Dénoncer la fausse science épidémiologique : réquisitoire contre l'article "Estimating the burden of SARS-CoV-2 in France" : 17 chercheurs de 10 instituts ne comprennent ni les probabilités ni les mathématiques et inventent " l'équation générale de la vérité " qu'ils résolvent en " double aveugle " avant d'en maquiller piteusement la présentation et de se suicider sur la théorie du R0

Vincent Pavan

Dénoncer la fausse science épidémiologique : réquisitoire contre l'article "Estimating the burden of SARS-CoV-2 in France" : 17 chercheurs de 10 instituts ne comprennent ni les probabilités ni les mathématiques et inventent « l'équation générale de la vérité » qu'ils résolvent en « double aveugle » avant d'en maquiller piteusement la présentation et de se suicider sur la théorie du R_0

Pour l'article des auteurs : HAL Id : pasteur-02548181 ; https ://hal-pasteur.archives-ouvertes.fr/pasteur-02548181
Preprint submitted on 20 Apr 2020

Par **Vincent PAVAN**

Maître de conférences Aix-Marseille Université
Auteur : Les algèbres extérieures, ISTE Editions

Mon intention n'était d'abord absolument pas dirigée contre les 17 auteurs, pas moins, de cet article, ni les 10 instituts auxquels ils appartiennent. Il s'agissait pour moi d'utiliser le matériau de l'article pour illustrer à mes étudiants les bienfaits des mathématiques, des probabilités, des statistiques et leur montrer que des questions brûlantes de société pouvaient effectivement trouver des éléments de réponse éclairée grâce à l'application, maîtrisée, de notions et d'outils formels. Lesquels, ce qui rendait le papier intéressant à exploiter, tombaient dans l'escarcelle de mes enseignements. Mon intention était donc pédagogique : je voulais expliciter les hypothèses, les méthodes employées, les modèles exposés et essayer de retrouver, avec mes étudiants, les conclusions des auteurs.

Car en effet, ces conclusions se sont rapidement retrouvées en première ligne des journaux télévisés, de la presse écrite (j'ai d'ailleurs pu télécharger l'article

en question grâce à Libération) du web (évidemment) et il n'est pas impossible de penser qu'elles auront pu remonter jusqu'aux conseillers du président lui-même. Cela étant, j'étais quand même un brin sceptique sur le résumé qu'en donnaient les auteurs :

The lockdown reduced the reproductive number from 3.3 to 0.5 (84% reduction). By 11 May, when interventions are scheduled to be eased, we project 3.7 million (range : 2.3-6.7) people, 5.7% of the population, will have been infected. Population immunity appears insufficient to avoid a second wave if all control measures are released at the end of the lockdown.
(pour ceux qui ne lisent pas l'anglais : le confinement a réduit le nombre de reproduction (de base) de 3.3 à 0.5 (84 % de réduction). Au 11 mai, lorsque les interdictions seront levées, nous estimons que 3.7 millions de personnes (intervalle de confiance 2.3-6.7), soit 5.7% de la population, auront été infectées. Insuffisant pour créer une immunité collective et pour éviter une seconde vague si toutes les mesures de contrôle sont relâchées à la fin du confinement).

L'idée d'une efficacité surpuissante du confinement me semblait néanmoins assez douteuse dans la mesure où les deux pays notoirement réfractaires au confinement strict (les Pays-Bas et la Suède) ne possédaient manifestement pas, lorsque l'on regardait leur taux de mortalité par million d'habitants, de résultat pire que ceux de la France, l'Italie ou l'Espagne, au contraire (mais il faut préciser d'emblée que les chiffres définitifs ne seront évidemment connus qu'à la fin de l'épidémie). D'autre part, l'idée que le confinement ait pu opérer un point anguleux sur les courbes de dépistage, d'hospitalisation ou de mort (dénotant ainsi une discontinuité dans la *vitesse* d'expansion de l'épidémie, c'est-à-dire dans la dérivée temporelle de sa progression) n'avait jamais été observée sur aucune mesure réelle, toutes semblant au contraire montrer que la *cinématique* de l'épidémie se poursuivait de façon *continue*. Pourtant, une figure désastreuse, empruntée à l'article, fit le tour des plateaux d'opinion et des commentateurs relayant cette information salutaire pour la politique : le confinement ça marche, et en plus il faudra songer à mettre en place des mesures de contrôle. Comme chacun le sait, une bonne figure (même sortie arbitrairement d'une affirmation sans *aucun* fondement scientifique) a toujours beaucoup plus d'effet qu'une série d'équations ou que la présentation rigoureuse d'un calcul. D'ailleurs, je suis bien placé, hélas, pour le savoir : les reviewers (i.e. ceux qui relisent les articles pour donner ou pas l'autorisation de publication dans un journal) ne lisent quasiment jamais les calculs. Ce qui fait qu'à la fin, les auteurs ne se donnent finalement plus la peine de les faire ni de les présenter. À part quelques mathématiciens papier-crayon (au bilan carbone favorable), dont je suis, et pour qui les calculs font toujours partie des démonstrations de sorte que la moindre faute entre deux lignes peut effectivement condamner 10 pages de résultats qui permettent de démontrer un théorème.

Les mauvais scientifiques, tels que ceux qui espèrent publier cet article, ont toujours deux caractéristiques : ils complexifient à outrance ce qui est simple - et en général sans intérêt - pour essayer de se donner du crédit, et ils simplifient à outrance ce qui est compliqué - mais primordial - car, de toute façon, ils n'ont pas les moyens de faire mieux. Nous allons illustrer ces deux points particuliers dans la désormais célèbre prépublication des soi-disant meilleurs instituts au

monde (INSERM, Cambridge, Institut Pasteur, Sorbonne, John Hopkins, etc.) qui, au nom de techniques qu'ils ne maîtrisent absolument pas, se permettent de conseiller des mesures - le confinement, le traçage - dont l'étude COCONEL (voir le vidéo sur le site de l'IHU Méditerranée) montre qu'elles ne tiennent en gros que parce que les médias et les "scientifiques" disent qu'elles sont efficaces. Dans quelle mesure peut-on faire confiance aux "mathématiciens" qui "modélisent" les réponses à ces questions ? L'étude mathématique détaillée de leur publication permet d'affirmer que les auteurs sont des tricheurs qui maquillent leurs résultats en essayant de faire croire à une compétence qu'ils n'ont pas.

1 Présentation de l'article - parties étudiées pour leur imposture scientifique

1.1 Présentation accélérée

L'article est rédigé comme la succession assez peu structurée d'un ensemble d'estimations statistiques associées à l'épidémie de COVID19 en France. Il semble assez évident, même s'il est précisé que chacun des auteurs a participé pour une part égale à la rédaction de l'article, que les contributeurs se sont réparti la tâche par question. Mais on peut largement imaginer qu'ensuite chaque auteur aura relu l'ensemble de l'article avant de donner son autorisation de le soumettre et de le diffuser sous forme de pré-print. On aurait largement apprécié des numérotations dans les sections, des sous-sections, etc. Mais les auteurs ne sont évidemment pas familiers des traitements de texte scientifique (Latex) qui ont au moins le mérite de toujours vous obliger à structurer les propos que vous tenez. D'une manière générale, l'article est - selon nous - assez mal présenté : les résultats regroupés en fin d'article ne sont pas toujours très clairement reliés aux parties qui en présentent les principes d'obtention. La présentation des résultats est séparée en deux groupes : tous les tableaux d'abord et toutes les figures ensuite, alors qu'il existe des tableaux et des figures qui se rapportent aux résultats d'une même section. D'autre part, les données qui permettraient de refaire les calculs ne sont pas transmises dans l'article. Les auteurs donnent des sites sur lesquels on doit pouvoir trouver les données, mais il n'est pas clair que tout le monde ait accès aux sites en question, sans même parler des différents logiciels ou des procédures numériques précises qui sont employés. Ce qui fait qu'il est impossible pour quiconque de vérifier si par hasard il n'y aurait pas eu une erreur dans l'utilisation de données ou dans les programmes informatiques ayant permis de les trouver. L'article peut donc se résumer à une gigantesque utilisation de boîtes noires auxquelles on vous demande de faire confiance. Certes, il s'agit là d'une tendance généralisée : il n'est jamais gratifiant pour un chercheur de repasser son temps à essayer de retrouver (ou pas d'ailleurs) les données publiées par d'autres chercheurs. Alors pourquoi les fournir ? *Sauf qu'ici l'article n'a pas vraiment d'intérêt scientifique : les problématiques dépassent rarement le niveau élémentaire et le travail principal consiste en fait à préparer les données pour les mouliner ensuite à ce que l'on appelle des "click" bouton : un logiciel quelconque (fait maison ou industriel) qui donne, en quelques heures au pire, la valeur d'un paramètre.* **Le seul intérêt de l'article est de vendre des paramètres "scientifiques" aux décideurs politiques qui leur permettront ensuite d'affirmer que leurs décisions sont prises sur la base d'études**

sérieuses et incontestables. Ainsi l'article est-il de nature totalement politique : convaincre les citoyens que le confinement a des effets spectaculaires sur la propagation de l'épidémie (c'est en réalité une affirmation totalement imaginaire de la part des auteurs, qui ne repose sur aucune vérification expérimentale) et qu'ensuite l'estimation (totalement arbitraire en vérité) du nombre de personnes infectées au moment du déconfinement est trop peu importante et donc qu'il faudra sans doute s'attendre à de nouvelles mesures de contrôle (on cherche ici évidemment à vendre l'application de traçage du gouvernement). L'article est donc un article de soutien aux mesures politiques du gouvernement. Il y a deux classes de "résultats" dans cet article :

1. Ceux qui proviennent de "fit" sur des données *mesurées*. Ceux-ci sont totalement élémentaires du point de vue scientifique. Ils ont essentiellement pour but de tenter de montrer la maîtrise par les auteurs des outils basiques des probabilités et statistiques. En réalité, on va voir que les auteurs ne comprennent hélas rien à la notion de modélisation en probabilité, rien à la notion de solution d'un problème, rien à l'efficacité des méthodes numériques, et que leur culture mathématique s'est probablement arrêtée au niveau du lycée.
2. Celles qui proviennent de "modèles" sur des données non mesurées et qui viennent soutenir les annonces du gouvernement : le confinement fonctionne de façon spectaculaire et il faudra en passer par un traçage généralisé. Comme il s'agit de données issues de modèles n'ayant aucune possibilité de vérification expérimentale, l'idée c'est de faire confiance aux auteurs en s'appuyant sur ce que l'on aurait estimé de leurs compétences dans les parties précédentes. Or le "modèle" utilisé sur l'efficacité de la mesure de confinement est parfaitement tautologique. Il consiste à dire : si on suppose que le confinement est efficace, alors on peut tracer une courbe (sur une donnée non mesurée : le taux d'infection de la population) qui montre que le confinement est efficace. Donc vous voyez bien que le confinement est efficace.

En fait, le passage par la première partie conditionne la seconde : en effet, la deuxième partie du problème va consister à essayer de faire des prédictions que le gouvernement pourra utiliser pour prendre ses décisions. Comme il s'agit dès lors d'un travail "sans filet", il faut montrer que sur les problèmes "avec filet" pour lesquels on a des données empiriques pour les vérifier, les "modèles" trouvés par les auteurs correspondent de façon très précise à ce qui est observé. C'est parce qu'ils auront fait la démonstration de leur capacité à reproduire, ***avec méthode***, ce que l'on mesure, que les auteurs pourront inspirer la confiance sur les modèles qu'on ne mesure pas. Il est donc vital de voir si effectivement les méthodes que les auteurs emploient sur les données observées sont si fiables que cela. Et la réponse tombe comme un couperet mathématique : les auteurs n'ont jamais obtenu, par le calcul méthodique et raisonné, relevant d'une équation précise, les "fit" merveilleux qu'ils ont prétendu obtenir. Leurs méthodes ont rendu des "modèles" dont les courbes ne satisfaisaient pas les données observées. Ils ont donc ***maquillé*** (c'est-à-dire finalement ajusté à la main un résultat qu'ils connaissaient à l'avance), dans les présentations, les faiblesses extrêmes de leurs prétentions pour essayer de faire croire à leur "science" de la modélisation.

1.2 Dénonciation du passage incriminé

Nous recopions ici en intégralité un passage de l'article dont nous allons montrer qu'il est caractéristique des insuffisances très graves des auteurs dans leur approche des modèles, des probabilités et des mathématiques. Le passage est en anglais pour ne pas trahir l'article.

Estimating delays from hospitalization to death and from hospitalization to ICU

In a growing epidemic, the time to death among individuals who have already experienced the outcome will be an underestimate of overall time to death, as many of those who take longer will not yet have died. We need to account for these delays when estimating the infection fatality ratio as otherwise we would underestimate the probability of death (8).

To capture the delay to death for the different age groups we use data from cases throughout France that had dates of hospitalization and dates of death. We assume that the delays follow a lognormal distribution as this has previously been shown to work well for SARS-CoV-2 infections (17).

We use the number of hospitalizations on a given day to account for the state of the epidemic at that time, similar to what has previously been used (5). We note that a subset of individuals die within a short period of time after entering hospital. We therefore use a mixture distribution composed of an exponential distribution for those that die within a short delay and a lognormal distribution for those that die after longer delays (Figure S3).

$$\pi_i^{true} = (1-\rho) * exponentielle\,(m) + \rho * lognormal\left(\mu, \sigma^2\right) \tag{1}$$

We denote by π_i^{true} *the true probability density function (pdf) of the delay, and* π^{obs} *the observed density, which will be biased to be right skewed as most individuals will not have had their outcome. We denote by* Π_i^{true} *and* Π_i^{obs} *their cumulative density functions (cdf), respectively. We can approximate the expected delay distribution* π_i^{exp}*, for a given age group* i*, at a given time* T *during the epidemic, thereby adjusting for the stage of the epidemic, given the true pdf for the delay* Π_i^{true} *using the following adjustment :*

$$\pi_i^{exp}(k) = \frac{\sum_{j=1}^{j=T} H_{i,j-k} \times \left(\Pi_{i,k+1}^{true} - \Pi_{i,k}^{true}\right)}{\sum_{j=1}^{j=T} \sum_{l=0}^{l=j-1} H_{i,j-k} \times \left(\Pi_{i,l+1}^{true} - \Pi_{i,l}^{true}\right)}, \quad forall\ delays\ k\ \in [0,T] \tag{2}$$

where $H_{i,j}$ *is the total number of hospitalized cases of age* i *at time* j*.*

For the correct pdf Π_i^{true}*, we should have :*

$$\pi^{exp} = \pi^{obs} \tag{3}$$

We estimate parameters of the true delay from hospitalization to death distribution π^{true} *for each group in turn by minimizing the sum of squared error (SSE) of the distribution* π_i^{exp} *to the observed data* π^{obs}*. Given the small number of*

deaths in younger age groups, we consider three age groups : <70y, 70-80, 80+. To get an overall estimate, we also repeat the calculation using all individuals across all age groups.

To fit the delays from hospitalization to ICU admission we use the same approach, however, we consider the delays are constant across age groups and that they follow an exponential distribution (Figure S1).

1.3 La partie étudiée et le maquillage scientifique

Dans le point qu'ils traitent dans leur article, les auteurs entendent résoudre cette question : sachant que l'on a fini par mourir, comment estimer la loi de probabilité donnant le nombre de jours après lesquels on meurt une fois hospitalisé ? Cette question est tout à fait légitime. Cependant, la manière avec laquelle les auteurs vont y répondre va illustrer de façon sidérante l'absence totale de maîtrise d'outils mathématiques pourtant élémentaires. Ainsi :

1. Les auteurs vont montrer qu'ils ignorent à peu près tout de la façon avec laquelle on construit une probabilité sur un ensemble discret fini (ce qui est pourtant le b.a. ba.) ; qu'ils ignorent comment il faut recourir au formalisme de la théorie des ensembles et des variables aléatoires discrètes ; qu'ils jonglent avec des raisonnements qui n'ont absolument aucune rigueur mathématique, engendrant ainsi des fautes majeures dans les problèmes alors étudiés.
2. Les auteurs vont montrer qu'ils ne comprennent rien aux fondamentaux des systèmes linéaires. Mieux, afin de se donner de la consistance, ils vont transformer un système linéaire en système non linéaire pour essayer de faire croire à la complexité du problème. Trouver plus grotesque approche dans l'histoire des sciences reviendrait à se donner comme objectif de démontrer que la Terre est plate et que le Soleil lui tourne autour selon des cercles parfaits dont elle est le centre.
3. Dans cette partie les auteurs vont exposer exactement tous les défauts des méthodes qu'ils utilisent. Cette partie nous permettra ainsi de mettre en évidence la philosophie implicite des épidémiologistes : « *Nous sommes capable de trouver une solution à tous les problèmes même - et surtout - ceux qui n'en ont pas. La seule condition à cela c'est en gros d'affirmer que* $0 = -1$ *est une approximation acceptable de la vérité et de généraliser cette conception à l'ensemble de notre travail.* »

L'étude de la partie que nous traitons est sans doute la plus intéressante car elle montre exactement ce qui semble en être d'une sorte de culture du pauvre des épidémiologistes : le problème traité se situant à un niveau L1 ou L2, il est très facile, exactement comme lorsque l'on corrige les étudiants, d'analyser les erreurs et de les rectifier. On y voit tout simplement que ce sont les concepts de base qui ne sont absolument pas assimilés. Au final, les résultats obtenus sont donc excessivement mauvais et il ne reste plus qu'à opérer à la mise à mort des méthodes pour reconstruire "à la main" des résultats ***présentés*** donc ***de façon falsifiée***. Nous verrons que la justification de cette falsification, les auteurs l'énoncent de la façon suivante :

"*In a growing epidemic, the time to death among individuals who have already experienced the outcome will be an underestimate of overall time to death, as many of those who take longer will not yet have died. We need to account for these delays when estimating the infection fatality ratio as otherwise we would underestimate the probability of death.*"
(traduction (très libre pour redonner le sens du texte) : Dans une épidémie en cours, pour les individus qui sont déjà décédés, la durée séparant l'hospitalisation de la mort est nécessairement sous-estimée puisque ceux qui vont mourir après une longue durée n'apparaissent pas encore dans les données disponibles. Nous avons besoin cependant de donner une idée de ces délais (qui donc n'apparaissent pas dans les données mesurées, nda) sans quoi la prédiction du taux de mortalité sera nécessairement sous-estimée.)

En gros, ce que disent les auteurs, c'est que de toute façon les données observées sont biaisées (ça ne s'invente pas) parce que :

1. les données définitives ne seront connues qu'à la fin de l'épidémie ;
2. nous ne sommes pas à la fin de l'épidémie.

Si cette affirmation est vraie, elle ne s'applique pas cependant pas au problème traité par les auteurs. En effet, dans l'observation du temps d'hospitalisation avant le décès, on remarque, sur les données disponibles, que plus de 50% des personnes meurent moins de 4 jours après leur hospitalisation et que moins de 3% des personnes meurent après plus de 10 jours d'hospitalisation. Ainsi, si vous avez 40 jours d'observation (ce qui est le cas des auteurs), il est assez faux de dire que les personnes qui vont mourir après plus de 40 jours d'hospitalisation (et qu'effectivement vous ne pouvez pas observer) vont venir perturber les fréquences établies : jusqu'à présent, de toute façon, vous n'avez quasiment jamais vu de décès au-delà de 23 jours (les probabilités observées de mourir après 23 jours sont totalement négligeables, à peine quelques décimales de pourcentages). Donc l'argument est totalement irrecevable. Le fait que vous ayez 80% des personnes qui meurent en moins de 10 jours, vous avez pu l'observer sur 30 dates successives d'enregistrement d'hospitalisation : cela semble déjà une série très représentative et la forme des fréquences ainsi établies a peu de chance d'évoluer significativement si vous rajoutez 40 jours d'observation (si l'on table en gros sur une épidémie qui dure de début mars à mi-mai). Bien sûr, tout cela devrait faire l'objet d'une modélisation statistique plus précise (et ça peut parfaitement être le cas), mais notre propos n'est pas d'aller jusque-là. On pourra, si besoin, une fois l'épidémie terminée, revenir sur ce biais complètement irréaliste annoncé par les auteurs et qui n'a sans doute aucune chance d'être observé - sauf par exemple à ce que l'on dépiste et que l'on traite massivement - auquel cas les gens viendront sans doute à l'hôpital plus tôt et donc n'attendront pas d'être au seuil de la mort pour venir se faire hospitaliser (ce qui se traduit par le fait que 50% meurent moins de 4 jours après leur hospitalisation). Comme on peut le voir sur la figure que nous avons reprise des auteurs (voir figure 1) , il semblerait qu'il y ait ainsi un accord quasi parfait entre le "modèle" (la courbe en rouge de la figure (A)) et l'observation (le diagramme bâton sur la même figure (A)). *Bien sûr, si vous lisez rapidement le papier, vous êtes très impressionné par la faculté des auteurs à pouvoir modéliser un phénomène connu.* ***Mais en réalité, cette***

courbe est un trucage : les auteurs n'ont jamais eu aucune méthode de calcul pour l'obtenir. En fait, les auteurs vont jouer d'une présentation grossièrement manipulatrice. Voyons cela.

1. Leur "modèle " (nous verrons qu'il s'agit d'un festival d'incohérences) leur a donné une décomposition du phénomène en deux phénomènes : ceux qui meurent rapidement "*Rapid death*" (en moyenne moins d'un jour après hospitalisation) et les autres "*Slow death*". Ils ont modélisé ces deux catégories de populations par les deux courbes de la figure de droite. Comme on le voit, il existe dans la seconde distribution ("*Slow death*") une partie assez significative de personnes qui, selon le "modèle", doivent mourir largement au-delà des 20 jours, voire des 30 jours. (de l'ordre de 15%).
2. Or, cela ne colle évidemment pas avec les données observées (l'histogramme en bâtons représentant les fréquences collectées auprès des hôpitaux). Dans ces données, on a effectivement du mal à mesurer des personnes qui mourraient 23 jours après leur hospitalisation (et c'est ce que nous avons pris comme hypothèse quand nous avons reconstruit leurs données).
3. Donc les auteurs ont inventé l'histoire des données mesurées qui en fait n'étaient pas vraiment réellement celles que l'on devrait mesurer (!) car en fin d'épidémie les données expérimentales feront apparaître beaucoup de cas de mort après 23 jours d'hospitalisation (ce qui est, comme nous l'avons souligné, une hypothèse totalement irréaliste). Ce que disent les auteurs, c'est que si l'on continue à observer encore 40 jours, on va voir apparaître de façon significative des proportions de gens mourant plus de 20 jours après hospitalisation, et ce, au détriment des personnes qui meurent moins de 20 jours avant.
4. Dès lors, puisque leur courbe théorique ne colle pas avec l'observation, ils vont s'autoriser à modifier cette courbe "à la main". En fait, leur modèle sous-estime de façon assez significative la durée de mort comprise en 1 et 5 jours (ce qui est gênant car 50% des personnes meurent effectivement dans cette durée) : c'est le prix à payer pour un "modèle" qui voit mourir des gens de façon non négligeable après plus de 20 jours alors que ce n'est pas le cas en réalité. Dans le commentaire de la fameuse courbe "rouge", les auteurs indiquent ainsi que "*models fitted to take into account that in a growing epidemic, observed deaths will be biased towards ones that die quickly*", ce qui signifie "les modèles ont été fittés pour prendre en compte le fait que dans une épidémie en cours, l'observation du nombre de morts est biaisée vers les durées les plus courtes". Dit autrement, ils ont expliqué le fait que leur modèle ne collait bien avec les données avec le fait que... les données étaient nécessairement approximatives.

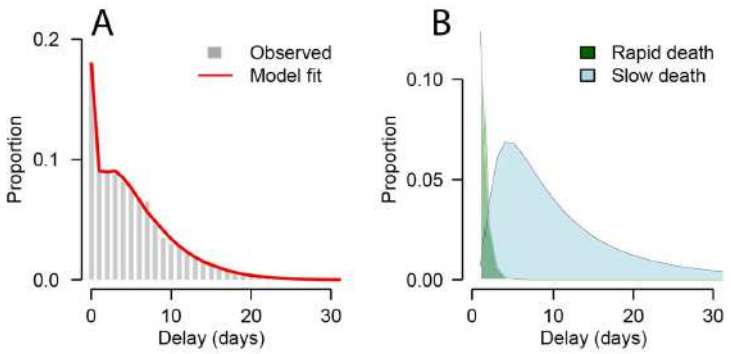

Figure S3. (A) Observed and fitted distribution of delays between hospital admission and death. **(B)** Model estimates of distribution of rapid decline and slow decline. Models fitted to take into account that in a growing epidemic, observed deaths will be biased towards ones that die quickly.

FIGURE 1 – données (diagramme bâtons) vs "modèle" proposé par les auteurs (courbes en rouge). En théorie, la courbe en rouge sur le dessin de gauche (A) devrait être la somme des deux courbes sur le dessin de droite (B). Ce n'est pas le cas. La courbe en rouge a été "fittée", c'est-à-dire ici maquillée sous un prétexte fallacieux.

Age group	Parameters				Overall Mean (days)
	P(short delay)	Exponential (for short delay)	Lognormal (for longer delays)		
		Mean (days)	Mean (days)	Median (days)	
<70	0.11	0.67	21.2	12.4	14.0
70-80	0.13	0.67	12.6	8.5	10.3
80+	0.18	0.67	10.5	7.5	8.6
Mean	0.15	0.67	13.2	8.6	10.1

FIGURE 2 – paramètres permettant de reconstruire les deux courbes modélisant la mort rapide et la mort plus lente. Les paramètres des deux courbes du (B) de la figure (1) sont ceux de la dernière ligne du tableau

.

Remarquons tout de suite que, dans les données du tableau précédent, il y a déjà un problème assez inquiétant de calcul des moyennes. En effet, si 15% des personnes hospitalisées qui vont décéder le font en moyenne 0.67 jour après leur hospitalisation et que 85% des autres le font en moyenne 13.2 jours après leur hospitalisation, alors il faut que la moyenne (générale) vérifie :

$$\texttt{Mean}_D = 0.15 * 0.67 + 0.85 * 13.2 = 11.32 \quad (4)$$

là où les auteurs annoncent une durée moyenne d'hospitalisation de 10.1 jours. Quand

on sait que la valeur moyenne observée sur les données est d'ailleurs de l'ordre de 6 jours, on est évidemment très inquiet sur les capacités des auteurs.

5. Dès lors, ils ont "fitté" les données. Normalement, "fitter" des données, cela consiste à les faire passer par des méthodes mathématiques, précisément justifiées dans leurs hypothèses et dans leur réalisation, de sorte que vous extirpiez au final une information intéressante. Mais rien de cela dans le cas qui nous intéresse. La seule "méthode" qu'ils avaient (en réalité une succession d'inepties mathématiques), les auteurs s'en sont servis pour les 2 courbes de la figure B. Pour passer de droite à gauche, ils n'ont appliqué aucune méthode : ils ont fait les corrections *à la main* c'est-à-dire en utilisant une cuisine spécifique au cas qui est traité. C'est pour cette raison, qu'au final, la courbe colle aussi bien aux résultats. Toutes les personnes qui font sérieusement de la modélisation savent qu'on ne peut jamais obtenir une coïncidence aussi spectaculaire, et ce, d'autant que les outils utilisés par les auteurs n'ont aucune consistance mathématique. Mais surtout, il est impossible que le modèle proposé par les auteurs ait une raison quelconque d'exister.

2 Étude théorique du modèle

2.1 Mathématique de $\boldsymbol{\pi}^{obs}$ et $\boldsymbol{\pi}^{true}$

Dans cette partie, nous montrons que l'équation (2) posée par les auteurs admet une solution analytique dont l'établissement ne pose aucune difficulté, et que la solution de cette équation montre que le rapport entre $(\pi_k^{true} : k \in [\![0T]\!])$ et $\left(\pi_k^{obs} : k \in [\![0T]\!]\right)$ n'est pas du tout celui entre "théorique" (vivant dans le monde des Idées) et son "observation" (vivant dans le monde des choses réelles).

Nous énonçons ici le premier théorème, d'un niveau L1-L2, qui résout de manière très rapide un problème que les auteurs ont manifestement trouvé insurmontable, utilisant une méthode d'optimisation par moindres carrés manifestement programmée sur une fonction non linéaire, au lieu d'utiliser un crayon de papier et une petite feuille :

Proposition 1. *Soit T un entier. On suppose que les hypothèses suivantes sont vérifiées*

1. *Soit $\left(\pi_k^{obs} : k \in [\![0;T]\!]\right)$ une densité de probabilité strictement positive, c'est-à-dire une famille de réels qui vérifie :*

$$\forall k \in [\![0;T]\!], \quad \pi_k^{obs} > 0, \quad \sum_{k=0}^{k=T} \pi_k^{obs} = 1 \tag{5}$$

2. *Soit $\alpha_k : k \in [\![0,T]\!]$ une famille de coefficients, telle que :*

$$\forall k \in [\![0;T]\!], \quad \alpha_k > 0 \tag{6}$$

Alors il existe une unique famille $(\pi_k^{true} : k \in [\![0;T]\!])$ *qui vérifie :*

$$\forall k \in [\![0;T]\!], \quad \pi_k^{obs} = \frac{\alpha_k \pi_k^{true}}{\sum_{l=0}^{l=T} \alpha_l \pi_l^{true}} \tag{7}$$

$$\forall k \in [\![0;T]\!] \ \pi_k^{true} > 0, \quad \sum_{k=0}^{k=T} \pi_k^{true} = 1 \tag{8}$$

de plus on peut calculer de façon explicite les expressions de $(\pi_k^{true} : k \in [\![0;T]\!])$ *par la formule :*

$$\frac{1}{\alpha} = \sum_{l=0}^{l=T} \frac{1}{\alpha_l} \pi_l^{obs}, \quad \pi_k^{true} = \frac{\alpha}{\alpha_k} \pi_k^{obs} \tag{9}$$

Avant de faire la démonstration de cette proposition assez élémentaire, notons que nous l'avons présentée de la manière avec laquelle elle apparaît dans la deuxième section de l'article. La présentation du problème à traiter est assez peu convaincante, car elle présente de façon a priori non linéaire un problème qui est en fait linéaire. Il s'agit là d'une remarque importante dans la suite : les auteurs de l'article ne s'intéressent jamais à l'idée de savoir si les problèmes mathématiques qui apparaissent dans leurs problèmes sont bien posés ou pas. Ainsi, n'importe quelle personne ayant fait des mathématiques sait bien qu'il existe deux classes de problèmes, s'agissant de résoudre des équations : soit on a affaire à un système linéaire et on est rassuré car il y a une théorie très riche ; soit on a affaire à du non-linéaire et là c'est immédiatement la jungle car il y a assez peu de résultats généraux. Un algorithme célèbre, l'algorithme de Newton, permet cependant de résoudre une équation non linéaire par une succession de résolutions de systèmes linéaires, preuve que ces derniers sont le point cardinal de toute l'analyse numérique. Ici, le problème tel qu'il est proposé est non linéaire, car il utilise la fonction vectorielle $\mathbf{f} := (f_k(\mathbf{x}) : k \in [\![0,T]\!])$ de la variable vectorielle $\mathbf{x} = (x_k : k \in [\![0,T]\!])$ définie par :

$$\forall k \in [\![0,T]\!], \quad f_k(x_0, \cdots, x_T) := \frac{\alpha_k x_k}{\sum_{l=0}^{l=T} \alpha_l x_l} \tag{10}$$

Une telle fonction est évidemment non linéaire. Plus inquiétant, le fait que si l'on peut trouver une solution $(\pi_k^{true} : k \in [\![0,T]\!])$ à l'équation $f_k(\mathbf{x}) = \pi_k^{obs}$, alors pour n'importe quel $\lambda \neq 0$ la famille $(\lambda \pi_k^{true} : k \in [\![0,T]\!])$ est encore une solution de l'équation puisque l'on remarquera immédiatement que :

$$\pi_k^{obs} = \frac{\alpha_k \pi_k^{true}}{\sum_{l=0}^{l=T} \alpha_l \pi_l^{true}} \Rightarrow \pi_k^{obs} = \frac{\lambda}{\lambda} \frac{\alpha_k \pi_k^{true}}{\sum_{l=0}^{l=T} \alpha_l \pi_l^{true}} = \frac{\alpha_k (\lambda \pi_k^{true})}{\sum_{l=0}^{l=T} \alpha_l (\lambda \pi_l^{true})} \tag{11}$$

l'absence d'unicité est en général un problème car, si vous lancez un algorithme qui, de plus, est formulé en utilisant une fonction non linéaire, vous ne savez absolument pas quelle solution l'ordinateur va vous retourner. Pour le problème des auteurs, on pouvait lever cette indétermination en imposant que la solution que l'on cherche, puisqu'elle devait impérativement elle-même être une densité de probabilité, devait vérifier que tous les $(pi_k^{true} : k \in [\![0,T]\!])$ soient positifs et leur somme égale à un. Mais cela n'était pas encore suffisant, car il fallait être capable de démontrer, en incluant les contraintes sur la solution recherchée, qu'il existait bien une unique solution. Or, sur les problèmes non linéaires, montrer

qu'il existe une solution unique est un exercice atrocement compliqué. Il existe très très peu de cas où l'on sait le faire. Pourquoi donc les auteurs ne sont-ils pas entrés dans des considérations de linéarité, les seules à même de montrer l'existence et l'unicité de la solution désirée ? Parce que, de toute façon, les épidémiologistes se moquent de ce genre de considérations : ils trouvent toujours une solution à tous les problèmes, y compris ceux qui n'en ont pas (nous expliquerons comment et pourquoi plus en détail). Donc pourquoi venir les embêter ?

