémie.

Par la suite, nous effectuons donc plusieurs scénarii où ce paramètre de distance va influencer la construction de notre réseau d'individus.

6.1. Disposition des individus sur le territoire

Avant même la construction du réseau, le premier enjeu lorsque nous intégrons le paramètre géographique réside dans la disposition des individus sur le territoire. Le périmètre de notre modèle est la France métropolitaine (par abus de langage nous utiliserons le terme « France » pour désigner ce territoire) et nous allons attribuer des coordonnées GPS à chacun de nos individus.

6.1.1. Répartition par région ou département

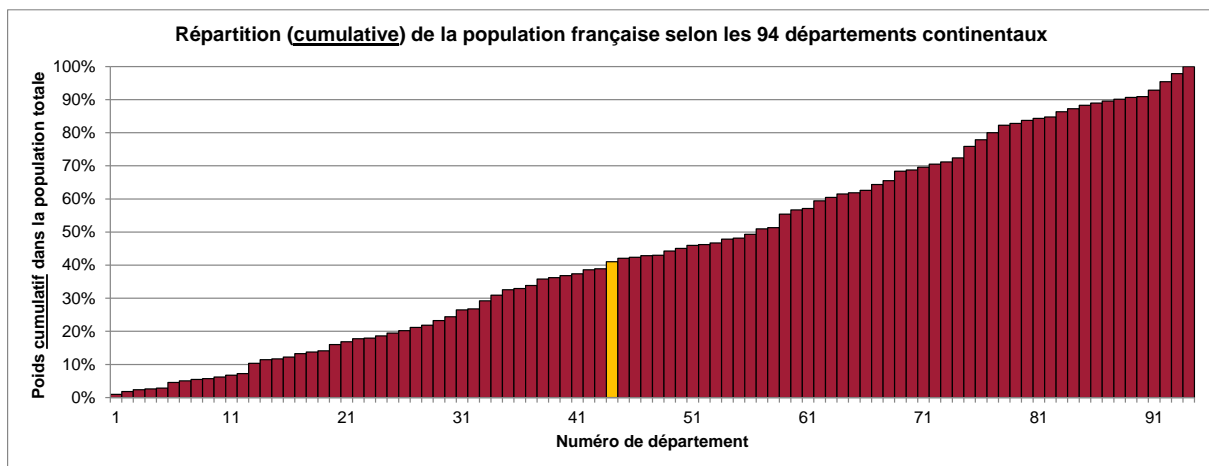
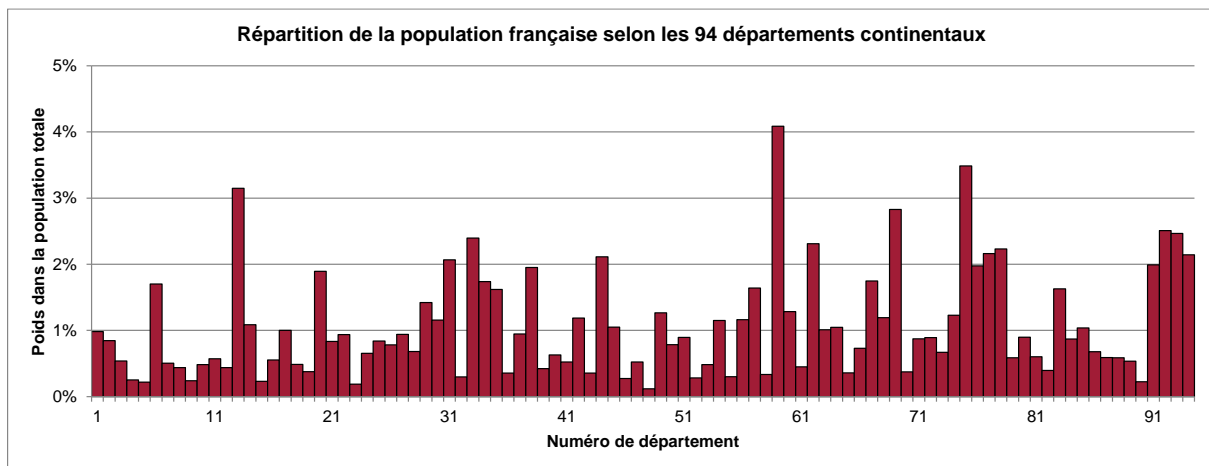
6.1.1.1. Modélisation de la France

Notre volonté de départ a été de répartir les individus par région ou par département en modélisant les 12 régions ou les 94 départements métropolitains (hors Corse) par des blocs rectangulaires. Nous

avons rapidement exclu l'option des régions dans la mesure où leur nombre est trop faible pour permettre une modélisation suffisamment fine. En effet, nous aurions dû représenter la France par 12 blocs rectangulaires, ce qui nous paraît trop approximatif et nous a amené à étudier plutôt l'option des départements.

Pour cela, nous conservons le découpage de la France en 94 départements³¹ et récupérons entre autres leur superficie, leur population³² ainsi que les coordonnées GPS de leurs chefs-lieux respectifs^{33 34}.

La première étape consiste à répartir nos N individus dans les 94 départements. Afin d'être le plus proches de la réalité, nous choisissons de respecter la répartition réelle des habitants en France. Nous créons ainsi une variable aléatoire à valeurs discrètes (entre 1 et 94) dont la densité et la fonction de répartition sont construites à partir des populations réelles de chaque département et leur rapport à la population totale.



Ensuite, nous réalisons pour chaque individu un tirage selon la loi uniforme sur $[0 ; 1]$ et lui attribuons le département correspondant grâce à la fonction de répartition précédemment construite. Par

³¹ Parmi les 95 départements métropolitains, nous excluons la Corse (n°20) et la remplaçons par le Val d'Oise (n°95) afin de ne pas décaler tous nos indices

³² Il s'agit de données datant de 2014, ce qui est suffisamment récent pour pouvoir être utilisé

³³ <https://www.data.gouv.fr/fr/datasets/listes-des-communes-geolocalisees-par-regions-departements-circonscriptions-nd/#>

³⁴ Wikipédia, articles sur les régions, départements, chefs-lieux et préfectures françaises. Se reporter à la bibliographie pour plus de détails

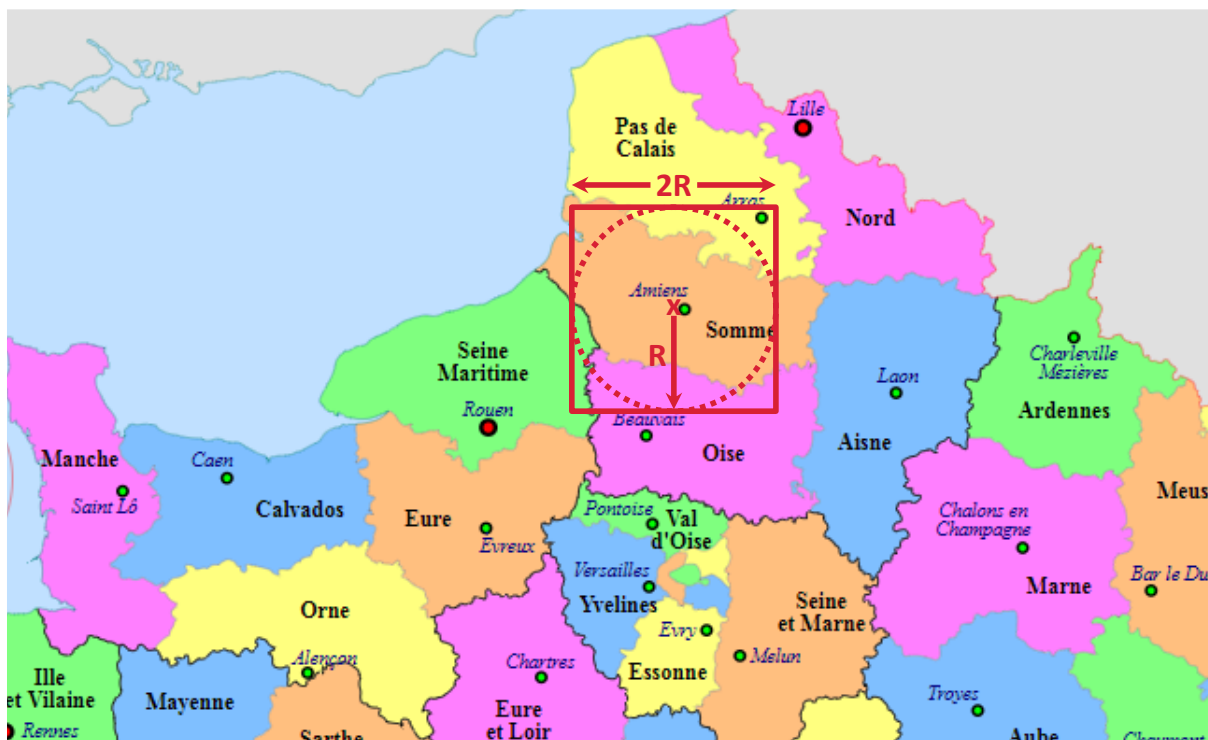
exemple, si le tirage selon la loi uniforme donne la valeur 0,4 (=40%), c'est la région 44 (en orange dans le graphique ci-dessus) qui lui sera attribuée.

Nous aurions pu opter pour la solution plus simple (déterministe) et figer la répartition de nos individus en appliquant les 94 poids respectifs à notre population totale N. Néanmoins, nous avons choisi cette solution stochastique afin d'éprouver la rigidité de notre modèle à cette répartition (nos résultats sont stables à chaque run que nous effectuons) et car cela ne consomme pas un temps machine significatif.

Maintenant que les individus sont répartis par département, la deuxième étape consiste à leur affecter des coordonnées GPS au sein de ce département. Pour chaque département, nous commençons par calculer le rayon R du cercle qui a une aire égale à la superficie du département.

$$\pi \times R^2 = \text{Superficie}_{\text{département}}$$

Nous modélisons ensuite le département par un carré de côté 2R (qui englobe donc le cercle précédent) et centré sur le chef-lieu. En effet, le codage d'une disposition aléatoire dans un carré est moins compliqué que pour un cercle, dans la mesure où il faut manier à la fois des coordonnées GPS (latitude / longitude) et des distances euclidiennes.



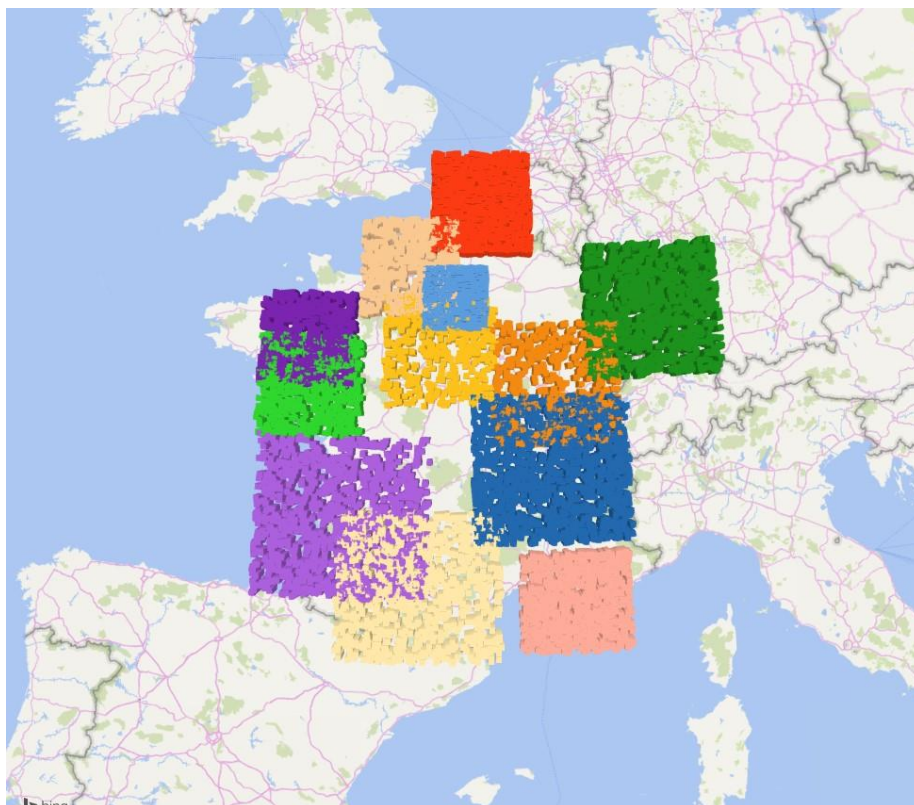
Par ailleurs, ce choix de disposition dans un carré au lieu d'un cercle n'altère pas notre modèle et permet au contraire de prendre une certaine marge de prudence car cela revient à considérer une zone un peu plus grande que la superficie du département :

$$\text{Aire du carré} = 4 \times R^2 > \pi \times R^2 = \text{Superficie}_{\text{département}}$$

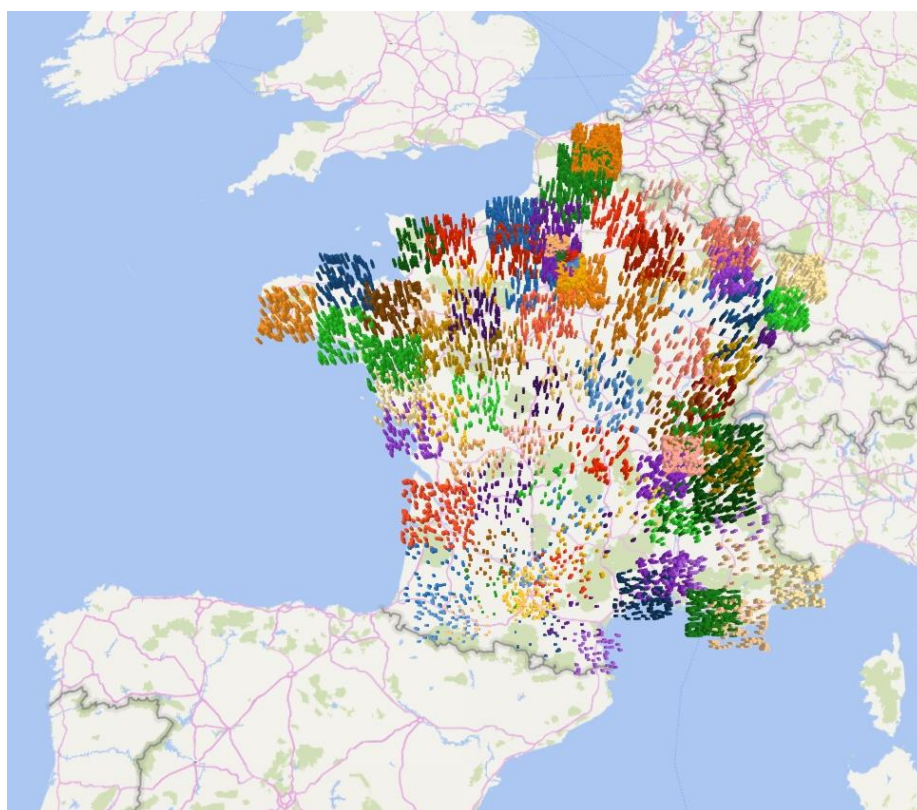
Cette méthode permet ainsi de corriger dans une certaine mesure les cas où les chefs-lieux ne sont pas tout à fait situés au centre géographique de leur département.