Démonstration. La preuve repose essentiellement sur le théorème de Perron-Frobenius (voir par exemple sur le Net l'excellent document de Jean-Etienne Rombaldi). Nous l'énonçons :

Théorème 1 (Perron-Frobenius)**.** *Soit $A := A_{ij}$ une matrice carrée de taille n dont tous les coefficients sont strictement positifs. On note par $\rho(A)$ son rayon spectral. Soit $\alpha > 0$. Alors il existe un unique vecteur $\mathbf{x}(\alpha) := (x_k(\alpha), k :\in [\![1, n]\!])$ tel que :*

$$A\mathbf{x}(\alpha) = \rho(A)\mathbf{x}(\alpha), \quad \forall k \in [\![1, n]\!]\ x_k(\alpha) > 0, \quad \sum_{k=1}^{k=n} x_k(\alpha) = \alpha \tag{12}$$

notre travail consiste donc simplement à remettre le problème des auteurs dans la forme de Perron-Frobenius. On a :

$$\pi_k^{obs} = \frac{\alpha_k \pi_k^{true}}{\sum_{l=0}^{l=T} \alpha_l \pi_l^{true}} \Rightarrow \pi_k^{obs} \sum_{l=0}^{l=T} \alpha_l \pi_l^{true} = \alpha_k \pi_k^{true} \tag{13}$$

Posons pour tout $k \in [\![0, T]\!]$ posons $p_k = \alpha_k \pi_k^{true}$ et considérons la matrice $A_{kl} = \pi_k^{obs}$. Le problème se ré-écrit sous sa forme vectorielle (linéaire) selon $A\mathbf{p} = \mathbf{p}$.

- Notons que tous les coefficients A_{kl} de la matrice sont strictement positifs.
- On a $\forall l \in [\![0, T]\!]$ le fait que $\sum_{k=0}^{k=T} A_{kl} = \sum_{k=0}^{k=T} \pi_k^{obs} = 1$ (On dit alors que la matrice A est une matrice stochastique (voir J.E. Rombaldi). Cela suffit à prouver que $\rho(A) = 1$.

On se trouve dans les condition du théorème : pour $\alpha > 0$ fixé, il existe un unique vecteur $\mathbf{p}(\alpha)$ tel que :

$$\forall k \in [\![0, T]\!], \quad p_k(\alpha) \geq 0, \quad \sum_{k=0}^{k=T} p_k(\alpha) = \alpha, \quad A\mathbf{p}(\alpha) = \mathbf{p}(\alpha) \tag{14}$$

Comme la matrice A est de rang 1 (toutes ses colonnes sont identiques, égales au vecteur colonne $\boldsymbol{\pi}^{obs}$ dont la somme des compsantes vaut 1), alors le vecteur $\mathbf{p}(\alpha)$ que l'on recherche est nécessairement $\alpha\boldsymbol{\pi}^{obs}$. On en déduit alors que :

$$\forall k \in [\![0, T]\!], \quad \pi_k^{true} = \frac{\alpha}{\alpha_k} \pi_k^{obs} \tag{15}$$

et donc les π_k^{true} sont déjà positifs. Il ne reste plus qu'à trouver α pour imposer la condition de normalisation :

$$\sum_{k=0}^{k=T} \pi_k^{true} = 1 \Rightarrow \sum_{k=0} \frac{\alpha}{\alpha_k} \pi_k^{obs} = 1 \Leftrightarrow \frac{1}{\alpha} = \sum_{k=0}^{k=T} \frac{1}{\alpha_k} \pi_k^{obs} \tag{16}$$

Ainsi, si une solution du problème initial existe, elle est unique et s'écrit sous la forme précédente. Mais on vérifie réciproquement que la famille précédente vérifie l'équation puisque :

$$\sum_{l=0}^{l=T} \alpha_l \pi_l^{true} = \sum_{l=0}^{l=T} \alpha_l \frac{\alpha}{\alpha_l} \pi_l^{obs} = \alpha \tag{17}$$

de sorte que l'on vérifie immédiatement :

$$\frac{\alpha_k \pi_k^{true}}{\sum_{l=0}^{l=T} \alpha_l \pi_l^{true}} = \frac{1}{\alpha} \alpha_k \pi_k^{true} = \frac{1}{\alpha} \alpha_k \frac{\alpha}{\alpha_k} \pi_k^{obs} = \pi_k^{obs} \tag{18}$$

□

Comme on peut le voir, la preuve reste très élémentaire une fois connu le théorème de Perron-Frobenius. Mais il va de soi que pour des personnes normalement censées être des spécialistes mondiaux des questions d'épidémiologie, un théorème si utile en théorie des probabilité ne peut pas leur être inconnu. Surtout à 17 auteurs des laboratoires les plus prestigieux. Encore que, pour qu'ils en arrivent à retrouver des réponses élémentaires, au moins aurait-il fallu qu'ils se posent effectivement des questions élémentaires.

Afin de mieux comprendre le lien mathématique entre la distribution d'arrivée $\boldsymbol{\pi}^{true}$ en fonction de la distribution de départ π^{obs}, nous allons démontrer un certain nombre de propositions mathématiques, assez faciles, qui nous permettront de dissiper, pour toujours nous l'espérons, les illusions des auteurs quant à l'interprétation absurde qu'ils ont bien voulu nous servir de ces distributions. En effet, pourquoi les exposants π^{obs} et π^{true} ? Parce que dans la tête des auteurs :

1. la densité de probabilité π^{true} est comprise comme la distribution représentant l'idéal du phénomène que l'on souhaite observer ;
2. la quantité π^{obs} ne constitue qu'une forme d'observation empirique de cet idéal.

On peut, à ce titre, véritablement reprendre la célèbre métaphore de la caverne pour essayer de comprendre les inconscients épistémologiques des auteurs. Il y aurait l'Idée de la densité de probabilité, le fameux $\boldsymbol{\pi}^{true}$ et son image déformée que l'on perçoit dans la réalité, le $\boldsymbol{\pi}^{obs}$. Ainsi les auteurs prétendent-ils cependant, grâce à un bête système linéaire (celui que l'on vient d'étudier), retrouver l'Idée de la probabilité à partir de son observation déformée. Pour vous donner une représentation : imaginer que vous jouiez aux dés. Vous réalisez 1000 lancers. Vous observez la fréquence d'apparition des faces

$$\{0.14, 0.18, 0.16, 0.17, 0.13, 0.22\}$$

Vous vous dites en fait que ce sont là les valeurs "observées" des "vraies" probabilités qui doivent être, dans l'hypothèse "d'un dé équilibré" :

$$\{1/6, 1/6, 1/6, 1/6, 1/6, 1/6\}$$

Il y a l'Idée de la probabilité et son image déformée dans la réalité. Bon, voilà en gros l'esprit des auteurs lorsque, dans leur modèle, ils ont introduit les deux

variables $\boldsymbol{\pi}^{obs}$ et $\boldsymbol{\pi}^{true}$. Retrouver le monde des Idées de Platon : voilà bien le programme "scientifique" qu'ils espèrent atteindre grâce aux "mathématiques". En réalité, ils ne comprennent strictement rien aux outils mathématiques. Pour être un peu plus concret, dans la tête des auteurs, la différence entre $\boldsymbol{\pi}^{obs}$ et $\boldsymbol{\pi}^{true}$ doit être en quelque sorte la réalisation d'un bruit de mesure, sans doute aléatoire et que l'on pourrait pourquoi pas modéliser comme la réalisation d'un bruit blanc (c'est-à-dire un processus stochastique gaussien).

Ils espéraient que le système d'équations qu'ils ont posé entre les deux distributions permette de reconstruire la densité "théorique" par rapport à la densité "observée". Autrement dit, ils ont pris leur système linéaire comme un processus de régularisation, consistant à enlever d'une observation ce que l'on estime être un bruit gaussien purement stochastique. Notez qu'il existe effectivement, dans les outils de traitement du signal, des procédés pour analyser puis éliminer les bruits gaussiens dans les mesures observées. Évidemment qu'elles n'ont rien à voir avec ce qu'ont fait les auteurs. Donc nous allons prouver deux résultats sur la différence entre $\boldsymbol{\pi}^{obs}$ et $\boldsymbol{\pi}^{true}$.

Proposition 2. *Quelle que soit la distribution $\boldsymbol{\pi}^{obs}$ alors il existe $k_0 \in [\![0, T-1]\!]$ tel que :*

1. *pour tout k de l'intervalle entier $[\![0, k_0]\!]$ la différence*

$$\left(\pi^{obs}(k) - \pi_k^{true} : k \in [\![0, T]\!]\right)$$

   est positive et pour tout cas de $[\![k_0 + 1, T]\!]$ cette différence est négative.

2. *Si la distribution $\pi^{obs}(k)$ est décroissante, alors la suite*

$$\left(\pi^{obs}(k) - \pi_k : k \in [\![0, T]\!]\right)$$

   est positive et décroissante sur $[\![0, k_0]\!]$

Démonstration. calculons la différence

$$\pi^{obs}(k) - \pi_k^{true} = \left(1 - \frac{\alpha}{\alpha_k}\right) \pi_{obs}(k) \tag{19}$$

La suite des $1/\alpha_k$ est croissante et $1/\alpha$ est l'espérance (i.e. la moyenne pondérée pour la distribution des $\left(\pi^{obs}(k) : k \in [\![0, T]\!]\right)$ des $1/\alpha_k$. Par conséquent, on a :

$$\frac{1}{\alpha} = \sum_{k=0}^{k=T} \frac{1}{\alpha_k} \pi_k^{obs} \tag{20}$$

$$\Rightarrow \quad \frac{1}{\alpha} \leq \frac{1}{\alpha_T} \sum_{k=0}^{k=T} \pi_k^{obs} = \frac{1}{\alpha_T} \tag{21}$$

$$\Rightarrow \quad \frac{1}{\alpha} \geq \frac{1}{\alpha_0} \sum_{k=0}^{k=T} \pi_k^{obs} = \frac{1}{\alpha_0} \tag{22}$$

on a donc très facilement :

$$\frac{1}{\alpha_0} \leq \frac{1}{\alpha} \leq \frac{1}{\alpha_T} \tag{23}$$

Par conséquent, comme la suite $\frac{1}{\alpha_k}$ est croissante, il existe $k_0 \in [0, T-1]$ tel que :

$$\frac{1}{\alpha_{k_0}} \leq \frac{1}{\alpha} \leq \frac{1}{\alpha_{k_0+1}} \tag{24}$$

Donc on voit immédiatement que l'on a :

$$\forall k \in [\![0, k_0]\!],\ \left(1 - \frac{\alpha}{\alpha_k}\right) \pi_{obs}(k) \geq 0 \tag{25}$$

$$\forall k \in [\![k_0 + 1, T]\!],\ \left(1 - \frac{\alpha}{\alpha_k}\right) \pi_{obs}(k) \leq 0 \tag{26}$$

ce qui répond à la première partie de la proposition. Plaçons-nous maintenant dans l'intervalle entier $[\![0, k_0]\!]$. On a :

$$sur\ \ [\![0, k_0]\!],\ \ \left(\frac{1}{\alpha} - \frac{1}{\alpha_k}\right)\ \ est\ positive\ et\ décroissante \tag{27}$$

$$sur\ \ [\![0, k_0]\!],\ \ \pi^{obs}(k)\ \ est\ positive\ et\ décroissante \tag{28}$$

par conséquent, sur $[\![0, k_0]\!]$ le produit des deux suites précédentes nous donne bien une suite décroissante positive. □

Ainsi, ce résultat nous donne-t-il effectivement l'intuition que la différence $\boldsymbol{\pi}^{obs} - \boldsymbol{\pi}^{true}$ ne pourra en aucun cas être la réalisation d'un bruit blanc : car si $\boldsymbol{\pi}^{obs}$ est une observation "empirique" de $\boldsymbol{\pi}^{true}$, alors la distribution des signes (sur la différence des valeurs) devrait être complètement "aléatoire", c'est-à-dire n'avoir aucune structure particulière. L'exact opposé de ce que nous dit le théorème. De plus, si la distribution $\boldsymbol{\pi}^{obs}$ est décroissante (ce qui a toutes les chances d'être le cas), on ne comprendrait pas pourquoi la différence est d'abord strictement positive et décroissante.

Un autre résultat va maintenant enfoncer définitivement le clou :

Proposition 3. *On a $\boldsymbol{\pi}^{obs} = \boldsymbol{\pi}^{true}$ si et seulement si $\forall k \in [\![0, T]\!],\ \alpha_k = \alpha$*

Nous verrons à quel type de données cela pourrait correspondre pour les auteurs de l'étude. Encore une fois, ici, si les $\boldsymbol{\pi}^{obs}$ étaient une observation des $\boldsymbol{\pi}^{true}$, il ne pourrait y avoir aucun cas où les deux distributions sont parfaitement égales.

Démonstration. Évident puisque la relation donnée entre $\boldsymbol{\pi}^{true} = \boldsymbol{\pi}^{obs}$ est telle que l'on a :

$$\forall k \in [\![0, T]\!],\ \pi_k^{true} = \frac{\alpha}{\alpha_k} \pi^{obs} \tag{29}$$

□

En fait, tous ces résultats sont assez superfétatoires s'agissant de montrer que le lien unissant $\boldsymbol{\pi}^{obs}$ à $\boldsymbol{\pi}^{true}$ n'est pas un lien de "théorie" (ou vérité) à "expérience" (ou observation). En effet, comme n'importe qui le remarque de façon immédiate, le lien entre $\boldsymbol{\pi}^{obs}$ et $\boldsymbol{\pi}^{true}$ est un système linéaire sous contraintes qui admet une unique solution. Ce qui veut dire qu'il y a un déterminisme parfait entre les deux : c'est la distribution des $(\alpha_k : k \in [\![0, T]\!])$ qui détermine complètement ce lien. Bien sûr, si l'on imagine ensuite que les α_k sont des réalisations d'une certaine loi de probabilité (par exemple qu'ils constituent eux-mêmes un bruit blanc), alors, effectivement, on pourra sûrement vérifier que la différence entre les $\boldsymbol{\pi}^{obs}$ et les $\boldsymbol{\pi}^{true}$ a quelque chose de stochastique. Mais ce n'est évidemment pas ce que veulent signifier les auteurs. Nous allons comprendre que cette dénomination entre $\boldsymbol{\pi}^{obs}$ et $\boldsymbol{\pi}^{true}$ vient en fait d'une erreur liée à leur incapacité profonde de raisonner sur un problème parfaitement élémentaire de probabilité.

2.2 Présentation du faux raisonnement des auteurs

Et lorsque nous disons élémentaire, c'est vraiment élémentaire : cela s'inscrit parfaitement dans le cadre des programmes du lycée. Et les plus sérieux de nos élèves n'auraient pas fait l'erreur qui a été commise par les auteurs. Nous allons pour cela présenter le raisonnement utilisé dans l'article. Reprenons le problème : les auteurs disposent d'un certain nombre d'observations. Pour chaque jour, ils connaissent les personnes hospitalisées. Leur objet d'intérêt c'est de définir "la probabilité de subir un événement (la mort D ou l'admission en service de réanimation intensif ICU après k jours d'hospitalisation"). Ils notent alors par π_k^{true} cette "probabilité". Grâce à cet "objet" (la "probabilité"), les auteurs se proposent de dénombrer, parmi leurs données, le nombre de personnes hospitalisées qui ont effectivement subi l'issue $\beta \in \{D, ICU\}$ (notons que pour chaque issue, il y a un problème séparé de probabilité même si les deux problèmes sont analogues au début). Ils proposent le raisonnement suivant :

1. soit le patient est arrivé le jour 0 et il a subi β après k jours ;
2. soit le patient est arrivé le jour 1 et il a subi β après k jours ;
3. etc.
4. soit le patient est arrivé le jour $m-k$ (on n'a pas de données après le jour m : on imagine que m est le dernier jour pour lequel on a les informations, ce qui est le cas lorsque l'on étudie une épidémie qui est toujours en cours de déroulement) et il a subi β après k jours.

Si l'on note pour chaque jour $j \in [\![0, m]\!]$ par H_j le nombre de personnes hospitalisées ce jour-là, on voit donc, en utilisant la "probabilité" π_k^{true} et le dénombrement précédent sur les dates d'hospitalisation, que l'on peut déduire le nombre total de personnes ayant eu à subir l'événement β par :

$$N_\beta(k) = \sum_{j=0}^{j=m-k} H_j \pi_k^{true} \tag{30}$$

autrement dit, pour chaque date d'hospitalisation, on peut connaître le nombre de personnes qui subiront l'issue β au bout de k jours : il suffit de multiplier le nombre de personnes hospitalisées à la date indiquée par la probabilité de subir l'issue β au bout de k jours. En sommant alors sur toutes les dates possibles, on a le nombre total de personnes ayant subi l'issue k. À partir de la connaissance des $(N(k) : k \in [\![0, T]\!])$, on peut compter *toutes* les personnes qui ont eu à subir l'issue β depuis le jour $j = 0$: il suffit de sommer tous les individus $N_\beta(k)$, pour tous les jours k possibles séparant l'entrée à l'hôpital et l'expérience de l'issue β. On suppose que si un patient doit subir l'issue k, cela est possible entre $k = 0$ et $k = T$: autrement dit, les patients ayant expérimenté l'issue β (mort, soin intensif) ne l'ont jamais fait plus de T jours après leur hospitalisation et peuvent l'expérimenter dès qu'ils arrivent à l'hôpital. Si N_β désigne le nombre total de personnes ayant eu l'issue β, lorsque l'on fait le bilan au jour m, on obtient alors un nombre N_β donné par :

$$N_\beta = \sum_{k=0}^{k=T} N_\beta(k) = \sum_{k=0}^{k=T} \sum_{j=0}^{j=m-k} H_j \pi^{true} \tag{31}$$

Or, que nous disent nos auteurs ? Qu'ils sont capables, par l'étude de tous les dossiers faisant l'objet d'un enregistrement dédié, de déterminer $N_\beta, N_\beta(k)$. Ils

appelleront alors $\left(\pi_k^{obs} : k \in [\![0, T]\!]\right)$ ces rapports d'"observations empiriques" des choses, liés à l'objet idéalisé $\boldsymbol{\pi}^{true}$. Ainsi, ils en déduisent que l'on doit former le système :

$$\frac{N_\beta(k)}{N_\beta} := \pi_k^{obs} = \frac{\sum_{j=0}^{j=m-k} H_j \pi_k^{true}}{\sum_{k=0}^{k=T} \sum_{j=0}^{j=m-k} H_j \pi_k^{true}} \tag{32}$$

Ainsi, si l'on pose par définition :

$$\alpha_k = \sum_{j=0}^{j=m-k} H_j \tag{33}$$

on voit que l'on retombe exactement sur l'équation que nous avons étudiée au début de ce document. En effet, on doit vérifier :

$$\sum_{k=0}^{k=T} \pi_k^{obs} = \sum_{k=0}^{k=T} \frac{N_\beta(k)}{N_\beta} = 1, \quad \pi_k^{obs} = \frac{N_\beta(k)}{N_\beta} > 0 \tag{34}$$

d'autre part, par hypothèse même sur π_k^{true}, puique c'est une densité de probabilité, on doit évidemment vérifier :

$$\pi_k^{true} > 0, \quad \sum_{k=0}^{k=T} \pi_k^{true} = 1 \tag{35}$$

avant d'aller plus loin, interprétons la manière avec laquelle on pourrait avoir $\boldsymbol{\pi}^{obs} = \boldsymbol{\pi}^{ture}$. On sait que la condition nécessaire et suffisante pour qu'il en soit ainsi c'est que :

$$\forall k \in [\![0, T]\!], \quad \alpha_k = \alpha \tag{36}$$

or, on voit ici que α_k représente le nombre cumulé de personnes ayant été hospitalisées à la date $m - k$. Par conséquent, pour que cette suite soit constante, il est donc nécessaire que l'on ait :

$$H_0 = N_\beta, \quad \forall j > 0, H_j = 0 \tag{37}$$

Autrement dit, on aura égalité, pour les auteurs, entre $\boldsymbol{\pi}^{obs}$ et $\boldsymbol{\pi}^{true}$ si et seulement si tous les patients arrivent au jour 0 et ensuite plus aucun patient n'arrive. On voit donc encore une fois ici que le $\boldsymbol{\pi}^{true}$ n'a pas grand-chose à voir avec une donnée "régularisée" (non bruitée) d'un $\boldsymbol{\pi}^{obs}$. La différence entre $\boldsymbol{\pi}^{true}$ et $\boldsymbol{\pi}^{obs}$ est alors liée de façon déterministe à la manière avec laquelle les patients subissant l'issue β arrivent à l'hôpital.

2.3 Le défaut de compréhension des probabilités en tant que langage mathématique

Le "raisonnement" précédent, celui "comptant" le nombre de patients subissant l'événement β, c'est celui que ferait à peu près n'importe quel élève de lycée, l'homme de la rue, n'importe qui sachant manipuler l'idée de proportionnalité. Évidemment, cela n'a aucun sens mathématique. Et c'est précisément pour éviter d'avoir à faire ce genre de raisonnement totalement faux que depuis plusieurs

siècles on développe la théorie mathématique des probabilités. Les probabilités ont deux inconvénients majeurs pour le commun des mortels : d'une part, ce sont des objets mathématiques QUI NE DOIVENT RIEN À L'INTUITION, d'autre part ces objets parlent de MESURE SUR DES SOUS-PARTIES D'UN ENSEMBLE. Ainsi, les probabilités ont réellement commencé à devenir précises lorsqu'on les a introduites dans un formalisme mathématique via des fonctions définies sur des sous-parties d'un ensemble. Ce que je rappelle toujours à mes étudiants :

ON NE PARLE JAMAIS DE PROBABILITÉS SI L'ON N'A PAS D'ABORD DIT À QUEL ENSEMBLE Ω CETTE PROBABILITÉ ÉTAIT LIÉE. SINON, CELA REVIENT À CONSIDÉRER DES PROBABILITÉS QUI SERAIENT UNIVERSELLES INDÉPENDANTES DE TOUTE AUTRE FORME DE CONSIDÉRATION.

Lorsque Ω est discret (c'est-à-dire lorsque l'on peu dire qu'il y a un nombre fini d'individus sur lesquels on veut travailler, c'est le cas de toutes les considérations que l'on fait lorsque l'on s'intéresse à des populations), on note par $\mathcal{P}(\Omega)$ l'ensemble de toutes les sous-parties de Ω. De fait, la définition mathématique d'une probabilité se fait en considérant une application :

$$p : \mathcal{P}(\Omega) \to [0,1] \tag{38}$$

qui vérifie (à peu près) les deux propriétés suivantes :

$$p(\Omega) = 1, \quad A \bigcap B = \emptyset \Rightarrow p\left(A \bigcup B\right) = p(A) + p(B) \tag{39}$$

(en réalité, dans la deuxième propriété, on doit plutôt, dans la définition, s'intéresser à une réunion infinie d'ensembles deux à deux disjoints). On peut voir ainsi une probabilité comme la mesure d'un sous-ensemble rapportée à la mesure de l'ensemble total. Si deux sous-ensembles n'ont pas d'éléments en commun, alors la mesure de leur réunion, c'est la somme de leur mesure (propriété de cardinalité).

Le détour par ces rappels via le formalisme mathématique est là pour appeler chacun à la prudence : *il n'existe pas de "probabilité en soi"*. Si vous commencez à parler d'une probabilité sans dire à quel ensemble elle se rattache, vous êtes hors du champ de la science. C'est comme si vous disiez : il existe une probabilité naturelle, intrinsèque, d'avoir 40 ans. Ça n'a pas de sens. En revanche, dans une population donnée, cela peut faire sens de définir une probabilité d'avoir 40 ans. On la définira comme le nombre de personnes ayant 40 ans divisé par le nombre total de personnes. Cela n'a aucun sens pour les personnes qui font proprement des mathématiques de parler de probabilité sans parler de l'ensemble sur lequel on définit cette probabilité. Si vous ne faites pas d'abord ça, tout ce que vous ferez par la suite, si vous ne posez pas les choses formellement, n'aura aucun sens. C'est exactement ce qu'ont fait les auteurs : leur approche des probabilités ne dépasse pas celle de l'homme de la rue (pour lequel j'ai beaucoup d'estime mais pas forcément en tant que scientifique). **Ainsi, l'objet "$\boldsymbol{\pi}^{true}$" des auteurs est-il une chimère : il n'a aucune existence scientifique car il n'a aucune existence mathématique**. C'est exactement pour cela que les

raisonnements qui vont être utilisés n'ont aucune validité scientifique : ce ne sont que de grossiers discours sans rigueur. Pour finir, remarquons que dans le "raisonnement" des auteurs, même en acceptant l'idée qu'il y aurait existence d'une probabilité en soi, le dénombrement était faux, car si l'on voulait alors compter les individus qui avaient subi l'événement β, il ne fallait pas s'intéresser au nombre total d'individus qui sont hospitalisés le jour j, mais seulement à la proportion de ces individus qui ont subi par la suite l'issue β. Autrement dit, il fallait remplacer le H_j (nombre de patients hospitalisés au jour j) par un H_j^β (nombre d'hospitalisations du jour j qui vont ensuite expérimenter l'événement β).

Comme nous l'avons dit, les probabilités s'attachent à donner des mesures "normalisées"' (c'est-à-dire en quelque sorte une valeur que l'on peut voir comme une proportion) de certains sous-ensembles d'un ensemble Ω. Mais encore faut-il pouvoir décrire proprement les sous-ensembles qui ont un intérêt. Pour cela, il faut introduire la notion de variable (aléatoire) associée aux individus qui composent votre ensemble. Qu'est-ce que cela ?

En fait, dans une population donnée, chaque individu possède sa "fiche" caractéristique, par exemple : son année de naissance, l'argent qu'il gagne chaque année, le nombre d'enfants qu'il a, etc. Ces données chiffrées, une fois stockée l'information dans un fichier, peuvent être ressorties pour chaque individu de la population qui nous intéresse. On définit une variable aléatoire comme étant une application qui, à chaque individu, associe une valeur numérique intéressante (âge, revenus, enfants, etc.). Remarquez que dans la locution "variable aléatoire"', le mot aléatoire n'a aucune interprétation intuitive. On peut plutôt dire qu'il s'agit d'un vocabulaire historique. D'un point de vue mathématique, une variable aléatoire (v.a. en notation abrégée) est exactement une application (on dit aussi une fonction). Comme pour toute fonction, il faut dire à quel ensemble image appartient le résultat. S'agissant par exemple de votre année de naissance, on sait qu'il s'agit d'un nombre entier que l'on peut raisonnablement situé après 1900. On dit que la v.a. est entière lorsque elle ne prend que des valeurs entières. Dans ce qui nous intéresse, les seules variables aléatoires seront entières. Non seulement elles ne pourront pas prendre de valeur négative mais pour chacune d'entre elles, elles ne pourront pas non plus dépasser une valeur limite (ainsi, si on s'intéresse à votre âge, manifestement cela ne peut pas dépasser disons 120). Les variables aléatoires sont des fonctions que l'on notera par des lettres latines en lettres capitales $(X, Y, Z, \cdots)$. Ainsi, on utilisera la notation :

$$X : \Omega \to [\![0, M]\!] \tag{40}$$

Ainsi, pour $\omega \in \Omega$ (pour un individu - un élément - de la population - de l'ensemble - Ω), la notation $X(\omega)$ désigne la valeur d'intérêt associée à ω. Une telle valeur est entière et vérifie l'inégalité $0 \leq X(\omega) \leq M$.

Maintenant que l'on est armé de la notion de variable aléatoire, on va parler d'une chose importante, celle d'image inverse. L'image directe c'est : je me donne un individu, et je lui associe sa valeur d'intérêt (i.e. son âge dans l'ensemble d'arrivée). L'image réciproque pose le problème inverse : si je me donne une valeur spécifique de l'ensemble d'arrivée (par exemple 40 ans), je vais asso-

cier à cette valeur l'ensemble des individus de la population qui ont précisément, pour la variable d'intérêt, la valeur donnée. Si $X : \Omega \mapsto [\![0, M]\!]$, alors on a :

$$\forall q \in [\![0, M]\!], \quad X^{-1}(q) = \{\omega \in \Omega \ tel \ que \ \ X(\omega) = q\} \subset \Omega \tag{41}$$

autrement dit, l'image réciproque *prend une valeur de l'espace d'arrivée et renvoie un sous-ensemble de l'espace* Ω. Ainsi, il est important de noter que l'image inverse $X^{-1}(\cdot)$ est définie sur $[\![0, M]\!]$ mais elle est à valeur dans $\mathcal{P}(\Omega)$ (et non pas dans Ω). On utilise souvent la notation suivante :

$$X^{-1}(k) \quad est \ aussi \ notée \ par \quad \{X = k\} \tag{42}$$

ainsi, pour le répéter une nouvelle fois, la notation $\{X = k\}$ désigne le sous-ensemble de la population Ω dont les individus partagent la valeur k vis-à-vis de la variable (aléatoire) X (âge, niveau de revenu, nombre d'enfants, etc.).

On peut, évidemment, pour une même population, avoir envie de considérer plusieurs types d'informations. On sera alors amené à considérer plusieurs variables aléatoires $X, Y, \cdots$. De sorte que l'on pourra avoir des sous-ensembles décrits de la façon suivante :

$$\{X = k\} \bigcap \{Y = l\} \tag{43}$$

qui sera ainsi l'ensemble des personnes ayant la valeur k pour la variable X (par exemple X désigne l'âge et k désigne la valeur 40) et la valeur l pour la variable Y (par exemple Y désigne le nombre d'enfants et $l = 2$). Alors on interprète $\{X = k\} \bigcap \{Y = l\}$ comme étant l'ensemble des personnes ayant 40 ans et deux enfants.

Pour finir, il faut maintenant se donner une manière concrète de calculer les probabilités sur l'ensemble $\mathcal{P}(\Omega)$. On choisira la probabilité dite uniforme, c'est-à-dire que pour une sous-partie de $B \subset \Omega$, on calculera :

$$p(B) := \frac{\texttt{Card}(B)}{\texttt{Card}(\Omega)} \tag{44}$$

où $\texttt{Card}(\cdot)$ compte le nombre d'éléments qu'il y a dans un sous-ensemble. C'est la manière la plus naturelle et évidente de définir la probabilité sur l'ensemble $\mathcal{P}(\Omega)$ quand on étudie les populations. Concrètement, cela signifie que chaque individu "a le même poids" statistique lorsque l'on cherche ensuite à établir certaine valeur moyenne par exemple sur l'âge, les revenus, le nombre d'enfants. Ainsi, pour faire l'âge moyen d'une population, on somme les âges de chacun des individus et on le divise par le nombre total des individus. Ainsi, on a :

$$\forall \omega \in \Omega, \quad p(\{\omega\}) = \frac{1}{N}, \quad N = \texttt{Card}(\Omega) \tag{45}$$

où $\{\omega\}$ désigne le singleton associé à ω, c'est-à-dire le sous-ensemble n'ayant qu'un seul élément qui est ω.

2.4 Comprendre l'erreur des auteurs - essayer interpréter leur équation

Si nous avons pris le temps de rappeler tous ces éléments parfaitement évidents du point de vue des probabilités et des mathématiques (en principe largement accessibles pour les étudiants et les élèves), c'est que, hélas, nos auteurs

ne sont même plus capables de faire proprement ce genre de travaux préliminaires de modélisation. Ils pensent faussement les comprendre et les maîtriser, ou alors ils sont sincèrement convaincus que c'est trop facile pour eux. En réalité, ils en sont incapables. Incapables donc de modéliser proprement une situation qui est le degré zéro de la modélisation en probabilité. Pour commencer donc, définissons les deux ensembles d'intérêts. Il y a deux problèmes, parfaitement identiques dans leurs modélisations mais pas dans les ensembles auxquels ils sont attachés :

1. un problème concernant le nombre de jours après lesquels on subit le placement en soin intensif (ICU) ;
2. un problème concernant le nombre de jours après lesquels on meurt (D).

Ω_D désignera l'ensemble des personnes qui sont mortes à l'hôpital et l'ensemble Ω_{ICU} désignera l'ensemble des personnes qui ont été placées en soins intensifs. On notera par $\Omega_\beta,\ \beta \in \{D, ICU\}$ ces ensembles. On note $N_\beta = \texttt{Card}\,(\Omega_\beta)$ le nombre total d'individus dans l'ensemble Ω_β. On note par $j = 0$ le jour pour lequel un premier patient a été hospitalisé. On suppose que l'on a des observations jusqu'à $j = m$. Pour chacun de ces ensembles, on définira les variables aléatoires suivantes :

1. la variable aléatoire $E : \Omega_\beta \to [\![0, m]\!]$ désigne le jour où le patient subit l'événement β ;
2. la variable aléatoire $X : \Omega_\beta \to [\![0, T]\!]$ désigne le nombre de jours après hospitalisation au bout desquels l'individu subit l'événement β.

Ainsi, grâce à ces deux variables aléatoires, on peut reconstruire la date d'entrée à l'hôpital par $A = E - X$. Il est clair, par définition, que l'on a :

$$\forall \omega \in \Omega, \quad 0 \leq A\,(\omega) \leq m \tag{46}$$

On va maintenant essayer de comprendre l'erreur des auteurs dans leur calcul. Avant cela, on rappelle la notion de famille totale de sous-ensembles.

Definition 1. *Soit $B_1, \cdots, B_n$ une famille de sous-parties de Ω_β. On dit que c'est une famille totale dans Ω_β lorsque les sous-ensembles n'ont aucun élément en commun et que leur réunion vaut Ω entier. Autrement dit :*

$$\forall i \neq j \in [\![1, n]\!],\ B_i \bigcap B_j = \emptyset, \quad B_1 \bigcup B_2 \bigcup \cdots \bigcup B_n = \Omega \tag{47}$$

Énonçons maintenant la loi des probabilités totales :

Proposition 4. *Soit $B_1, \cdots, B_n$ une famille totale. Alors pour tout sous-ensemble $A \subset \Omega_\beta$ on a :*

$$P\,(A) = \sum_{i=1}^{i=n} P\left(A \bigcap B_i\right) \tag{48}$$

On va maintenant pouvoir à peu près essayer de recouvrer le calcul des auteurs. C'est-à-dire comprendre le sens de $\boldsymbol{\pi}^{true}$ en repartant du comptage qu'ils ont essayé d'effectuer. Il est tout à fait clair que l'on a par définition :

$$P\,(\{X = k\}) = \frac{N_\beta\,(k)}{N_\beta} \tag{49}$$

c'est en réalité ce que les auteurs appellent $\left(\pi_k^{obs} : k \in [\![0, T]\!]\right)$. Or, il est également évident que la famille :

$$\{E = j\} : j \in [\![0, m]\!] \tag{50}$$

constitue une famille totale de Ω. Concrètement, cela signifie qu'un patient ayant subi l'issue β, l'a forcément subie un certain jour compris entre 0 et m et que s'il subit l'événement au jour j, il ne peut pas l'avoir subi au jour $j' \neq j$. On applique maintenant la loi des probabilités totales :

$$P(\{X = k\}) = \sum_{j=0}^{j=m} P\left(\{X = k\} \bigcap \{E = j\}\right) \tag{51}$$

À ce stade, on ne sait pas calculer la probabilité du sous-ensemble suivant :

$$\{X = k\} \bigcap \{E = j\}$$

Cependant, une chose est certaine : si vous subissez l'événement β à la date j et que l'on sait que vous subissez cet événement k jours après votre date d'entrée à l'hôpital, c'est que vous êtes rentré à la date $j - k$. On a donc :

$$\{X = k\} \bigcap \{E = j\} \subset \{E - X = j - k\} \tag{52}$$

notons que l'on a en outre $j - k < 0 \Rightarrow \{E - X = j - k\} = \emptyset$. De l'inclusion précédente, on tire alors :

$$P\left(\{X = k\} \bigcap \{E = j\}\right) \leq P(\{E - X = j - k\}) \tag{53}$$

(en effet, d'une façon générale, les probabilités sont des fonctions croissantes pour la relation d'inclusion, c'est-à-dire que si $A \subset B$ alors $P(A) \leq P(B)$). On va donc poser :

$$\frac{P(\{X = k\} \bigcap \{E = j\})}{P(\{E - X = j - k\})} = \eta_{jk} \ si \ \ P(\{E - X = j - k\}) > 0 \tag{54}$$

$$\eta_{jk} = 0 \ \ pour \ \ P(\{E - X = j - k\}) = 0 \tag{55}$$

où η_{jk} est un paramètre de l'intervalle réel $[0, 1]$. La "subtilité" vient ici du fait que si $E(\omega) = j$ et $X(\omega) = k$, alors on a forcément $(E - X)(\omega) = j - k$, mais que l'on peut avoir $(E - X)(\omega) = j - k$ sans que nécessairement on ait $E(\omega) = j$ et $X(\omega) = k$. Ainsi, si l'on trouve ω tel que $E(\omega) = j + 1$ et $X(\omega) = k = k + 1$, la différence reste égale à $j - k$, mais la date de l'issue et la durée sont différentes. Notons maintenant, pour faire comme les auteurs, par H_{j-k}^{β} le nombre de personnes ayant subi l'issue β et hospitalisées à la date $j - k \geq 0$. Par définition des probabilités sur Ω_β, on a alors :

$$\forall j - k \geq 0, \ \ P(\{E - X = j - k\}) = \frac{H_{j-k}^{\beta}}{N_\beta} \tag{56}$$

On en tire anisi :

$$P\left(\{X = k\} \bigcap \{E = j\}\right) = \eta_{jk} \frac{H_{j-k}^{\beta}}{N_\beta}, \ \ j - k \geq 0 \tag{57}$$

d'où l'on tire finalement :

$$P(X=k) = \sum_{j=k}^{j=m} \eta_{jk} \frac{H_{j-k}^{\beta}}{N_\beta} \tag{58}$$

en effet, seuls les jours $j \geq k$ interviennent dans le comptage. Or il est clair que :

$$\sum_{j=k}^{j=m} H_{j-k}^{\beta} = \alpha_k > 0 \tag{59}$$

la quantité α_k s'interprète comme le nombre cumulé de personnes hospitalisé entre les jours 0 et $m-k$ (cette suite est donc décroissante) α_0 représente le nombre total de personnes hospitalisées, tandis que $\alpha_T = H_0$ est le nombre de personnes rentrées au jour $j=0$. Ainsi, la durée k étant fixée, on sait que l'on peut trouver un nombre π_k^{true} tel que l'on ait :

$$\sum_{j=k}^{j=m} \eta_{jk} \frac{H_{j-k}^{\beta}}{N_\beta} = \left(\sum_{j=k}^{j=m} \frac{H_{j-k}^{\beta}}{N_\beta} \right) \pi_k^{true} = \frac{\alpha_k}{N_\beta} \pi_k^{true} \tag{60}$$

Or, il est clair également que la famille des $(\{X=k\} : k \in [\![0,T]\!])$ est une famille totale de Ω. Ainsi on a :

$$\sum_{k=0}^{k=T} P(\{X=k\}) = 1 \tag{61}$$

ce qui veut dire bien entendu que l'on a :

$$\sum_{k=0}^{k=T} \frac{\alpha_k}{N_\beta} \pi_k^{true} = 1 \Leftrightarrow \sum_{k=0}^{k=T} \alpha_k \pi_k^{true} = N_\beta \tag{62}$$

On peut ainsi remplacer N_β par la somme précédente. On obtient au final :

$$\forall k \in [\![0,T]\!], \quad P(\{X=k\}) = \frac{\alpha_k \pi_k^{true}}{\sum_{l=0}^{l=T} \alpha_l \pi_l^{true}} \tag{63}$$

ainsi, en reprenant leur notation $\pi_k^{obs} = P(\{X=k\})$, on retrouve bien l'équation qu'ils ont eux-mêmes posée.

La question que l'on se pose est alors de savoir si les $(\pi_k : k \in [\![0,T]\!])$ sont eux-mêmes une densité de probabilité. La réponse est non en général. Plus précisément, on a la proposition suivante :

Proposition 5. *La famille $(\pi_k^{true}, k \in [\![0,T]\!])$ est une densité de probabilité si et seulement si les α_k sont tous égaux à N_β. C'est-à dire :*

$$\sum_{k=0}^{k=T} \pi_k^{true} = 1 \Leftrightarrow \forall k \in [\![0,T]\!] \ \alpha_k = N_\beta \Leftrightarrow H_0 = N_\beta \quad H_j = 0, \ j > 0 \tag{64}$$

Démonstration. C'est élémentaire puisque l'on a de façon immédiate :

$$\pi_k^{true} = \frac{N_\beta}{\alpha_k} P(X=k) \tag{65}$$

or la suite $(\alpha_k : k \in [\![0,T]\!])$ étant décroissante, supposons qu'il existe k_0 tel que $\alpha_{k_0} < N_\beta$. Alors pour tous les indices entiers $k \geq k_0$, on a $\pi_k^{true} > P(X=k)$. Comme on a forcément en outre $\pi_k^{true} \geq P(X=k)$, on en tire alors :

$$\sum_{k=0}^{k=T} \pi_k^{true} = \sum_{k=0}^{k=T} \frac{N_\beta}{\alpha_k} P(X=k) > \sum_{k=0}^{k=T} P(X=k) = 1 \tag{66}$$

En réalité, la seule possibilité pour que $\boldsymbol{\pi}^{true}$ soit une densité de probabilité, c'est que la suite $(\alpha_k : k \in [\![0,T]\!])$ soit constante, égale à N_β. On retrouve alors une situation déjà mise en avant, c'est-à-dire que $H_0^\beta = N_\beta$ et $H_j^\beta = 0, j > 0$. Autrement dit, tous les patients sont bien rentrés le même jour $j = 0$. □

La conclusion de cette étude, c'est qu'il existe effectivement un cadre probabiliste pour comprendre l'équation qui est proposée par les auteurs. Cependant, ce cadre ne permet aucune interprétation en termes de nouvelles probabilités , puisqu'il consiste juste à poser $\pi_k^{true} = N_\beta/\alpha_k \pi_k^{obs}$. *Pourtant, jouant de la nature de l'équation qui est invariante par homothéties sur les distributions, on peut toujours s'arranger pour trouver une solution probabiliste au problème.*

2.5 Comprendre la différence entre prénotion et science : le retour aux dés

Pour voir à quel point la probabilité introduite par les auteurs n'est qu'une pure chimère, il faut revenir à la situation des dés, que les rédacteurs ont intégrée de façon délirante dans leurs considérations. Tant que l'on ne fait rien de sérieux, au niveau du formalisme, on peut toujours dire qu'à chaque dé est associé une densité de probabilité "réelle" : celle de réaliser, lors d'un lancer, un chiffre annoncé à l'avance avec une probabilité "idéelle" de 1/6. Ainsi, cette probabilité existerait "dans la nature" et si l'on savait lancer les dés de façon "parfaite", il va de soi que c'est ce que rendrait comme conclusion n'importe quelle expérience s'attachant à "mesurer" ce qu'il en est de cette distribution de probabilité. Nous allons montrer sur cet exemple élémentaire que le $\boldsymbol{\pi}^{true}$ des dés est évidemment impossible à retrouver avec le système linéaire absurde que les auteurs essayent de reconstruire. Mieux, nous allons construire avec ces dés et leur $\boldsymbol{\pi}^{true}$ exactement le même raisonnement que les auteurs avec le temps d'hospitalisation avant de subir l'événement β. Pour cela, on imagine qu'une personne réalise des expériences de lancers de dés.

1. Chaque jour pour j (j variant de 0 à 9), le joueur réalise H_j lancers. (cela correspond au fait que chaque jour il y a des patients qui arrivent à l'hôpital et qui vont subir l'issue β).
2. Pour chaque jour j, et pour chaque numéro de face k, le joueur note par $N_j(k)$ le nombre de fois où le chiffre k est apparu le jour j.

Ainsi, à la fin des 10 jours, on a observé, pour chaque numéro de face k, un nombre d'apparitions $N(k)$ qui vaut simplement :

$$\forall k \in [\![1,6]\!],\ N(k) = \sum_{j=0}^{j=9} N_j(k) \tag{67}$$

Le nombre total de lancers, $N = 100$, on peut l'obtenir en disant que c'est la somme, pour tous les chiffres possibles compris entre 1 et 6, du nombre de fois où ces chiffres sont apparus. Ainsi, on retrouvera que l'on a :

$$N = \sum_{k=1}^{k=6} N(k) \tag{68}$$

Pour chaque chiffre k, on peut alors définir sa fréquence d'apparition (le fameux π^{obs}) par le rapport entre $N(k)$ et N. On a ainsi :

$$\pi_k^{obs} = \frac{N(k)}{N} \tag{69}$$

il est évident que π_k^{obs} est ainsi une densité de probabilité car tous les termes sont positifs et, par définition même de N, la somme sur ces valeurs vaut 1. Or, maintenant, que vont dire nos auteurs ? Puisqu'il existe une "vraie" probabilité d'obtenir le chiffre k, on peut s'attendre à ce que, chaque jour, le nombre $N_j(k)$ soit en fait donné par :

$$N_j(k) = H_j \pi_k^{true} \tag{70}$$

On reconstruit alors les données par :

$$N(k) = \sum_{j=0}^{j=10} H_j \pi_k^{true} = \left(\sum_{j=0}^{j=10} H_j \right) \pi_k^{true}, N = \sum_{k=1}^{k=6} N(k) = \sum_{k=1}^{k=6} \sum_{j=0}^{j=10} H_j \pi_k^{true} \tag{71}$$

Si l'on pose maintenant :

$$\forall k \in [\![1, 6]\!], \quad \alpha_k = \sum_{j=0}^{j=10} H_j = \alpha \tag{72}$$

on peut donc retrouver exactement l'équation des auteurs qui est :

$$\pi_k^{obs} = \frac{\sum_{k=1}^{k=6} \alpha_k \pi_k^{true}}{\sum_{l=1}^{l=6} \alpha_l \pi_k^{true}} \tag{73}$$

comme alors les α_k sont constants, on est dans la situation évidente où $\pi_k^{true} = \pi_k^{obs}$. Ainsi, si vous avez effectivement observé une distribution π^{obs} sous la forme

$$\{0.14, 0.18, 0.16, 0.17, 0.13, 0.22\}$$

et que dans votre idéal la densité de probabilité est effectivement

$$\{1/6, \cdots, 1/6\}$$

vous en arrivez immédiatement à la conclusion que :

$$0.14 = 1/6,\ 0.18 = 1/6,\ 0.16 = 1/6,\ 0.17 = 1/6,\ 0.13 = 1/6\ 0.22 = 1/6 \tag{74}$$

et donc par transitivité de l'égalité, vous en arrivez à :

$$0.14 = 0.18 = 0.16 = 0.17 = 0.13 = 0.22 \tag{75}$$

Mais, vous diront les auteurs de l'article : certes, ces égalités ne sont pas exactes, mais tout de même elles sont bien vérifiées de manière approchée... L'erreur dans ce raisonnement est évidente : le nombre $N_j(k)$ n'est jamais le nombre $H_j\pi_k^{true}$. Ce que l'on peut juste dire, c'est que $H_j\pi_k^{true}$ constitue un nombre rêvé : cela revient à prendre vos rêves pour des réalités. Quoi qu'il en soit, leur $\boldsymbol{\pi}^{true}$ est une grandeur sortie tout droit de leur imagination. Elle ne représente absolument rien. Elle n'est donc pas susceptible d'une modélisation mathématique. Faire des mathématiques, c'est sortir des prénotions (au sens de Durkheim) pour faire entrer la discussion dans la sphère scientifique. Soit l'exact opposé de ce que font les auteurs. Donc, lorsque vous partez d'un monde imaginaire, vous écrivez, sans aucune forme de rigueur, des équations qui n'ont plus aucun sens. C'est pour cette raison qu'il faut toujours repartir du formalisme. C'est ce que nous avons fait avec un ensemble précis, des variables aléatoires précises, et des formules dûment démontrées de la théorie des probabilités. Pour en arriver à la conclusion que π^{true} n'avait effectivement aucune consistance. Ainsi, il est parfaitement établi que les auteurs ne comprennent absolument rien à la modélisation et aux calculs des probabilités.

3 À la recherche d'un modèle

Dans la partie précédente, nous avons donc dissipé l'idée que les $\boldsymbol{\pi}^{true}$ puissent avoir une quelconque interprétation probabiliste précise. D'ailleurs, si pour la suite de la discussion nous allons quand même travailler avec le distribution $\boldsymbol{\pi}^{true}$, c'est juste pour respecter le parti des auteurs. Nous verrons *en pratique* (numérique) que la différence entre $\boldsymbol{\pi}^{true}$ et des $\boldsymbol{\pi}^{obs}$ est d'ailleurs assez marginale (tout de même de l'ordre de 9%). Évidemment, nous nous interdirons formellement d'utiliser, quant à nous, les $\boldsymbol{\pi}^{true}$.

3.1 À propos de l'estimation paramétrique

Une fois donc les $\boldsymbol{\pi}^{true}$ en poche - mais cela les auteurs n'ont pas même cherché à savoir si on pouvait les obtenir facilement car leur culture théorique est nulle -, on peut maintenant essayer de leur donner un "modèle". Cela revient à dire en gros que l'on cherche une fonction "continue" dont π_k^{true} serait une approximation discrète. Qu'est-ce que cela ?

En fait, dans les données fournies par les hôpitaux, seul le jour de l'issue (mort D ou admission en soins intensifs ICU) est manifestement pris en compte. Il est possible de penser que l'heure de la mort est donnée, car il faut manifestement prononcer le décès. S'agissant de l'entrée en soins intensifs, il est moins évident que l'heure d'entrée dans le service apparaisse dans les données. Bref, de toute façon, connaître le jour paraît largement suffisant. Mais pour nos auteurs, cela est d'une imprécision insupportable : ce qu'ils veulent savoir, ce n'est pas seulement le jour auquel vous êtes mort, mais bien l'heure, la minute et même la seconde à laquelle l'issue β vous est arrivée. La réponse à la question qu'ils cherchent c'est : combien d'issues β entre $10h30$ et $11h45$ après k jours d'hospitalisation ? Il y a peu de chance que cela puisse être pertinent pour comprendre le phénomène qui est décrit, mais bon, encore une fois, il s'agit pour les auteurs d'essayer de faire la démonstration de leur maîtrise des techniques ma-

thématiques. Il y a toujours un certain "prestige" à manipuler les probabilités continues plutôt que les probabilités discrètes. Surtout, on a beaucoup plus de "moyens" de modéliser le continu (plus de lois sont disponibles) que de le discret. Là encore, les auteurs vont faire preuve d'une manière sidérante de considérer les problèmes, montrant la pauvreté inouïe de leur culture mathématique.

L'idée des auteurs est la suivante : pour eux, chaque densité observée π_k^{true} est en fait la valeur moyenne d'une loi de probabilité observée du jour k au jour $k+1$. Ils supposent donc qu'il existe une "densité continue" de probabilités, qu'ils vont noter $f(t)$ telle que : la probabilité de subir l'événement β entre les instants $t > 0$ et $t + dt$ est $f(t)\,dt$ (dt désigne ici une durée très courte, que l'on dit souvent "infinitésimale" (au sens de Leibnitz)). Ainsi, le nombre de personnes subissant l'événement β entre t et $t + dt$ est donné par :

$$dN_\beta(t) = N_\beta f(t)\,dt \tag{76}$$

N_β étant toujours le nombre de personnes ayant subi l'événement β. Bien sûr, pour qu'une définition d'une telle sorte fasse sens, il faut au moins trouver une personne subissant l'événement β à chaque intervalle de temps dt. Ce qui signifie en gros que vous devez avoir un flux quasi continu de morts (D) ou de mises en soins intensifs (ICU) : heureusement que ce n'est quand même pas tout à fait le cas dans les hôpitaux. Mais à nouveau, peu importe aux auteurs la pertinence de ce qu'ils posent : il s'agit d'en imposer. Sachant que l'on a alors, par un formalisme d'intégration :

$$N_\beta = \int_{t=0}^{+\infty} dN_\beta(t) = N_\beta \int_{t=0}^{+\infty} f(t)\,dt \Leftrightarrow \int_{t=0}^{+\infty} f(t)\,dt = 1 \tag{77}$$

En principe, le comptage doit s'arrêter à $t = T$, et donc l'intégration aussi. Il faut supposer que, pour le phénomène étudié, ce qui pourrait provenir d'événements entre T et $+\infty$ est effectivement négligeable... Reprenons maintenant notre variable aléatoire X qui donne la durée d'hospitalisation avant de subir l'issue β. Cette durée est maintenant comptée en temps "continu" (c'est-à-dire avec une précision "infinie" et cette variable n'est plus bornée par T (on imagine qu'il y a des gens qui vont rester une durée jusqu'à l'infini avant de subir l'événement β). On va donc écrire :

$$X : \Omega_\beta \to \mathbb{R}^+ \tag{78}$$

où $\mathbb{R}^+$ désigne l'ensemble des réels positifs, (ici les durées d'hospitalisation avant l'événement β). Pour que ce soit intéressant, on suppose que Ω_β a un nombre très grand d'individus... On va considérer dans cet ensemble $\mathbb{R}^+$ des "plages" de durée. C'est-à-dire que l'on va se donner deux instants $0 \leq s \leq t$ et on va définir :

$$\{s \leq X \leq t\} = \{\omega \in \Omega_\beta, s \leq X(\omega) \leq t\} \tag{79}$$

comme étant l'ensemble des personnes ayant subi l'événement β pour une durée comprise entre s et t, après leur hospitalisation. La probabilité qu'un tel événement arrive est donnée, selon le formalisme intégral par la quantité :

$$P(\{s \leq X \leq t\}) = \int_s^t f(\tau)\,d\tau \tag{80}$$

Les auteurs cherchent donc une fonction $f : \mathbb{R}^+ \to \mathbb{R}^+$ (c'est une densité de probabilité continue) qui vérifie alors pour chaque jour $k \in [\![0, T]\!]$

$$\forall k \in [\![0, T-1]\!], \quad \pi_k^{true} = \int_{t=k}^{t=k+1} f(t)\, dt \tag{81}$$

$$k = T \quad \pi_T^{true} = \int_{t=T}^{+\infty} f(t)\, dt \tag{82}$$

Notez que la dernière condition est faite pour assurer le fait que la famille $(\pi_k^{true} : k \in [\![0, T]\!])$ vérifie bien la condition de normalisation :

$$\sum_{k=0}^{k=T} \pi_k^{true} = 1 \tag{83}$$

(ce qui est bien le minimum lorsque l'on traite de probabilité). Dans leur article, les auteurs ne se sont pas même donné cette peine puisqu'ils ont défini :

$$\pi_T^{true} = \int_{\tau=T}^{\tau=T+1} f(\tau)\, d\tau \tag{84}$$

mais il y a longtemps que l'on n'est plus à ça près... Donc il faut trouver une fonction f, positive, qui vérifie les conditions précédentes. Sauf que : l'ensemble des fonctions f qui potentiellement vérifient de telles équations est tout à fait gigantesque. Il n'y a donc évidemment pas unicité de la solution du problème précédent et même très loin de là. L'interprétation mathématique est évidente : vous essayez de recréer une information ultra fine sur la durée d'hospitalisation avant l'événement β, alors que vous n'avez pas les moyens d'une telle précision. Par conséquent, la seule conclusion à laquelle vous pouvez parvenir, c'est que sans autre information supplémentaire, l'affaire est sans espoir. Que faire donc pour trouver la fameuse fonction f ? Il faut que vous injectiez vous-même de l'information supplémentaire pour faire en sorte que le système possède une solution unique. Le problème, c'est que la plupart du temps, cette information est arbitraire.

1. Pour l'événement D, les auteurs ont remarqué que la décroissance des π_k^{true} avait l'air plus complexe à décrire. Dans leurs considérations, ils écrivent qu'il existe manifestement deux types d'individus mourant à l'hôpital : une fraction d'entre eux meurent très rapidement après l'hospitalisation, tandis qu'une autre fraction meurt de façon plus étalée dans le temps. Reprenons leurs propos :
   "Nous remarquons qu'une partie des individus meurent en dessous d'une durée très courte après leur entrée à l'hôpital. Nous utilisons donc un mélange d'une distribution exponentielle pour ceux qui meurent avant cette durée courte et une distribution lognormale pour ceux qui meurent après cette durée courte."

   $$\pi^{true} \sim (1-\rho) * exponentielle(m) + \rho * lognormal\left(\mu, \sigma^2\right) \tag{85}$$

   ce qui est remarquable, avec cette formulation qui a l'air anodine, c'est qu'elle montre exactement en quoi les auteurs ne maîtrisent absolument pas le vocabulaire des probabilités. Nous allons analyser dans le détail ce qu'ils entendent par cette décomposition, et comment on peut y accéder de façon élémentaire.

2. Pour l'événement ICU, les auteurs ont choisi de chercher la fonction f_{ICU} sous la forme :

$$\forall t \geq 0,\ f_{ICU}(t) = \lambda \exp(-\lambda t) \tag{86}$$

C'est la loi dite exponentielle. Pourquoi un tel choix ? En fait, cette loi modélise ce qu'il est convenu d'appeler les phénomènes sans mémoire. En pratique cependant, ce qui vous fait choisir une telle loi comme modèle, c'est le fait que vous constatez que la suite des π_k^{true} décroît très rapidement. Et donc, la décroissance rapide, c'est ce qui suggère d'utiliser une fonction exponentielle. Pas plus. On aurait pu choisir plein d'autres types de fonctions... Cela reste assez arbitraire. Donc peu intéressant. Une chose importante à signaler dans la suite : la loi exponentielle ne dépend que d'un seul paramètre λ. Ce paramètre est lié à la valeur moyenne de la variable aléatoire par la relation :

$$\lambda = \frac{1}{\mathbb{E}(X)} \tag{87}$$

où $\mathbb{E}(X)$ désigne la valeur moyenne de X, c'est-à-dire pour nous le nombre de jours moyen qu'un individu reste hospitalisé avant de subir l'événement ICU. C'est une valeur très facile à obtenir dès que vous connaissez la densité de probabilité $(\pi_k^{true} : k \in [\![0, T]\!])$. Nous y reviendrons.

3.2 Estimer les paramètres de la distribution associée à la mort (D)

En réalité, la décomposition que proposent les auteurs peut être formulée de façon simplissime grâce à la notion de probabilité conditionnelle. En effet, grâce à la variable aléatoire $X : \Omega_\beta \to [\![0, T]\!]$ que nous avons introduite et qui associe à chaque individu ω la durée d'hospitalisation avant de subir l'événement β, on peut, pour une "courte durée" k_0 fixée, s'intéresser aux seuls patients qui subissent l'événement avant k_0. Par définition, cet ensemble d'individus c'est :

$$\{X \leq k_0\} \subset \Omega_\beta \tag{88}$$

Ainsi, on veut maintenant définir deux probabilités : celle associée aux individus subissant leur issue avant k_0 (la fameuse courte durée) et ceux subissant leur issue après k_0. Ces nouvelles probabilités, ce sont des probabilités conditionnelles. Elles disent : **Sachant** d'abord que j'ai subi mon événement avant la durée k_0, quelle est la probabilité que j'ai de subir cet événement à un instant k ? Il est clair que pour $k > k_0$ une telle probabilité est nulle. On rappelle la définition mathématique de la probabilité conditionnelle :

Definition 2. *Soit $B \subset \Omega_\beta$ un sous-ensemble tel que $P(B) > 0$. On définit sur $\mathcal{P}(\Omega)$ la probabilité conditionnelle $P(\cdot \mid B)$ par la formule suivante :*

$$\forall A \subset \Omega_\beta,\quad P(A \mid B) = \frac{P(A \bigcap B)}{P(B)} \tag{89}$$

Pour les auteurs, les deux événements intéressants sont donnés par :

$$B = \{X \leq k_0\},\quad \overline{B} = \{X > k_0\} \tag{90}$$

on dit que $\overline{B}$ est le complémentaire de B, c'est-à-dire que $B \bigcup \overline{B} = \Omega_\beta$ et $B \bigcap \overline{B} = \emptyset$:

1. la probabilité conditionnelle selon B, c'est celle des gens qui meurent avant la durée k_0 ;
2. la probabilité conditionnelle selon $\overline{B}$, c'est celle des gens qui meurent après la durée k_0.

Pour l'une, ils proposent ensuite un modèle selon une loi de type exponentielle, pour l'autre, un modèle de type lognormal. Puisque $B, \overline{B}$ est une partition, on a très facilement :

$$\begin{aligned} \forall A \subset \Omega_\beta, \quad P(A) &= P\left(A \bigcap B\right) + P\left(A \bigcap \overline{B}\right) && (91) \\ &= P(A \mid B)\, P(B) + P\left(A \mid \overline{B}\right) P\left(\overline{B}\right) && (92) \\ &= \left(1 - P\left(\overline{B}\right)\right) P(A \mid B) + P\left(\overline{B}\right) P\left(A \mid \overline{B}\right) && (93) \end{aligned}$$

Ainsi, on peut maintenant très clairement établir ce que les auteurs cherchent en réalité :

1. le facteur ρ est en fait $P\left(\overline{B}\right) = 1 - P(B)$;
2. la loi $\mathit{exponentielle}(m)$ est en fait la loi $P(\cdot \mid B)$;
3. la loi $\mathit{lognormale}\left(\mu, \sigma^2\right)$ est en fait la loi $P\left(\cdot \mid \overline{B}\right)$.

Le dernier problème à résoudre, qui n'est pas donné par les auteurs, c'est que l'on ne connaît pas k_0. Il est possible de penser que n'ayant pas formalisé suffisamment leur problème, ils attendaient de retrouver ρ grâce à l'algorithme des moindres carrés. Il est tout à fait possible que ce soit le cas car, dans les résultats qui sont donnés, les auteurs indiquent alors :

$$P(\mathit{short\ delays}) \tag{94}$$

Il semble donc raisonnable de penser que pour les auteurs on a :

$$P(\mathit{short\ delays}) = P(X \leq k_0^*) = (1 - \rho) \tag{95}$$

Cependant, comme aucun détail n'est fourni dans l'article, il est impossible, même pour le reviewer le plus sérieux, de savoir ce qu'il en est...