La méthode ainsi définie peut être simulée à une « maille région » (12 carrés autour des 12 chefs-lieux de région) ou à une maille plus fine : la maille département (94 chefs-lieux en métropole).

En synthèse, voici la représentation géographique de nos individus selon que nous retenons une maille de répartition régions ou une maille départements :



Répartition des assurés avec une maille région



Répartition des assurés avec une maille département

Nous constatons que la maille région est trop imprécise pour obtenir une répartition satisfaisante de nos assurés sur le territoire française et choisissons donc de retenir la maille département.

6.1.1.2. Modélisation des pays frontaliers

Le deuxième enjeu lorsque nous intégrons le paramètre géographique réside dans la prise en compte des pays frontaliers. En effet, si l'on considérait la France de façon isolée sans pays adjacents, tous les départements périphériques (Nord, Bas Rhin, Pyrénées Atlantiques, Alpes Maritimes, Finistère, etc.) seraient naturellement les moins touchés en moyenne par les épidémies puisque par définition, ils ont moins de contacts avec les autres départements, donc moins d'exposition aux virus.

Or ceci n'est pas réaliste puisque certains de ces départements périphériques sont en première ligne et donc davantage exposés aux épidémies provenant des pays voisins. C'est par exemple le cas de la région Hauts-de-France, qui sera davantage sinistré que les autres départements en cas d'épidémie provenant de la Belgique.

En revanche, on peut conserver l'approximation de ne pas modéliser de zone frontalière aux départements situés en bordure de mer ou d'océan, ce qui écart les scénarii de pandémies ayant traversé les mers ou océans (pour les virus qui se transmettent par voie aérienne). Cette approximation peut être retirée et le modèle peut être affiné pour tenir compte de ces épidémies de provenance plus lointaine. Pour cela, il suffit par exemple de rajouter et dimensionner (tout l'enjeu est dans le dimensionnement) un ou plusieurs paramètres η_m dans la construction du réseau.

Nous choisissons donc de modéliser trois régions frontalières à la France par trois blocs de populations :

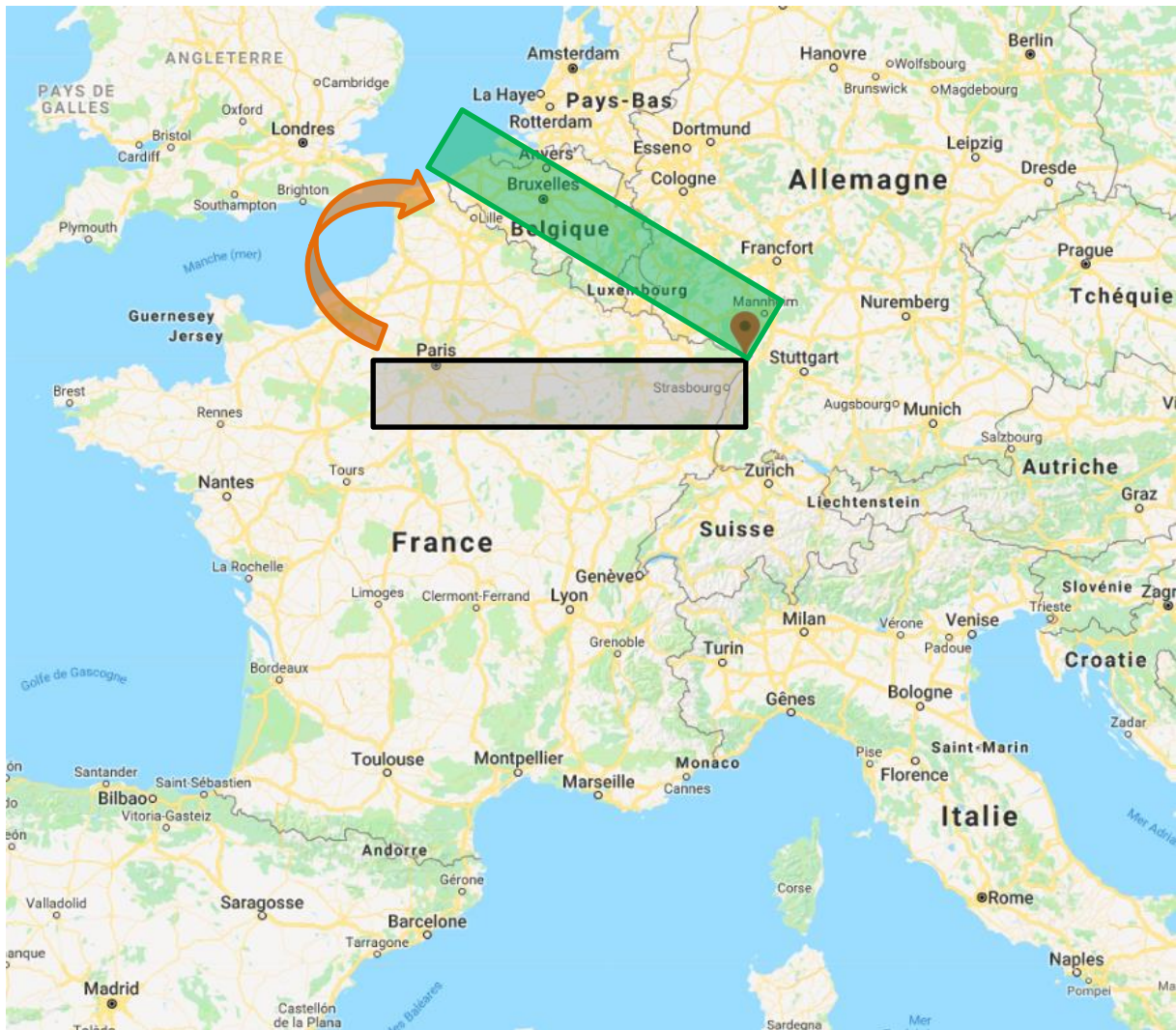
- Un bloc à la frontière sud-ouest, représentant l'Espagne et le Portugal
- Un bloc à la frontière nord-est, représentant la Belgique, le Luxembourg, les Pays-Bas et le nord de l'Allemagne
- Un bloc à la frontière Est, représentant la Suisse, l'Autriche, le nord de l'Italie et le sud de l'Allemagne

Le bloc nord-est est le plus long à coder et à modéliser car il s'agit de représenter un rectangle qui est incliné dans le plan. Nous avons donc d'abord représenté un rectangle, auquel nous avons ensuite appliqué une rotation dans le plan. En termes d'algorithme, il est en effet plus simple de générer aléatoirement des coordonnées dans un rectangle « droit » (c'est-à-dire défini par des intervalles de latitude et longitudes) puis d'appliquer une rotation à cet ensemble de coordonnées, que de générer des coordonnées dans un rectangle incliné dès le début.

Extrait du code R de cette opération :

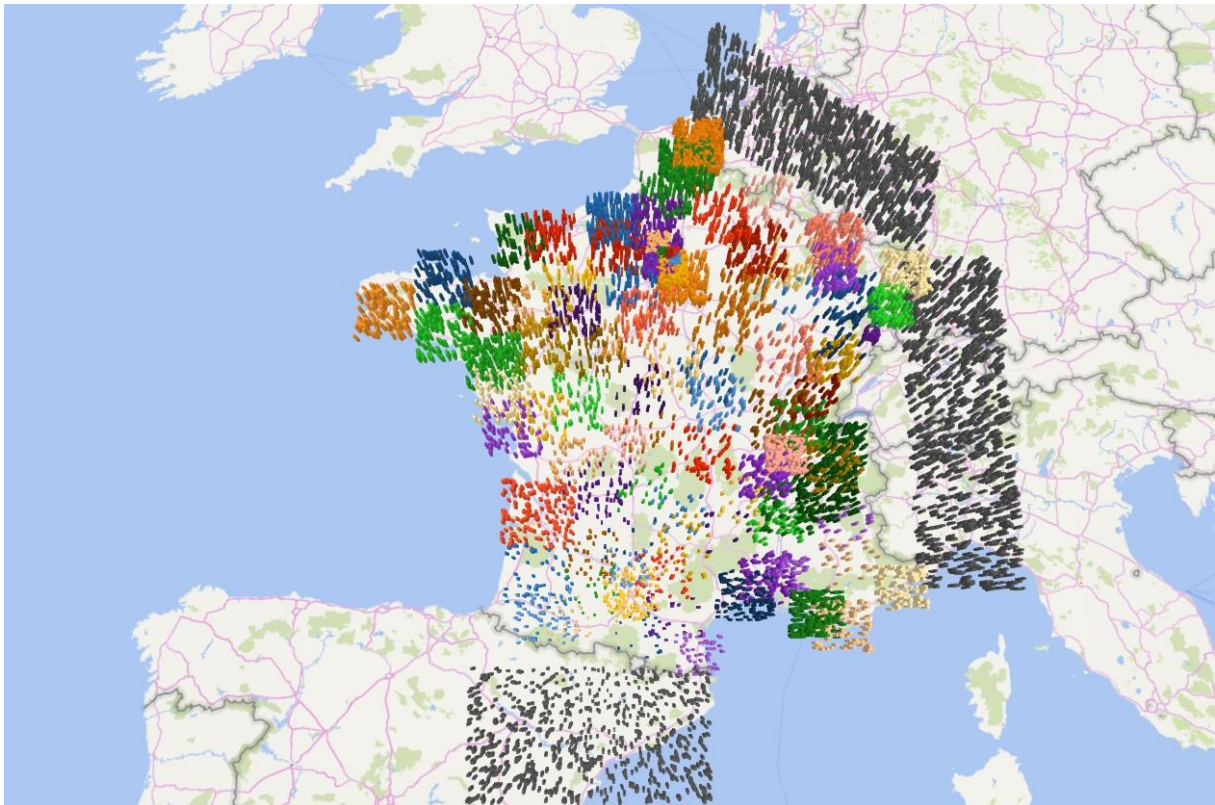
```
#on applique une rotation sur les latitudes/longitudes des assurés de la frontière nord-est
for(i in seq(1,N)) { #boucle sur les N assurés
  if (IDregion[i,1] != 95) #la région nord-est est identifiée "95"
    next
  #centrage du repère en la point la plus à l'est du pays
  lat0 = IDregion[i,2] - 48.9101
  lon0 = IDregion[i,3] - 8.1173
  #application de la rotation d'angle theta
  thet = -0.392699082
  lat1 = lat0*cos(thet)+lon0*sin(thet)
  lon1 = -lat0*sin(thet)+lon0*cos(thet)
  #stockage des coordonnées après rotation et "décentrage" du repère
  IDregion[i,2] = lat1 + 48.9101
  IDregion[i,3] = lon1 + 8.1173
}
```

Illustration graphique de cette opération :



Nous nous sommes ensuite interrogés sur le dimensionnement de la population dans ces zones frontalières et avons estimé qu'un choix raisonnable et réaliste était de placer 10% de notre population totale française dans chacun des trois blocs.