3.3 La méthode employée par les auteurs

Dans ce que nous avons présenté, nous avons prouvé que l'estimation des π_k^{true} se faisait de façon unique et analytique par rapport à la donnée des π^{obs}. Mais ça, bien sûr, les auteurs l'ignorent de façon sidérante. La raison pour laquelle ils ignorent cela, c'est qu'ils n'ont jamais pris la peine d'analyser leur équation. Tout mathématicien qui se respecte essaye, au minimum, de savoir s'il y a existence d'une solution et s'il n'y a pas d'unicité, comment on peut rendre le problème unique. Une fois que vous essayez de répondre à cette question, dans le cas qui nous intéresse, le théorème dont vous avez besoin vous donnera en prime la forme exacte de la solution permettant de calculer π^{true} (dont on rappelle qu'il n'a cependant aucune interprétation probabiliste vis-à-vis du phénomène étudié). Donc, pour les auteurs, il y a un souci qui est qu'ils ne savent absolument pas ce qu'il en est de leur problème s'agissant de l'existence ou non d'une solution. Ils savent qu'ils cherchent une densité de probabilité, mais ils ne savent pas

comment la trouver. D'autre part, une fois trouvé le "$\boldsymbol{\pi}^{true}$", les auteurs veulent trouver les densités continues dont ils pensent que ces $(\pi_k^{true} : k \in [\![0, T]\!])$ sont une moyenne dans l'intervalle $[k, k+1)$ (et pour $k = T$ on pose $k + 1 = +\infty$).

Plutôt que de traiter chaque problème séparément, ce qui permet en fait, comme nous allons le voir, de gagner en simplicité et en efficacité, les auteurs vont en fait traiter deux problèmes en même temps, sans aucunement savoir si chacun d'eux, pris séparément, a une solution. Je vous explique.

Imaginons d'abord que vous ayez les $(\pi_k^{true} : k \in [\![0, T]\!])$. Vous cherchez une densité qui a une forme particulière (exponentielle, lognormale, etc.), c'est-à-dire qui dépendent dépend de certains paramètres, mais que vous ignorez effectivement. Vous notez par $(f[\theta](t), \theta \in \Theta)$ ces fonctions. Comme vous vous dites de vous connaissez les valeurs moyennes de la densité sur certains intervalles, vous vous dites qu'il suffit de résoudre l'équation des intervalles, c'est-à-dire trouver $\theta \in \Theta$ tel que l'on ait :

$$\forall k \in [\![0, T-1]\!], \quad \pi_k^{true} = \int_{t=k}^{t=k+1} f[\theta](t)\, dt \tag{96}$$

$$k = T, \quad \pi_k^{true} = \int_{t=T}^{+\infty} f[\theta](t)\, dt \tag{97}$$

vous avez donc a priori une équation avec $T+1$ équations, pour un nombre de paramètres n (pour nous $n = 4$ dans le cas de l'événement $\beta = D$ - la mort - et $n = 1$ dans le cas de l'événement $\beta = ICU$). Il s'agit de ce que l'on appelle un *problème inverse*. Qu'est-ce que cela ?

Si vous connaissez θ, vous connaissez la fonction $f[\theta](t)$ et il est très facile de produire les $(\pi_k^{true} : k \in [\![0, T]\!])$ via un calcul d'intégrale : c'est le problème direct. En revanche, si on vous donne les $(\pi_k^{true} : k \in [\![0, T]\!])$, il est fantastiquement difficile de remonter à θ : c'est le problème inverse. Commençons par faire une remarque : supposons que les fonctions $f[\theta](t)$ soient paramétrées par n paramètres $(\theta = (\theta_1, \cdots, \theta_n))$. Alors, le système que l'on a écrit précédemment se traduit par la résolution d'un système de $T+1$ équations pour n inconnues. Qui plus est ces équations sont gravement *non linéaires* (pour le cas linéaire, on sait étudier la nature du système grâce au théorème de Rouché-Fontené). Lorsque $n < T$ (ce qui va être toujours le cas pour nous : par exemple, pour l'événement β, on aura 4 paramètres inconnus pour 23 valeurs de $(\pi_k^{true} : k \in [\![0, T]\!])$), cela pose un problème très délicat car il y a des problèmes de compatibilité des données : il est tout à fait clair que n'importe quelle série de $(\pi_k^{true} : k \in [\![0, T]\!])$ ne pourra pas donner lieu à une solution en θ. Dit autrement, il nous manque une donnée essentielle qui est la suivante. Définissons l'application suivante :

$$\mathbf{L} : \Theta \rightarrow \mathbb{R}_+^{T+1} \tag{98}$$

$$\theta \rightarrow \left(\int_0^1 f[\theta](t)\, dt, \cdots, \int_{T-1}^{T} f[\theta](t)\, dt, \int_T^{+\infty} f[\theta](t) \right) \tag{99}$$

La question qui se pose est la suivante : si je me donne l'ensemble Θ, puis-je décrire l'ensemble image $\mathbf{L}(\Theta) \subset \mathbb{R}_+^{T+1}$? La réponse est totalement négative a

priori : c'est beaucoup trop dur. Autrement dit, si vous vous donnez a priori des valeurs de $(\pi_k^{true} : k \in [\![0,T]\!])$, vous n'avez aucune manière de savoir s'il existe bien des paramètres θ qui vont vous permettre de résoudre le système que vous avez posé. Mieux : quand vous calculez dans le sens direct, il y a en général une grande régularité dans les fonctions $f[\theta](\cdot)$ et les images que vous calculez (les $\mathbf{L}(\theta)$) ont elles aussi une forme de "régularité" (absence de bruit). Or les $(\pi_k^{true} : k \in [\![0,T]\!])$ que vous avez récupérées (et qui sont déterministes par rapport aux $\left(\pi_k^{obs} : k \in [\![0,T]\!]\right)$, sont elles-mêmes bruitées : il n'y a donc aucune chance pour que le système que vous posiez ait une solution en θ. Comment faire alors ?

L'une des méthodes les plus efficaces va permettre :

1. à la fois de "débruiter" le signal $(\pi_k^{true} : k \in [\![0,T]\!])$;
2. à la fois de n'utiliser que l'information dont on a besoin, c'est-à-dire d'avoir autant d'équations que d'inconnues.

Cette méthode, c'est la méthode des moments. Il s'agit là de l'un des problèmes les plus célèbres et les plus étudiés des mathématiques. Le problème des moments, comme on l'appelle, est associé à de grands noms, en particulier celui de Hilbert, car il est lié au XVIIe problème du mathématicien allemand sur les polynômes positifs. Le cas qui nous intéresse est le plus accessible a priori. Résumons : plutôt que de chercher les paramètres qui vont nous rendre toutes valeurs moyennes sur $[k, k+1]$, on va utiliser ces valeurs moyennes pour retrouver les moments associés aux densités $f[\theta](t)$.

Definition 3. *Soit $g : \mathbb{R}^+ \to \mathbb{R}^+$ une densité de probabilité, c'est-à-dire une fonction positive vérifiant :*

$$\int_{t=0}^{+\infty} g(t)\, dt = 1 \tag{100}$$

On dit qu'elle admet un moment d'ordre q, que l'on note par $\mu_q(g)$ lorsque l'on a la condition suivante :

$$\int_{t=0}^{+\infty} t^q g(t)\, dt = \mu_q(g) < +\infty \tag{101}$$

Grâce à la connaissance des $(\pi_k^{true} : k \in [\![0,T]\!])$, on va pouvoir former leur moment empirique d'ordre q par la formule suivante :

Definition 4. *Pour une distribution $(p_k : k \in [\![0,T]\!])$, on calcule son moment (empirique) d'ordre q sous la forme :*

$$\overline{\mu}_q(\mathbf{p}) = \sum_{k=0}^{k=T} \left[\frac{(k+1)^{q+1} - k^{q+1}}{q+1} \right] p_k \tag{102}$$

En fait, pour trouver une approximation des moments d'ordre q, nous avons besoin d'une formule de quadrature. Nous voulons en outre que cette formule de quadrature pour le moment d'ordre q soit exacte pour les densités de probabilité qui sont constantes sur les intervalles de la forme $[k, k+1], k \in [\![0,T]\!]$.

Supposons donc que f vaille f_k sur un intervalle de la forme $[k, k+1]$. On a alors :

$$\int_{t=k}^{t=k+1} t^q f_k dt = \left[\frac{(k+1)^{q+1} - k^{q+1}}{q+1}\right] f_k \tag{103}$$

Notons que cette technique est "régularisante" : s'il existe des "bruits" de mesure sur les $(p_k : k \in [\![0, T]\!])$, alors le fait de réaliser une somme aura tendance à gommer les bruits dans le résultat. Les "erreurs" ont ainsi tendance à se compenser. Ainsi, si l'on cherche n paramètres, on calculera les moments de 1 à n (sachant que le moment d'ordre 0 dit juste que la famille est une densité de probabilité : autrement dit, la somme de tous les termes est égale à 1). Ainsi, on va chercher à résoudre maintenant le système :

$$\forall i \in [1, n], \quad \overline{\mu}_i\left(\boldsymbol{\pi}^{true}\right) = \int_{t=0}^{+\infty} t^i f\left[\theta\right](t)\, dt \tag{104}$$

Ce problème est beaucoup mieux posé. Il y a autant d'équations que d'inconnues. De plus, on a, grâce à aux matrices de Hankel, une caractérisation du fait qu'il existe au moins une densité de probabilité qui vérifie l'équation aux moments (cela ne voudra pas dire qu'elle sera forcément de la forme $f\left[\theta\right](t)$) mais au moins on ne résoudra pas un problème dont on sait à l'avance qu'il n'a pas de solution.

1. Si l'on cherche une loi exponentielle qui vérifie une certaine équation aux moments, c'est très facile : cette loi ne dépend que du premier moment. Ainsi, pour $\overline{\mu}_1\left(\boldsymbol{\pi}^{true}\right) > 0$ positif donné, il existe une unique loi exponentielle de la forme $\lambda \exp\left(-\lambda t\right)$ qui donne comme moyenne $\overline{\mu}_1\left(\boldsymbol{\pi}^{true}\right)$. Elle vérifie la relation :
$$\lambda = \frac{1}{\overline{\mu}_1\left(\boldsymbol{\pi}^{true}\right)} \tag{105}$$

2. Si l'on cherche une loi lognormale de paramètre μ (qui n'est pas la moyenne) et σ^2 (qui n'est pas la variance), là encore, c'est facile car il suffit de résoudre le système linéaire suivant :
$$\ln\left(\overline{\mu}_1\left(\boldsymbol{\pi}^{true}\right)\right) = \mu + \frac{1}{2}\sigma^2 \quad \ln\left(\overline{\mu}_2\left(\boldsymbol{\pi}^{true}\right)\right) = 2\mu + 2\sigma^2 \tag{106}$$
ce qui donne exactement :
$$\mu = \frac{1}{2}\left(4\ln\left(\overline{\mu}_1\left(\boldsymbol{\pi}^{true}\right)\right) - \ln\left(\overline{\mu}_2\left(\boldsymbol{\pi}^{true}\right)\right)\right) \tag{107}$$
$$\sigma^2 = \ln\left(\overline{\mu}_2\left(\boldsymbol{\pi}^{true}\right)\right) - 2\ln\left(\overline{\mu}_1\left(\boldsymbol{\pi}^{true}\right)\right) \tag{108}$$
On en déduit une condition nécessaire et suffisante pour que des moments empiriques soit effectivement réalisables contre une densité de type loi lognormale : il faut et il suffit que l'on ait $\overline{\mu}_1 > 0,\ \overline{\mu}_2 > \overline{\mu}_1^2$

On voit que l'on a simplifié les problèmes au maximum, *en gardant les informations nécessaires qui sont optimales pour leur mise en pratique*, tout en contrôlant l'existence et l'unicité des solutions à chaque étape : nous avons donc un problème très bien posé pour sa résolution mathématique et numérique. Si le modèle proposé a un fondement, alors il est certain que nous trouverons la

solution. Qui plus est de façon totalement élémentaire : exactement à l'inverse de ce que font les auteurs. Qu'ont donc fait les auteurs ?

D'abord, ils n'ont jamais employé la méthode des moments (pourtant à utiliser de façon prioritaire, elle est aussi simple qu'efficace). Pourquoi ? Tout part du fait qu'ils ne savent pas résoudre le système de Frobenius. Ainsi, ils posent effectivement l'équation :

$$\pi_k^{true} = \frac{\alpha_k \pi_k^{true}}{\sum_{l=0}^{l=1} \alpha_l \pi_l^{true}} \tag{109}$$

et ils imposent qu'ils cherchent une densité de probabilité, c'est-à-dire, une famille $(\pi_k^{true} : k \in [\![0, T]\!])$ d'éléments positifs dont la somme vaut 1. Mais ils étaient incapables de trouver la formule :

$$\pi_k^{true} = \frac{\alpha}{\alpha_k} \pi_k^{obs}, \quad \frac{1}{\alpha} = \sum_{l=0}^{l=T} \frac{1}{\alpha_l} \pi_l^{obs} \tag{110}$$

Donc, comme ils sont ignorants, ils vont faire d'une pierre deux coups :

1. Ils vont paramétrer la solution qu'ils cherchent, les $(\pi_k^{true} : k \in [\![0, T]\!])$, par les fameuses fonctions $f[\theta](t)$: de cette manière en posant pour chaque $k \in [\![0, T]\!]$ le fait que :

$$\pi_k^{true} = \int_{t=k}^{t=k+1} f[\theta](t)\, dt \tag{111}$$

   on est sûr alors que les π_k^{true} sont positifs et que leur somme vaut effectivement 1. (à condition toutefois, comme nous l'avons signalé que pour le dernier élément π_T^{true} , l'intégrale se fasse sur l'intervalle $[T, +\infty)$)
2. Maintenant ils vont faire jouer leurs paramètres $\theta \in \Theta$ de sorte que leur objectif, à savoir faire coller les $(\pi_k^{true} : k \in [\![0, T]\!])$ avec leur "observation" $\boldsymbol{\pi}^{obs}$), soit réalisé

VOILÀ CE QUE L'ON PEUT APPELER UNE RÉSOLUTION "EN DOUBLE AVEUGLE" : VOUS NE CONNAISSEZ PAS LES $(\pi_k^{true} : k \in [\![0, T]\!])$ ET VOUS NE CONNAISSEZ PAS LES PARAMÈTRES θ MAIS VOUS ALLEZ RESOUDRE LES DEUX D'UN COUP.

Enfin bon, sauf que, vous n'allez rien résoudre du tout car comme nous l'avons déjà dit : si vous ne passez pas par les méthodes des moments, il est impossible que le système des équations suivant :

$$\left(\int_0^1 f[\theta](t)\, dt, \cdots, \int_{T-1}^{T} f[\theta](t)\, dt, \int_T^{+\infty} f[\theta](t) \right) = \left(\pi_0^{true}, \cdots, \pi_T^{true}\right) \tag{112}$$

ait une solution sur les paramètres θ. Donc, par conséquent, il est impossible que les auteurs trouvent la solution du système linéaire (lequel d'ailleurs a été mis sous forme non linéaire, ce qui est le pire crime numérique que l'on puisse connaître) en utilisant leur approche. Pourtant, ils vont annoncer avoir trouvé une solution comment ? C'est l'objet de la section suivante.

3.4 Comment toujours trouver des solutions aux équations impossibles

On arrive doucement au cœur de la méthode des épidémiologistes (en tout cas des auteurs de cette étude). Avoir une méthode qui donne toujours des solutions, même quand il n'y en a pas. Cela s'appelle "la méthode des moindres carrés". C'est la fameuse phrase des auteurs :
We estimate parameters of the true delay from hospitalization to death distribution $\boldsymbol{\pi}^{true}$ for each group in turn by minimizing the sum of squared error (SSE) of the distribution $\boldsymbol{\pi}^{exp}$ to the observed data $\boldsymbol{\pi}^{obs}$.
(traduction : nous donnons une estimation des paramètres associés à la distribution $\boldsymbol{\pi}^{true}$ du délai véritable de la durée d'hospitalisation avant la mort en minimisant la somme des erreurs au carré de la distribution $\boldsymbol{\pi}^{exp}$ aux observations $\boldsymbol{\pi}^{obs}$).

Je vous explique sur un exemple simple. Imaginons que vous cherchiez à résoudre l'équation :

$$\textit{Touver } x \in \mathbb{R}, \quad \textit{tel que } \quad x^2 = -1 \tag{113}$$

bien sûr, chacun sait qu'une telle équation n'a pas de solution dans les ensembles réels car le carré d'un quelconque nombre réel est toujours positif. Les mathématiciens sont formels. Ils ont d'ailleurs inventé les nombres complexes pour résoudre cette équation. Invention ô combien salutaire pour l'histoire de l'humanité. Mais pour les épidémiologistes, jamais rien d'impossible. Voyez plutôt ce qu'ils vont faire :

1. ils vont former la différence entre l'objectif (-1) et la fonction qui s'applique à leur inconnue. Ils forment donc la quantité $x^2 + 1$;
2. ensuite ils vont dire que résoudre l'équation revient à résoudre l'équation $x^2 + 1 = 0$ (ils sont très forts) ;
3. ils disent alors, en élevant les deux membres au carré, que cela revient à résoudre l'équation $\left(x^2 + 1\right)^2 = 0$;
4. puisque la fonction carré est toujours positive ou nulle, ils disent alors que pour que cette fonction soit égale à 0, eh bien il suffit qu'elle atteigne son minimum ;
5. donc ils affirment qu'en trouvant le minimum de la fonction $\left(x^2 + 1\right)^2$, ils vont résoudre l'équation $x^2 = -1$.

Évidemment, l'arnaque dans cette méthode, c'est que l'on peut trouver le minimum de la fonction $\left(x^2 + 1\right)^2$ sans que ce minimum soit égal à 0, c'est-à-dire sans que le x qui réalise le minimum ne résolve l'équation. Le drame, en outre, c'est que sous des hypothèses très faibles (c'est-à-dire qui sont quasiment toujours vérifiées), les problèmes d'optimisation que l'on pose par la méthode des moindres carrés ont quasiment toujours des solutions optimales. Mais notons deux choses :

1. il est parfois très difficile de trouver les minima car on peut souvent trouver ce qu'il est convenu d'appeler des optima locaux, mais trouver les minima globaux est une autre paire de manche ;

2. évidemment, il n'y a pas *d'équivalence* entre la solution de la minimisation et le fait d'avoir trouvé la solution de l'équation : s'il y a une solution à l'équation, la résolution effective du problème d'optimisation vous donne une solution. En revanche, il y a toujours une solution au problème d'optimisation même si l'équation n'a pas de solution.

Ainsi, si l'on développe la méthode des moindres carrés pour l'équation $x^2 = -1$, on transforme cela en un problème de minimisation, c'est-à-dire trouver le minimiseur de la fonction $\left(x^2+1\right)^2$, qui existe toujours de façon unique : c'est $x = 0$. Ainsi, les épidémiologistes en sont certains : l'unique solution de l'équation $x^2 = -1$, c'est $x = 0$. On a donc à ce titre $0 = -1$. Certes, on en reste un peu pantois, mais ce n'est pas grave. Après tout, 0 n'est pas strictement égal à -1, mais c'est sans doute une approximation acceptable de la vérité... Donc $x = 0$ n'est *pas loin* de résoudre l'équation : on a fait du bon boulot.

C'est exactement ce qu'on fait nos épidémiologistes : il est évidemment impossible de trouver des paramètres θ tels que l'on ait simultanément les égalités :

$$\forall k \in [\![0, T]\!], \quad \pi_k^{true} = \int_{t=k}^{t=k+1} f[\theta](t)\, dt \tag{114}$$

$$\forall k \in [\![0, T]\!], \quad \pi_k^{obs} = \frac{\alpha_k \pi_k^{true}}{\sum_{l=0}^{l=T} \alpha_l \pi_l^{true}} \tag{115}$$

(sauf à commettre ce que l'on appelle *le crime inverse* : c'est-à-dire calculer directement le membre de gauche à partir du membre de droite), mais qu'à cela ne tienne, on va mettre tout mettre dans un problème de moindres carrés . Ainsi va-t-on former la fonction d'erreur :

$$E(\theta) = \sum_{k=0}^{k=T} \left(\pi_k^{obs} - \frac{\alpha_k \int_{t=k}^{t=k+1} f[\theta](t)\, dt}{\sum_{l=0}^{l=T} \alpha_l \int_{t=l}^{t=l+1} f[\theta](t)\, dt} \right)^2 \tag{116}$$

et on va, par un algorithme d'optimisation, trouver la (les) valeur(s) de θ qui rende(nt) cette expression minimum... Un tel algorithme va effectivement rendre une solution : mais il n'y a aucune chance qu'elle donne effectivement les $\boldsymbol{\pi}^{true}$ du système de Frobenius, et donc cela va distordre complètement la recherche des paramètres θ. En outre, les algorithmes d'optimisation qui règlent les problèmes de moindres carrés ont souvent la particularité de rendre des solutions qui dépendent de l'initialisation choisie pour essayer d'avoir une solution numérique au problème.

3.5 Un modèle « prédicitif » de mort (D) qui a toutes les chances ne de rien prédire

Nous rappelons que le parti-pris des auteurs dans la recherche d'une distribution continue (i.e. permettant de rendre compte de la durée d'hospitalisation avant la mort avec une précision infinie) tendrait à modéliser les données moyennes journalières (celles qui sont remontées par les données fournies par les hôpitaux). Il s'agit de « mixter » une loi de décroissance exponentielle et une loi lognormale selon la formule de l'équation (1) :

$$\boldsymbol{\pi}^{true} \sim (1-\rho) * exponentielle(m) + \rho * lognormal\left(\mu, \sigma^2\right) \tag{117}$$

Nous allons voir que, sous cette forme, la modélisation a quasiment toujours le grave défaut de présenter une structure en « siphon », c'est-à-dire que la densité de probabilité $f[\theta](t)$ qui va résulter du modèle proposé sera d'abord décroissante, atteindra un minimum (local), remontera vers un maximum (local) et finira ensuite par décroître vers 0 (à l'infini). Or un tel modèle a toutes les chances de ne pas être pertinent. Si l'on se réfère en effet à l'histogramme des mesures (les bâtons du graphique (A) des auteurs (voir la figure 1)), il semblerait que la caractéristique de la courbe c'est d'être fondamentalement toujours décroissante, mais avec une sorte de plateau sur les jours $j \in \{1,2,3,4\}$. Donc l'idée de vouloir approcher une telle distribution avec une courbe en forme de siphon, voilà qui n'est pas vraiment pertinent...

Lemme 1. *Soit $f(t), t \in [0,+\infty)$ une densité de probabilité de la forme*

$$f \sim (1-\rho) * exponentiel(m) + \rho * lognormale\left(\mu, \sigma^2\right), \quad \rho \in (0,1) \tag{118}$$

Supposons que l'on vérifie la condition :

$$F(\rho, m, \sigma, \mu) < 0 \tag{119}$$

où $F(\rho, m, \sigma, \mu)$ est la fonction définie par :

$$t(\mu, \sigma) = \exp(\mu)\exp(\sigma T_-(\sigma)) \tag{120}$$

$$T_-(\sigma) = -\frac{1}{2}\left(\sqrt{4+\sigma^2}+\sigma\right), \quad T_+(\sigma) = \frac{1}{2}\left(\sqrt{4+\sigma^2}-\sigma\right) \tag{121}$$

$$F = \frac{1-\rho}{\rho}\sqrt{2\pi}\sigma^2\frac{t^2(\sigma,\mu)}{m^2}\exp\left(-\frac{t(\sigma,\mu)}{m}\right) - T_+(\sigma)\exp\left(-\frac{1}{2}T_-^2(\sigma)\right) \tag{122}$$

alors il existe $0 < t_1 < t_2 < +\infty$ tel que :

1. *la fonction f est strictement décroissante sur $[0, t_1]$;*
2. *la fonction f est strictement croissante sur $[t_1, t_2]$;*
3. *la fonction f est décroissante sur un voisinage ouvert de t_2 dans $(t_2, +\infty)$.*

Ainsi, on voit que la fonction f admet un minimum local en t_1 et un maximum local en t_2

A priori, sur les exemples que nous traiterons, on verra que la fonction f reste décroissante sur $[t_2, +\infty)$, mais nous n'avons pas besoin de démontrer ce résultat (plus délicat à réaliser) pour ce dont nous aurons besoin dans la suite.

Démonstration. D'après les hypothèses qui sont fournies, on sait que la fonction f s'écrit pour tout $t > 0$

$$f(t) = (1-\rho)\frac{1}{m}\exp\left(-\frac{t}{m}\right) + \rho\frac{1}{t\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\ln(t)-\mu)^2}{2\sigma^2}\right) \tag{123}$$

Une dérivation immédiate nous retourne l'expression

$$f'(t) = -(1-\rho)\frac{1}{m^2}\exp\left(-\frac{t}{m}\right) - \rho\left(\frac{\sigma^2+(\ln(t)-\mu)}{t^2\sigma^2\sqrt{2\pi\sigma^2}}\right)\exp\left(-\frac{(\ln(t)-\mu)^2}{2\sigma^2}\right) \tag{124}$$

On voit immédiatement que l'on a :

$$\lim t \to 0^{+} f'(t) = -(1-\rho)\frac{1}{m^2} < 0 \tag{125}$$

Ce qui veut dire que la densité de probabilité commence toujours par être décroissante. Pour montrer qu'elle atteint un minimum local, il faut démontrer que l'on peut trouver une valeur t_1 telle que $f'(t_1)$ s'annule et change de signe (de négatif devient ensuite positif). On va d'abord chercher à résoudre :

$$f'(t) = 0 \tag{126}$$

Bien sûr, il semble délicat de trouver, par le calcul analytique, une solution à cette équation. Il est évident que l'on a $f'(t) = 0$ si et seulement si :

$$\frac{1-\rho}{\rho}\sqrt{2\pi}\sigma^2\frac{t^2}{m^2}\exp\left(-\frac{t}{m}\right) + (\sigma + T)\exp\left(-\frac{1}{2}T^2\right) = 0 \tag{127}$$

où l'on a posé $T = (\ln(t) - \mu)/\sigma$. Comme on sait que l'on a toujours :

$$\frac{1-\rho}{\rho}\sqrt{2\pi}\sigma^2\frac{t^2}{m^2}\exp\left(-\frac{t}{m}\right) > 0 \tag{128}$$

il est nécessaire que l'on puisse trouver $T \in \mathbb{R}$ (en effet la fonction $T(t)$ est bijective de $\mathbb{R}^{+*}$ dans $\mathbb{R}$) telle que le terme suivant :

$$(\sigma + T)\exp\left(-\frac{1}{2}T^2\right) \tag{129}$$

soit négatif. Une étude de variation nous indique que l'on doit étudier le signe du trinôme $-T(\sigma + T) + 1$

$$-T(\sigma + T) + 1 = 0 \Leftrightarrow T_{\pm}(\sigma) = -\frac{1}{2}\left(\sigma \pm \sqrt{4 + \sigma^2}\right) \tag{130}$$

En remarquant que la fonction $(\sigma + T)\exp\left(-\frac{1}{2}T^2\right)$ a une limite nulle en $\pm\infty$, il est clair que la fonction $(\sigma + T)\exp\left(-\frac{1}{2}T^2\right)$ possède :

1. un minimum global en $T_{-}(\sigma) = -1/2\left(\sigma + \sqrt{4 + \sigma^2}\right) < 0$ et que la valeur de ce minimum est négative. Elle vaut :

$$\left(\sigma - 1/2\left(\sigma + \sqrt{4 + \sigma^2}\right)\right)\exp\left(-\tfrac{1}{2}T_{-}^2(\sigma)\right) \tag{131}$$

$$= \quad -T_{+}(\sigma)\exp\left(\tfrac{1}{2}T_{-}^2(\sigma)\right) \tag{132}$$

2. un maximum global $T_{+}(\sigma) = -1/2\left(\sigma - \sqrt{4 + \sigma^2}\right) > 0$ et que la valeur de ce maximum global est positive. Elle vaut :

$$\left(\sigma + 1/2\left(\sqrt{4 + \sigma^2}\right) - \sigma\right)\exp\left(-\tfrac{1}{2}T_{+}^2(\sigma)\right) \tag{133}$$

$$= \quad -T_{-}(\sigma)\exp\left(-\tfrac{1}{2}T_{+}^2(\sigma)\right) \tag{134}$$

L'espoir, c'est de pouvoir démontrer que pour la valeur de t telle que la fonction $(\sigma + T)\exp\left(-1/2T^2\right)$ atteigne sa valeur négative la plus petite, on peut obtenir une valeur négative pour la dérivée $f'(t)$. Pour cela, il suffit que l'on ait :

$$\frac{1-\rho}{\rho}\sqrt{2\pi}\sigma^2\frac{t^2(\sigma,\mu)}{m^2}\exp\left(-\frac{t(\sigma,\mu)}{m}\right) - T_{+}(\sigma)\exp\left(-\frac{1}{2}T_{-}^2(\sigma)\right) < 0 \tag{135}$$

$$t(\sigma,\mu) = \exp(\mu)\exp(\sigma T_{-}(\sigma)) \tag{136}$$

d'où le critère que l'on a posé dans le lemme. Supposons donc que ce critère soit réalisé. Alors puisque f est continue, par le théorème des valeurs intermédiaires elle va passer par un t_1 tel qu'à gauche elle sera négative et à droite elle sera positive. Elle atteint donc un minimum local. Cependant, une fois ce minimum atteint, elle devra repasser par un maximum local. En effet, on sait qu'après t_1, on a f strictement croissante. Si elle ne redevient pas décroissante, on a alors :

$$\forall t \geq t_1, \quad f(t) \geq f(t_1) > 0 \tag{137}$$

et donc on ne peut plus obtenir la condition de normalisation de la probabilité, à savoir :

$$\int_{t=0}^{+\infty} f(t)\, dt = 1 \tag{138}$$

□

Ainsi, le résultat de cette démonstration est le suivant : si les paramètres $\left(\rho, m, \mu, \sigma^2\right)$ vérifient la condition que nous avons donnée dans le lemme

$$F(\rho, m, \mu, \sigma) < 0$$

alors la fonction $f\left[\rho, m, \mu, \sigma^2\right](t)$ qui modélise la distribution des $\boldsymbol{\pi}^{true}$ passe forcément d'abord par un minimum puis ensuite doit avoir un maximum. Elle a une forme en « siphon ». En particulier, si la condition est vérifiée, en aucun cas la fonction $f\left[\rho, m, \mu, \sigma^2\right](t)$ ne peut être monotone, c'est-à-dire, ici, simplement décroissante.

4 Application numérique

4.1 Préparation des données numériques

Pour étudier toutes les questions numériques, nous avons donc besoin d'avoir les π_k^{obs}, ainsi que les $H_j, j \in [\![0, m]\!]$. Or cela est difficile car les auteurs ne donnent rien de ces valeurs. Ce qui fait, comme nous l'avons déjà signalé, qu'il est de fait très difficile d'essayer de reproduire leurs calculs. Il n'y a que deux cas que nous pouvons traiter eu égard aux seules informations disponibles dans le papier :

1. des données de π_k^{obs} pour la durée d'hospitalisation avant la mort (D) pour l'ensemble des patients hospitalisés. Celles-ci ont été présentées dans la figure (A) des auteurs (voir figure (1)) sous forme d'histogramme à lire soi-même ;
2. des données de π_k^{obs} pour la durée d'hospitalisation avant la mise en réanimation (ICU) pour l'ensemble des patients hospitalisés. Elles sont données dans la figure (3) ci-dessous.

En aucun cas donc les valeurs numériques utilisées par les auteurs ne sont fournies. Ce qui fait qu'il faut les reconstruire par des lectures graphiques, ce qui manque évidemment de précision. Nous allons expliquer rapidement comment nous avons procédé. Nous sommes désolé d'en passer par cette cuisine, mais cela est rendu nécessaire par le manque de données de l'article.

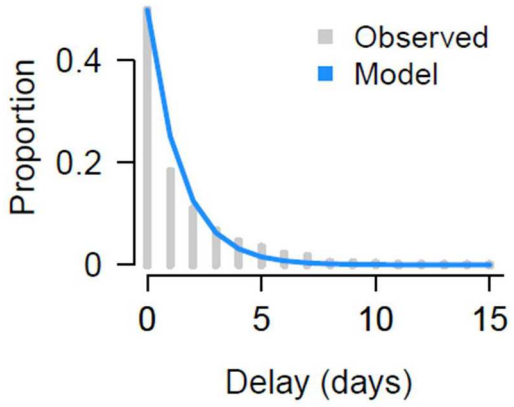

FIGURE 3 – données sur la durée d'hospitalisation avant de subir une admission en service de réanimation. À nouveau, selon les auteurs eux-mêmes, la courbe a été « fittée » (i.e. maquillée) pour tenir compte du fait que les auteurs obtenaient des courbes donnant des probabilités trop élevées de subir une admission en service de réanimation après des temps longs.

1. On commence par les données concernant la durée d'hospitalisation avant de subir un décès à l'hôpital.
   (a) Pour les valeurs $k = \{0, 1, 2, 3\}$, on fait une lecture directe approximative :
   $$\begin{array}{cccccc} k & 0 & 1 & 2 & 3 \\ \pi_k^{obs} & 0.17 & 0.09 & 0.09 & 0.09 \end{array} \tag{139}$$
   dont la probabilité cumulée vaut 0.440
   (b) pour les valeurs de 4 à 10, on effectue une approximation linéaire en partant de $\pi_4^{obs} = 0.08$ et l'on va jusqu'à $\pi_{10}^{obs} = 0.03$. On a donc :
   $$\forall k \in [\![4, 10]\!] \quad \pi_k^{obs} = \frac{(0.08 - 0.03)}{6} (10 - k) + 0.03 \tag{140}$$
   ce qui nous donne alors :
   $$\begin{array}{ccccccccc} k & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ \pi_k^{obs} & 0.08 & 0.072 & 0.063 & 0.055 & 0.047 & 0.038 & 0.03 \end{array} \tag{141}$$
   dont la probabilité cumulée vaut 0.385.
   (c) On considère que $\pi_k^{obs}, k \geq 23 = 0$. Finalement, pour k allant de 11 à 22, on fixe une décroissance linéaire sur les probabilités, de sorte qu'au final les probabilités cumulées valent 1. On a donc :
   $$\pi_k^{obs} = \Delta (23 - k), \quad \sum_{k=11}^{22} \pi_k^{obs} = 1 - (0.44 + 0.385) \tag{142}$$
   L'équation sur Δ nous donne alors :
   $$78\Delta = 0.175 \Rightarrow \Delta \approx 0.00224 \tag{143}$$

on reconstruit alors les données en arrondissant pour que la somme totale vaille alors exactement 1.

k	11	12	13	14	15	16	17	18
π_k^{obs}	0.027	0.025	0.022	0.020	0.018	0.016	0.013	0.011
k	19	20	21	22				
π_k^{obs}	0.009	0.007	0.005	0.002				

2. On peut réaliser, pour le cas $\beta = ICU$, le même travail approximatif (nous y sommes hélas obligés) de reconstruction des valeurs numériques de $\boldsymbol{\pi}^{obs}$.
   (a) On effectue pour $k = 0$ jusqu'à $k = 5$ une lecture directe des données

$$\begin{array}{cccccccc} k & 0 & 1 & 2 & 3 & 4 & 5 \\ \pi_k^{obs} & 0.5 & 0.19 & 0.12 & 0.065 & 0.050 & 0.045 \end{array} \tag{144}$$

   (b) ensuite on considère à partir de $k = 9$ (inclus) que l'observation est nulle, et que la décroissance est linéaire entre $k = 6$ et $k = 9$. On cherche donc Δ tel que :

$$\pi_k^{obs} = \Delta(9 - k), k \in [\![6, 9]\!] \tag{145}$$

   de sorte que la normalisation de π_k^{obs} soit respectée. Ce qui nous donne le système :

$$6\Delta = 1 - (0.5 + 0.19 + 0.12 + 0.065 + 0.05 + 0.045) = 0.03 \tag{146}$$

   Finalement, les dernière valeurs s'écrivent :

$$\begin{array}{ccccc} k & 6 & 7 & 8 \\ \pi_k^{obs} & 0.015 & 0.010 & 0.005 \end{array} \tag{147}$$

Les nouvelles hospitalisations quotidiennes $H_j, j \in [0, m]$ (pour coller aux indications données par les auteurs de l'article), l'article ne les donne absolument pas. Il faut donc qu'on puisse les reconstruire. En particulier, on ne connaît pas le jour de référence en deçà duquel on considère que $H_j = 0$. On ne connaît pas non plus le jour m au-delà duquel on arrête le comptage. Faute de mieux, on ira donc sur la page de Wikipedia France qui donne ces valeurs : on choisit un $j = 0$ pour le 19 mars 2020 et $m = 38$ pour le 26 avril 2020. On les donne sous forme d'histogramme (voir la figure 4), les valeurs numériques étant reportées en annexe.

4.2 À propos de $\boldsymbol{\pi}^{true}$ et $\boldsymbol{\pi}^{obs}$: illustration des propriétés mathématiques

On rappelle que la formule qui lie π^{true} à π^{obs} s'écrit sous la forme :

$$\forall k \in [\![0, T]\!]\ \alpha_k = \sum_{j=0}^{m-k} H_j,\ \frac{1}{\alpha} = \sum_{k=0}^{k=T} \frac{1}{\alpha_k} \pi_k^{obs},\ \forall k \in [\![0, T]\!]\ \pi_k^{true} = \frac{\alpha}{\alpha_k} \pi_k^{obs} \tag{148}$$

Il est maintenant très facile d'avoir les valeurs des $\boldsymbol{\pi}^{true}$. Les valeurs numériques de $\boldsymbol{\pi}^{true}$ sont données en annexe, nous les représentons graphiquement sur la

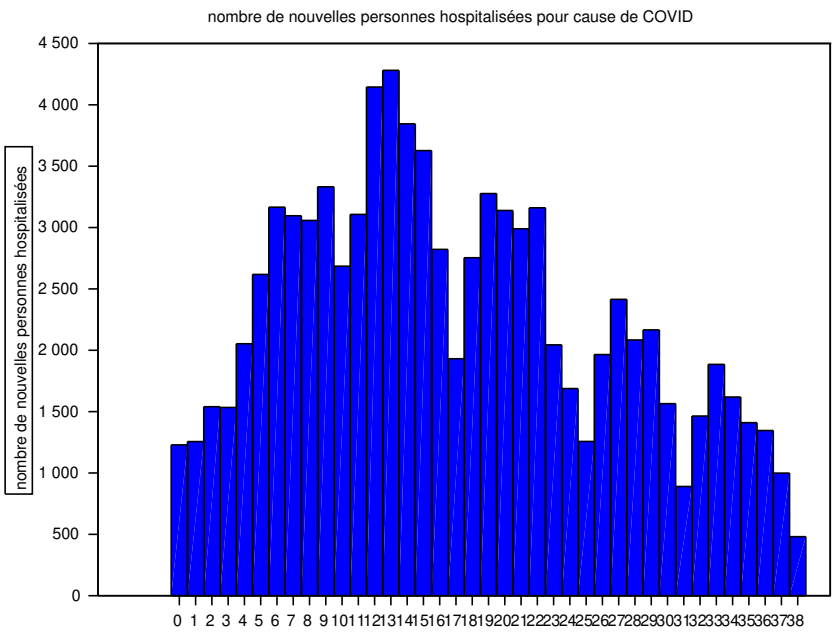

FIGURE 4 – nouvelles personnes hospitalisées selon les jours. la valeur j = 0 correspond au 19 mars 2020. Données d'après la page COVID France de Wikipedia. Les valeurs numériques sont reportées en annexe.

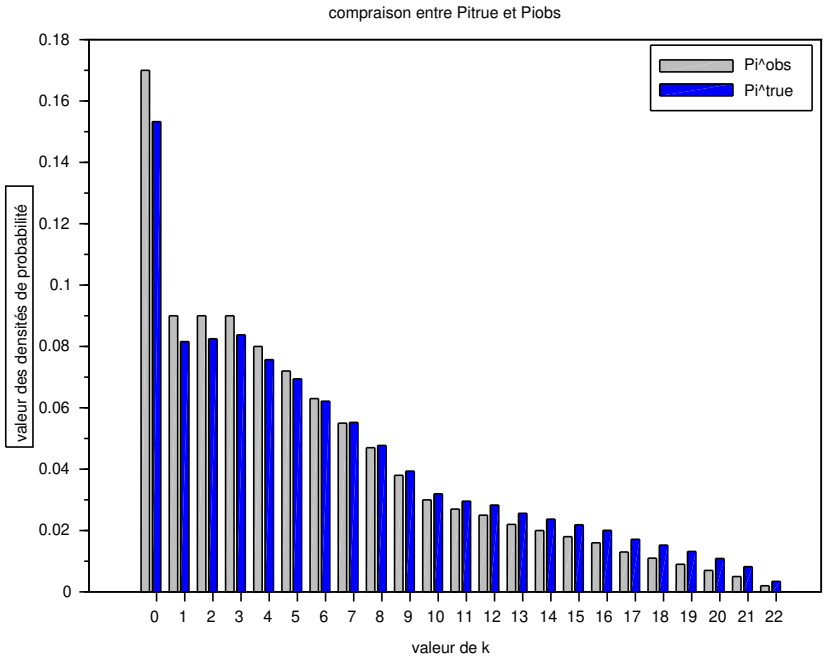

FIGURE 5 – comparaisons entre les valeurs de $\boldsymbol{\pi}^{obs}$ et $\boldsymbol{\pi}^{true}$

figure (5). Comme on le voit sur la figure, les deux distributions sont proches. On peut même calculer leur différence relative. On obtient alors :

$$\frac{\left\|\boldsymbol{\pi}^{true}-\boldsymbol{\pi}^{obs}\right\|_1}{\left\|\boldsymbol{\pi}^{obs}\right\|_1}=\left\|\boldsymbol{\pi}^{true}-\boldsymbol{\pi}^{obs}\right\|_1=0.09 \tag{149}$$

ce qui signifie que la différence entre les deux distributions est de l'ordre de 9%. Comme les π^{true} sont systématiquement plus petits que les π^{obs} dans les valeurs significatives, cela fait que, tout à fait logiquement, lorsque l'on va lancer un algorithme d'optimisation qui force les $\boldsymbol{\pi}^{true}$ plutôt que les $\boldsymbol{\pi}^{obs}$, on risque de dévier naturellement le modèle vers des probabilités artificiellement plus élevées aux grandes valeurs de k.

Le fait que les deux distributions $\boldsymbol{\pi}^{obs}$ et $\boldsymbol{\pi}^{true}$ soient proches, comme nous l'avons dit, n'est certainement pas lié au fait que l'une serait l'observation empirique de l'autre. Il s'agit simplement d'une propriété particulière liée à la distribution des $(\alpha_k : k \in [\![0,T]\!])$ et des $\left(\pi_k^{obs} : k \in [\![0,T]\!]\right)$. Il est possible qu'avec d'autres distributions l'erreur relative soit plus élevée.

On peut maintenant illustrer la propriété de la suite $\left(\pi_k^{obs}-\pi_k^{true} : k \in [\![0,T]\!]\right)$ (voir figure 6) qui veut que :

1. en dessous d'un $k_0 \in [\![0,T]\!]$, cette suite est positive (pour notre exemple on aura $k_0 = 6$) ;

2. à partir de $k_0 + 1$, cette suite est négative ;
3. dans le cas où la suite $\left(\pi_k^{obs} : k \in [\![0, T]\!]\right)$ est elle-même décroissante (ce qui est notre cas ici), alors les termes qui sont inférieurs à k_0 s'organisent selon un mode décroissant.

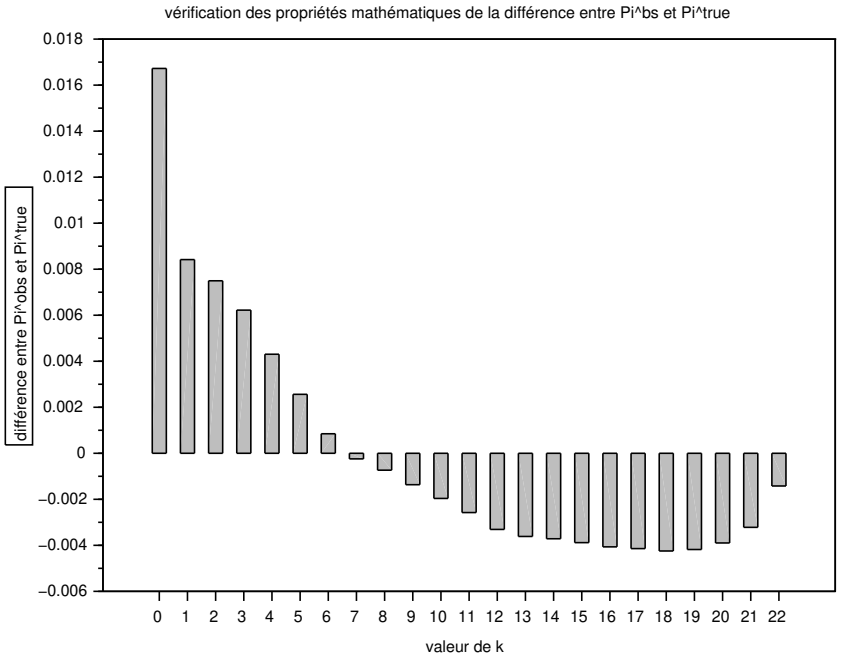

FIGURE 6 – histogramme des différences entre $\boldsymbol{\pi}^{obs}$ et $\boldsymbol{\pi}^{true}$.

À nouveau ces propriétés doivent effectivement nous faire penser qu'entre les densités $\boldsymbol{\pi}^{true}, \boldsymbol{\pi}^{obs}$, il ne peut absolument pas y avoir d'interprétation d'une qui serait l'observation et l'autre la "vérité" de cette observation : si c'était le cas, la suite $\left(\pi^{obs}(k) - \pi_k^{true} : k \in [\![0, k_0]\!]\right)$ n'aurait aucune forme de régularité : les écarts seraient distribués positivement ou négativement selon un certain aléa : vraisemblablement une loi variable aléatoire gaussienne.

4.3 Établir les paramètres du modèle et le comparer aux données observées

Donc nous l'avons dit : les auteurs se sont lancés dans la résolution d'un problème mal posé. La conséquence mathématique, c'est qu'ils ne peuvent pas espérer des résultats qui soient précis. En particulier, nous avons insisté sur le fait que la méthode proposée, outre qu'elle passait par une extravagante (et non interprétable) distribution $(\pi_k^{true} : k \in [\![0, T]\!])$, que la méthode proposée donc ***ne contrôlait pas la moyenne de la distribution***. Et de fait, les résultats

de modélisation obtenus par les auteurs seront décevants, exactement comme il fallait s'y attendre eu égard au festival d'absurdités mathématiques. En particulier, comme leur moyenne n'est pas contrôlée, l'algorithme qu'ils vont lancer va pouvoir « dériver » vers des valeurs de probabilités non négligeables pour les longues durées d'hospitalisation avant la mort (et manifestement c'est la même chose pour les durées d'hospitalisation avant mise en réanimation car, là encore, les auteurs précisent qu'ils ont dû « tenir compte du fait que l'épidémie était encore en cours »et donc que, selon eux, cela avait tendance à sur-représenter les distributions vers les faibles durées (ce que nous avons fermement contesté)).

4.3.1 Le modèle élémentaire de probabilité conditionnelle et les paramètres associés

Pour l'instant, expliquons comment nous avons produit le modèle imposé par les auteurs (et dont nous avons déjà expliqué, via la forme en siphon qu'il allait produire, qu'il serait de toute façon peu représentatif de la distribution observée π^{obs}). Ce modèle se décompose selon :

1. une loi exponentielle (pour les durées courtes) ;
2. une loi lognormale (pour les durées plus longues).

Bien évidemment, au regard du non-sens de $(\pi_k^{true} : k \in [\![0, T]\!])$, nous nous sommes abstenus de les utiliser dans la modélisation.

1. Nous avons sélectionné d'abord le jour k_0 permettant de définir la durée courte. Puisque les auteurs ont voulu introduire une loi exponentielle (et donc décroissante), il faut au moins $k_0 = 1$ pour observer une décroissance (on a alors deux données π_0^{obs}, π_1^{obs}). Cependant, sur les données suivantes, on voit que $\pi_2^{obs} = \pi_1^{obs}$ et donc la décroissance s'arrête et il n'y a plus lieu de penser que l'on est dans la densité exponentielle. Donc $k_0 = 1$.
2. À partir des $\left(\pi_k^{obs} : k \in [\![0, T]\!]\right)$ et de l'événement $B = \{X \leq k_0\}$, on reconstruit les deux probabilités $P(\cdot \mid B)$ et $P\left(\cdot \mid \overline{B}\right)$ de la manière suivante :
   (a) pour $P(\cdot \mid B)$, on part du vecteur $\left(\pi_k^{obs} : k \in [\![0, T]\!]\right)$, on garde les deux premières valeurs, et on met ensuite toutes les autres à 0. On fait la somme des valeurs qu'il reste en on divise ensuite chacune des valeurs par cette somme ;
   (b) pour $P\left(\cdot \mid \overline{B}\right)$, on part du vecteur $\left(\pi_k^{obs} : k \in [\![0, T]\!]\right)$ on met les deux premières valeurs à 0, et on garde toutes les autres valeurs. On fait la somme des valeurs qu'il reste et on divise ensuite chacune des valeurs par cette somme.

Les résultats que l'on obtient pour les distributions $P(\cdot \mid B)$ et $P\left(\cdot \mid \overline{B}\right)$ sont donnés sur le graphique suivant (voir la figure 7) (les valeurs calculées sont reportées en annexe).

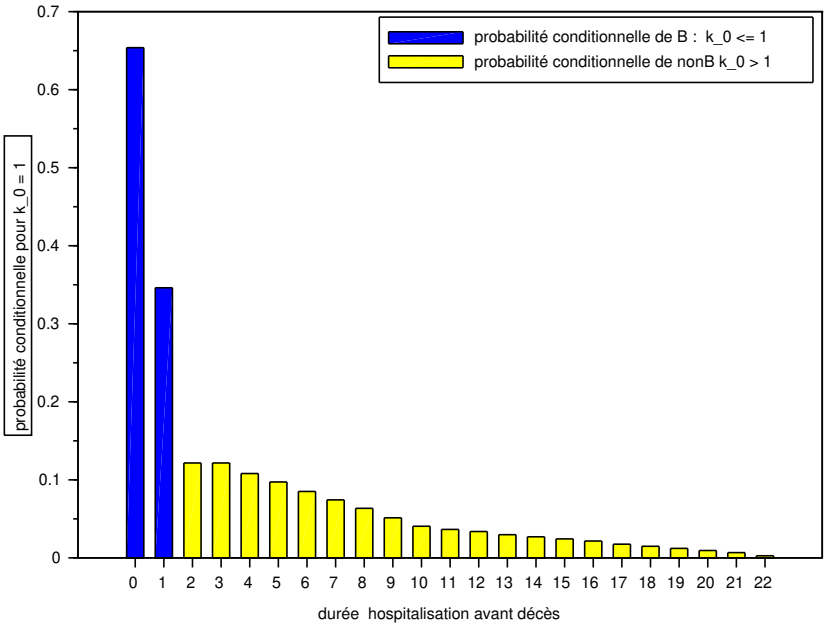

FIGURE 7 – Probabilité conditionnelle : en bleu $P(\cdot \mid B)$ et en jaune $P\left(\cdot \mid \overline{B}\right)$.

On peut alors maintenant trouver, pour chaque probabilité et selon son modèle, les paramètres m, μ, σ^2. On rappelle dans un premier temps que l'on a :

$$(1-\rho) = P(X \leq 1) = 0.17 + 0.09 = 0.26, \quad \rho = 0.74 \tag{150}$$

Ensuite, pour avoir le paramètre m de la loi exponentielle modélisant la loi $P(\cdot \mid B)$, il suffit d'effectuer la moyenne empirique de cette loi. On trouve alors :

$$\overline{\mu}_1 \left(P(\cdot \mid B)\right) = \sum_{k=0}^{k=1} \left(k + \frac{1}{2}\right) P(X = k \mid B) = 0.8461538 \; jour \tag{151}$$

On a alors :

$$P(\cdot \mid B) \sim \frac{1}{m} \exp\left(-\frac{t}{m}\right), \quad m = 0.8461538 \; jour \tag{152}$$

On peut maintenant s'intéresser aux paramètres μ, σ^2 d'une loi lognormale se rapportant à la probabilité $P\left(\cdot \mid \overline{B}\right)$ en calculant, là encore, ses moments :

$$\overline{\mu}_1 \left(P\left(\cdot \mid \overline{B}\right)\right) = \sum_{k=2}^{k=T} \left(k + \frac{1}{2}\right) P\left(X = k \mid \overline{B}\right) = 7.9040541 \; jour \tag{153}$$

$$\overline{\mu}_2 \left(P\left(\cdot \mid \overline{B}\right)\right) = \sum_{k=2}^{k=T} \left(k\,(k+1) + \frac{1}{3}\right) P\left(X = k \mid \overline{B}\right) = 85.430631 \; (jour)^2 \tag{154}$$

On utilise maintenant la formule 107 donnant $\left(\mu, \sigma^2\right)$ en fonction de $\overline{\mu}_1\left(P\left(\cdot \mid \overline{B}\right)\right)$ et de $\overline{\mu}_2\left(P\left(\cdot \mid \overline{B}\right)\right)$. On obtient ainsi :

$$\mu = \frac{1}{2}\left(4 * \ln\left(\overline{\mu}_1\left(P\left(\cdot \mid \overline{B}\right)\right)\right) - \ln\left(\overline{\mu}_2\left(P\left(\cdot \mid \overline{B}\right)\right)\right)\right) = 1.9108992 \tag{155}$$

$$\sigma^2 = \ln\left(\overline{\mu}_2\left(P\left(\cdot \mid \overline{B}\right)\right)\right) - 2 * \ln\left(\overline{\mu}_1\left(P\left(\cdot \mid \overline{B}\right)\right)\right) = 0.3129531 \tag{156}$$

ce qui nous renvoie ainsi une moyenne (déjà connue évidemment) et une valeur médiane :

$$\texttt{Moy} = \exp\left(\mu + \frac{\sigma^2}{2}\right) = 7.9040541\ jour, \quad \texttt{Med} = \exp\left(\mu\right) = 6.7591642\ jour \tag{157}$$

On peut finalement calculer le critère du « siphon » qui est :

$$F\left(\rho, m, \mu, \sigma\right) = -0.2301254 < 0 \tag{158}$$

ainsi le critère est validé et la courbe aura bien un minimum et un maximum locaux. On peut bien sûr maintenant calculer l'écart entre le modèle et l'observation. Pour cela, nous avons une formule analytique pour calculer la moyenne d'une loi exponentielle et la moyenne d'une loi lognormale sur un intervalle de la forme $[k, k+1]$ pour $k \in [\![0, T-1]\!]$:

$$I_k = \int_k^{k+1} \frac{1}{m} \exp\left(-\frac{t}{m}\right) dt = \frac{1}{m^2}\left[\exp\left(-\frac{t}{m}\right)\right]_k^{k+1} \tag{159}$$

$$J_k = \int_k^{k+1} \frac{1}{t\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\ln\left(t\right) - \mu\right)^2}{2\sigma^2}\right) dt = \frac{1}{2}\left[\texttt{Erf}\left(\frac{\ln\left(t\right) - \mu}{\sqrt{2}\sigma}\right)\right]_k^{k+1} \tag{160}$$

On posera dans la suite par $\left(\pi_k^{mod} = \left(1-\rho\right) I_k + \rho J_k : k \in [\![0, T]\!]\right)$. On rappelle que pour la fonction $f\left[\rho, m, \mu, \sigma\right](t)$, le terme π_k^{mod} dénote l'événement probabiliste $P\left(k \leq X \leq k+1\right)$, c'est-à-dire la probabilité de subir un décès le jour k (c'est-à-dire entre $24 * k$ et $24 * (k+1)$ heures après l'hospitalisation). Et pour finir, afin de préserver la normalisation des probabilités, on pose :

$$I_T = 1 - \sum_{k=0}^{k=T-1} I_k, \quad J_T = 1 - \sum_{k=0}^{k=T-1} J_k \tag{161}$$

On effectue alors l'erreur relative en norme 1 selon

$$\epsilon_1 = \sum_{k=0}^{k=T} \frac{\left|\delta_k\right|}{\left\|\boldsymbol{\pi}^{obs}\right\|_1} = \sum_{k=0}^{k=T} \left|\delta_k\right|, \quad \delta_k = \pi_k^{obs} - \left(\left(1-\rho\right) I_k + \rho J_k\right) \tag{162}$$

On obtient alors une erreur relative en norme 1 qui vaut en fait $\epsilon_1 = 0.22$, c'est-à-dire une erreur relative de 22%. On peut représenter sur le même graphique 3 choses :

1. la donnée des $\left(pi_k^{obs} : k \in [\![0, T]\!]\right)$;
2. la donnée des $\left(\pi_k^{mod}\left(1-\rho\right) I_k + \rho J_k : k \in [\![0, T]\!]\right)$. On répète ici que pour chaque entier $k \in [\![0, T-1]\!]$ cette valeur représente en fait la quantité

$P(k \leq X \leq k+1)$ tandis que l'on a $\pi_T^{mod} = P(X > T)$. Ainsi, eu égard à la définition choisie, la dernière barre de π_k^{mod} représente, puisque le modèle continu peut avoir des durées d'hospitalisation qui vont jusqu'à un temps infini, la probabilité $P(X > 23)$;

3. la densité de probabilité $f[\rho, m, \mu, \sigma](t)$.

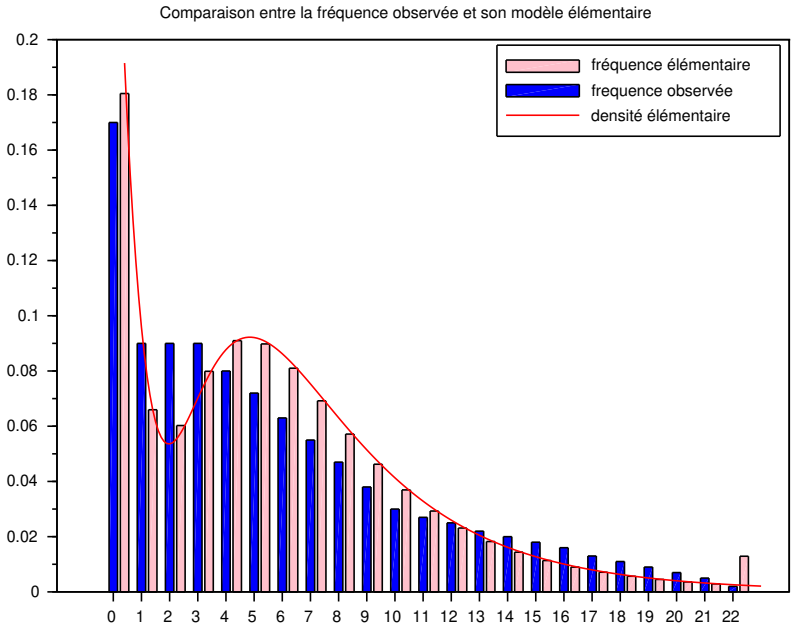

FIGURE 8 – Comparaison entre les données observées et les données modélisées de façon "élémentaire" avec un formalisme rigoureux.

On voit dans cette figure que dans le modèle continu la probabilité $P(X \geq 23)$ n'est pas très importante (elle vaut 0.013) : le modèle continu a donc une fréquence cumulée qui est très faible pour l'événement $\{X > 23\}$. De fait, et puisque cela est contrôlé par la méthode des moments, on a exactement :

$$\mathbb{E}_{obs}(X) = \mathbb{E}_{mod}(X) = 6.069 \tag{163}$$

On voit effectivement que c'est la forme en « siphon » de la densité qui provoque, sur les fréquences moyennes, d'abord une sous-estimation puis une surestimation par rapport aux fréquences observées. Ainsi, sur les jours $k \in \{1, 2, 3\}$, on observe $\{0.9, 0.9, 0.9\}$ alors que le modèle rend $\{0.66, 0.61, 0.79\}$.

4.3.2 Le modèle des auteurs

Nous allons d'abord retrouver les paramètres (ρ, m, μ, σ). Dans le tableau S_3 que nous avons reproduit, on donne comme paramètres (pour l'ensemble des

personnes hospitalisées ayant subi un décès) :

$$\rho = 1 - P\left(short\ delay\right) = 0.85 \quad m = 0.67\ \ (jour) \tag{164}$$

Les auteurs donnent également la moyenne $\texttt{Mean}_D$ et la médiane $\texttt{Med}_D$ associées à la distribution lognormale :

$$\texttt{Mean}_D = 13.2\ \ (jours)\,, \quad \texttt{Med}_D = 8.6 \tag{165}$$

On rappelle que la moyenne et la médiane d'une loi lognormale sont liées aux paramètres μ, σ^2 par la relation :

$$\texttt{Mean}_D = \exp\left(\mu + \frac{\sigma^2}{2}\right), \quad \texttt{Med}_D = \exp\left(\mu\right)\ \ jours \tag{166}$$

On en tire donc les valeurs des auteurs :

$$\mu = \ln\left(\texttt{Med}_D\right) = \ln\left(8.6\right) = 2.1517622 \tag{167}$$

$$\sigma^2 = 2 * \left(\ln\left(13.2\right) - \ln\left(8.6\right)\right) = 0.8569093 \tag{168}$$

On peut donc maintenant évaluer :

$$F\left(\rho, m, \mu, \sigma\right) = -0.018 \tag{169}$$

Donc, on a bien un minimum puis un maximum : la courbe présente également un aspect en « siphon »même si l'écart entre le minimum local et le maximum local n'est pas très important. Les résultats sont regroupés dans la figure 9 ci-dessous.

Notons que l'erreur relative générée par ce modèle vaut

$$\epsilon_2 = \frac{\left\|\pi_k^{mod} - \pi_k^{obs}\right\|_1}{\left\|\pi_k^{obs}\right\|_1} = \left\|\pi_k^{mod} - \pi_k^{obs}\right\|_1 = 0.33 \tag{170}$$

Soit 33% d'erreur relative au lieu de 22% dans le modèle précédent : on voit à quel point les auteurs se moquent du monde. Ce qui frappe dans cette figure, c'est la surestimation exécrable de l'événement $P\left(X > 23\right)$: de 1% dans le modèle rigoureux, on arrive à 15% dans le modèle des auteurs. Cela a ainsi une conséquence très importante sur l'estimation de la valeur moyenne. De fait, alors que dans le modèle précédent la valeur moyenne était parfaitement contrôlée, dans le modèle des auteurs, la valeur moyenne de X devient deux fois ce qu'elle est *en réalité* :

$$\mathbb{E}_{mod}\left[X\right] = 11.32 \sim 2 * \mathbb{E}_{obs}\left[X\right] \tag{171}$$

Il est évident que cela est totalement inacceptable. Il y a trois crimes qu'il ne faut jamais commettre lorsque l'on cherche des lois "continues" qui viennent modéliser des phénomènes observables :

1. trouver une densité qui peut prendre des valeurs négatives ;
2. trouver une densité dont le moment d'ordre 0 n'est pas égal à 1 ;
3. trouver une densité qui ne respecte pas la valeur moyenne.