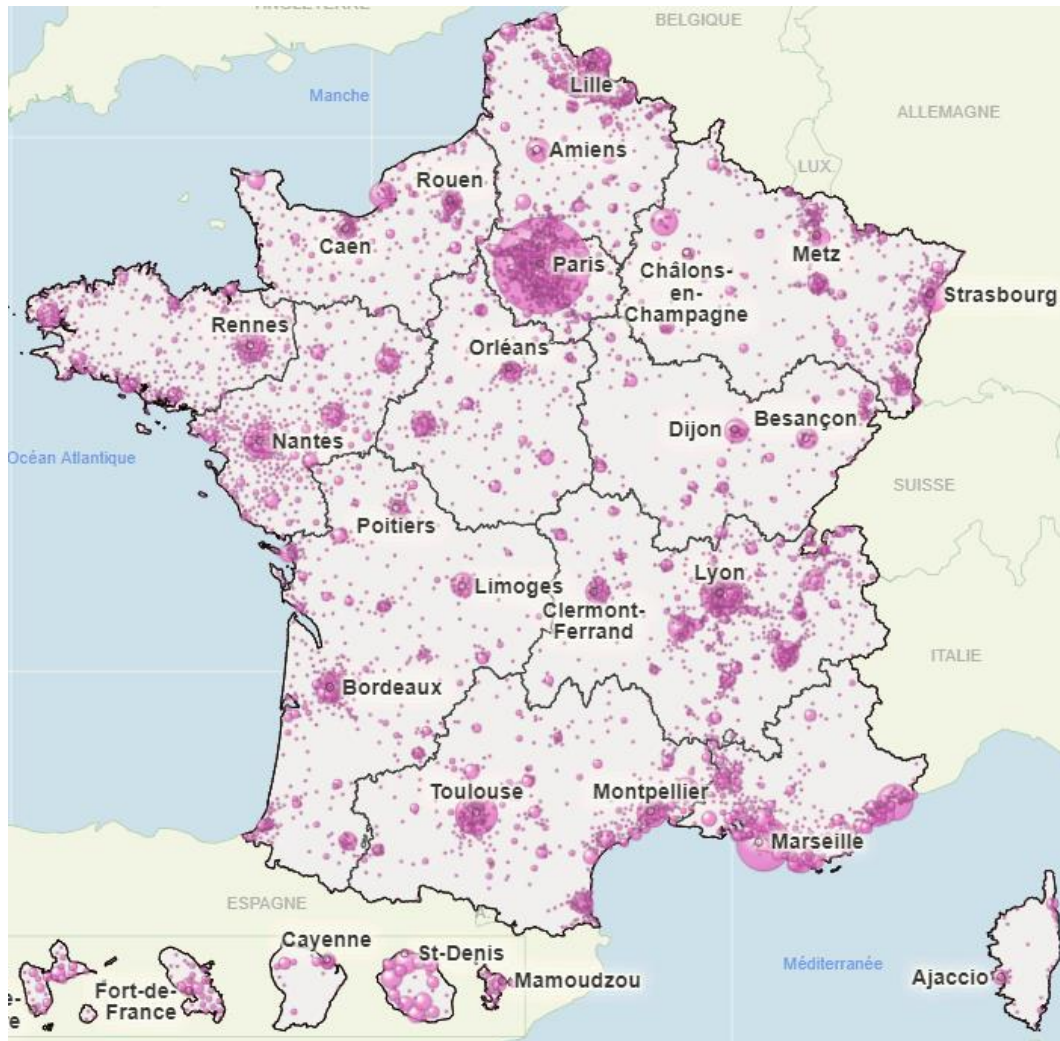
In fine, voici la représentation graphique de la France et de nos trois régions frontalières (en gris) :



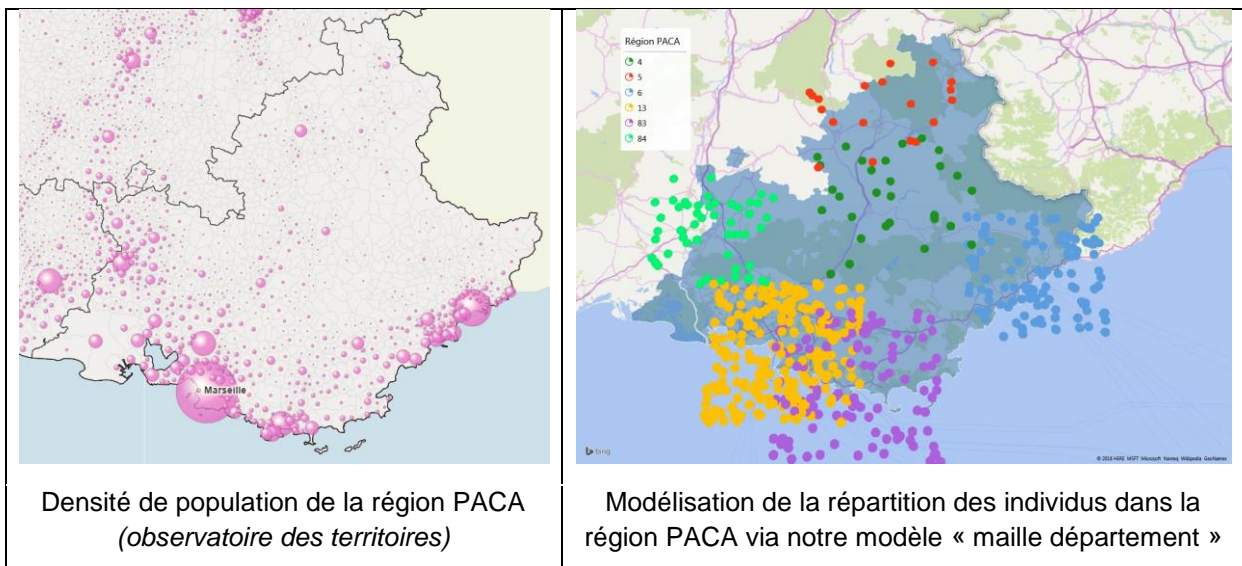
6.1.2. Répartition par commune

Ces deux représentations précédentes nous montrent que la maille de modélisation par départements est globalement satisfaisante et s'approche suffisamment de la réalité démographique française puisqu'on y retrouve l'essentiel des concentrations de population et également des régions moins denses.

Toutefois, cela ne nous satisfait pas complètement dans la mesure où certaines zones paraissent trop approximatives. Si l'on compare notre carte à la carte de la répartition de la population française établie par l'Observatoire des territoires :



Nous constatons que la modélisation de la région PACA n'est pas satisfaisante :



En plus de positionner des individus en dehors des frontières de la région, la densité de population de la région n'est pas restituée de manière fidèle.

Cela s'explique simplement par notre méthode qui consiste à modéliser six départements de la région PACA (04 - Alpes-de-Haute-Provence, 05 - Hautes-Alpes, 06 - Alpes-Maritimes, 13 - Bouches-du-Rhône, 83 - Var, 84 - Vaucluse) par des blocs carrés et à disposer la population de ces départements de façon uniforme dans ces blocs.

En effet, pour un département donné, les populations sont très généralement concentrées dans son chef-lieu et parfois quelques autres villes de taille importante aux alentours. Si nous approchons cette réalité en centrant notre bloc sur le chef-lieu, c'est ensuite la répartition uniforme autour de ce centre qui nous fait dévier de la réalité. Une façon de parer à ce problème serait par exemple de ne pas réaliser une distribution uniforme mais une distribution normale, centrée sur le chef-lieu évidemment et avec une variance à calibrer. Cette méthode conserverait toutefois le défaut de ne pas modéliser les autres villes du département à forte population.

C'est pourquoi nous avons opté pour une maille de répartition à un niveau plus fin, en les disposant par commune. Pour cela, nous avons utilisé une base de données publique³⁵ recensant les plus grandes villes du monde entier avec leurs coordonnées GPS et populations respectives. Pour la France par exemple, il n'y a pas l'intégralité des communes de France (environ 35 000) mais les 860 plus importantes.

Puis, comme précédemment, nous avons procédé en deux temps en procédant d'une part à la répartition des individus sur la France d'une part et dans les trois régions frontalières (telles que définies ci-dessus) d'autre part.

6.1.2.1. Modélisation de la densité de population en France

Concernant la France, nous retenons l'ensemble des 860 villes recensées par la base de données (la plus peuplée étant bien sûr Paris avec plus de 2 millions d'habitants) qui représentent une population totale de 29 millions habitants, soit les 45 premiers pourcents des villes les plus peuplées de France métropolitaine (65 millions d'habitants). Puis, nous faisons l'hypothèse que cet échantillon (ensemble des villes de France de plus de 7 000 habitants) est représentatif de la densité de population de l'ensemble du pays.

Autrement dit, cela revient à ne modéliser que les 29 millions d'habitants des villes les plus importantes et de faire abstraction des 36 autres millions. Si cette hypothèse peut paraître trop forte, nous verrons par la suite que cela est amplement suffisant pour nos besoins et permet déjà de se rapprocher davantage de la réalité par rapport à la modélisation par département. En effet, les 36 millions d'habitants non modélisés sont tellement diffus sur le territoire français que cela représente finalement une « couche » uniforme de population et que cela n'affecte donc pas les disparités de densités, or c'est justement ce que nous souhaitant capter dans notre modèle.

6.1.2.2. Modélisation de la densité de population des pays frontaliers

Nous construisons ensuite la densité de population de nos trois blocs frontaliers (Sud, Nord-Est, et Est).

Nous illustrons notre méthodologie sur le bloc Sud (Espagne et Portugal) :

L'objectif est de construire le sous-ensemble des villes espagnoles et portugaises les plus au Nord (c'est-à-dire les plus proches de la frontière avec la France). Pour ce faire, nous construisons une fonction qui en partant de la latitude de la frontière avec la France (latitude environ égale à 43), décrémente cette latitude (nous nous déplaçons ainsi vers le sud) et ajoute au fur et à mesure les

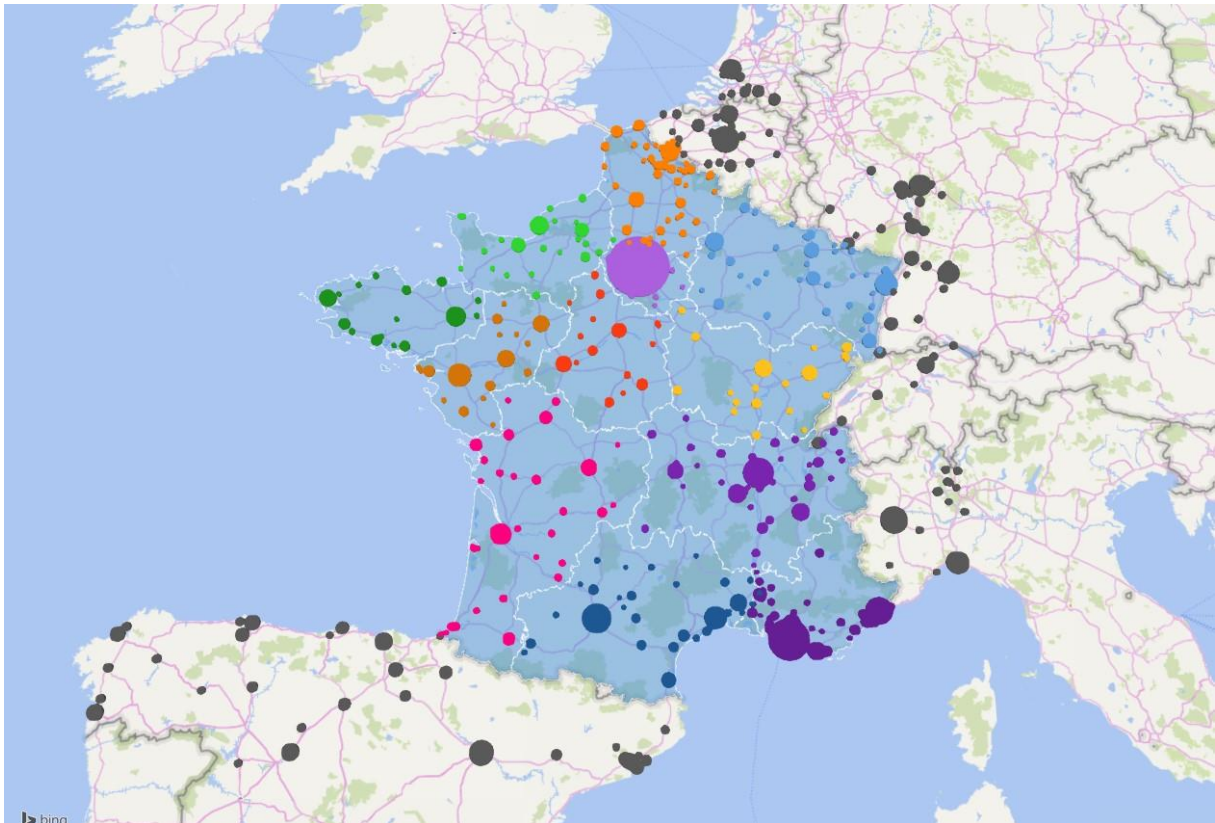
³⁵ <https://www.maxmind.com/fr/free-world-cities-database>, la date des données n'est pas précisée mais cela n'a pas d'impact significatif sur notre modèle.

viles espagnoles et portugaises ainsi parcourues en comptant la population cumulée de ces villes. La fonction s'arrête lorsque la population de cet échantillon atteint 6 millions de personnes (10% de la population Française).

Nous procédons de manière identique pour les blocs frontaliers Nord-Est et Est, en nous déplaçons cette fois vers l'est (via des variations de longitudes).