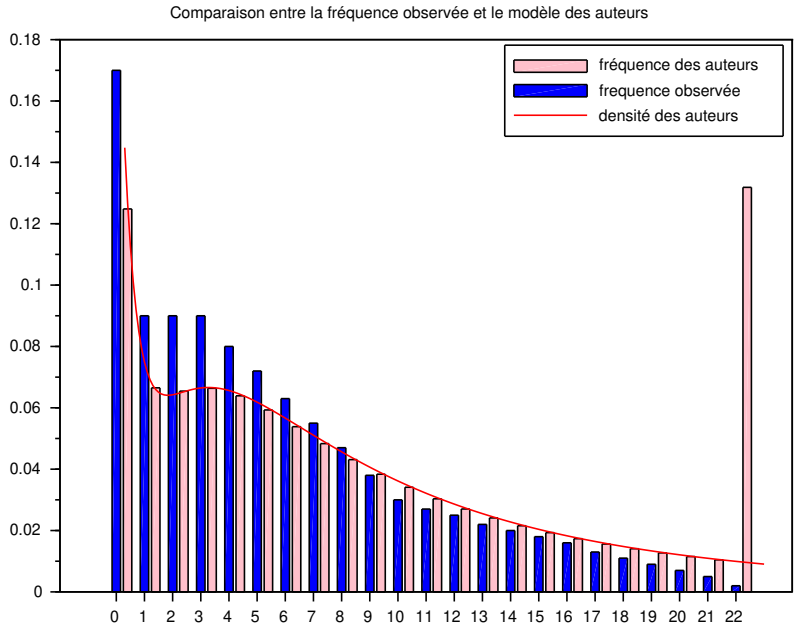

FIGURE 9 – Comparaison entre les fréquences observées et les fréquences du modèle trouvé par les auteurs.

De ce point de vue le "modèle" des auteurs est un crime abominable sur les modélisations en probabilité. D'où vient cette incurie du modèle ? Tout simplement de leur algorithme des moindres carrés. Comme nous l'avons dit, en définissant la fonction d'erreur 116

$$E(\theta) = \sum_{k=0}^{k=T} \left(\pi_k^{obs} - \frac{\alpha_k \int_{t=k}^{t=k+1} f[\theta](t)\, dt}{\sum_{l=0}^{l=T} \alpha_l \int_{t=l}^{t=l+1} f[\theta](t)\, dt} \right)^2 \tag{172}$$

les auteurs se sont lancé comme objectif de résoudre une équation qui n'avait pas de solution. Dès lors, ils ne savent plus vraiment où l'algorithme va les emmener (c'est pour cela que les mathématiciens sérieux ne se lancent jamais dans ce genre de problèmes). En particulier, dans cette équation sans solutions, la fonction d'erreur ne cherche pas à minimiser la différence sur la valeur moyenne entre l'observation et le modèle. Ainsi, le risque c'est de voir dériver le modèle vers une fonction qui ne vérifie pas la contrainte de la moyenne. En fait, pour comprendre comment va se comporter l'algorithme des moindres carrés, il faut noter qu'il va essayer de retrouver les valeurs de fréquence pour $k \in \{1, 2, 3\}$, c'est-à-dire qu'il va essayer *de reconstruire un plateau* sur ces trois valeurs. Or, on l'a dit : la fonction qui modélise possède la plupart du temps d'abord un minimum local puis un maximum local. Donc la recherche d'un plateau va se faire vraisemblablement en essayant de remonter le minimum local et simultanément en descendant le maximum local de sorte que les deux soient

le plus proche possible. De fait, la courbe est effectivement faiblement variable dans la zone $k \in \{1,2,3\}$ ce qui a l'air satisfaisant. Cependant, cela n'évite pas la sous-estimation des fréquences à ce niveau puisque les auteurs trouvent $\{0.066513 \quad 0.0654426 \quad 0.0663611\}$ au lieu de $\{0.9, 0.9, 0.9\}$, mais surtout cela se paye très chèrement ensuite sur l'estimation des fréquences pour l'événement $\{X > 23\}$. En effet, en forçant le « plateau »pour $k \in \{1,2,3\}$, on empêche le modèle de « replonger »significativement pour les jours suivants. Ce qui fait que la distribution devient catastrophique pour les $k > 23$, surestimant ainsi de 15% le nombre de patients qui vont mourir après 23 jours d'hospitalisation. Et c'est précisément là que les auteurs vont inventer l'incroyable excuse explicative de la « growing epidemy ».

Les auteurs sont déçus effectivement que leur modèle sous-estime systématiquement les faibles valeurs de k (la valeur modélisée est effectivement systématiquement en dessous de la valeur observée pour les $k \leq 8$), tandis que le modèle surestime les valeurs de $k \geq 23$ et de façon très significative puisque dans le modèle on a $P(X > 23) = 0.15$ (alors que dans l'observation c'est totalement négligeable). D'où la figure de style imposée aux auteurs :
« En fait notre modèle est parfaitement juste, mais ce sont les observations qui sont complètement fausses. En effet, l'épidémie n'étant pas finie, il y a des tas de gens qui vont mourir après des dates très longues d'hospitalisation, mais évidemment, ceux-là on ne peut pas encore les voir. De fait, à la fin de l'épidémie, une partie des fréquences pour les $\{X < 10\}$ sera forcément reportée sur les $\{X > 23\}$. CQFD. »

Elle est pas belle la vie du monde des Idées ? On a quand même beaucoup de peine pour les auteurs car :

1. nous avons dissipé l'idée qu'avec plus de temps d'observation, la distribution $\boldsymbol{\pi}^{obs}$ puisse effectivement se déplacer significativement vers la droite ;
2. si la distribution change, il faudra relancer leur algorithme et ils trouveront encore des paramètres différents avec leur modèle et donc encore une courbe qui risque de ne pas coller aux données ;
3. nous avons en fait expliqué que la principale raison du très mauvais comportement de leur modèle vient essentiellement du fait qu'ils ne comprennent rien à rien, ni à la modélisation, ni aux probabilités, ni aux méthodes numériques.

En fait, ce qu'ont choisi alors de faire les auteurs, c'est de **DÉCALER À LA MAIN** la fonction de modélisation. En gros, cela revient à poser un changement dans la fonction lognormale de la façon suivante :

$$f\left[maquillée\right](t) = \frac{\eta(t)}{t\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\ln(t) - \mu\right)^2}{2\sigma^2}\right) \tag{173}$$

voilà donc le "fit" que les auteurs ont introduit dans leur "modélisation". Connaissant à l'avance le résultat et ne pouvant croire que leur modélisation et méthode numérique étaient en fait inconsistantes, ils ont fini par maquiller leurs résultats. Il s'agit là bien entendu d'une escroquerie scientifique majeure : les auteurs veulent absolument masquer leur incapacité à trouver des modèles. En fait, ils

vivent dans un monde où ils supposent que tout phénomène à une modélisation parfaite.

Nous donnons ici quelques indications sur la manière avec laquelle on peut construire la fonction $\eta(t)$. D'abord, il est clair que les auteurs n'ont cherché la fonction $\eta(t)$ que pour les $t \geq 1$, ce qui explique le point anguleux dans leur courbe (voir figure (A) sur la figure (1)) (alors que le modèle étant régulier, il est impossible d'avoir un tel point anguleux). Ensuite, vous effectuez le rapport :

$$k \in [\![1, T]\!], \quad \eta_k = \frac{\pi_k^{obs}}{\pi_k^{mod}}, \quad \pi_k^{mod} = \int_k^{k+1} f\left[\rho, m, \mu, \sigma\right](t)\, dt \tag{174}$$

Puis vous cherchez une fonction régulière $\eta(t)$ (disons de classe C^1, c'est-à-dire continûment dérivable pour éviter justement les points anguleux) telle que :

$$\forall k \in [0, T], \quad \eta\left(k + \frac{1}{2}\right) = \eta_k \tag{175}$$

On peut faire une telle construction avec des splines d'interpolation : une telle construction est parfaitement documentée. Alors, dans ce cas, la fonction $\eta(t)$ devrait parfaitement faire l'affaire... D'ailleurs plutôt que de passer par une telle succession d'incohérences grossières, scandaleuses et totalement non scientifiques, il aurait été sûrement beaucoup plus efficace, si l'on cherchait effectivement à trouver une courbe qui permettait de passer par tous les points de la courbe, de procéder directement à l'emploi d'une interpolation par des splines.

4.4 Conclusion sur le modèle de la durée d'hospitalisation avant décès

Mais au final, on pourra se demander à quoi peut bien servir le fait de trouver un modèle continu à une observation discrète. Dans ce cas précis, la réponse est simple : ABSOLUMENT À RIEN. Comme nous l'avons dit, le passage au continu revient, dans cet exercice, à essayer d'aller chercher une information que l'on n'a pas : établir avec une précision infinie l'instant exact où une personne va décéder, alors même que cette donnée n'est que journalière. Rien, aucun intérêt. Alors pourquoi les auteurs se sont-ils intéressés à cette question ?

1. D'abord, comme ils comprennent rien aux probabilités et à la modélisation, ils n'ont même pas remarqué que le problème qu'ils s'étaient posé relevait du niveau de la classe de Terminale. Ainsi, la superposition des lois qu'ils cherchaient n'était qu'une bête application des probabilités conditionnelles. Le premier exemple que l'on donne dans n'importe quel TD en classe au lycée.
2. Ensuite, comme ils ne comprennent rien à la résolution des problèmes inverses (problème linéaire de Frobenius et méthode des moments), ils ont cru que le problème qu'ils se posaient était complexe à résoudre. Ils l'ont d'ailleurs rendu complexe en le posant de telle sorte que l'on soit sûr qu'il n'ait pas de solution. Mais comme la méthode des moindres carrés permet toujours de trouver une réponse à tous les problèmes, y compris ceux qui n'ont pas de solution, cela ne les a pas dérangés. Au contraire, en rendant compliqué un problème simple, ils se donnaient à bon compte, c'est-à-dire

pour ceux qui ne comprennent rien aux mathématiques, l'illusion d'une compétence sérieuse.

3. **Surtout, les auteurs ont essayé de faire croire qu'ils avaient un talent exceptionnel en modélisation. Qu'ils savaient poser des équations, les résoudre. Et qu'avec leur technique, ils parvenaient trouver des « modèles » qui fittaient parfaitement les données expérimentales lorsque celles-ci étaient connues. Il était important pour les auteurs d'en passer par une équation et par un modèle utilisant des lois probabilistes connues, car, sur les questions où précisément l'on a pas de données empiriques, seules la consistance des équations que l'on pose et l'habilité que l'on a à les résoudre numériquement fait gage de compétences pour aborder les problèmes posés.**

C'EST EN CELA QUE L'ON PEUT PARLER DE MAQUILLAGE : non ce n'est pas *un modèle* qui leur a donné le résultat, mais un bête passage par la théorie de l'interpolation. L'interpolation est une théorie mathématique très utile. Cependant L'INTERPOLATION NE FONCTIONNE QUE SI L'ON CONNAÎT DÉJÀ CE QUE L'ON VEUT « MODÉLISER ». Or, ce n'est pas là ce que prétendent faire les auteurs. *Un modèle, cela consiste à poser par un raisonnement, une équation sur une grandeur intéressante dont la résolution mathématique nous permettra de trouver un résultat* ***sans le connaître à l'avance***. Il y a donc dans l'idée de modèle le fait d'être capable de prédire un résultat avec une méthode (ou une équation) qui est toujours la même pour toutes les situations qui relèvent d'un même phénomène. Si on entend par modèle le fait de devoir trouver des courbes qui passent par des points que l'on connaît à l'avance, alors il y a bien longtemps que la théorie de l'interpolation a trouvé des réponses remarquablement efficaces. Mais ce faisant, il n'y a plus d'interprétation physique à donner aux fonctions que l'on manipule et, grave inconvénient pour nos auteurs, il n'y a plus non plus de moyen de dire que l'on *sait raisonner pour trouver ensuite des résultats qui ne seraient pas déjà connus à l'avance.* Ainsi, cet exercice de « modélisation » avait pour but :

1. de montrer que les auteurs savaient mettre un problème en équation ;
2. De montrer ensuite qu'ils savaient le résoudre.

Évidemment, nous avons parfaitement établi qu'ils ne savaient faire ni l'un ni l'autre, et ce, dans des proportions absolument abyssales. Sans passer par la théorie de l'interpolation, *leur « modèle » devenait ainsi largement inférieur à celui que peut construire n'importe quel élève de Terminale, pour peu que son enseignant le dirige en TD.*

Ainsi, il apparaît une chose remarquable : en essayant de montrer qu'ils possédaient un savoir qu'ils n'avaient pas, les auteurs ont fait preuve d'une sidérante révélation : ils ont dévoilé qu'ils étaient prêts à tout, même à mentir sur les méthodes, pour forcer un résultat dont ils avaient envie qu'il soit vrai. Ils révèlent en fait qu'ils ne maîtrisent rien à la modélisation ni aux méthodes de résolution. Ainsi, il faut que vous le sachiez : lorsque les auteurs ont prétendu que le confinement était efficace, c'est simplement qu'ils **l'ont décrété de façon arbitraire**, exactement comme à la fin de leur « modèle » de durée d'hospitalisation, ils ont fini par ajuster à leur guise le résultat qu'ils voulaient obtenir. Nous allons voir

dans la section suivante, plus rapidement, en quoi leur prétendue modélisation de l'efficacité du confinement EST UNE AFFIRMATION AUSSI GRATUITE QUE STUPIDE.

5 Confinement et nombre de reproduction de base

L'efficacité du confinement sur le développement de l'épidémie, voilà bien une affirmation dont le gouvernement a besoin qu'on lui fournisse des bases « scientifiques » : la dureté de l'épreuve est telle en effet que, s'agissant comme il se doit d'évaluer un rapport bénéfice/risque, il s'agit évidemment de montrer que le bénéfice est largement supérieur à toute forme de risque à tendance humaine non sanitaire (économique, psychologique, juridique, etc.). Or, on sait déjà, à l'instar de deux pays nordiques, les Pays-Bas et la Suède, qu'il y a peu de chances, hélas, au final que ce soit le cas. Ainsi, l'appui de la communauté scientifique paraît indispensable. Au moins pour un moment. Il s'agit donc de montrer « scientifiquement » que le confinement ça marche. Cette « démonstration scientifique », les auteurs de l'article espèrent la placer sur la discussion du R_0 : le nombre de reproduction de base. Nous allons présenter rapidement cette théorie, avant de voir en quoi la manipulation de ce R_0 par les auteurs de l'article est aussi arbitraire que consternante. Avant d'aller plus loin, sachez une chose :

Il est certain que les auteurs de l'article ne comprennent rien à la théorie du R_0 car cette théorie est précisément basée sur l'étude des opérateurs positifs, exactement celle qui donne lieu au théorème de Perron-Frobenius, précisément celui que les auteurs ont totalement manqué dans l'étude du système qui devait les mener de $\boldsymbol{\pi}^{obs}$ à $\boldsymbol{\pi}^{true}$.

Il s'agit, comme on le on dit, de planter le décor. Sachez que la théorie des opérateurs positifs est l'une des plus magnifiques qui soient. J'ai eu à l'utiliser il y a quelques années lors de la rédaction d'un article. La littérature sur le sujet est aussi dense que riche et inépuisable. Cela dit, n'ayez pas peur, nous présenterons juste les résultats sans démonstration. Il s'agit juste de mettre en place le vocabulaire et de se représenter les choses. Eu égard au niveau mathématique dramatiquement faible des auteurs, ce vocabulaire leur sera manifestement utile, à eux aussi, pour affiner leur perception de la chose. Comme on a vu qu'ils ne comprenaient pas grand-chose aux probabilités, il y a peu de chance qu'ils maîtrisent parfaitement la théorie du R_0, laquelle, pour le coup, se situe plutôt à un niveau M_1 ou M_2, et sans doute dans les formations de mathématiques relativement théoriques.

5.1 Présentation du modèle structurant la population et des équations correspondantes

D'abord, posons le cadre de l'étude. On travaille toujours au sein d'une population (c'est le propre d'ailleurs de l'épidémiologie). Mais ici la population qui va nous intéresser est la population totale d'un pays. On note par Ω cet ensemble. L'une des caractéristiques, c'est que son cardinal (i.e. le nombre d'individus de cette population) est effectivement très grand. Dans la suite, on le

notera par N. Ce qui autorise effectivement pas mal de facilité et en particulier de travailler de façon satisfaisante dans un cadre continu. Cependant, dans la mesure où le formalisme continu est plus difficile à se représenter et à manipuler pour ceux qui n'y sont pas habitués, nous allons présenter les choses dans le cadre discret. En particulier, on rappelle que l'on munit Ω de la probabilité uniforme, c'est-à-dire que :

$$\forall B \subset \Omega, \quad \mathbb{P}(B) = \frac{\texttt{card}(B)}{\texttt{card}(\Omega)} = \frac{N(B)}{N} \tag{176}$$

où $N(B)$ désigne le nombre de personnes appartenant au sous ensemble B.

À nouveau, on va parler des sous-ensembles grâce aux variables aléatoires. Dans ce qui va nous servir le plus, il y a la variable d'âge. Mais également pour le COVID19, il semblerait que la variable du genre soit aussi significative : ainsi, non seulement l'on observe des différences entre les âges, mais également entre les sexes. Il y a donc deux variables aléatoires qui nous intéressent :

$$A : \Omega \to [\![0, M]\!], \quad S : \Omega \to \{0, 1\} \tag{177}$$

où A désigne la variable aléatoire de l'âge (M étant l'âge constaté de la plus vieille personne) et S la variable du sexe : 0 pour un homme et 1 pour une femme. On peut alors découper la population totale en classes d'âge, et en classes de genre. Ainsi on se donne une suite :

$$k_0 = 0 < k_1 < k_2 \cdots < k_L < M \tag{178}$$

et on considère ainsi le sous-ensemble de Ω défini par :

$$\forall l \in [\![1, L]\!], \quad A_l = \{k_{l-1} \leq X < k_l\} = \{\omega \in \Omega,\ k_{l-1} \leq X(\omega) < k_l\} \tag{179}$$

Pour les auteurs de l'article, on a $k_l = 10 * l$: autrement dit, on considère des classes d'âge par tranche de 10 ans et on s'arrête à $L = 8$, ce qui signifie que toutes les personnes de plus de 80 ans sont regroupées dans la dernière classe d'âge. On définit maintenant la variable aléatoire du genre par la formule suivante :

$$\forall g \in \{0, 1\},\ S_g = \{S = g\} = \{\omega \in \Omega \textit{ tel que } S(\omega) = g\} \tag{180}$$

De fait, l'ensemble $A_3 \bigcap S_0$ désigne l'ensemble des personnes dont l'âge est compris entre 30 et 40 ans et qui sont de sexe masculin. Il existe évidemment une autre variable aléatoire d'intérêt, c'est celle qui va associer à chaque individu son état d'infection. La difficulté de cette variable, c'est qu'elle dépend du temps. Il s'agit donc de ce que l'on convient d'appeler un processus stochastique. L'occasion de définir cette notion (mais de façon totalement particulière pour les questions qui nous intéressent : en vérité la définition est plus générale que cela) :

Definition 5. *Soit l'ensemble $\mathbb{R} \times \Omega$. On appelle processus stochastique toute application*

$$I : \mathbb{R} \times \Omega \mapsto \{0, 1\} \tag{181}$$

telle que pour tout $t \in \mathbb{R}$, l'application $I(t, \cdot)$ soit une variable aléatoire sur l'ensemble Ω à valeur dans $\{0, 1\}$.

Pour nous, la fonction $I(t,\omega)$ renvoie la valeur 0 si l'individu ω n'est pas infecté à l'instant t et renvoie la valeur 1 si l'individu est infecté à l'instant t. Il est clair que ce que l'on cherche à connaître c'est la dynamique de ce processus stochastique.

Dans la suite, on adoptera le vocabulaire suivant, dit vocabulaire des états, qui va permettre de structurer notre discours sur l'épidémie et sur les personnes infectées.

Definition 6. *Soit $l \in [\![1, L]\!]$ et $g \in \{0,1\}$. On dit que l'individu ω est dans l'état (s,l) si on a $\omega \in A_l \bigcap S_g$*

On peut définir maintenant, par un formalisme tout à fait précis, les quantités dont on cherche à évaluer la dynamique. En particulier le nombre d'individus dans Ω qui sont infectés à l'instant t. On désignera par $\mathcal{I}(t,(l,g))$ le nombre de personnes de l'ensemble $A_l \bigcap S_g$ qui sont infectées à l'instant t. On a donc :

$$\begin{aligned} \mathcal{I}(t,(l,g)) &= \mathtt{card}\left(\{I(t,\cdot)=1\} \bigcap A_l \bigcap S_g\right) && (182)\\ &= N\mathbb{P}\left(\{I(t,\cdot)=1\} \bigcap A_l \bigcap S_g\right) && (183) \end{aligned}$$

Si besoin, dans la suite, on désignera par $\mathbf{e} := (l,g) \in [\![1, L]\!] \times \{0,1\}$ un état. Au besoin on pourra munir l'ensemble des états de l'ordre lexicographique (i.e. l'ordre du dictionnaire), c'est-à-dire que l'on notera :

$$\forall \mathbf{e} = (l,g)\,,\ \mathbf{f} = (l^*,g^*)\,,\quad \mathbf{e} \leq \mathbf{f} \Leftrightarrow [l < l^*] \quad ou \ \ [l = l^* \ et \ g \leq g^*] \tag{184}$$

ainsi, on classe d'abord les personnes par leur âge, et ensuite, pour un âge donné, on les classe par leur sexe. On a donc dans cet ordre-là la classification suivante :

$$(1,0) < (1,1) < (2,0) < (2,1) < \cdots \tag{185}$$

l'ordre lexicographique est un ordre total. D'une façon synthétique, on notera alors par :

$$\mathbf{1} = (1,0)\,,\ \mathbf{2} = (1,1)\,,\ \mathbf{3} = (2,0)\,, \cdots, \mathbf{2L} = (L,1) \tag{186}$$

l'algorithme étant alors que si on écrit $\mathbf{n} = (l,g)$, on a « symboliquement » sous forme typographique $\mathbf{n} = \mathbf{bold}\{(2 * l - 1) + g\}$. Cela nous permettra d'écrire des sommes portant sur les deux indices en même temps tout en ayant un ordre pour organiser nos calculs si besoin. S'il le faut, nous reviendrons sur des doubles sommes avec les indices (l,g). Pour la suite, l'état $\mathbf{e} \in [\![1, L]\!] \times \{0,1\}$ étant donné, on notera par $N(\mathbf{e})$ le nombre d'individus de l'état $\mathbf{e}$. C'est-à-dire que l'on a :

$$\forall \mathbf{e} = (l,g)\,,\ N(\mathbf{e}) = \mathtt{card}\left(A_l \bigcap S_g\right) \tag{187}$$

Voilà donc pour les notations et la notion d'état associé à un individu : un état est donc à la fois une classe d'âge et un sexe. Avec nos nouvelles notations, on peut donc écrire que l'on a :

$$N = \sum_{\mathbf{e}=\mathbf{1}}^{\mathbf{e}=\mathbf{2L}} N(\mathbf{e}) \tag{188}$$

Maintenant nous allons définir l'OBJET ESSENTIEL DE LA THÉORIE DU NOMBRE BASIQUE DE REPRODUCTION. Il s'agit de ce que l'on peut appeler le taux d'infectivité (en anglais « *infectivity* ») que l'on note par $\mathcal{A}(\tau, \mathbf{e}, \mathbf{f})$ et qui va permettre de dénombrer le nombre de personnes de l'état $\mathbf{e}$ qu'un individu, qui appartient à l'état $\mathbf{f}$, va contaminer. Plus précisément :

1. on suppose que l'on est à l'instant 0 et que l'on a choisi un temps de translation $\tau \geq 0$ qui dit que l'on se place dans le passé au temps $t = 0 - \tau$.
2. On considère un temps « infinitésimal » $d\tau$. Entre l'instant $t = -\tau$ et $t + d\tau$, on suppose que le nombre de personnes de l'état $\mathbf{e}$ infectées par le contaminant de l'état $\mathbf{f}$ se calcule selon :

$$N(\mathbf{e})\,\mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\,d\tau \tag{189}$$

   Cela justifie que $\mathcal{A}$ soit qualifié de « taux ». En effet, pour avoir le nombre total d'individus infectés, il faut multiplier $\mathcal{A}$ par une durée et par l'ensemble des individus de la classe $\mathbf{e}$, c'est-à-dire par $N(\mathbf{e})$.
3. Ainsi, on peut déduire le nombre total de *nouvelles* personnes de la classe $\mathbf{e}$ ayant été contaminées jusqu'à l'instant $t = 0$ par une personne de la classe $\mathbf{f}$ par :

$$N(\mathbf{e}) \int_{\tau=0}^{\tau=+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\,d\tau \tag{190}$$

4. Il y a évidemment une contrainte que l'on doit vérifier, c'est que :

$$\forall \mathbf{e}, \mathbf{f}, \quad N(\mathbf{e}) \int_{\tau=0}^{\tau=+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\,d\tau \leq N(\mathbf{e}) \tag{191}$$

   vous ne pouvez pas contaminer plus de personnes de l'état $\mathbf{e}$ qu'il n'existe, dans la population, de personnes de l'état $\mathbf{e}$ (!). En réalité, on travaillera dans l'hypothèse dite de *linéarisation*. C'est-à-dire que l'on aura :

$$\forall (\mathbf{e}, \mathbf{f}), \quad N(\mathbf{e}) \int_{\tau=0}^{\tau=+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\,d\tau << N(\mathbf{e}) \tag{192}$$

   Les Anglais diraient que l'on a besoin que la quantité ci-dessus (le membre de gauche dans l'égalité) soit « sizable », c'est-à-dire d'un ordre de grandeur ni trop grand (et donc très loin de $N(\mathbf{e})$) ni trop petit. En pratique, la quantité précédente sera donc de l'ordre de quelques unités. On va voir comment être plus précis sur cette quantité. La littérature, pour parler de cette hypothèse, parle de régime linéarisé. Il semble plus parlant, pour reprendre ce qui se fait en physique, de parler de ***régime dilué***.
5. Pour les gens qui aiment la physique et qui sont amateurs de physique statistique sur les particules, la quantité

$$N(\mathbf{e}) \int_{\tau=0}^{\tau=+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\,d\tau \tag{193}$$

   est assimilable à une section de collision. Reprenons ainsi l'expérience de Rutherford. Cela consiste à envoyer un grand nombre de particules avec

une certaine vitesse (disons l'état $\mathbf{f}$) sur une surface S qui contient une particule cible et à compter toutes celles qui ressortent avec une autre vitesse (appelons cela l'état $\mathbf{e}$) (car elles ont été déviées à cause de la particule cible). Alors notera par $N(\mathbf{e})\sigma(\mathbf{e},\mathbf{f})/S$ le nombre de particules déviées. $\sigma(\mathbf{e},\mathbf{f})$ s'appelle la section (différentielle) de collision. On voit que l'on est ici, avec le taux d'infectivité, dans la même démarche : introduire un opérateur de comptage qui permet de dénombrer des individus subissant un changement d'état à cause d'une « rencontre ». On pourra même, comme pour la physique, appeler le terme $N(\mathbf{e})\mathcal{A}(\tau,\mathbf{e},\mathbf{f})$ un ***taux différentiel d'infectivité*** , tandis que la quantité :

$$N(\mathbf{e})\int_{\tau=0}^{+\infty}\mathcal{A}(\tau,\mathbf{e},\mathbf{f})\,d\tau \tag{194}$$

pourrait se concevoir comme ***un taux efficace d'infectivité***.

6. Le nombre total de personnes infectées par une personne de l'état $\mathbf{f}$ à l'instant $t=0$ se calcule donc sous la forme :

$$\sum_{\mathbf{e}=\mathbf{1}}^{\mathbf{e}=\mathbf{2L}} N(\mathbf{e})\int_{\tau=0}^{+\infty}\mathcal{A}(\tau,\mathbf{e},\mathbf{f})\,d\tau \tag{195}$$

cet opérateur s'appelle « l'opérateur de la génération suivante » (en anglais « *next gen operator* »). Il est aussi connue sous le nom de WAIFW-matrix « Who Acquires Infection From Whom »(*la matrice de qui se fait contaminer par qui*).

DANS LA SUITE, POUR ALLEGER LES NOTATION, NOUS UTILISERONS LA NOTATION

$$\mathcal{A}(\tau,\mathbf{e},\mathbf{f})$$

EN LIEU ET PLACE DE LA NOTAION

$$N(\mathbf{e})\mathcal{A}(\tau,\mathbf{e},\mathbf{f})$$

Il faut maintenant noter deux choses très importantes, qui vous nous permettre de mieux comprendre comment se construit la théorie du R_0 :

1. d'abord, pour chaque translation chronologique τ, et pour chaque couple d'état $(\mathbf{e},\mathbf{f})\in[\mathbf{1},\mathbf{2L}]\times[\mathbf{1},\mathbf{2L}]$, le taux différentiel d'infectivité $\mathcal{A}(\tau,\mathbf{e},\mathbf{f})$ est une MATRICE ;
2. chaque élément de cette matrice est un nombre réel POSITIF (OU NUL) ;
3. ainsi, le taux efficace d'infectivité

$$N(\mathbf{e})\int_{\tau=0}^{+\infty}\mathcal{A}(\tau,\mathbf{e},\mathbf{f})\,d\tau \tag{196}$$

est lui aussi une matrice n'ayant que des termes positifs ou nuls.

AINSI LA THÉORIE DU R_0 EST UNE THÉORIE LIÉE AUX MATRICES POSITIVES. Il est important de noter cela car la théorie des matrices positives va nous permettre de comprendre ce qu'il en est de la définition du R_0 et de ce que l'on peut tirer comme information. Le document de Jean-Etienne Rombaldi sur la théorie de Perron-Frobenius que nous avons déjà cité et que l'on peut trouver facilement sur la Toile nous sera encore d'un secours inestimable. Nous ne pouvons que remercier les mathématiciens qui rendent gracieusement disponible leur travail. Avant cela, nous terminons cette partie en établissant l'équation d'évolution du taux d'infectivité par unité de temps en utilisant le taux différentiel d'infectivité.

Rappelons que pour $(t, \mathbf{e}) \in \mathbb{R} \times [\![\mathbf{1}, \mathbf{2L}]\!]$, on note $\mathcal{I}(t, \mathbf{e})$ le nombre total d'individus dans l'état $\mathbf{e}$ qui sont infectés (et donc infectants) à l'instant t. Plaçons-nous à un instant t. On veut déterminer le nombre de *nouvelles* personnes qui sont infectées à l'instant t. Notons par $J_{new}(t)$ ce nombre de personnes. Pour cela, on va se translater d'un temps $\tau > 0$ dans le passé. À $t - \tau$ il y a :

1. pour chaque état $\mathbf{f}$ un nombre $\mathcal{I}(t - \tau, \mathbf{f})$ de personnes infectées.
2. Entre τ et $\tau + d\tau$, chaque personne de l'état $\mathbf{f}$ contamine un nombre de personnes de l'état $\mathbf{e}$ donné par :
$$\mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\, d\tau. \tag{197}$$
ce qui fait, puisque l'on a $\mathcal{I}(t - \tau, \mathbf{f})$ personnes de l'état $\mathbf{f}$ qui sont contaminantes, que le nombre total de personnes de l'état $\mathbf{e}$ qui sont infectées par l'état $\mathbf{f}$ s'écrit :
$$\mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\, \mathcal{I}(t - \tau, \mathbf{f})\, d\tau. \tag{198}$$
3. En intégrant sur le passé ($\tau \in [0, +\infty)$) et en sommant sur tous les états $\mathbf{f}$, on obtient ainsi, à l'instant t, le nombre total de nouvelles personnes de l'état $\mathbf{e}$ qui sont infectées à l'instant t. On a ainsi :
$$J_{new}(t, \mathbf{e}) = \sum_{\mathbf{f}=\mathbf{1}}^{\mathbf{f}=\mathbf{2L}} \int_{\tau=0}^{+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\, \mathcal{I}(t - \tau, \mathbf{f})\, d\tau \tag{199}$$
4. En effectuant le même raisonnement à l'instant $t + dt$, on obtient rapidement :
$$J_{new}(t + dt, \mathbf{e}) = \sum_{\mathbf{f}=\mathbf{1}}^{\mathbf{f}=\mathbf{2L}} \int_{\tau=0}^{+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\, \mathcal{I}(t + dt - \tau, \mathbf{f})\, d\tau \tag{200}$$
En faisant alors la différence et en divisant par dt, il vient :
$$\frac{J_{new}(t + dt, \mathbf{e}) - J_{new}(t, \mathbf{e})}{dt} = \sum_{\mathbf{f}=\mathbf{1}}^{\mathbf{f}=\mathbf{2L}} \int_{\tau=0}^{+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\, \frac{\mathcal{I}(t + dt - \tau, \mathbf{f}) - \mathcal{I}((t - \tau, \mathbf{f}))}{dt} d\tau \tag{201}$$
On pose maintenant par $i(t, \mathbf{f})$ le taux (par unité de temps) d'infectivité, c'est-à-dire le nombre de nouvelles personnes de l'état $\mathbf{f}$ infectées par unité de temps. Il est clair que l'on a :
$$J_{new}(t + dt, \mathbf{e}) - J_{new}(t, \mathbf{e}) = i(t, \mathbf{e})\, dt \tag{202}$$
$$\mathcal{I}(t + dt - \tau, \mathbf{f}) - \mathcal{I}((t - \tau, \mathbf{f})) = i(t - \tau, \mathbf{f})\, dt \tag{203}$$

5. on en déduit alors l'équation du taux d'infectivité par état qui est :

$$i(t,\mathbf{e}) = \sum_{\mathbf{f=1}}^{\mathbf{f=2L}} \int_{\tau=0}^{+\infty} \mathcal{A}(\tau,\mathbf{e},\mathbf{f})\, i(t-\tau,\mathbf{f})\, d\tau \tag{204}$$

Quelques remarques sur cette équation :

1. Il est tout à fait naturel que l'équation porte sur le taux d'infectivité et non pas sur le nombre de personnes infectées $\mathcal{I}(t,\mathbf{e})$ car l'opérateur $\mathcal{A}(\tau,\mathbf{e},\mathbf{f})$ compte le nombre de *nouveaux* cas à partir de ceux existants.
2. On peut connaître le nombre de personnes de l'état $\mathbf{e}$ qui ont été infectées entre $(-\infty,t)$ par le relation

$$\mathcal{I}(t,\mathbf{e}) = \int_{-\infty}^{s=t} i(s,\mathbf{e})\, ds \tag{205}$$

en supposant simplement que l'on a comme condition « initiale » :

$$\forall \mathbf{e} \in [\![\mathbf{1},\mathbf{2L}]\!], \quad \mathcal{I}(-\infty,\mathbf{e}) = 0 \tag{206}$$

Certes, l'objet que l'on manipule, le taux différentiel d'infectivité, est un objet théorique remarquable. Mais la vraie question est plutôt la suivante : EXISTE-T-IL UN MOYEN DE CONSTRUIRE $\mathcal{A}(\tau,\mathbf{e},\mathbf{f})$? Et c'est bien là que les choses se gâtent. En fait, pour espérer avoir une *construction* a priori et utilisable de ce taux différentiel, il FAUT CONNAÎTRE LA PHYSIQUE NEWTONIENNE (!) DE LA TRANSMISSION DU VIRUS. Ainsi, on peut par exemple distinguer deux cas diamétralement opposés :

1. celui des maladies sexuellement transmissibles. Dans ce cas, tout semble être réuni pour que le modèle soit pertinent et constructible. En particulier tous ceux qui ne sont pas déjà infectés sont des personnes totalement contaminables car, sauf vaccin, il n'existe pas forcément de défense naturelle. D'autre part, les personnes infectées le restent en l'absence de traitement et donc sont totalement contaminantes indépendamment du temps : ce qui signifie que l'opérateur $\mathcal{A}(\tau,\mathbf{e},\mathbf{f})$ ne dépend pas du temps. Il "suffit" alors de construire la matrice des contacts sexuels pour avoir la structure de celle de $\mathcal{A}$ et si l'on donne une probabilité de transmettre le virus par unité de rapport sexuel, on peut ainsi avoir, a priori, une construction efficace du taux différentiel.
2. Celui des virus respiratoires. Dans ce cas en général, la physique (newtonienne) de la transmission est en général très difficile à déterminer. le virus peut se transmettre par voix respiratoire, être manuporté, c'est-à-dire qu'il se transmet par contact des mains avec des surfaces contaminées (d'où l'importance souvent rapportée du lavage des mains), passer par les sprays, c'est-à -dire les gouttelettes émises par les toux, les éternuements, etc. Bref, contrairement au cas des maladies sexuellement transmissibles, il est parfaitement chimérique de vouloir construire le taux différentiel d'infectivité a priori par on ne sait quelle matrice des contacts. Dit autrement, il est possible qu'il existe un taux différentiel d'infectivité $\mathcal{A}(\tau,\mathbf{e},\mathbf{f})$ qui permette de modéliser la propagation du virus, mais pour le physicien (mathématique) que je suis, il semble clair qu'un tel objet ne peut être

identifié que de façon "inverse" : étant connu dans la phase de croissance exponentielle la dynamique par état du virus, alors on doit pouvoir essayer de reconstruire la forme du taux différentiel d'infectivité grâce à l'observation. Certes, le problème sera mathématiquement mal posé (ce qui veut dire que vous ne pourrez pas trop déduire d'informations de vos observations), mais les mathématiciens savent traiter cet écueil (sans vouloir toutefois jamais donner des solutions à des problèmes qui n'en n'ont pas).

5.2 Rapides rappels sur la théorie du R_0 : le nombre de reproduction de base.

Le nombre de reproduction de base, dit aussi R_0, a une interprétation assez simple dans la modélisation des épidémies : sous des hypothèses idéalisées, il indique combien de personnes vont être contaminées par une personne qui est elle-même infectée. Ce nombre appartient en fait au *taux efficace d'infectivité* , c'est-à-dire à la matrice suivante :

$$\forall\,(\mathbf{e},\mathbf{f}) \in [\![\mathbf{1},\mathbf{2L}]\!] \times [\![\mathbf{1},\mathbf{2L}]\!], \quad \mathbb{A}\,[\mathbf{e},\mathbf{f}] := \int_{\tau=0}^{\infty} \mathcal{A}\,(\tau,\mathbf{e},\mathbf{f})\,d\tau \tag{207}$$

comme nous l'avons déjà dit, cette matrice est positive, c'est-à-dire que tous ces coefficients sont positifs ou nuls. La raison fondamentale à cela c'est que cette matrice, si vous lui donnez en entrée un vecteur de type $(C_{\mathbf{f}} : \mathbf{f} \in [\![\mathbf{1},\mathbf{2L}]\!])$ contenant, pour chaque état $\mathbf{f}$, un nombre $C_{\mathbf{f}} \geq 0$ de personnes contaminées, alors, en retour, le produit matrice vecteur renverra :

$$D = \mathbb{A}C, \ i.e \ \ \forall \mathbf{e} \in [\![\mathbf{1},\mathbf{2L}]\!], \ D_{\mathbf{e}} = \sum_{\mathbf{f}=\mathbf{1}}^{\mathbf{f}=\mathbf{2L}} \mathbb{A}\,[\mathbf{e},\mathbf{f}]\,C_{\mathbf{f}} \tag{208}$$

est toujours tel que $D_{\mathbf{e}} \geq 0$. Introduisons la définition suivante :

Definition 7. *Soit $C \in \mathbb{R}^{\mathbf{2L}}$ un vecteur. On dit que ce vecteur est positif, et l'on note $C \succeq 0$ lorsque*

$$\forall \mathbf{f} \in [\![\mathbf{1},\mathbf{2L}]\!], \quad C_{\mathbf{f}} \geq 0 \tag{209}$$

On note $C \succ 0$ lorsque $C \neq 0$ et $C \succeq 0$. Finalement on note $C > 0$ lorsque

$$\forall \mathbf{f} \in [\![\mathbf{1},\mathbf{2L}]\!], \quad C_{\mathbf{f}} > 0 \tag{210}$$

un vecteur $C \succ 0$ est ainsi un vecteur qui a toutes ses composantes positives ou nulles et donc au moins une est non nulle. On peut étendre les notations à la matrice $\mathbb{A}$.

Definition 8. *Soit $\mathbb{A} \in \mathcal{M}_{\mathbf{2L}}(\mathbb{R})$ une matrice carrée de taille $\mathbf{2L}$. On dit que $\mathbb{A} \geq 0$ lorsque tous ses coefficients sont positifs ou nuls et on dit que $\mathbb{A} > 0$ lorsque tous les coefficients de $\mathbb{A}$ sont strictement positifs. Dans le premier cas, on dit que $\mathbb{A}$ est positive ou nulle, dans le second cas on dit que $\mathbb{A}$ est strictement positive.*

On remarque par exemple que $C \succ 0$ et $\mathbb{A} > 0$ entraînent $\mathbb{A}C > 0$. Les matrices positives ont des propriétés structurelles importantes. C'est la théorie de Perron-Frobenius (dont fait partie le théorème que l'on a déjà utilisé). Pour comprendre cette théorie, nous avons besoin d'un peu de précision supplémentaire.

Definition 9. *Soit $\mathbb{A} \in \mathcal{M}_n(\mathbb{R})$ une matrice (carrée de taille n) à coefficients réels (pas nécessairement positifs). On rappelle que le polynôme caractéristique de $\mathbb{A}$, noté $\chi[\mathbb{A}](Z)$, est obtenu formellement par :*

$$\chi(\mathbb{A})(z) = \det(\mathbb{A} - z\mathbb{I}) \tag{211}$$

où $\mathbb{I}$ désigne la matrice identité et $\det(\cdot)$ la fonction de déterminant. Le polynôme caractéristique de $\mathbb{A}$ est un polynôme de degré n.

Spectre d'une matrice réelle :

Definition 10. *On appelle spectre de $\mathbb{A}$, noté $\mathcal{Sp}(\mathbb{A})$, l'ensemble des racines (prises dans l'ensemble des complexes $\mathbb{C}$) du polynôme caractéristique de $\mathbb{A}$, c'est-à-dire que :*

$$z \in \mathbb{C} \in \mathcal{Sp}(\mathbb{A}) \Leftrightarrow \chi[\mathbb{A}](z) = 0 \tag{212}$$

finalement, on appelle rayon spectral de $\mathbb{A}$, que l'on note $\rho(\mathbb{A})$, le plus grand module des éléments de $\mathcal{Sp}(\mathbb{A})$. Autrement dit, on a :

$$\rho(\mathbb{A}) = \max\{|z|,\ z \in \mathcal{Sp}(\mathbb{A})\} \tag{213}$$

Nous allons voir en fait que le R_0 n'est rien d'autre que le rayon spectral de la matrice $\mathbb{A}$ du taux effectif d'infectivité. Mais avant de le définir brutalement ainsi, nous allons d'abord essayer de comprendre pourquoi le rayon spectral est un concept intéressant pour les matrices positives. Introduisons maintenant la notion de valeur propre d'une matrice :

Definition 11. *Soit $\mathbb{A} \in \mathcal{M}_n(\mathbb{R})$ une matrice (carrée de taille n) à coefficients réels (pas nécessairement positifs). On dit que $\lambda \in \mathbb{R}$ est une valeur propre de $\mathbb{A}$ lorsqu'il existe un vecteur $C \neq 0$ tel que l'on ait :*

$$AC = \lambda C \tag{214}$$

Un réel λ est une valeur propre de $\mathbb{A}$ si et seulement si c'est une racine réelle du polynôme $\chi[\mathbb{A}](z)$

A priori, il n'y a pas forcément de lien entre rayon spectral et valeur propre. Dans le cas où la matrice $\mathbb{A}$ est diagonalisable, il y a égalité entre la plus grande valeur propre de $\mathbb{A}$ et le rayon spectral. Mais une matrice peut même n'avoir aucun vecteur propre. Hormis le cas assez simple de la diagonalisation, il y a un autre cas où l'on peut relier le rayon spectral avec la plus grande valeur propre (qui alors existe), c'est le cas des certaines matrices positives ou nulles. Nous aurons besoin des définition suivantes (voir Jean-Étienne Rombaldi).

Definition 12. *Soit $\mathbb{A} \geq 0 \in \mathcal{M}_n(\mathbb{R})$ une matrice positive ou nulle de taille n. On dit que :*

1. *la matrice $\mathbb{A}$ est irréductible si et seulement si la matrice $(\mathbb{I} + \mathbb{A})^{n-1}$ est strictement positive ;*
2. *la matrice $\mathbb{A}$ est primitive si et seulement si la matrice $\mathbb{A}^{n^2-2n+2}$ est strictement positive.*

On voit que dans tous les cas, une matrice qui est strictement positive est tout à la fois irréductible et primitive. Notons qu'une matrice primitive est

toujours irréductible mais que l'inverse est parfaitement faux. Par exemple, si vous considérez $n = 2$ et que vous choisissiez la matrice

$$\mathbb{A} = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}, \quad a > 0, \quad b > 0 \tag{215}$$

c'est une matrice positive, carrée, de taille $n = 2$. Alors, on a facilement :

$$(\mathbb{I} + \mathbb{A})^{n-1} = \begin{bmatrix} 1 & a \\ b & 1 \end{bmatrix}, \quad \mathbb{A}^{n^2-2n+2} = \mathbb{A}^2 = \begin{bmatrix} ab & 0 \\ 0 & ab \end{bmatrix} \tag{216}$$

Ainsi, on voit bien que la matrice $\mathbb{A}$ est irréductible mais certainement pas primitive puisque $\mathbb{A}^{n^2-2n+2}$ possède des valeurs nulles.

Annonçons maintenant les deux théorèmes principaux de la théorie de Perron-Frobenius

Théorème 2 (Perron-Frobenius Irréductible)**.** *Soit $\mathbb{A} \geq 0 \in \mathcal{M}_n(\mathbb{R})$ une matrice positive ou nulle de taille n qui est irréductible. Alors :*

1. *$\rho(A) > 0$ est une valeur propre simple de $\mathbb{A}$;*
2. *l'espace propre associé à cette valeur propre est engendré par un vecteur $C > 0$ (i.e. toutes ses composantes sont strictement positives).*

On va maintenant arriver enfin au but ultime de la théorie du R_0, celui sur le comportement asymptotique des matrices primitives.

Théorème 3 (Perron-Frobenius Primitive)**.** *Soit $\mathbb{A} \geq 0 \in \mathcal{M}_n(\mathbb{R})$ une matrice positive ou nulle de taille n qui est primitive. Soit $\rho(\mathbb{A})$ son rayon spectral et $C_0 > 0$ son unique vecteur propre de norme 1 (pour la norme $\|\cdot\|_1$ justement). Alors :*

1. *il existe une unique vecteur $C_*^0 > 0$ tel que :*
2. *Pour tout vecteur $C \succeq 0$ on a*

$$\mathbb{A}^m C \sim \left\langle C_*^0, C \right\rangle \rho^m(\mathbb{A}) C^0, \quad m \to +\infty \tag{217}$$

où $\langle \cdot, \cdot \rangle$ désigne le produit scalaire entre vecteur.

Ce théorème dit la chose suivante : si la matrice est primitive, alors quel que que soit votre vecteur de départ C, vous aurez

$$: \lim_{m \to +\infty} \frac{1}{\rho^m(\mathbb{A})} \mathbb{A}^m C = \left\langle C_*^0, C \right\rangle C^0 \tag{218}$$

Autrement dit, indépendamment de votre point de départ, votre point d'arrivée sera toujours le même (à un facteur près). Vous finirez par être colinéaire à C^0, mais cette colinéarité va s'exprimer de différentes manières :

1. si $\rho(\mathbb{A}) < 1$, vous serez colinéaire à C^0 et vous allez également tendre vers 0 ;
2. si $\rho(\mathbb{A}) = 1$, vous tendrez vers $\left\langle C_*^0, C \right\rangle C^0$;
3. si $\rho(\mathbb{A}) > 1$, vous serez colinéaire à C^0 mais votre norme va finir par devenir infinie.

On en vient maintenant à la définition précise de R_0.

Definition 13. *Soit* $(\mathbb{A}[\mathbf{e},\mathbf{f}], \mathbf{e} \in [\![1,\mathbf{2L}]\!],\ \mathbf{f} \in [\![1,\mathbf{2L}]\!])$ *la matrice représentant le taux efficace d'infectivité. Cette matrice est nécessairement positive ou nulle. On suppose qu'elle est* ***primitive****. Alors on pose :*

$$R_0 = \rho(\mathbb{A}) \tag{219}$$

le nombre basique de reproduction.

Pour interpréter ce taux de reproduction, on va passer en mode « next gen ». Supposons, pour simplifier, que l'on ait :

$$\forall \tau > T_i, \quad \mathcal{A}[\tau,\mathbf{e},\mathbf{f}] = 0 \tag{220}$$

autrement dit, une personne infectée cesse d'être infectante au bout d'un temps T_i. Supposons qu'à $t=0$ on note par $(\mathcal{I}(0,\mathbf{f}),\ \mathbf{f} \in [\![1,\mathbf{2L}]\!])$ le nombre de personnes infectées (et donc infectantes) par état, à l'instant $t=0$. Le nombre de personnes qui auront été infectées à l'instant T_i sera donné par :

$$\mathcal{J}(T_i,\mathbf{e}) = \sum_{\mathbf{f=1}}^{\mathbf{f=2L}} \mathbb{A}[\mathbf{e},\mathbf{f}]\,\mathcal{I}(0,\mathbf{f}) \tag{221}$$

Mais à leur tour, les personnes $(\mathcal{J}(T_i,\mathbf{f})\,\mathbf{f} \in [\![1,\mathbf{2L}]\!])$, celles de la première génération, vont pouvoir infecter des personnes. On produit ainsi la génération à $2T_i$ selon :

$$\mathcal{J}(2T_i,\mathbf{e}) = \sum_{\mathbf{f=1}}^{\mathbf{f=2L}} \mathbb{A}[\mathbf{e},\mathbf{f}]\,\mathcal{J}(T_i,\mathbf{f}) \tag{222}$$

et ainsi de suite donc. Si bien qu'on peut calculer la m-ième génération par

$$\mathcal{J}(mT_i,:) = \mathbb{A}^m \mathcal{I}(0,:) \tag{223}$$

comme la matrice $\mathbb{A}$ est supposée être primitive, alors, lorsque m devient suffisamment grand, on a :

$$\mathcal{J}(mT_i,:) \sim \left\langle C_*^0, \mathcal{I}(0,:)\right\rangle R_0^m \mathcal{I}(0,:) \tag{224}$$

comme on a $C_*^0 > 0$ et $\mathcal{I}(0,:) \succ 0$, il est clair que $\left\langle C_*^0, \mathcal{I}(0,:)\right\rangle > 0$. Par conséquent, on a trois cas :

1. ou bien $R_0 < 1$: dans ce cas, l'épidémie disparaît car la génération m voit le nombre de personnes tendre vers 0. On a en effet :

$$\sum_{\mathbf{f=1}}^{\mathbf{f=2L}} \mathcal{J}(mT_i,\mathbf{f}) = R_0^m \left(\left\langle C_*^0, \mathcal{I}(0,:)\right\rangle \sum_{\mathbf{f=1}}^{\mathbf{f=2L}} \mathcal{I}(0,\mathbf{f})\right) \to 0 \tag{225}$$

2. Ou bien $R_0 = 1$: alors dans ce cas l'épidémie est contrôlée. L'ordre de grandeur de $\mathcal{J}(mT_i,:)$ est le même que celui de la génération initiale (infectante) $\mathcal{I}(0,:)$. On a :

$$\sum_{\mathbf{f=1}}^{\mathbf{f=2L}} \mathcal{J}(mT_i,\mathbf{f}) = \left(\left\langle C_*^0, \mathcal{I}(0,:)\right\rangle \sum_{\mathbf{f=1}}^{\mathbf{f=2L}} \mathcal{I}(0,\mathbf{f})\right) \tag{226}$$

3. Ou bien $R_0 > 1$: dans ce cas l'épidémie va s'accroître rapidement. On a en effet :

$$\sum_{\mathbf{f=1}}^{\mathbf{f=2L}} \mathcal{J}\left(mT_i, \mathbf{f}\right) = R_0^m \left(\left\langle C_*^0, \mathcal{I}\left(0,:\right) \right\rangle \sum_{\mathbf{f=1}}^{\mathbf{f=2L}} \mathcal{I}\left(0,\mathbf{f}\right) \right) \rightarrow +\infty \tag{227}$$

Ajoutons ici quelques remarques :

1. L'interprétation de R_0 est la suivante : R_0 représente quelque part le nombre moyen de personnes qui vont être contaminées par une personne infectante. Ainsi, pour donner une image, si une personne en contamine trois, les trois vont en container neuf et ainsi de suite, conduisant à une croissance exponentielle du nombre de personnes infectées.
2. Avec un $R_0 < 1$ et avec le même raisonnement, l'épidémie doit disparaître de façon exponentielle !
3. On retrouve ici le fameux $R_0 > 1$ et $R_0 < 1$. Ce que prétendent les auteurs de l'article, c'est que grâce au confinement, on est passé miraculeusement d'un $R_0 > 1$ (les auteurs le mesurent à 3), à un $R_0 < 1$. Autrement dit, SELON LES AUTEURS, GRÂCE AU CONFINEMENT, L'ÉPIDÉMIE AURAIT DISPARU AUSSI VITE QU'ELLE EST APPARUE ; qu'un tel bobard soit repris par tous les médias nous paraît absolument scandaleux. IL absolument faux que le confinement aurait pour effet de faire décroître exponentiellement l'épidémie.
4. La théorie du R_0 qui repose sur les matrices (opérateurs) primitives, voilà bien une caractéristique qui semble avoir échappé aux inventeurs eux-mêmes de cette notion. Ainsi, dans le papier « *On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations* » O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, J. Math. Biol. (1990) 28 :365-382, le premier exemple que prennent les auteurs pour illustrer l'intuition du R_0 est une matrice de la forme :

$$\mathbb{A} = \left[\begin{array}{cc} 0 & a \\ b & 0 \end{array} \right] \tag{228}$$

qui est effectivement irréductible mais pas primitive. De fait, il ne peut y avoir de convergence indépendamment de la valeur initiale $\mathcal{I}\left(0,:\right)$ vers le vecteur propre de norme maximum. Il est assez étonnant que les auteurs de l'article en question - une référence absolue dans le milieu - se soient eux-mêmes laissés aller à ne pas distinguer irréductible et primitive. Pourtant, je tiens cet article pour un article de très très grande qualité (et je suis en général plutôt exigeant). On peut d'ailleurs développer « à la main »l'absence de convergence. Notons d'abord que cette matrice possède deux valeurs propres distinctes, et deux vecteurs propres :

$$\lambda = \pm\sqrt{ab}, \quad C_+^0 = \left[\sqrt{a}, \sqrt{b}\right], \quad C_-^0 = \left[-\sqrt{a}, \sqrt{b}\right] \tag{229}$$

si l'on choisit $\mathcal{I}\left(0,:\right) = \left[0,1\right]$, alors on voit que l'on a :

$$\mathbb{A}^{2p}\mathcal{I}\left(0,:\right) = (ab)^p \left[0,1\right], \quad \mathbb{A}^{2p+1}\mathcal{I}\left(0,:\right) = (ab)^p \left[a,0\right] \tag{230}$$

Ainsi, on voit que les suites $\left(\mathbb{A}^{2p}\mathcal{I}\left(0,:\right)\right)$ et $\left(\mathbb{A}^{2p+1}\mathcal{I}\left(0,:\right)\right)$ sont orthogonales : elles ne peuvent donc pas converger vers la même limite.

Pour finir, il faut donner le lien, comme cela est indiqué dans *Diekmann* & al. (et lorsque la matrice $\mathbb{A}$ est primitive) et l'unique vecteur propre (de norme unité) associé à la valeur propre $R_0 = \rho(\mathbb{A})$. Souvenons-nous de l'équation du taux d'infectivité $i(t, \mathbf{e})$:

$$i(t, \mathbf{e}) = \sum_{\mathbf{f}=\mathbf{1}}^{\mathbf{2L}} \int_{\tau=0}^{\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\, i(t-\tau, \mathbf{f})\, d\tau \tag{231}$$

Cherchons des solutions de cette équation sous la forme $i(t, \mathbf{e}) = \exp(\lambda t)\, \boldsymbol{\phi}(\mathbf{e})$. Alors, cela revient à trouver $(\lambda, \phi(\mathbf{e}))$ sous la forme :

$$\exp(\lambda t)\, \boldsymbol{\phi}(\mathbf{e}) = \sum_{\mathbf{f}=\mathbf{1}}^{\mathbf{2L}} \int_{\tau=0}^{+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f}) \exp(\lambda(t-\tau))\, \boldsymbol{\phi}(\mathbf{f})\, d\tau \tag{232}$$

$$\boldsymbol{\phi}(\mathbf{e}) = \sum_{\mathbf{f}=\mathbf{1}}^{\mathbf{2L}} \int_{\tau=0}^{+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f}) \exp(-\lambda\tau)\, \boldsymbol{\phi}(\mathbf{f})\, d\tau \tag{233}$$

Ainsi, on peut trouver à cette équation une solution de la forme désirée, à condition de montrer qu'il existe λ tel que la matrice

$$\hat{\mathcal{A}}(\lambda, \mathbf{e}, \mathbf{f}) = \int_{\tau=0}^{+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f}) \exp(-\lambda\tau)\, d\tau \tag{234}$$

possède un vecteur propre (strictement positif) pour la valeur 1 (au passage la matrice $\hat{\mathcal{A}}(\lambda, \mathbf{e}, \mathbf{f})$ est la transformée de Laplace du taux différentiel d'infectivité $\mathcal{A}(\tau, \mathbf{e}, \mathbf{f})$). Selon le signe de λ, on aura :

$$\lambda > 0 \Rightarrow \int_{\tau=0}^{+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f}) \exp(-\lambda\tau)\, d\tau < \int_{\tau=0}^{+\infty} \mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\, d\tau \tag{235}$$

Ainsi, on pourra écrire, au sens de l'inégalité matricielle (i.e. répétée sur toutes les composantes), que l'on a :

$$0 < \hat{\mathcal{A}}(\lambda) < \mathbb{A} \Rightarrow \rho\left(\hat{\mathcal{A}}(\lambda)\right) \leq \rho(\mathbb{A}) \tag{236}$$

le rayon spectral va être ici une fonction continue de λ (cela tient au fait que les racines d'un polynôme sont continues avec les coefficients et que les coefficients sont calculés par les déterminants qui sont aussi des fonction continues. Finalement prendre le max revient à prendre la norme infinie sur un vecteur et c'est aussi continu). De plus, il est clair que ce rayon spectral est une fonction strictement décroissante de λ qui tend vers 0 à l'infini. Donc :

1. si $R_0 < 1$, il n'est pas possible de trouver une solution ;
2. en revanche, si $R_0 > 1$, on peut toujours trouver un unique $\lambda_0 > 0$ tel que le rayon spectral de $\hat{\mathcal{A}}(\lambda)$ soit exactement égal à 1 ;
3. dans ce cas, on voit que par Perron-Frobenius, il y a un unique vecteur positif $\boldsymbol{\phi}_0$ qui est vecteur propre strictement positif, de norme unité (en norme 1) pour la matrice $\hat{\mathcal{A}}(\lambda_0)$.

Donc on arrive à la conclusion suivante :

Proposition 6. *Si $R_0 > 1$, alors il existe un unique $\lambda_0 > 0$ et un unique vecteur $\boldsymbol{\phi}_0 > 0$ normalisé tel que :*

$$\forall (t, \mathbf{e}), \ i(t, \mathbf{e}) = \exp(\lambda_0 t) \boldsymbol{\phi}_0(\mathbf{e}) \tag{237}$$

soit solution de l'équation du taux d'infectivité.

Remarque : avec $R_0 < 1$ on a exactement le même résultat d'existence et d'unicité, sauf qu'alors $\lambda_0 < 0$.

Les solutions exponentielles sont très importantes en pratique car en général on observe effectivement ces solutions en début d'épidémie. Cela rejoint en effet l'approche « next gen » pour laquelle un vecteur initial vient effectivement se caler sur un vecteur propre. On peut citer ici Diekmann & al :
« *Note that λ_0 and $\boldsymbol{\phi}_0$ describe the growth and the* **e***-state distribution in the exponential phase of an epidemic, when the influence of the precise manner in which the epidemic started has died off and the influence of the nonlinearity is not yet perceptible.* »
Ce qui veut dire qu'en début d'épidémie, écartés les premiers jours d'épidémie non significatifs, on a forcément $R_0 > 1$ et on doit observer, a priori, une croissance exponentielle du nombre d'individus nouvellement infectés par unité de temps.

5.3 Exploitation des données numériques - comparaison avec le modèle des auteurs

Il faut noter que le modèle utilisant le taux différentiel d'infectivité - qui permet donc de modéliser la dynamique de la propagation du virus - est écrit sur la population globale du pays. Il permet de suivre, avec le temps, le nombre de personnes qui sont infectées. Or, il arrive évidemment que l'on soit parfaitement incapable d'avoir accès à cette info à l'échelle de la population. Seules les personnes ayant été testées et dont le test s'est révélé positif peuvent ainsi - après en outre un long travail administratif de traitement des données - se retrouver dans les chiffres. La première question que l'on se pose alors c'est de savoir si les données disponibles sur un échantillon d'individus reflètent effectivement la situation sur l'ensemble de la population. Rien n'est moins sûr pour un tas de raisons. Mais l'hypothèse la plus raisonnable, la plus simple et la plus efficace c'est de supposer que ce que l'on voit dans les chiffres représente une fraction de ce que l'on observe dans la réalité : autrement dit, il y a un facteur de proportionnalité directe. Bien sûr, vous ne connaissez pas le facteur et donc connaître l'état réel de la population infectée à partir de celui observé est sans doute sans espoir. Cependant, eu égard à la linéarité du modèle, cela est sans conséquence sur un certain nombre d'informations que l'on pourra tirer des observations faites sur les gens détectés.

À nouveau, nous devons nous débrouiller seul s'agissant de trouver les données des personnes infectées. À nouveau, c'est sur le site de Wikipedia que nous sommes allé chercher ces données. Les deux données intéressantes sont les suivantes :

1. le nombre de nouveaux cas détectés par jour : l'histogramme quotidien ;

2. le nombre total de personnes détectées jusqu'à un certain jour : l'histogramme cumulé quotidien.