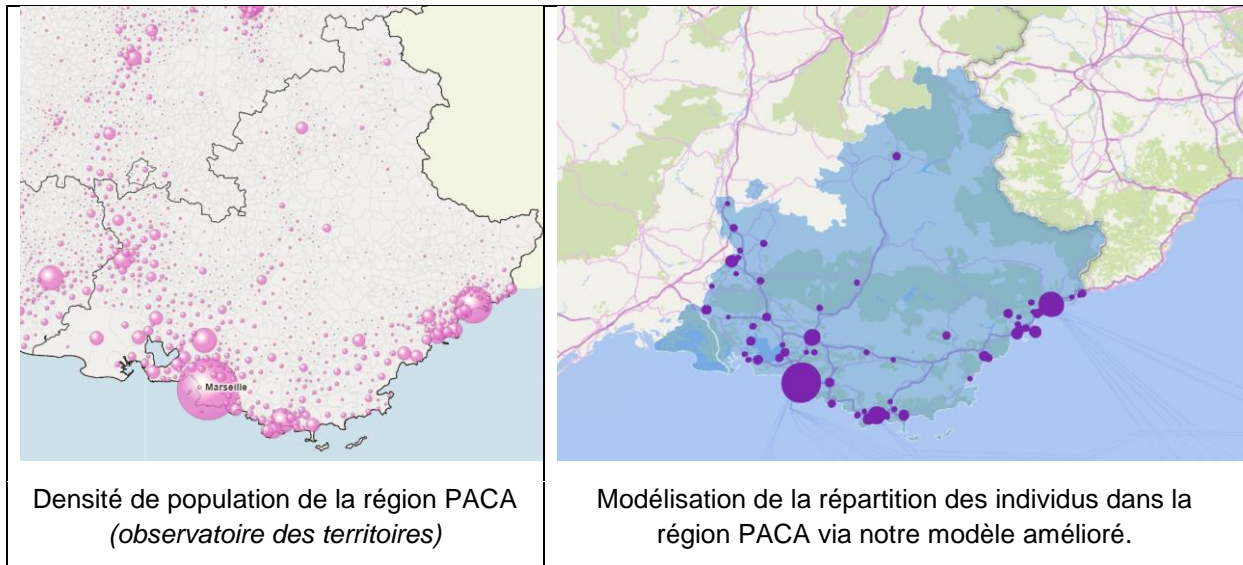
6.1.2.3. Résultat : densité de population ainsi obtenue

Le résultat de l'application de la méthode de répartition par commune est nettement plus satisfaisant que celui de la méthode par département, nous obtenons ainsi la carte suivante :



Cette carte est beaucoup plus fidèle la densité de population réelle des pays considérés.

Si nous reprenons l'exemple de la région PACA, nous pouvons constater cette amélioration :



6.2. Prise en compte de la distance

6.2.1. Rappel du fonctionnement de base

Nous avons vu précédemment que la formule générale de construction d'un graphe d'Erdős-Rényi est décrite par la formule :

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \sum_m \eta_m X_{\{i,j\},m}$$

ou de façon équivalente :

$$p_{ij} = \frac{\exp(\sum_m \eta_m X_{\{i,j\},m})}{1 + \exp(\sum_m \eta_m X_{\{i,j\},m})}$$

où :

- les η_m sont des paramètres réels
- les $X_{\{i,j\},m}$ sont les termes (i,j) de m matrices dyadiques

Dans le package *epinet*, ces informations sont résumées dans une matrice dyadique à $[N \times (N-1)/2]$ lignes (le nombre de paires parmi les N individus) ayant ce format (exemple avec $m=3$) :

<i>Individu i</i>	<i>Individu j</i>	$X_{\{i,j\},1}$	$X_{\{i,j\},2}$	$X_{\{i,j\},3}$
1	2	$X_{\{1,2\},1}$	$X_{\{1,2\},2}$	$X_{\{1,2\},3}$
1	3	$X_{\{1,3\},1}$	$X_{\{1,3\},2}$	$X_{\{1,3\},3}$
1	4	$X_{\{1,4\},1}$	$X_{\{1,4\},2}$	$X_{\{1,4\},3}$
...
N-2	N-1	$X_{\{N-2,N-1\},1}$	$X_{\{N-2,N-1\},2}$	$X_{\{N-2,N-1\},3}$
N-2	N	$X_{\{N-2,N\},1}$	$X_{\{N-2,N\},2}$	$X_{\{N-2,N\},3}$
N-1	N	$X_{\{N-1,N\},1}$	$X_{\{N-1,N\},2}$	$X_{\{N-1,N\},3}$

Les paramètres η_m sont réunies dans un vecteur à m entrées, qui est une autre donnée d'entrée du package *epinet*.

Dans le cas standard (celui de notre scénario 1), il n'y a qu'un seul paramètre $\eta_1 = \eta$ et tous les $X_{\{i,j\},1}$ sont égaux à 1, de sorte que toutes les probabilités p_{ij} ne dépendent pas de i et j et soient toutes égales à la même valeur p :

$$\forall (i,j) \quad p_{ij} = \frac{\exp(\eta \times 1)}{1 + \exp(\eta \times 1)} = \frac{\exp(\eta)}{1 + \exp(\eta)} = p = 3/N$$

La matrice dyadique est donc dans une forme très simple

<i>Individu i</i>	<i>Individu j</i>	$X_{\{i,j\},1}$
1	2	1
1	3	1
1	4	1
...
N-2	N-1	1
N-2	N	1
N-1	N	1

6.2.2. Comment s'assurer du correct dimensionnement des p_{ij} ?

Notre travail consiste à garder toujours un seul paramètre η mais à rendre le paramètre $X_{\{i,j\},1}$ variable et dépendant de la distance entre i et j , il en sera alors de même pour la probabilité p_{ij} qui intervient dans la construction du réseau (rappelons qu'il s'agit du paramètre de la loi de Bernoulli qui simule le tirage pour établir ou non un lien entre i et j).

- **Etape 1 : mesure des distances entre individus**

Après avoir placé les individus de façon aléatoire dans nos 97 zones (voir précédemment), soit 94 départements français et trois régions frontalières, nous mesurons pour chaque paire (i,j) la distance les séparant (en km), donnée par la formule suivante :

$$\begin{aligned} \text{distance}(i,j) &= d_{ij} \\ &= \arccos[\sin(\text{lat}_i) \times \sin(\text{lat}_j) + \cos(\text{lat}_i) \times \cos(\text{lat}_j) \times \cos(\text{lon}_j - \text{lon}_i)] \times 6371 \text{ km} \end{aligned}$$

avec :

$$\left\{ \begin{array}{l} (\text{lat}_i, \text{lon}_i) \text{ et } (\text{lat}_j, \text{lon}_j) \text{ les latitudes et longitudes (en radians) des individus } i \text{ et } j \\ 6371 \text{ km le rayon moyen de la Terre} \end{array} \right.$$

Nos coordonnées étant en degrés, nous avons une étape intermédiaire pour convertir les degrés en radians :

$$\text{radians} = \text{degrés} \times \frac{\pi}{180}$$

- **Etape 2 : normalisation des distances**

Nous transposons cet échantillon de distances de façon linéaire sur un intervalle plus réduit, ce qui permettra de réduire le temps machine lors de notre résolution numérique par la suite. Nous choisissons ainsi de ramener ces distances sur l'intervalle [0,5 ; 1,5] :

$$A_{ij} = \frac{(d_{ij} - \min_{i,j} d_{ij})}{(\max_{i,j} d_{ij} - \min_{i,j} d_{ij})} + 0,5$$

Nous effectuons une deuxième normalisation de ces X_{ij} afin de vérifier des propriétés de cas standards, en définissant :

$$B_{ij} = A_{ij} \times \frac{N(N-1)/2}{\sum A_{ij}}$$

Ainsi on retrouve l'égalité suivante qui est vérifiée dans le cas standard lorsqu'on somme tous les termes de la matrice dyadique égaux à 1 :

$$\sum_{i,j} B_{ij} = \frac{N(N-1)}{2} = \sum_{i,j} 1$$

- **Etape 3 : propriété à vérifier**

Avant de continuer, revenons un instant sur quelques propriétés des variables aléatoires avec lesquelles nous agissons.

Nous avons vu que, dans le cas standard, la construction du réseau se fait par le biais de $M = \frac{N(N-1)}{2}$ tirages de Bernoulli de probabilité p. Or ces M variables aléatoires sont également indépendantes entre elles. On sait alors que la somme de ces M tirages indépendants de Bernoulli est une variable aléatoire qui suit une loi binomiale de paramètres (M,p).

Rappelons que pour une loi $S \sim \text{Binomiale}(M,p)$:

$$\left\{ \begin{array}{l} M \geq 0 \text{ et } p \in [0,1] \\ E[S] = Mp \\ V[S] = Mp(1-p) \end{array} \right.$$

Dans le scénario que nous sommes en train de construire, le réseau va se construire via M tirages indépendants de Bernoulli, chacun de paramètre respectif p_{ij} . Une condition nécessaire pour que le dimensionnement de ces p_{ij} soit correct est que le réseau construit avec ces M tirages soit semblable au réseau construit dans le cas standard, sans paramètre géographique et avec M tirages de Bernoulli de même paramètre p .

En effet, rappelons que le réseau est fondamental et déterminant dans l'impact de l'épidémie qui va lui être appliquée. L'idée est donc de dire qu'à partir d'un modèle où on ne tient pas compte du paramètre géographique, le fait de rajouter cet élément dans le modèle ne doit pas modifier l'impact global d'une épidémie (à commencer par le réseau sous-jacent) mais doit seulement permettre d'être plus précise et d'affiner la vision par département, région etc.

Nous avons donc d'un côté le cas standard avec la loi binomiale de paramètres (M, p) et de l'autre côté le cas que nous construisons avec la somme, notée S_M , de M lois de Bernoulli de paramètre p_{ij} . Nous cherchons à nous servir de ce parallèle pour imposer une condition sur les p_{ij} et assurer un bon dimensionnement de notre modèle.

Dans un premier temps, nous avons pensé au paradigme de Poisson qui nous rappelle que la somme S_n d'un grand nombre de variables aléatoires de Bernoulli indépendantes de petit paramètre suit approximativement la loi de Poisson de paramètres $E[S_n]$.

Nous nous sommes posé la question d'approcher notre somme S_M par une loi de Poisson, mais cela aurait imposé les conditions suivantes :

$$\left\{ \begin{array}{l} \text{Egalité des espérances : } E[S_M] = \sum_{i,j} p_{ij} = E[\text{binomiale}(M, p)] = Mp \\ \text{Egalité des variances : } V[S_M] = E[S_M] = \sum_{i,j} p_{ij} = V[\text{binomiale}(M, p)] = Mp(1 - p) \end{array} \right.$$

Cela impose que $p = 0$ et nous conduit à ne pas opter pour cette solution, dont les conditions ci-dessus sont trop fortes.

Néanmoins, nous retenons la condition sur l'espérance puisqu'en effet, notre objectif principal est de vérifier qu'en moyenne, les probabilités p_{ij} soient égales à p , afin d'avoir la stabilité des réseaux.

Ainsi, la condition que nous imposons sur les p_{ij} est :

$$\sum_{i,j} p_{ij} = E[\text{binomiale}(M, p)] = Mp = \frac{N(N-1)}{2} \times p$$

En reprenant l'expression des p_{ij} , la condition devient sur les $X_{\{i,j\}}$ ($= X_{\{i,j\},1}$) :

$$\sum_{i,j} p_{ij} = \sum_{i,j} \frac{\exp(\eta \times X_{ij})}{1 + \exp(\eta \times X_{ij})} = Mp = \frac{N(N-1)}{2} \times \frac{\exp(\eta)}{1 + \exp(\eta)}$$

- **Etape 4 : résolution numérique**

Il n'est pas trivial de solution à cette équation, encore moins sous forme analytique. Nous avons donc choisi de passer par une méthode de résolution numérique, toujours dans le logiciel R. nous commençons par définir la fonction centrale de cette équation définie sur l'ensemble des réels :

$$g: x \mapsto \frac{\exp(\eta x)}{1 + \exp(\eta x)}$$

Notre équation devient :

$$(E2) \quad \sum_{i,j} g(X_{ij}) = M \times g(1)$$

Précédemment, nous avons déjà construit les B_{ij} , qui vérifient une égalité similaire :

$$\sum_{i,j} B_{ij} = \frac{N(N-1)}{2} = \sum_{i,j} 1 = M$$

Ces grandeurs B_{ij} semblent être de bons candidats pour être solutions de l'équation (E2). Pour nous assurer du bon dimensionnement de ces B_{ij} et que la transposition sur le segment $[0,5 ; 1,5]$ ne perturbe pas la résolution de notre équation, nous effectuons une dernière normalisation en définissant $X_{ij} = B_{ij}/\Delta$ et cherchons la solution numérique à :

$$\sum_{i,j} g\left(\frac{B_{ij}}{\Delta}\right) = M \times g(1)$$

Pour cela, nous utilisons la fonction `optimize()` dans R, cette fonction permet de trouver une solution numérique à une équation :

```
taille <- N*(N-1)/2
g <- function(z) exp(eta0*z)/(1+exp(eta0*z))

d0 <- optimize(function(delta) abs(taille*g(1)-sum(g(dcm[,4]/delta))), c(-
100,100), tol=.Machine$double.eps^0.5)$mini
```

6.3. Scénario 2 : avec facteur géographique / assurés répartis de la même façon que la population totale

Dans ce scénario 2, nous plaçons donc les individus selon la répartition actuelle de la population française et nous tenons compte des distances entre eux dans la construction de notre réseau. L'objectif principal est ici de vérifier que nos paramètres p_{ij} ont été correctement dimensionnés et que, outre la répartition géographique des sinistres, l'impact global de l'épidémie (en termes de nombre d'infectés, temps d'épidémie, etc.) est similaire au scénario 1. Cela nous permet de nous assurer de la robustesse de notre modèle géographique, avant de créer d'autres scénarii.

Voici la synthèse des paramètres du scénario 2 :

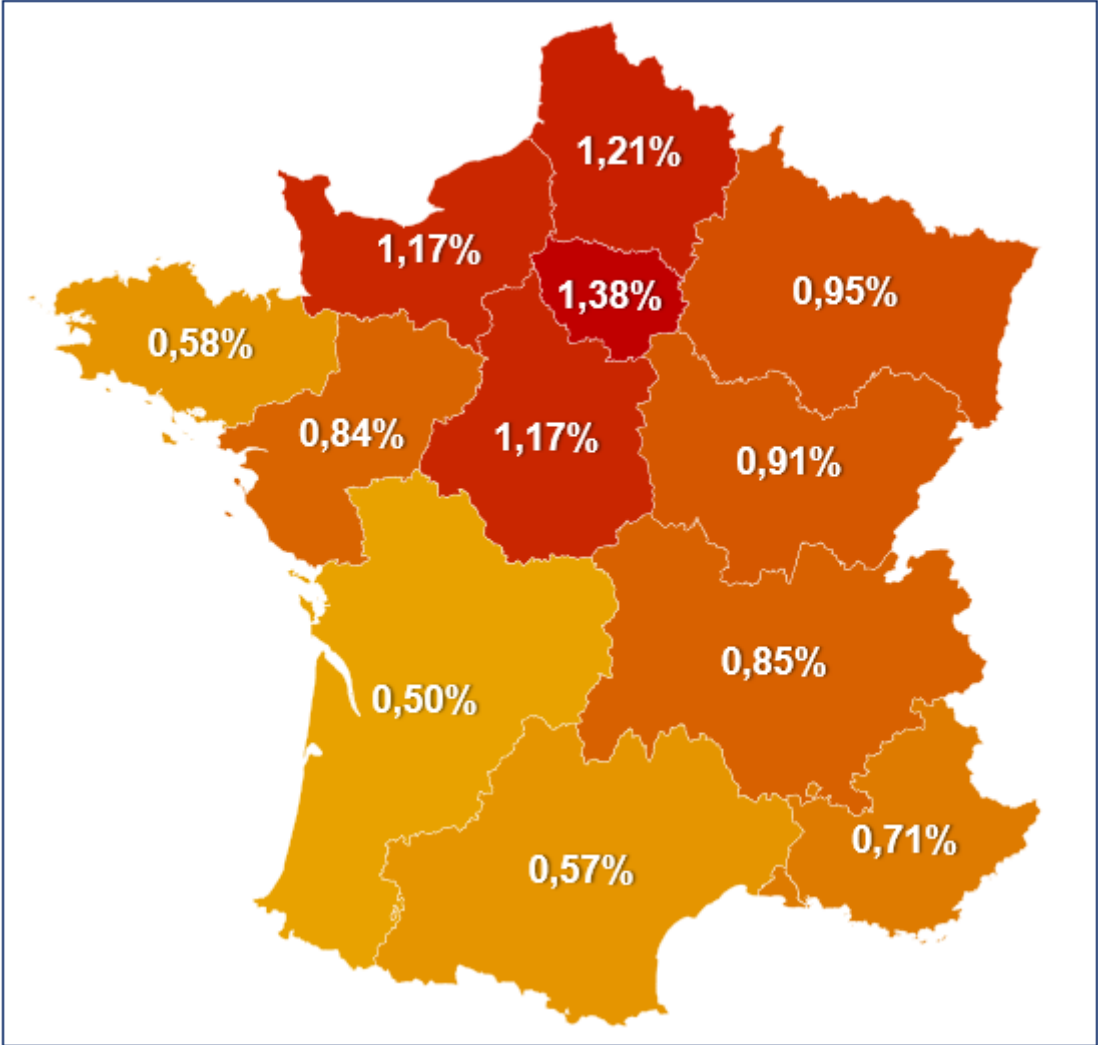
Scénario 2	Valeurs
Nombre d'individus N	10 000 (France + zones frontalières)
Disposition de la population globale	Sur les 94 départements français selon la répartition réelle de la population française Dans les trois zones frontalières : 10% de la population française dans chacun des trois blocs
Répartition de la population assurée	Selon la même répartition que la population française
Paramètre géographique	Pris en compte
Nombre d'épidémies	100
Nombre de runs par épidémie	50
Paramètres des épidémies (chaque épidémie est modélisée selon le SEIR stochastique avec un quintuplet spécifique, déterminé comme suit) :	
β	Aléatoire selon loi exponentielle $Exp(10)$
$k_e = k_I = k$	Constant égal à 2
$\theta_e = \theta_I = \theta$	Aléatoire selon loi exponentielle $Exp(k/5,5) = Exp(2/5,5)$
Taux de létalité	1 000 valeurs aléatoires selon loi log-normale de paramètres $\mu = \log(0,5\%)$ et $\sigma = 0,8$

Les résultats de ce scénario n°2 sont les suivants :

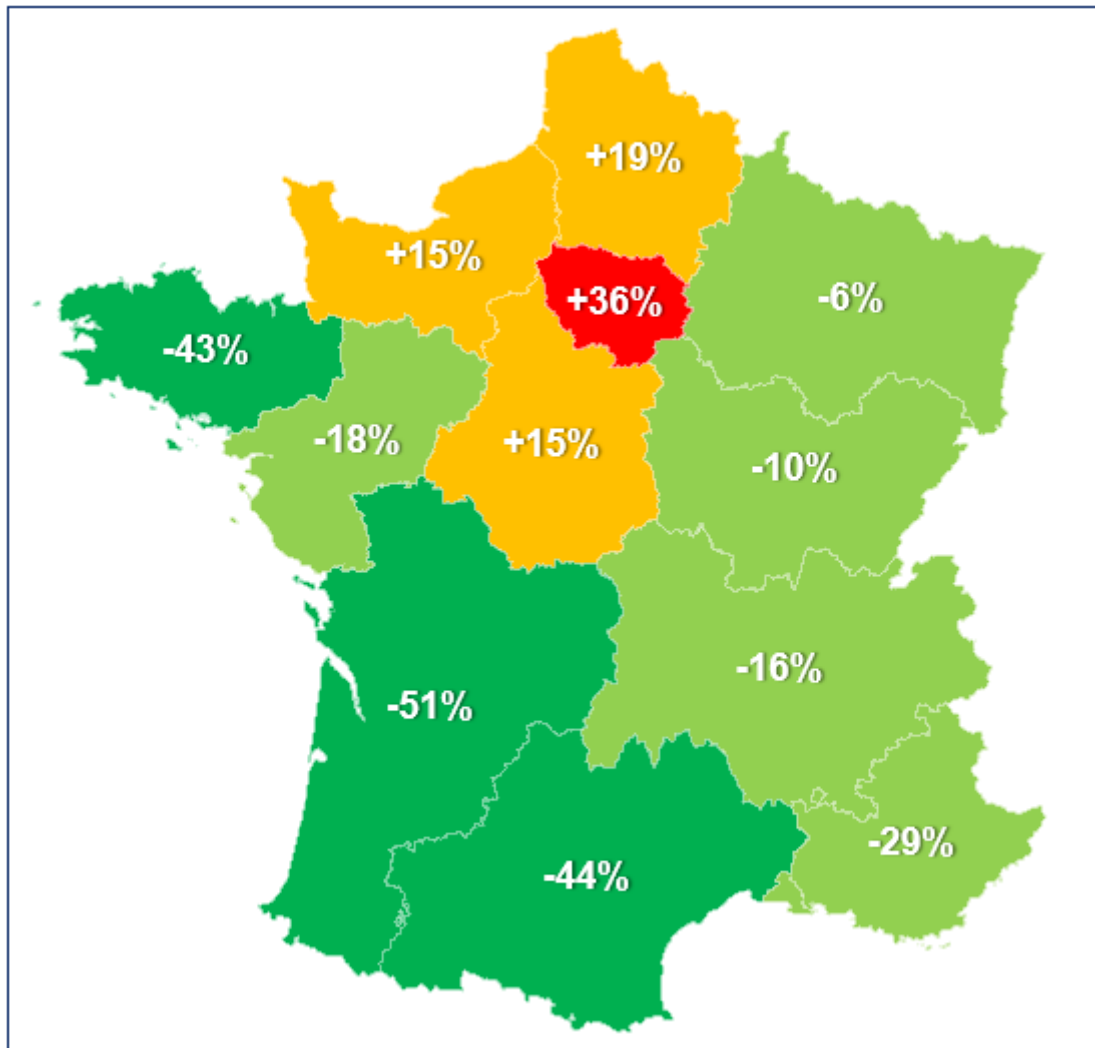
Sur les métriques illustrées en 5.3.1 pour les résultats du scénario n°1 (taux de surmortalité, fonction de répartition de la charge de sinistre décès), nos résultats vont être similaires (par construction du modèle).

L'intérêt de notre modèle réside dans l'ajout de la dimension géographique et donc l'exploration de ces résultats à une maille plus fine : région, département.

Dit autrement, nous pouvons observer comment le taux de surmortalité à 99,5% de 1,01% de la population se construit, région par région :



Nous pouvons comparer ce taux de surmortalité par région au taux national, pour en déduire la déformation régionale du taux de surmortalité nationale :



L'Île-de-France est la région la plus touchée avec un taux de surmortalité de plus de 36% supérieur à la moyenne nationale, la Nouvelle Aquitaine est de son côté la région la moins touchée, avec un taux de surmortalité égal à la moitié de la moyenne nationale.

6.4. Scénario 3 : avec facteur géographique / assurés selon des répartitions particulières

Pour un assureur qui dispose d'un portefeuille dont la répartition est très proche de celle de la France, le modèle avec paramètre géographique n'a que peu d'intérêt pour mesurer le risque global de pandémie. En effet, il lui suffit d'utiliser le modèle standard sans paramètre géographique (celui du scénario 1), l'essentiel étant que celui-ci soit correctement calibré pour refléter la réalité des épidémies en France.

En revanche, un assureur qui dispose d'une répartition plus atypique avec des surreprésentations et sous-représentations selon les départements est beaucoup plus intéressé par l'impact du paramètre géographique que propose notre modèle. Il est important pour cet acteur de tenir compte de la réalité de son portefeuille, du fait qu'il se situe par exemple dans une région à forte densité et où les contacts entre individus sont plus nombreux.

Dans ce scénario, nous simulons donc 5 portefeuilles assurés qui représentent chacun une part de marché égale à 3% et qui sont chacun concentrés sur 5 zones distinctes en France :

- Portefeuille 1 : Ile de France²
- Portefeuille 2 : Nord / Nord-Ouest de la France (régions Hauts de France, Normandie, Bretagne, Centre Val de Loire et Pays de la Loire)
- Portefeuille 3 : Est de la France (régions Bourgogne-Franche-Comté et Grand Est)
- Portefeuille 4 : Sud-Ouest de la France (régions Nouvelle-Aquitaine et Occitanie)
- Portefeuille 5 : Sud-Est de la France (régions Auvergne-Rhône-Alpes et Provence-Alpes-Côte-d'Azur)

Voici la synthèse des paramètres du scénario 3 :

Scénario 3	Valeurs
Nombre d'individus N	10 000 (France + zones frontalières)
Disposition de la population globale	Sur les 94 départements français selon la répartition réelle de la population française Dans les trois zones frontalières : 10% de la population française dans chacun des trois blocs
Répartition de la population assurée	5 portefeuilles d'assurés concentrés sur 5 régions distinctes de la France (Ile de France + 4 cadrans de la France)
Paramètre géographique	Pris en compte
Nombre d'épidémies	100
Nombre de runs par épidémie	50

Paramètres des épidémies (chaque épidémie est modélisée selon le SEIR stochastique avec un quintuplet spécifique, déterminé comme suit) :

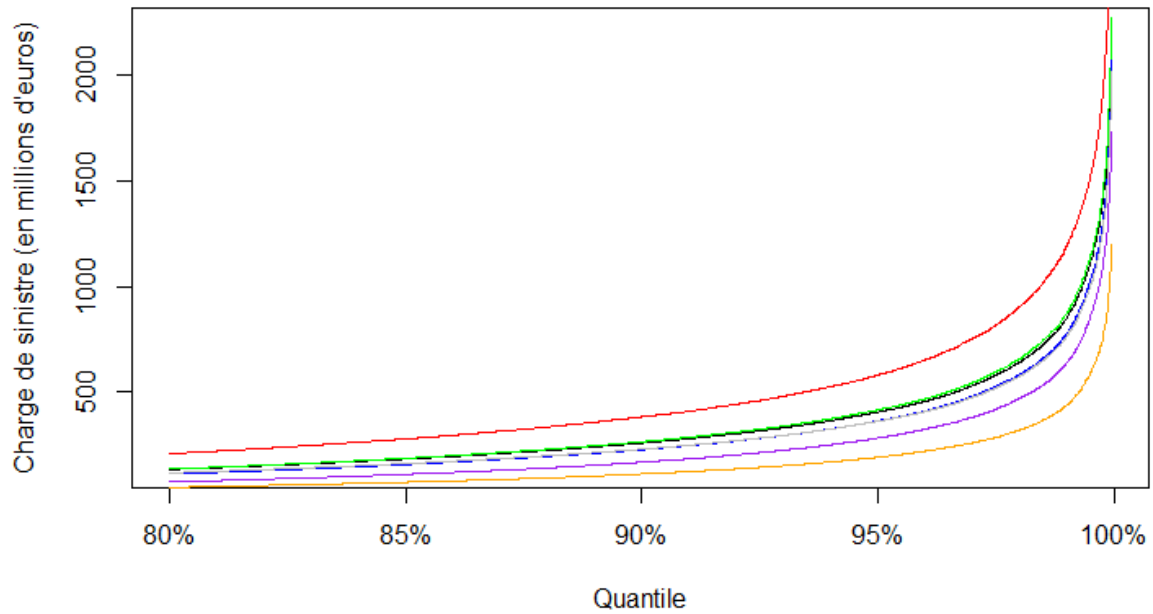
β	Aléatoire selon loi exponentielle $Exp(10)$
$k_e = k_I = k$	Constant égal à 2
$\theta_e = \theta_I = \theta$	Aléatoire selon loi exponentielle $Exp(k/5,5) = Exp(2/5,5)$
Taux de létalité	1 000 valeurs aléatoires selon loi log-normale de paramètres $\mu = \log(0,5\%)$ et $\sigma = 0,8$