D'un point de vue mathématique et pour revenir à nos notations :

1. le nombre de nouveaux cas quotidiens correspond au taux d'infectivité calculé sur toute la population, c'est-à-dire en sommant sur tous les états possibles :

$$i_s(t) = \sum_{\mathbf{e}=\mathbf{1}}^{\mathbf{e}=\mathbf{2L}} i(t, \mathbf{e}) \tag{238}$$

2. le nombre de cas total quotidien correspond au nombre de personnes infectées calculé sur toute la population, c'est-à-dire là encore en sommant sur tous les états possibles :

$$\mathcal{I}_s(t) = \sum_{\mathbf{e}=\mathbf{1}}^{\mathbf{e}=\mathbf{2L}} \mathcal{I}(t, \mathbf{e}) \tag{239}$$

La relation liant $\mathcal{I}_s(t)$ à $i(t)$ est celle d'un rapport de dérivation et d'intégration : plus précisément :

$$\forall t \;\; i_s(t) = \frac{d\mathcal{I}_s}{df}(t), \quad \forall t_0 \; \mathcal{I}_s(t) = \mathcal{I}_s(t_0) + \int_{s=t_0}^{s=t} i(s)\, ds \tag{240}$$

D'une manière générale, lorsque vous aller faire des problèmes inverses, il vaut mieux travailler sur les grandeurs intégrales que sur les grandeurs dérivées. Nous travaillerons donc sur la grandeur $\mathcal{I}(t)$, c'est-à-dire sur la fréquence cumulée. Supposons que l'on soit dans le régime dit « exponentiel », c'est-à-dire que l'on estime raisonnablement que l'on a :

$$\forall t \geq t_0, \; \forall \mathbf{e} \in [\![\mathbf{12L}]\!], \; i(t, \mathbf{e}) = \exp(\lambda t)\, \psi_0(\mathbf{e}) \tag{241}$$

Par sommation sur les états on en tire :

$$\forall t \geq t_0, \quad i_s(t) = i_s(t_0) \exp(\lambda t), \quad i_s(t_0) = \sum_{\mathbf{e}=\mathbf{1}}^{\mathbf{e}=\mathbf{2L}} \psi_0(\mathbf{e}) \tag{242}$$

D'où, par intégration :

$$\begin{aligned} \mathcal{I}_s(t) - \mathcal{I}_s(t_0) &= \int_{\tau=t_0}^{\tau=t} i_s(t_0) \exp(\lambda\tau)\, d\tau & (243) \\ &= i_s(t_0) \frac{1}{\lambda} [\exp(\lambda t) - \exp(\lambda t_0)] & (244) \end{aligned}$$

Il y a plusieurs manière d'exploiter cette relation. On va procéder de la manière suivante :

$$\mathcal{I}_s(t) - \mathcal{I}_s(t_0) = \frac{i_s(t_0) \exp(\lambda t_0)}{\lambda} [\exp(\lambda(t - t_0)) - 1] \tag{245}$$

Si l'on se restreint, alors pour les $t \geq t_{obs}$ tel que $\exp(\lambda(t - t_0)) >> 1$, on a avec une très bonne approximation :

$$\forall t \geq t_{obs}, \quad \ln(\mathcal{I}_s(t) - \mathcal{I}_s(t_0)) \approx \ln\left(\frac{i_s(t_0) \exp(\lambda t_0)}{\lambda}\right) + \lambda(t - t_0) \tag{246}$$

Si, pour les mêmes instants $t \geq t_{obs}$ pour lesquels on a $\exp(\lambda(t-t_0)) >> 1$ on a aussi $\mathcal{I}(t) >> \mathcal{I}(t_0)$, on en déduit alors :

$$\ln(\mathcal{I}_s(t)) \approx \ln\left(\frac{i_s(t_0)\exp(\lambda t_0)}{\lambda}\right) + \lambda(t-t_0) \tag{247}$$

On va alors procéder de la sorte :

1. d'abord, on choisit une période $[t_0, t_f]$ pour laquelle on estime que le régime exponentiel est bien atteint. On sait en effet d'après la théorie « next gen »et la théorie du R_0 qu'il faut attendre un certain temps avant le début de l'épidémie avant de rentrer dans ce régime.
2. On choisira donc de travailler avec $T_i = 8$, c'est-à-dire que l'on se place 8 jours après le début de l'épidémie. Ce chiffre est relativement arbitraire MAIS : si vous pensez effectivement que la modélisation par taux différentiel d'infectivité est correcte, alors vous avez le DEVOIR de procéder ainsi. Sur une épidémie qui va durer environ $\Delta = 70$ jours, il est honnête de penser qu'à $T = 0.1\Delta$, vous serez à la fois dans le début de l'épidémie, mais que vous aurez laissé suffisamment de temps pour atteindre le régime exponentiel.
3. Puisqu'il est censé y avoir un événement brutal qui change le mode de transmission de l'épidémie, à savoir le confinement, nous étudions alors le phénomène jusqu'à $T_t = 22$, soit 22 jours après l'épidémie, c'est-à-dire jusqu'au 17 mars 2020.
4. Dans cette période $[t_0, t_f]$, on choisit ensuite une date t_{obs} pour laquelle ensuite on fera la régression linéaire sur $[t_0, T_f]$. Il y a deux conditions pour que ce soit le cas :

   (a) que l'on ait $\mathcal{I}_s(t_{obs}) >> \mathcal{I}_s(T_i)$;

   (b) que l'on ait $\exp(\lambda t_{obs}) >> \exp(\lambda T_i)$.

   Avec une approximation assez grossière, on peut estimer que l'on a $\forall t \geq T_i,\ \ \mathcal{I}_s(t) \sim \exp(\lambda t)$ et donc les deux contions précédentes sont en générales vérifiées en même temps.

En ce qui nous concerne, rappelons quelques dates. Sur Wikipedia, les cas de COVID rapportés sont enregistrés à partir du 25 Février 2020. Comme on essaye ici de savoir ce que vaut l'affirmation de l'efficacité du confinement. Nous les reportons en annexe. Il est assez facile de sélectionner :

$$\begin{array}{cccc} t & t_0 = 8 & t_{obs} = 14 & t_f = 22 \\ \mathcal{I}_s(t) & 212 & 1372 & 7690 \end{array} \tag{248}$$

On trace alors la quantité $\ln(\mathcal{I}_s)$. Si tout se passe comme il le faut, on doit avoir une droite de la forme $d(t) = at + b$ pour laquelle on sait estimer les paramètres a, b via une régression linéaire (en ce qui nous concerne, la fonction *reglin* du logiciel (gratuit) SCILAB). On obtient alors la figure suivante (voir la figure 10)

Comme on peut le constater, l'accord est vraiment excellent : le coefficient de

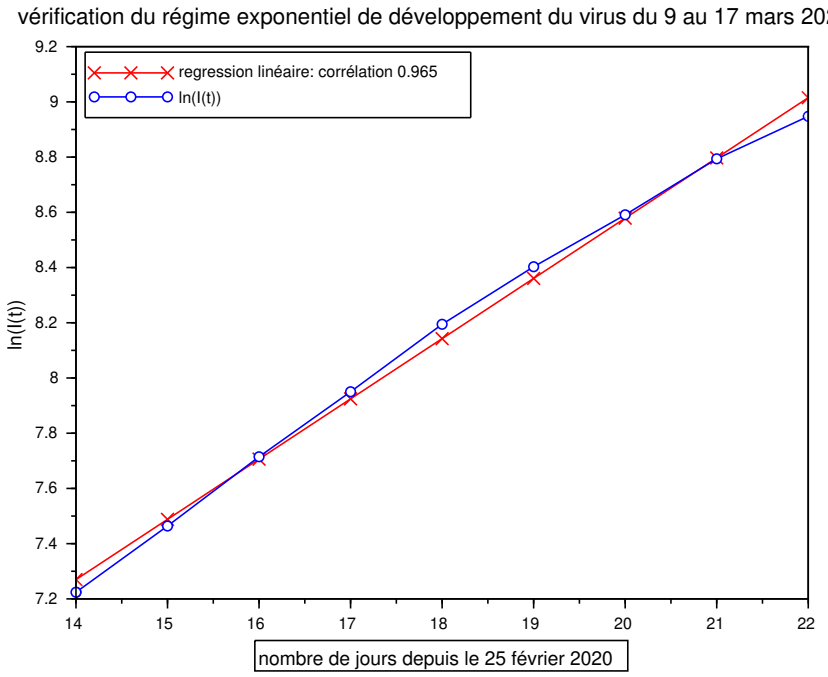

FIGURE 10 – comparaion entre la théorie du régime exponentiel et l'observation des cas de COVID19 en France du 9 au 17 mars 2020

corrélation linéaire est de 97%. Les coefficients trouvés sont :

$$\begin{aligned}
\forall t \in [t_{obs}, t_f], \quad y &= at + b, && (249)\\
&\rightarrow \quad a = 0.2181581 jour^{-1}, \quad b = 4.2156691 jours && (250)\\
&= \ln\left(\frac{i_s(t_0)\exp(\lambda t_0)}{\lambda}\right) + \lambda(t - t_0) && (251)\\
&\leftrightarrow \quad \lambda = a = 0.2181581 jour^{-1} && (252)
\end{aligned}$$

On remarque sur la courbe qu'il y a un point particulier, $t_v = 16$ [1] pour lequel la valeur donnée par la régression linéaire est exactement celle donnée par $\ln(\mathcal{I}_s(t_v))$. On peut alors se servir de ce point pour estimer la donnée qu'il nous manque, à savoir $i_s(t_0)$. On obtient ainsi :

$$i_s(t_0) = \lambda \mathcal{I}(t_v)\exp(-\lambda t_v) \tag{253}$$

En principe, si le calcul n'est pas idiot, on doit retrouver, à peu près, pour $i_s(t_0)$ la valeur mesurée du nombre de nouveaux cas à $t_0 = 8$. En réalité, il n'y a assez peu d'espoir que l'on trouve quelque chose de trop précis car les données mesurées sont très fluctuantes. D'autre part, il ne vous aura pas échappé que l'on a nous-même utilisé un modèle continu alors que les données sont discrètes.

1. il y a un décalage de 1 jour entre l'affichage et notre numérotation

Il y a quand même de nombreuses hypothèses théoriques dont on ne sait pas trop ce qu'elles valent. Cependant le calcul nous donne ainsi :

$$\begin{aligned} i_s\left(t_0 = 8\ jours\right) &= 0.2181581 * 1372 * \exp\left(-0.2181581 * 16\right) && (254) \\ &= 14.90\ individu/jour^{-1} && (255) \\ i_s^{obs}\left(t_0 = 8\ jours\right) &= 21\ \ individu/jour^{-1} && (256) \end{aligned}$$

Dans ces phénomènes exponentiels qui ont beaucoup de bruit et qu'il faut impérativement régulariser, c'est bien l'ordre de grandeur qui prime et ici on peut être très satisfait de l'accord qui est de 30%.

Fort de ses résultats qui nous permettent d'avoir une valeur fiable, semble-t-il dans le régime exponentiel, on peut croire que la théorie fonctionne. On va maintenant essayer d'exploiter au mieux le λ.

Pour cela, il faut revenir à un fait très important que vous ne pouvez pas comprendre si vous n'avez pas fait la théorie des états exponentiels. Nous avons vu que l'état exponentiel, c'est-à-dire la solution $i\left(t, \mathbf{e}\right) = \exp\left(\lambda t\right) \boldsymbol{\phi}_0\left(\mathbf{e}\right)$ existe pour un unique λ :

le réel λ est l'unique paramètre tel que la matrice

$$\hat{\mathcal{A}}\left(\lambda, \mathbf{e}, \mathbf{f}\right) := \int_{t=0}^{+\infty} \exp\left(-\lambda\tau\right) \mathcal{A}\left(\tau, \mathbf{e}, \mathbf{f}\right) \tag{257}$$

ait un rayon spectral égal à 1. On va faire maintenant une hypothèse, certes simplificatrice, mais qui va nous permettre de préciser un peu les choses. Reprenons la matrice efficace d'infectivité, à savoir :

$$\mathbb{A}\left(\mathbf{e}, \mathbf{f}\right) = \int_{\tau=0}^{\tau=+\infty} \mathcal{A}\left(\tau, \mathbf{e}, \mathbf{f}\right) d\tau \tag{258}$$

L'hypothèse la plus simple (et d'ailleurs que font les auteurs sans s'en apercevoir), c'est de supposer que l'on a en fait :

$$\mathcal{A}\left(\tau, \mathbf{e}, \mathbf{f}\right) = \chi\left(t \leq T_c\right) \underline{\mathbb{A}}\left(\mathbf{e}, \mathbf{f}\right) \tag{259}$$

où $\chi\left(t \leq T_c\right)$ est la fonction caractéristique de l'intervalle $[0, T_c]$, c'est-à-dire que cette fonction vaut 1 si $0 \leq t \leq T_c$ et qu'elle vaut 1 si $t > T_c$. Ici le temps T_c s'interpréter comme la durée durant laquelle une personne elle-même infectée est contagieuse. Dans ce cas, on a peut relier maintenant la matrice $\hat{\mathcal{A}}\left(\lambda, \mathbf{e}, \mathbf{f}\right)$ avec la matrice efficace d'infectivité $\mathbb{A}\left(\mathbf{e}, \mathbf{f}\right)$:

$$\mathbb{A}\left(\mathbf{e}, \mathbf{f}\right) = T_c \underline{\mathbb{A}}\left(\mathbf{e}, \mathbf{f}\right) \tag{260}$$

Or, on sait que l'on a également :

$$\begin{aligned}
\hat{\mathcal{A}}(\lambda, \mathbf{e}, \mathbf{f}) &= \int_{\tau=0}^{\tau=\infty} \exp(-\lambda\tau)\, \mathcal{A}(\tau, \mathbf{e}, \mathbf{f})\, d\tau && (261)\\
&= \int_{\tau=0}^{+\infty} \chi(t \leq T_c) \exp(-\lambda\tau)\, \underline{\mathbb{A}}(\mathbf{e}, \mathbf{f})\, d\tau && (262)\\
&= \left(\int_{\tau=0}^{\tau=T_c} \exp(-\lambda\tau)\, d\tau \right) \underline{\mathbb{A}}(\mathbf{e}, \mathbf{f}) && (263)\\
&= \frac{1}{\lambda} (1 - \exp(-\lambda T_c))\, \underline{\mathbb{A}}(\mathbf{e}, \mathbf{f}) && (264)
\end{aligned}$$

On en tire ainsi la relation :

$$\hat{\mathcal{A}}(\lambda, \mathbf{e}, \mathbf{f}) = \frac{1}{\lambda T_c} (1 - \exp(-\lambda T_c))\, \mathbb{A}(\mathbf{e}, \mathbf{f}) \quad (265)$$

Maintenant, c'est assez simple car, dans l'hypothèse où $R_0 > 1$ (ce qui revient à dire que $\lambda > 0$), l'on sait que $\lambda > 0$ est le seul réel tel que l'on ait :

$$\rho\left(\hat{\mathcal{A}}(\lambda)\right) = 1 \quad (266)$$

On peut alors facilement écrire maintenant, dans le modèle que nous avons posé, la relation entre λ, T_c, R_0 puisque l'on a (avec $\rho(\cdot)$ l'application qui donne le rayon spectral d'une matrice) :

$$\begin{aligned}
\rho\left(\hat{\mathcal{A}}(\lambda, \mathbf{e}, \mathbf{f})\right) &= \rho\left(\frac{1}{\lambda T_c} (1 - \exp(-\lambda T_c))\, \mathbb{A}(\mathbf{e}, \mathbf{f})\right) && (267)\\
1 &= \frac{1}{\lambda T_c} (1 - \exp(-\lambda T_c))\, \rho(\mathbb{A}(\mathbf{e}, \mathbf{f})) && (268)\\
1 &= \frac{1}{\lambda T_c} (1 - \exp(-\lambda T_c))\, R_0 && (269)\\
R_0 &= \frac{\lambda T_c}{(1 - \exp(-\lambda T_c))} && (270)
\end{aligned}$$

Comme on ne sait pas très quelle valeur donner à T_c, on peut fournir la courbe donnant le R_0 en fonction de T_c, dans le modèle que nous avons donné de $\mathcal{A}(\tau, \mathbf{e}, \mathbf{f})$. On obtient la figure :

La courbe est presque une droite pour les grandes valeurs de T_c puisque l'on a raisonnablement $1 - \exp(-\lambda T_c) \approx 1$. Dans cette approximation-là, si comme les auteurs l'affirment (et on ne sait pas trop comment que $R_0 = 3.3$), alors on en déduit que l'on a $T_c \approx 15$ *jours*. Ainsi, avec toutes ses hypothèses, on arrive à l'idée, si l'on accepte un R_0 de 3.3, que, d'après l'estimation que nous avons faite de λ, une personne infectée est contaminante pendant $T_c = 15$ `jours`. Il faut à ce stade noter une chose essentielle :

SI L'ON FAIT L'HYPOTHÈSE D'UN TAUX DIFFÉRENTIEL D'INFECTIVITÉ DE LA FORME $\chi(t \leq T_c)\, \mathbb{A}(\mathbf{e}, \mathbf{f})$, ALORS LE R_0 EST INDÉPENDANT DE LA FORME DE $\mathbb{A}(\mathbf{e}, \mathbf{f})$ SI L'ON ESTIME QUE λ EST UNE DONNÉE MESURÉE.

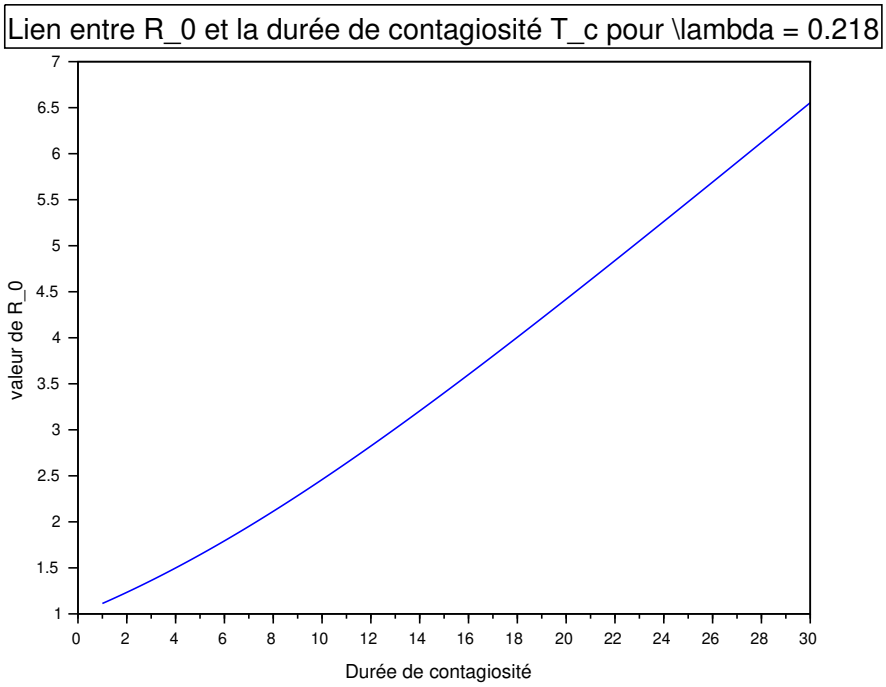

FIGURE 11 – relation entre le nombre basique de reproduction et le temps de contagiosité pour $\lambda = 0.218$ jour^{-1}

Ce qui est en pratique le cas : il est important de mesurer λ comme nous l'avons fait et ensuite de s'en servir pour avoir une relation entre R_0 et T_c. Cette relation est une conséquence immédiate de la théorie du taux différentiel d'infectivité, sous l'hypothèse - dirons-nous - du $\chi(t \leq T_c)$. En général, on peut avoir une bonne idée de T_c, ce qui rend alors la détermination de R_0 assez facile. Mais cela ne tient que si l'on a des observations qui permettent d'avoir une bonne confiance en λ. Comme pour notre plage d'observation c'est le cas (on l'a via une régression linéaire avec un coefficient de 97%), il semble que l'on peut faire relativement confiance aux chiffres obtenus.

5.4 le festival de nullités des auteurs quant au calcul de R_0

Dans le cadre de l'hypothèse du $\chi(t \leq T_c)$, le lien entre λ, T_c, R_0 (via la formule 270) est évident. Il est ainsi remarquable que la plupart des épidémiologistes cherchent impérativement à donner un R_0, quand bien même la seule mesure possible est celle du *taux de croissance exponentielle* λ. Ensuite, si l'on se donne une estimation de T_c, on arrive directement sur le R_0. Cela, bien sûr, nos auteurs l'ignorent : ils n'ont jamais été capables de comprendre l'article originel

sur la théorie du R_0 et donc de comprendre ce lien. Au lieu, donc, de travailler de manière précise, ils vont produire un festival d'hypothèses toutes aussi barbares les unes que les autres... Dans un premier temps, ils vont se donner une forme a priori de la *matrice efficace*. Il vont la choisir selon celle utilisée dans la référence N. Hens, G. M. Ayele, N. Goeyvaerts, M. Aerts, J. Mossong, J. W. Edmunds, P. Beutels, *Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries.* BMC Infect. Dis. 9, 187 (2009) :

$$\mathbb{A}_{aut}(\mathbf{e},\mathbf{f}) = \frac{ND}{L} q\mathbb{C} \tag{271}$$

Notons déjà qu'il y a un problème dans l'article cité par les auteurs car dans cet article on confond manifestement déjà matrice différentielle et matrice efficace. Originellement, la matrice « next gen » est une matrice efficace, c'est à dire que c'est un taux par unité de personne, mais plus par unité de temps :

1. dans le modèle des auteurs $\mathbb{C}$ est la matrice de « contacts »par jour et par habitant selon les classes d'âge. Autrement dit, $C(\mathbf{e},\mathbf{f})$ est le nombre de « contacts » qu'un individu de la classe $\mathbf{e}$ échange avec ceux de la classe $\mathbf{f}$ en une journée. C'est ce qui fait qu'il s'agit plus de la matrice différentielle d'infectivité que la matrice effective.
2. Le nombre q est ensuite un facteur de proportionnalité : celui de la « probabilité »que le virus a de se transmettre à chaque contact. Nous avons déjà dit ce que nous pensions de l'existence « en soi »d'une telle probabilité. Nous ne reviendrons pas sur cette hérésie. En outre, l'idée que le virus se propage uniquement via des contacts (qui ne sont pas sexuels) quand on ignore à peu près tout du mode exact de transmission du virus, voilà une idée exactement sans intérêt. D'autre part, pour être précis, q devrait vérifier a minima la condition de dilution, à savoir :

$$\int q\rho(\mathbb{C})\, d\tau << 1 \tag{272}$$

   Mais évidemment vous ne sauriez rien de tout cela. Le modèle des auteurs n'est pas fait pour être sérieux.
3. N est la population du pays.
4. D est le « temps moyen de transmission de l'infection » : il s'agit sans doute du temps
$$\frac{1}{\lambda}\left(1-\exp\left(-\lambda T_c\right)\right)$$
5. L est la « durée de vie ». Manifestement il s'agit du temps T_c.

Évidemment, rien de tout cela n'est spécifié ni dans l'article des auteurs ni dans la référence qu'ils utilisent. Cependant, si l'on résume et que l'on identifie proprement la prétention des auteurs avec le modèle du taux différentiel d'infectivité, dans l'hypothèse $\chi(t \leq T_c)$, on voit que l'on a :

$$\int Nq\mathbb{C}(\mathbf{e},\mathbf{f})\, d\tau \sim \mathbb{A}(\mathbf{e},\mathbf{f}) \tag{273}$$

(Si l'on estime que q joue le rôle d'une « probabilité en soi »). Si l'on accepte alors l'idée que le temps moyen de transmission de l'infection vaut précisément :

$$D = \int_{\tau=0}^{+\infty} \exp\left(-\lambda\tau\right) d\tau = \frac{1}{\lambda} \approx \frac{1}{\lambda}\left(1 - \exp\left(-\lambda T_c\right)\right) \tag{274}$$

on obtient alors immédiatement que l'on a :

$$\frac{ND}{L} q\mathbb{C} \sim \frac{1}{\lambda T_c}\left(1 - \exp\left(-\lambda T_c\right)\right)\mathbb{A} = \hat{\mathcal{A}}\left(\lambda\right) \tag{275}$$

Dit autrement, dans leur article, les auteurs confondent la matrice efficace $\mathbb{A}$ avec $\hat{\mathcal{A}}\left(\lambda\right)$ qui est la transformée de Laplace de $\mathcal{A}\left(\tau\right)$!!! Cela revient exactement à confondre :

$$\mathbb{A}\left(\mathbf{e},\mathbf{f}\right) = \int_{\tau=0}^{+\infty} \mathcal{A}\left(\tau,\mathbf{e},\mathbf{f}\right) d\tau \quad et \quad \int_{\tau=0}^{\tau=\infty} \mathcal{A}\left(\tau,\mathbf{e},\mathbf{f}\right) \exp\left(-\lambda\tau\right) d\tau = \hat{\mathcal{A}}\left(\lambda,\mathbf{e},\mathbf{f}\right)$$

Or, si le rayon spectral de l'une doit donner le fameux R_0, alors on sait que le lambda qui va permettre de retrouver les régimes exponentiels est précisément celui pour lequel (et il ne peut y en avoir qu'un seul) la matrice $\hat{\mathcal{A}}\left(\lambda\right)$ a effectivement un rayon spectral égal à 1. Quant à la manière avec laquelle ils vont tenter de « démontrer »que le confinement a effectivement permis de réduire le R_0, croyez-moi : même avec la meilleure volonté du monde - et j'en ai - c'est incompréhensible. D'ailleurs, il est inutile d'essayer d'aller plus loin car, en confondant la matrice efficace et la transformée de Laplace du taux différentiel d'infectivité, toute tentative de construire un raisonnement valable est sans espoir. On va maintenant démontrer très simplement que les affirmations des auteurs sont totalement fausses. Pour cela :

1. On va supposer, comme le prétendent les auteurs, que le confinement, dès le 17 mars, a un effet sur la transmission du virus, qui est capable de changer radicalement le taux différentiel d'infectivité. Notons par $\mathcal{A}_l\left(\tau,\mathbf{e},\mathbf{f}\right)$ un tel taux.
2. Cela revient à faire une sorte de « restart »dans le problème : on part de $\mathcal{I}_s\left(t_f\right)$ la population infectées à la date t_f et on écrit une nouvelle succession de « next gen ».
3. Au bout d'un certain temps - disons $\Delta_t = 8$ jours -, le nouveau problème entre dans son régime exponentiel. On sait alors que l'on doit avoir :

$$\forall t \geq t_f + \Delta_t, \;\; \mathcal{I}_s\left(t\right) - \mathcal{I}_s\left(t_f + \Delta_t\right) = \int_{\tau=t}^{\tau=t_f+\Delta_t} i_s\left(t_f + \Delta_t\right) \exp\left(\lambda_l \tau\right) d\tau, \tag{276}$$

   où selon les auteurs $\lambda_l < 0$ puisque selon eux on a $R_l < 1$ (le nouveau taux basique de reproduction).
4. Ce régime exponentiel doit au moins s'observer sur une dizaine de jours.

Les deux seules choses que l'on doit obtenir, c'est donc $i_s\left(t_f + \Delta_t\right)$ et λ_l. Pour cela :

1. On peut obtenir une estimation de λ_l en se donnant l'équation 270 qui est aussi valable pour $R_0 < 1$ et pour $\lambda < 0$. Une résolution numérique donne alors :

$$\lambda_l = -0.08435 \; \texttt{jour}^{-1} \tag{277}$$

2. Ne reste plus qu'à trouver maintenant la valeur de $i_s\left(t_f+\Delta_t\right)$ avec $t_f = 22$ et $\Delta_t = 8$. On a alors finalement $i_s\left(30\right) = 2931$. Ainsi que $\mathcal{I}_s\left(30\right) = 25193$

On peut donc tracer facilement la courbe pour les $\forall t \geq 30$

$$\mathcal{I}_s\left(t\right) = 25193 - \frac{2931}{0.08435}\left[\exp\left(-0.08435t\right) - \exp\left(-0.08435 * 30\right)\right] \tag{278}$$

On obtient alors la courbe suivante, qui se passe de commentaires :

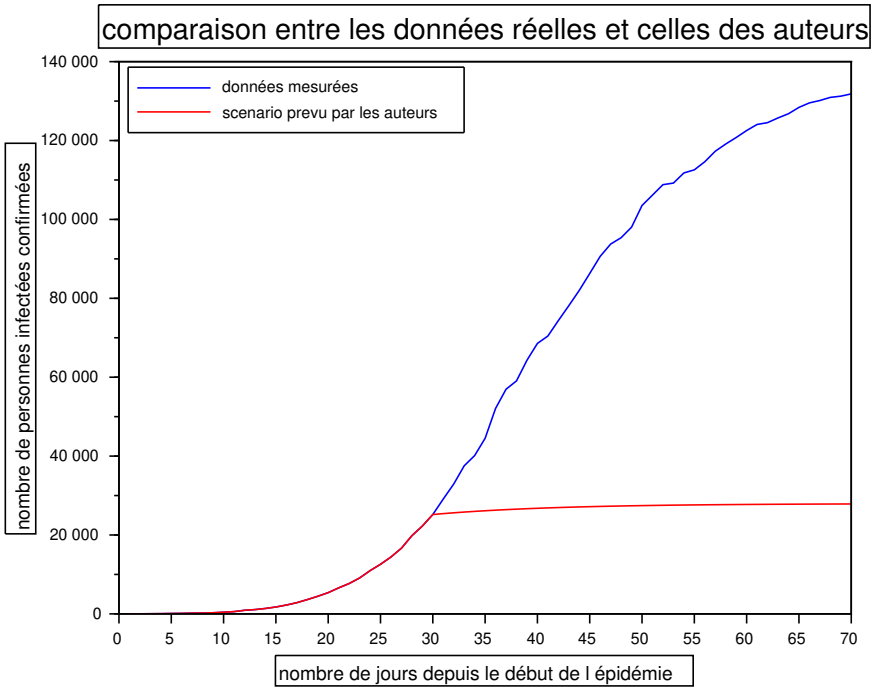

FIGURE 12 – Comparaison entre les données mesurées et le scenario prevu par les auteurs.

6 Conclusion

Au terme de cette étude mathématique, il apparaît ainsi que 17 auteurs de 10 laboratoires - dont les noms comptent parmi les plus connus au monde - n'ont aucune compétence pour traiter des questions d'épidémiologie.
En réalité, cet article n'est en aucun cas un article qui pourrait remplir un quelconque critère de qualité pour publication et encore moins pour des gouvernants avisés.
Totalement illisible, truffé d'erreurs grossières, d'incompréhensions notoires sur des notions mathématiques parmi les plus élémentaires, ce papier ne visait en fait qu'à servir un pouvoir avide de justifications « scientifiques » de sa politique : le confinement et la surveillance généralisée via des mesures de contrôle.

Il n'y a aucun scientifique dans cette liste d'auteurs. Seulement de nouveaux Lyssenko en puissance, dangereusement empressés de pouvoir jouer un rôle dans la conduite d'une politique liberticide du pays.
À vous les pseudos épidémiologistes qui vous permettez ainsi d'insulter la science tout en vous félicitant d'éclairer le monde de vos merveilleuses compétences, nous répondons ici que vous ne semez que l'ignorance barbare. Vous servez docilement, par des trucages scientifiques, les dirigeants qui prétendent museler la contestation de leur politique au nom de la science.
À jamais honte à vous.

7 annexe

Nombre de nouvelles personnes hospitalisées à partir du 18 mars 2020 pour cause de COVID

j	0	1	2	3	4	5	6	7	8
H_j	1229	1256	1540	1534	2053	2618	3166	3096	3058
j	9	10	11	12	13	14	15	16	17
H_j	3332	2685	3107	4145	4281	3845	3627	2822	1931
j	18	19	20	21	22	23	24	25	26
H_j	2754	3277	3139	2990	3161	2044	1688	1257	1965
j	27	28	29	30	31	32	33	34	35
H_j	2415	2084	2166	1565	890	1464	1885	1619	1410
H_j	36	37	38						
j	1346	999	481						

k	0	1	2	3	4	5	6	7
π_k^{obs}	0.17	0.09	0.09	0.09	0.08	0.072	0.063	0.055
π_k^{true}	0.1532748	0.0815819	0.0825034	0.0837783	0.075695	0.0694375	0.0621514	0.0552432
k	8	9	10	11	12	13	14	15
π_k^{obs}	0.047	0.038	0.03	0.027	0.025	0.022	0.02	0.018
π_k^{true}	0.0477341	0.0393653	0.0319624	0.0295761	0.028309	0.0256149	0.0237143	0.0218831
k	16	17	18	19	20	21	22	
π_k^{obs}	0.016	0.013	0.011	0.009	0.007	0.005	0.002	
π_k^{true}	0.0200667	0.0171424	0.0152465	0.0131818	0.0108977	0.0082186	0.0034214	

Les données du nombre de nouvelles personnes confirmées depuis le 25 février.

$jour$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
N_j	13	5	20	19	43	30	61	21	73	138	150	336	177	286	372	497

$jour$	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
N_j	595	785	839	923	1210	1097	1404	1861	1617	1847	2230	3167	2446	2931	3922

$jour$	32	33	34	35	36	37	38	39	40	41	42	43	44
N_j	3809	4611	2599	4376	7578	4861	2116	5233	4267	1873	3912	3777	3881

$jour$	45	46	47	48	49	50	51	52	53	54	55	56	57
N_j	4286	4342	3114	1613	2673	5499	2633	2641	405	2569	785	2051	2667

$jour$	58	59	60	61	62	63	64	65	66	67	68	69	70
N_j	1827	1653	1773	1537	461	1195	1065	1607	1139	604	794	308	576