Les résultats de ce scénario n°3 sont les suivants :

6.5. Mesure de la sinistralité décès de l'assureur

Nous calculons maintenant la charge de sinistre décès induite par ces épidémies sur nos 5 portefeuilles fictifs d'assureurs comprenant chacun 1 millions d'assurés :

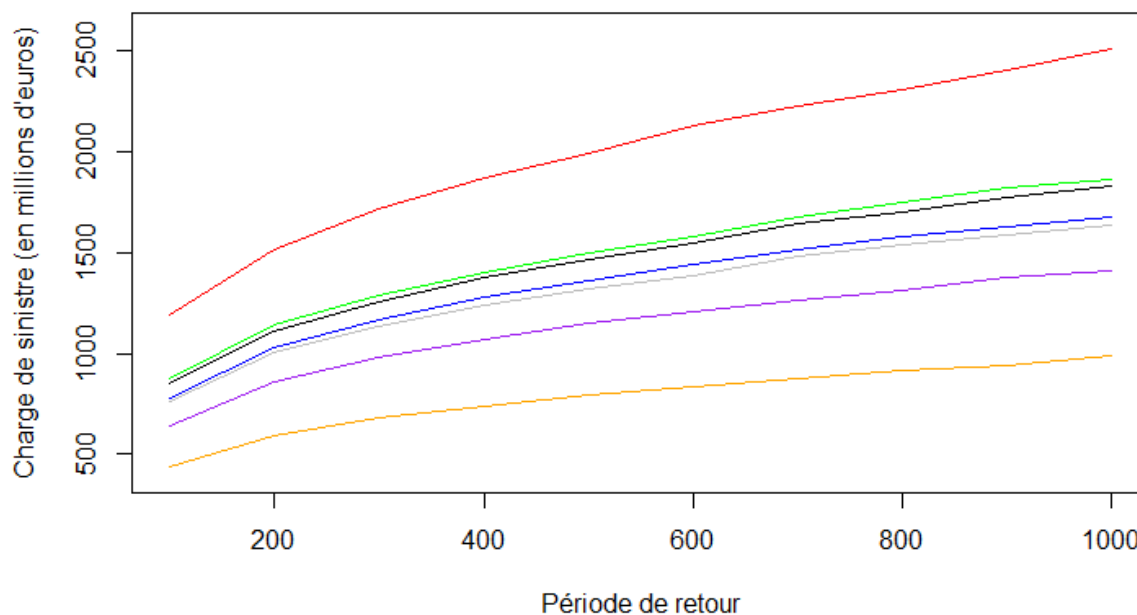
Nous comparons ces résultats au portefeuille « moyen français » du scénario 2.

Quantile de répartition de la charge de sinistre décès

Légende :

Couleur du graphique	Portefeuille
Rouge	1 : Ile de France
Vert	2 : Nord / Nord-Ouest
Bleu	3 : Est
Orange	4 : Sud-Ouest
Violet	5 : Sud-Est
Noir	6 : « Moyenne nationale »

Quantile de répartition de la charge de sinistre décès



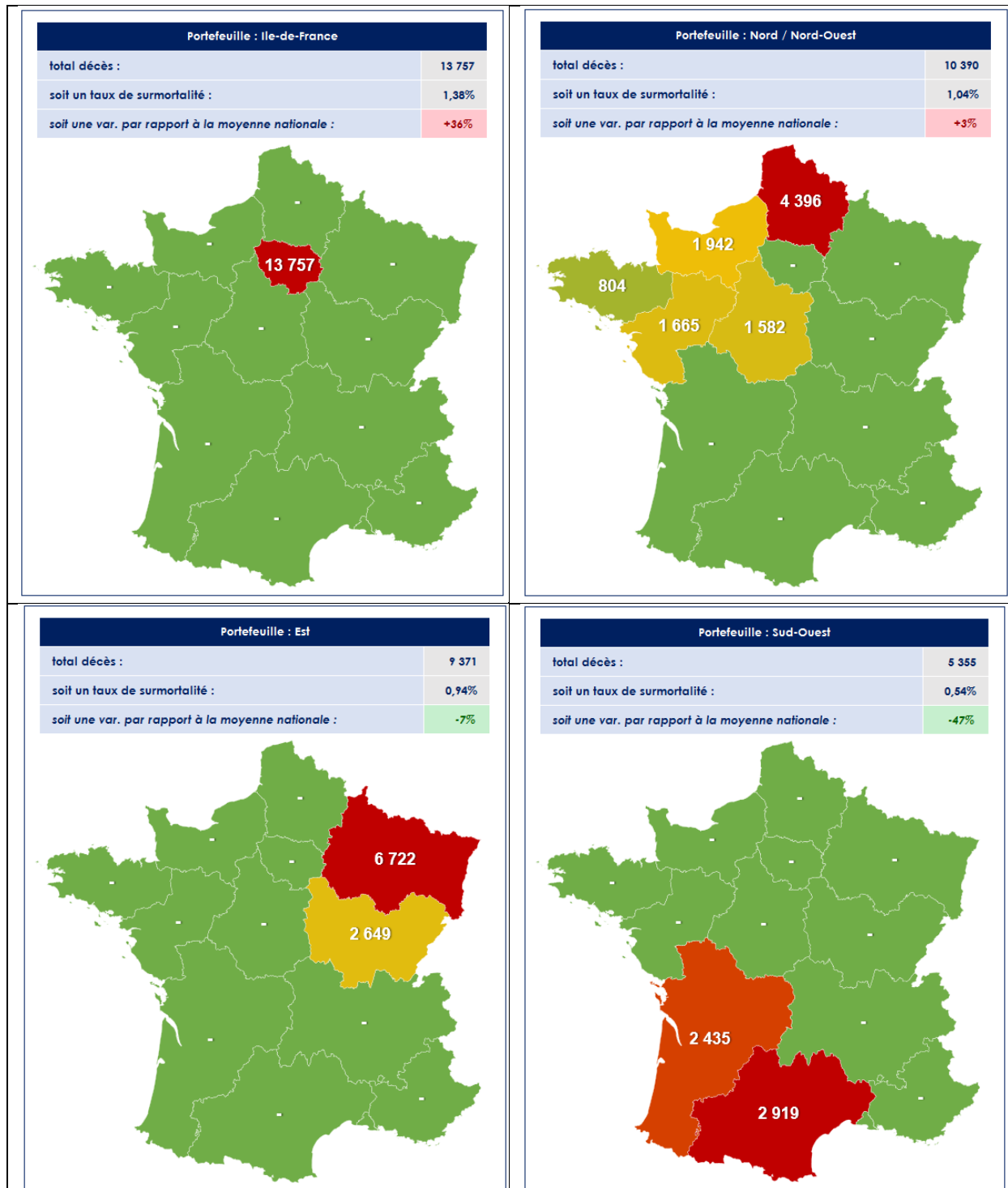
A 99,5%, la charge de sinistre décès est la suivante :

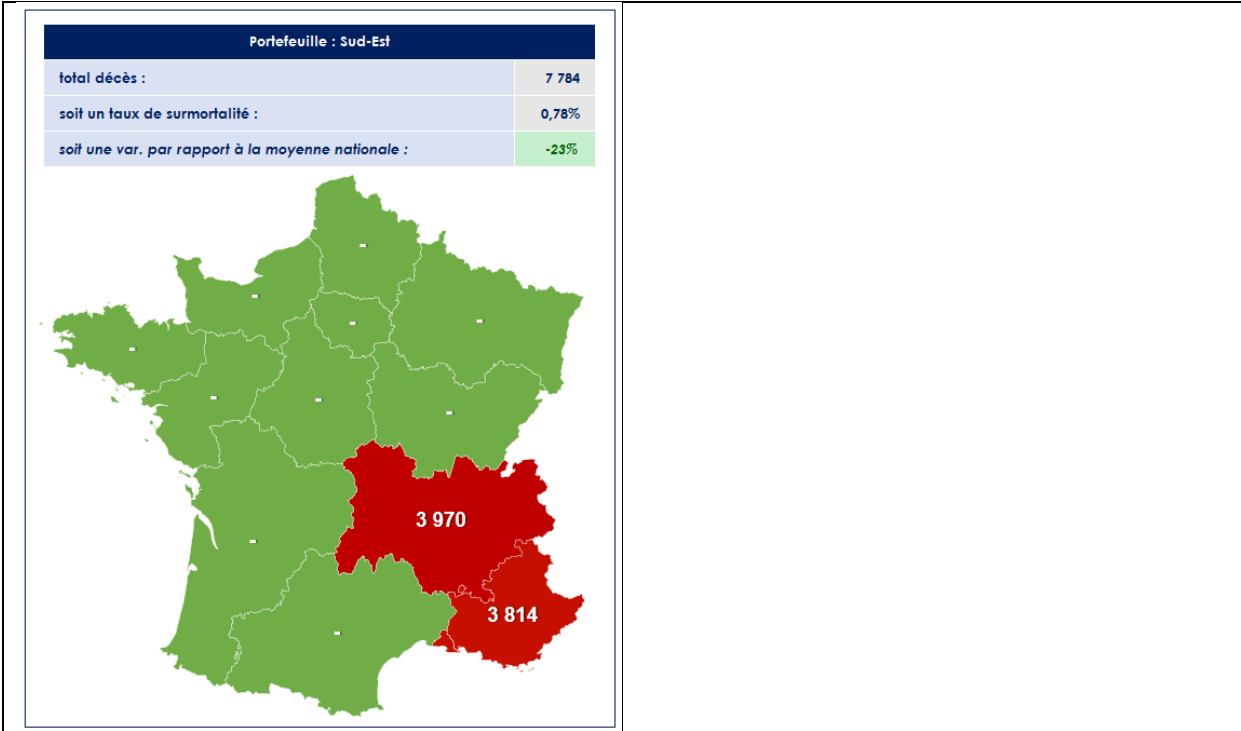
Portefeuille	Charge de sinistre (en M€)	Var. par rapport au ptf. n°6
1 : Ile de France	1 513	+36%
2 : Nord / Nord-Ouest	1 143	+3%
3 : Est	1 031	-7%
4 : Sud-Ouest	589	-47%
5 : Sud-Est	856	-23%
6 : « Moyenne nationale »	1 114	-

Ces résultats appellent les remarques suivantes :

- Un assureur avec un portefeuille exclusivement francilien, en cas de pandémie sévère (risque à 99,5%) verrait sa sinistralité s'élever à plus de 36% au-delà de celle d'un assureur avec un portefeuille d'assurés répartis partout en France,
- A l'inverse, le risque de pandémie sévère d'un assureur avec un portefeuille d'assurés exclusivement implantés dans le Sud-Ouest serait moitié moins touché que celui d'un assureur « national ».

Enfin, comme dans le scénario 2, nous pouvons nous intéresser à la répartition géographique des victimes pour ces 5 portefeuilles d'un million d'assurés.





6.6. Construction d'un module de « data visualisation »

Les sorties de nos différents scénarios proposées dans les paragraphes précédents sont « statiques » (ce sont des images, des tableaux), toutefois, pour bien illustrer le phénomène de propagation dans l'espace d'une pandémie, nous avons construit un module de « data visualisation ».

Ce module, part d'une simulation d'épidémie et construit une série temporelle qui trace à chaque instant l'état (S, E, I ou R) de nos 10 000 individus, et ce pour toute la durée de l'épidémie. Nous y avons ajouté les coordonnées GPS (latitude, longitude) de nos individus.

Nous avons ainsi créé un fichier .csv, que nous avons « lu » avec le module 3D Graph d'Excel. Ce module permet de représenter sur une carte des données géocodées (c'est-à-dire dotées de coordonnées GPS).

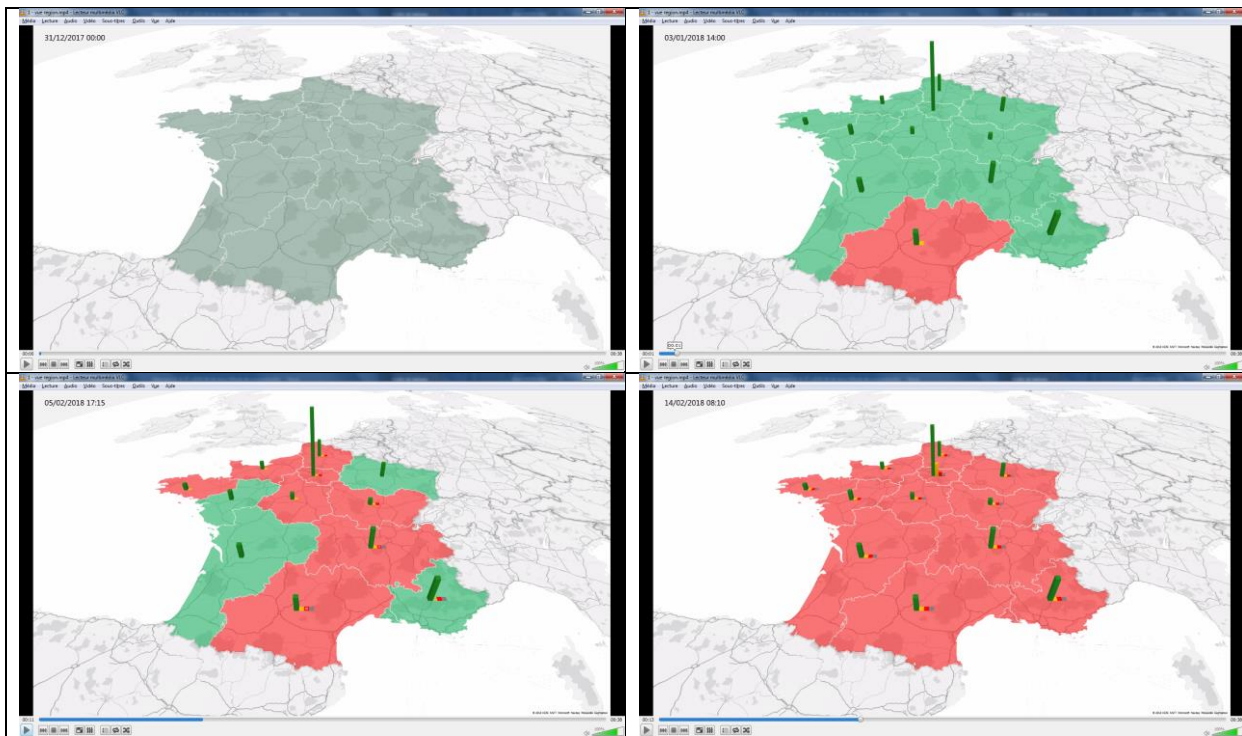
Nous représentons sur cette carte deux éléments :

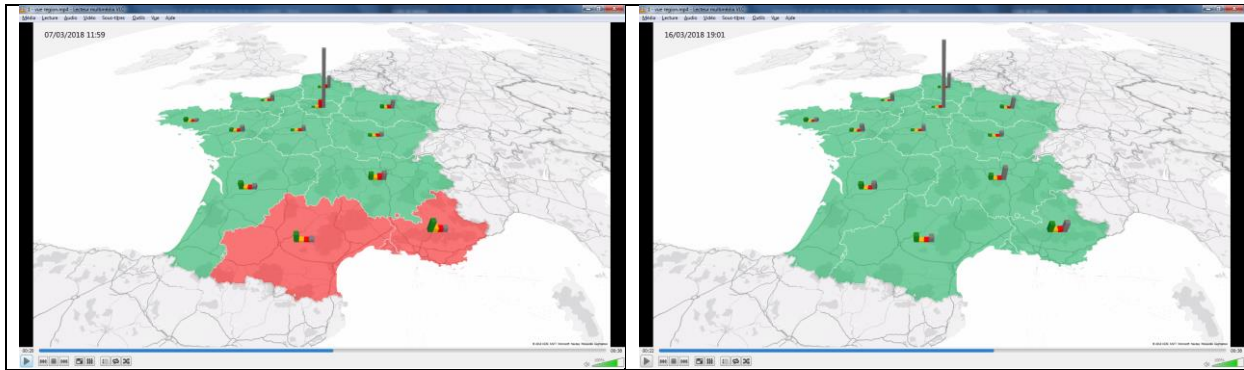
- 1) Quatre histogrammes (1 pour chaque état) comptant le nombre d'individus par région ou département dans chacun des 4 états.
- 2) Une couleur (vert ou rouge) indiquant si dans la région ou le département considéré, le nombre d'individus exposés ou infectieux (E ou I) est supérieur à celui des individus remis (état R) : lorsque c'est le cas, nous colorons la région ou le département en rouge, dans le cas contraire c'est du vert qui est retenu.

Les vidéos ainsi créées sont disponibles à l'adresse suivante : <https://goo.gl/QCUiwV>

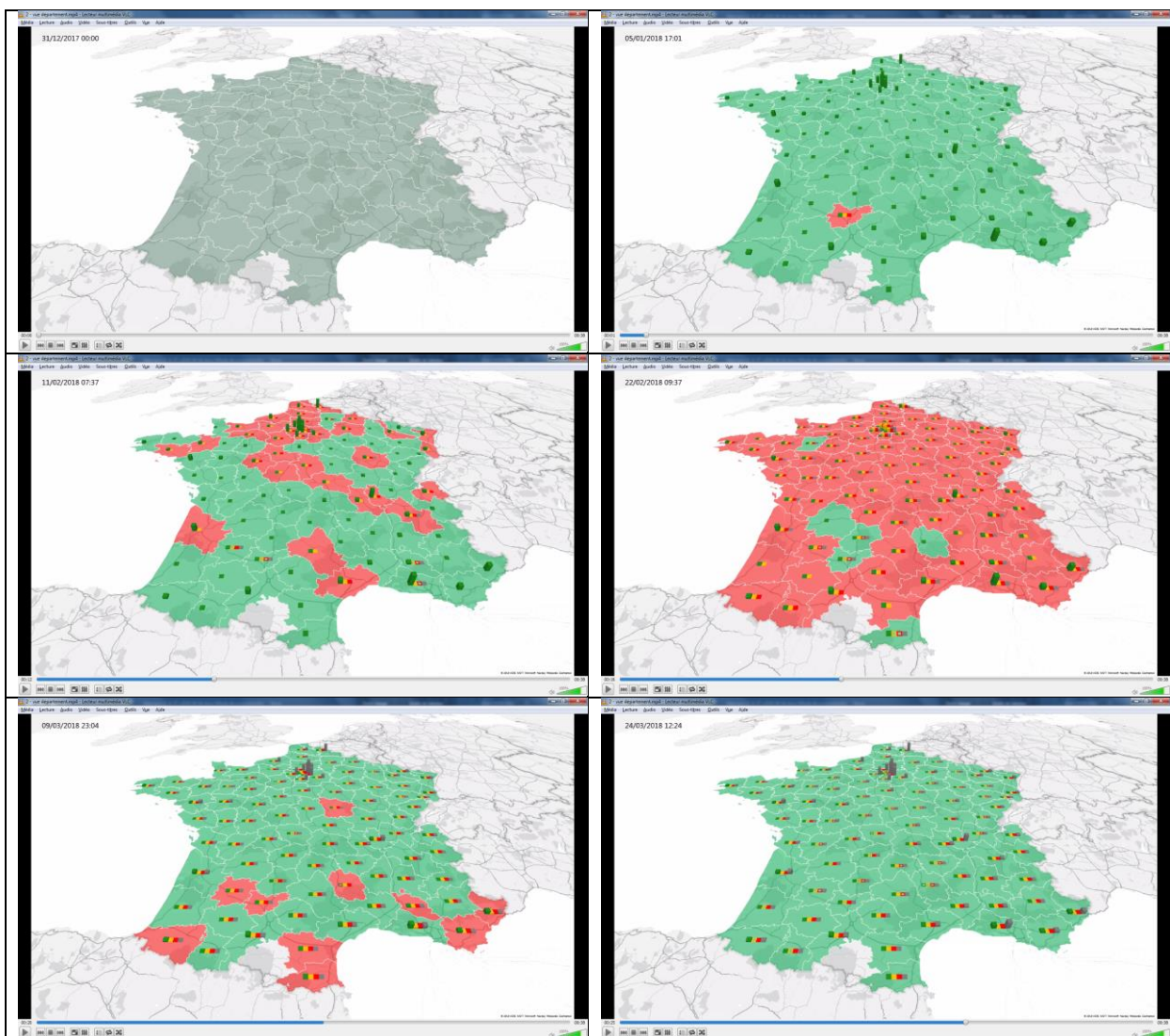
Ci-dessous, nous en extrayons quelques images :

1) Avec une maille de restitution : région





2) Avec une maille de restitution : département



Ce type de livrables permet de comprendre et d'interpréter visuellement (donc plus rapidement) les résultats de notre modèle. En regardant plus loin, c'est le type de livrables qui peut être souhaité par un assureur qui achèterait un tel modèle auprès d'un prestataire de marché.

7. Perspectives

7.1. Avantages et limites de notre modèle

7.1.1. Avantages

➤ Valeur ajoutée avec l'intégration du paramètre géographique

La principale valeur ajoutée de notre modèle, qui était l'objectif principal de notre étude, consistait en la prise en compte du paramètre géographique dans la propagation d'une épidémie entre individus. En effet, au travers de nos recherches et de notre bibliographie, le modèle SEIR stochastique est utilisé ou présenté à plusieurs reprises mais nous n'avons pas rencontré de modèle qui tenait compte de ce paramètre.

Nous sommes donc parvenus à construire un modèle plus précis dans la mesure où il tient compte de la réalité d'un territoire et de la disposition des individus sur ce territoire. Nous l'avons appliqué sur la France continentale mais il peut aisément s'appliquer ou être étendu à une autre région (cela peut être étendu jusqu'au monde entier), il suffit d'adapter la répartition des individus selon la réalité du terrain en question.

➤ Flexibilité du modèle et des paramètres

Outre la flexibilité évoquée ci-dessus sur le territoire auquel s'applique le modèle, notre code offre est construit et organisé de sorte à être compréhensible sans difficulté particulière, pour peu que le lecteur soit un minimum averti du langage R.

Le code permet ainsi à tout utilisateur qui souhaiterait le réutiliser de modifier aisément de nombreux paramètres du modèle : nombre d'individus, nombre de variables X_{ij} , lois utilisées pour modéliser β , k et θ , taux de létalité, etc.

➤ Adaptabilité à n'importe quel portefeuille d'assurés

Notre modèle démarre sur la construction d'un réseau de liens au sein de notre population de N individus puis applique à ce réseau un spectre d'épidémies. En sortie, on obtient ainsi les résultats (nombre d'infections moyen, temps d'épidémie, etc.) pour toute la population et de façon individuelle, ce qui permet d'avoir une vision extrêmement fine.

En outre, quand on s'intéresse ensuite à la sinistralité pour un assureur selon son propre portefeuille, il suffit finalement d'extraire une partie de la population qui correspond à sa répartition d'assurés sur le territoire français et de récupérer les résultats correspondant à ce sous-ensemble d'individus. Ainsi, on peut de façon équivalente récupérer les résultats de sinistralité pour un assureur A qui a un portefeuille réparti de façon uniforme en France et pour un assureur B qui dispose d'un portefeuille concentré dans une ou plusieurs régions en particulier.

7.1.2. Limites

➤ Restriction du modèle au territoire français continental

Notre modèle est restreint au territoire français continental et ne tient pas totalement compte des îles voisines, des autres pays européens et plus généralement du reste du monde, ce qui serait le scénario le plus complet pour modéliser un échantillon d'épidémies extrêmement varié, dont certaines sont locales et d'autres en provenance de pays européens ou encore plus lointains.

Nous avons contourné une partie de ces limites en modélisant les trois zones frontalières à la France, ce qui permet à la fois de tenir compte de la population voisine (facteur important pour la propagation) et de modéliser des épidémies naissantes dans des zones hors France et qui attaquent en premier lieu les régions frontalières (régions Sud, Nord et Est de la France).

La limite mentionnée ici est donc une limite de données et non une limite de notre modèle. En effet, pour pouvoir utiliser un modèle aussi complet, il faudrait intégrer les données de population et de répartition géographique. Bien évidemment, cela nécessiterait des processeurs et un temps machine beaucoup plus importants que ce que nous avons utilisé.

Si toutefois nous avons utilisé des données à l'échelle mondiale, se serait posé la question de paramétrer la construction du réseau. C'est une limite qui est relativement facile à contourner (dans la théorie mais qui demande un travail très conséquent en pratique), par l'introduction d'un ou plusieurs paramètres supplémentaires dans la formule de construction du réseau (E1). On peut par exemple construire une variable qui diminue la probabilité d'établir un contact entre deux individus si ceux-ci sont séparés par une mer. En reprenant les notations déjà vues avec le modèle de Groendyke et *al.* on peut ainsi penser à :

$$X_{\{i,j\},2} = \begin{cases} 0 & \text{si } i \text{ et } j \text{ sont séparés par une mer} \\ 1 & \text{sinon} \end{cases}$$

avec un η_2 à dimensionner en fonction des autres paramètres et du spectre de probabilités voulu dans le modèle (c'est bien là toute la difficulté de travail).

Le modèle permet d'ajouter autant de variables que voulues et de descendre à une maille encore plus fine. De façon similaire, on pourrait donc le complexifier en imaginant des variables qui tiennent compte de :

- La séparation par un océan
- La présence de montagnes entre deux zones
- L'appartenance de deux individus à un groupe commun (professionnel, amical, familial, etc.)

➤ **Dimensionnement des paramètres du R_0**

Dans notre modèle, nous avons choisi des lois pour modéliser les paramètres β , k et θ , de sorte à ce que R_0 soit entre 0 et 3, avec une distribution ciblée (plutôt concentrée sur la première moitié de l'intervalle). En parallèle, nous avons choisi de modéliser le taux de létalité par une loi log-normale de sorte qu'il soit plutôt concentré dans l'intervalle [0% ; 6%].

Ces dimensionnements sont cohérents et sont réalistes pour modéliser la majorité des gripes, ce qui est le périmètre choisi pour notre étude. Néanmoins un autre utilisateur de ce modèle pourra descendre à une maille plus fine sur la modélisation de ces paramètres, pour les confronter à d'autres études statistiques et documenter un R_0 le plus proche possible de la réalité.

➤ **Mesure des distances à vol d'oiseau**

Dans notre modèle, les distances entre individus sont mesurées à vol d'oiseau, i.e. en ligne droite. C'est une approche simplificatrice bien que suffisamment réaliste, néanmoins il serait intéressant d'affiner le modèle dans la mesure de cette distance, en tenant compte par exemple des moyens de transport (routes, trains, avions, pédestre, etc.) qui existent aujourd'hui.

Cela permet d'être encore plus précis dans la mesure des probabilités de contact entre individus, critère important pour la propagation du virus. Dans la mesure où notre modèle s'applique à l'échelle de la France, on peut considérer que cette complexification pour la mesure des distances n'ajouterait

pas forcément beaucoup de valeur ajoutée. En revanche, cela a nettement plus d'impact si le modèle était appliqué à l'échelle mondiale.

En effet, dans un tel cas, il y a un vrai intérêt à tenir compte notamment des liaisons aériennes. Comme nous l'avons vu au travers des exemples historiques, les épidémies se propagent d'autant plus vite travers le monde que les flux migratoires sont importants. Ainsi, un américain et un français ont plus de chance de se transmettre un virus aujourd'hui que ce n'était le cas il y a quelques siècles.

7.2. Mesure du risque

De par sa grande souplesse, ce modèle dispose de tous les éléments pour pouvoir être utilisé à la construction d'un modèle interne sous Solvabilité 2 ou même d'autres référentiels tels que celui des normes suisses (SST).

En effet, il permet d'intégrer les caractéristiques propres d'un portefeuille et donc à son assureur de mieux maîtriser son risque. Ainsi, le modèle avec paramètre géographique offrira la possibilité de mieux évaluer (par rapport à la formule standard) son exigence en capital relative au risque pandémique.

7.3. Tarification (assurance directe et réassurance)

L'atout principal que nous avons ajouté au modèle épidémiologique grâce au paramètre géographique est la possibilité d'évaluer le risque pandémique à une maille très fine puisqu'il permet une mesure de la sinistralité par commune, région et département.

C'est un élément qui va s'avérer très utile pour un assureur qui souhaitera tarifier des garanties relatives au risque pandémique, par exemple pour une reprise de portefeuille ou lors d'une réponse à un appel d'offre grands comptes : en fonction de la localisation des particuliers et/ou des entreprises à assurer, l'assureur saura si le risque pandémique est plus ou moins fort et donc adaptera son tarif en fonction de cela.

De même, un réassureur disposant de ce modèle pourra demander à un assureur qui le sollicite sur le risque pandémique la localisation de ses assurés, particuliers comme entreprises, de sorte à tarifier au plus juste et lui proposer la couverture de réassurance qui est la plus représentative de son risque. Ce serait une alternative forte et un élément important de démarcation par rapport aux autres réassureurs du marché (qui ne disposeraient pas de cet élément géographique dans leur modèle), donc un avantage concurrentiel.

7.4. Autres perspectives

En termes de perspectives, nous pensons que notre modèle est un bon point de départ pour un utilisateur qui aurait le temps, les données et les moyens informatiques pour :

- l'étendre à une région plus large que la France (l'Europe ou le monde entier par exemple)
- augmenter le nombre d'individus pris en compte dans l'échantillon, en essayant de converger vers la population réelle pour être le plus proche de la réalité (par exemple sur un échantillon de 60 millions d'individus pour le territoire français)

Ces perspectives sont complètement dans l'air du temps puisqu'elles peuvent s'envisager avec l'aide des différentes technologies du traitement de données en masse, qu'on appelle de façon générale « le Big Data ». Ces technologies sont en effet indispensables si on souhaite converger vers le

modèle extrême qui modéliserait des millions voire milliards d'individus sur un territoire aussi étendu que le monde. Indépendamment de la taille de l'échantillon et du territoire modélisé, cela permettrait d'augmenter le nombre d'épidémies et/ou le nombre de runs.

8. Conclusion

L'objectif de ce mémoire était de présenter un modèle épidémiologique permettant de représenter le plus fidèlement possible la diffusion d'une pandémie dans l'espace. Pour ce faire, nous avons donc procédé en deux temps.

Dans un premier temps, notre travail a consisté à coder le modèle épidémiologique le plus répandu dans la littérature : le modèle SEIR. Après en avoir détaillé les principaux paramètres, nous avons utilisé le package R « epinet » qui reproduit le modèle SEIR dans sa version déterministe. Nous avons alors modifié ce modèle déterministe pour développer une version stochastique.

Pour ce faire, nous avons construit un ensemble de 100 000 scénarios d'épidémie avec des paramètres de taux de reproduction (lié au nombre de personnes touchées par l'épidémie) et de taux de létalité (lié au nombre de décès causés par l'épidémie) permettant de couvrir aussi bien les épidémies à faible sévérité que les épidémies extrêmes.

Nous avons appliqué ce premier modèle à un portefeuille fictif de 10 000 assurés et avons ainsi pu produire des analyses sur les quantiles de répartition du nombre de victimes et de la charge de sinistre sur ce portefeuille. Ces résultats sont cohérents avec ceux produits par un logiciel de marché, véritable référence pour les modèles de pandémie : la solution RMS LifeRisk® Infectious Disease Model.

Pour autant nous avons identifié dans ce modèle une limite, qui nous semblait préjudiciable pour l'exploitation par un assureur des résultats produits : en effet ce modèle de place ne prend pas en compte la répartition géographique des individus du portefeuille. Ainsi, ce modèle donnera les mêmes résultats à un assureur national dont les assurés sont répartis partout en France (voire dans le monde) qu'à un assureur « régional » qui aurait par exemple l'ensemble de son portefeuille d'assurés résidant dans la même région.

Pour pallier cette limite, nous avons ainsi souhaité développer un second modèle, toujours basé sur le modèle épidémiologique SEIR, mais incluant un module de propagation dans l'espace. Notre modèle devait ainsi être capable de prendre en « input » la répartition géographique d'une population donnée, ainsi que de la sous-population des assurés au sein de cet ensemble.

Nous avons ainsi consacré une part importante de nos travaux à la modélisation la plus fine possible de la répartition sur le territoire français de notre population de référence.

Une fois ce travail réalisé, nous avons alors produit les mêmes analyses qu'avec notre premier modèle mais nous étions cette-fois ci capables de ventiler ces résultats par département ou par région.

Nous avons ainsi pu illustrer la pertinence de l'intégration de ce critère géographique en appliquant notre modèle à des portefeuilles fictifs d'assureur (aux répartitions géographiques d'assurés différentes) et constater des écarts significatifs dans les résultats de sinistralités obtenus.

Nous avons également consacré du temps à la restitution de ces résultats, que nous avons souhaité la plus « parlante » possible. Nous avons ainsi développé différentes sorties graphiques de notre modèle, telles que la carte de France des taux de surmortalité par région, mais également un module de « data visualisation » permettant d'illustrer sur une épidémie donnée, sa propagation spatio-temporelle, du patient zéro jusqu'au dernier individu touché. Ces résultats sont disponibles en ligne à l'adresse suivante : <https://goo.gl/QCUiWV>

Le modèle que nous avons développé présente toutefois des limites. Ainsi, la modélisation de la sinistralité du portefeuille nécessiterait d'être affinée en fonction de la pyramide des âges du portefeuille (une épidémie ne touchant pas toutes les classes d'âge de la même manière). Nous avons également dû nous limiter à un portefeuille de 10 000 assurés pour des raisons de capacité de calcul, il serait intéressant de pouvoir faire tourner notre modèle sur des portefeuilles plus importants.

Notre pourrait enfin être complété par des analyses sur d'autres risques assurés que le décès : l'arrêt de travail par exemple (même si les franchises présentes dans les contrats d'assurance sur les périodes d'incapacité nous mènent à penser que la sinistralité induite serait moins sévère que celle en décès), mais également en frais de santé (la fréquence des postes médecine de ville et hospitalisation notamment pouvant être fortement augmentées en présence d'une épidémie sévère).

Toutefois, malgré ces limites, notre mémoire nous a permis de vérifier notre intuition initiale, à savoir l'influence de la répartition géographique des assurés sur les niveaux de sinistralité observés sur des scénarios d'épidémie. Ces conclusions pourront être exploitées par nos employeurs respectifs pour mesurer aussi bien le risque de pandémie du portefeuille que celui d'une affaire nouvelle en prévoyance collective (très grand compte ou branche) : à la souscription d'un contrat collectif, notre modèle permet en effet de chiffrer le risque d'épidémie des assurés en fonction de leur répartition géographique sur le territoire.

9. Bibliographie

- 1 APREF (2016). *Pandémie – Assurances de personnes*. Groupe de travail « Pandémie », Récupéré le 7 décembre 2016 de <http://www.apref.org>
- 2 Britton, T., & O'Neill, P. (2002). *Bayesian Inference for Stochastic Epidemics in Populations with Random Social Structure*. Scandinavian Journal of Statistics
- 3 Demiris, N. & O'Neill, P. (2005). *Bayesian inference for stochastic multitype epidemics in structured populations via random graphs*. Journal of the Royal Statistical Society. Series B (Statistical Methodology)
- 4 Données publiques sur les communes françaises. Récupéré le 14 avril 2017 sur <https://www.data.gouv.fr/fr/datasets/listes-des-communes-geolocalisees-par-regions-departements-circonscriptions-nd/#>
- 5 Fontanet, A. (2016). *Conférence SCOR : "Faut-il craindre les pandémies ?"*
- 6 Groendyke, C., Welch, D., & Hunter, D. (2012). *A Network-based Analysis of the 1861 Hagelloch Measles Data*. Penn State Department of Statistics Technical Report #11-03
- 7 Groendyke, C., Welch, D., & Hunter, D. (2010). *Bayesian inference for contact networks given epidemic data*. Penn State Department of Statistics Technical Report #10-02
- 8 Günter Edenharter. *The SEIR model with births and deaths*. Récupéré le 25 février 2017 de <https://maple.cloud/#doc=6407056173039616>
- 9 Guy Carpenter (2016, 27 avril). *Etude Pandémie – Reconstitution des portefeuilles et résultats de l'étude pandémie*. Etude réalisée pour Malakoff Médéric
- 10 INSEE (2017). *T74 – Décès selon le sexe et l'âge atteint dans l'année (Séries depuis 1962 pour la France métropolitaine, 1994 pour la France entière)*. Récupéré le 8 mars 2017 de <https://www.insee.fr/fr/statistiques/3124970?sommaire=3053204&q=T74>
- 11 Limoge, F. (2015, 4 juin). *Prévoyance : les assureurs se couvrent contre le risque de pandémie*. Récupéré le 2 février 2016 de <http://www.argusdelassurance.com/acteurs/prevoyance-les-assureurs-se-couvrent-contre-le-risque-de-pandemie.94250>
- 12 Mandhouj, K. (2011). *Analyse du Risque Catastrophe d'une Pandémie en Assurance Prévoyance par une Approche Epidémiologique*. ENSAE : mémoire d'actuariat
- 13 Maxmind. *Base de données sur la population mondiale*. Récupéré le 16 avril 2017 sur <https://www.maxmind.com/fr/free-world-cities-database>
- 14 McCandless, D. (2014). *The Microbe-scope*. Sources: Centers for National Disease Control, World Health Organization, CIDRAP, studies. Récupéré le 8 décembre 2016 de <https://informationisbeautiful.net/visualizations/the-microbescope-infectious-diseases-in-context/>
- 15 Ray, J., Marzouk, Y. (2008). *A Bayesian method for inferring transmission chains in a partially observed epidemic*. Sandia National Laboratories, Livermore, CA 94550-0969
- 16 Speisser, R. (2013). *Evaluation du risque de pandémie et construction de deux modèles internes partiels en assurance de personnes dans le cadre de Solvabilité II*. ESSEC/UPMC
- 17 Swiss Re. (2007). *Pandemic influenza: A 21st century model for mortality shocks*.
- 18 Wikipédia. *Liste des préfectures de France*. Récupéré le 14 avril 2017 sur https://fr.wikipedia.org/wiki/Liste_des_pr%C3%A9fectures_de_France
- 19 Wikipédia. *Liste des départements français*. Récupéré le 14 avril 2017 sur https://fr.wikipedia.org/wiki/Liste_des_d%C3%A9partements_fran%C3%A7ais
- 20 Wikipédia. *Chef-lieu*. Récupéré le 14 avril 2017 sur <https://fr.wikipedia.org/wiki/Chef-lieu>
- 21 Wikipédia. *Région française*. Récupéré le 14 avril 2017 sur https://fr.wikipedia.org/wiki/R%C3%A9gion_fran%C3%A7aise

10. Glossaire

ACPR	Autorité de Contrôle Prudentiel et de Résolution
BCAC	Bureau Commun des Assurances Collectives
MCR	Minimum Capital Requirement
OMS	Organisation Mondiale de la Santé
ORSA	Own Risk and Solvency Assessment
SCR	Solvency Capital Requirement
SIDA	Syndrome d'immunodéficience acquise
SST	Swiss Solvency Test

