Sur le pic épidémique dans un modèle S-I-R

Nicolas Bacaër*

Quadrature 117 (2020) p. 9-12.

Résumé

On étudie le comportement asymptotique, lorsque la taille N de la population est grande, du temps T que met une épidémie modélisée par un système différentiel de type S-I-R pour atteindre son pic. On trouve que $T \sim (\ln N)/(a-b)$, où a est le taux de contact effectif et b le taux auquel les personnes contagieuses cessent de l'être.

1 Introduction

Considérons comme par exemple dans [1, 2, 6, 7, 8] le modèle S-I-R de Kermack et McKendrick pour une épidémie

$$\frac{dS}{dt} = -a\,\frac{S\,I}{N}, \quad \frac{dI}{dt} = a\,\frac{S\,I}{N} - b\,I, \quad \frac{dR}{dt} = b\,I,$$

où $S(t)$ est le nombre de personnes susceptibles d'être infectées, $I(t)$ le nombre de personnes infectées, $R(t)$ le nombre de personnes retirées de la transmission (par suite d'un confinement, d'une guérison ou d'un décès) et $N = S(t) + I(t) + R(t)$ la population totale. Cette population totale est supposée constante et suffisamment grande pour qu'il soit raisonnable de modéliser l'épidémie par un système différentiel plutôt que par un processus stochastique. Considérons les conditions initiales

$$S(0) = N - 1, \quad I(0) = 1, \quad R(0) = 0,$$

qui correspondent au démarrage de l'épidémie à partir d'un seul « patient zéro ». Le paramètre a (avec $a > 0$) est en fait le produit de deux nombres : le nombre de contacts par unité de temps et la probabilité de transmission lors d'un contact entre une personne non infectée et une personne infectée. Les contacts sont supposés avoir lieu au hasard, de sorte qu'un contact d'une personne non infectée est lui-même infecté avec une probabilité I/N, ce qui explique le terme quadratique en SI/N dans le système différentiel. Le paramètre b (avec $b > 0$) est le

*Institut de recherche pour le développement, Unité de modélisation mathématique et informatique des systèmes complexes, Les Cordeliers, Paris, France, nicolas.bacaer@ird.fr

taux auquel les personnes quittent le compartiment I pour se retrouver dans le compartiment R ; l'hypothèse sous-jacente est donc que le temps passé dans le compartiment I est distribué selon une loi exponentielle de paramètre b. On supposera $a > b$ pour qu'une véritable épidémie ait lieu.

Un exemple est détaillé à la figure 1. On a utilisé le logiciel Scilab et sa fonction `ode` de résolution des systèmes différentiels ; celle-ci implémente un « schéma prédicteur-correcteur d'Adams » pour les systèmes qui ne sont pas raides. On a choisi $N = 65 \times 10^6$ (la population de la France), $a = 1/2$ par jour et $b = 1/4$ par jour. Au début de l'épidémie, une personne infectée infecte donc en moyenne une personne tous les deux jours. Quant à la durée moyenne d'infection $1/b$, elle est de 4 jours. Cela correspond à une « reproductivité » $\mathcal{R}_0 = a/b = 2$ (le terme et la notation proviennent de [9, p. 102]). Ce $\mathcal{R}_0$ est donc le nombre moyen de cas secondaires qu'une personne infectée infecte au début de l'épidémie. Sans prétendre être réaliste, il s'agit de valeurs du même ordre de grandeur que ce qu'on peut trouver pour l'épidémie de coronavirus de 2020 dans le cas hypothétique où il n'y aurait aucune intervention pour limiter l'épidémie [5].

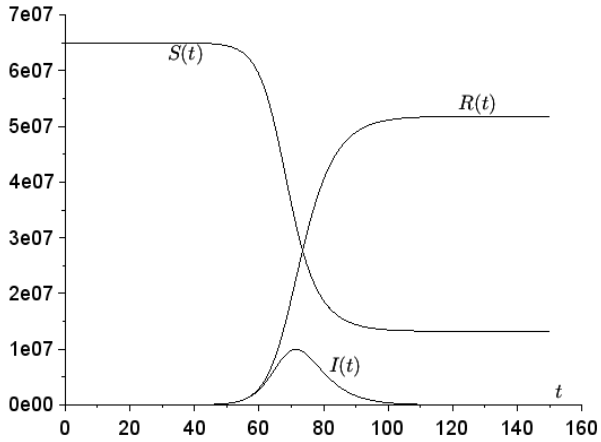


FIGURE 1 – Une simulation du modèle S-I-R avec en abscisse le temps t en jours.

2 Le pic épidémique

En combinant la première et la troisième équation du système S-I-R, il est bien connu [8, p. 76] qu'on trouve une intégrale première :

$$\frac{1}{S}\frac{dS}{dt} = -a\,\frac{I}{N} = -\frac{a}{bN}\frac{dR}{dt},$$

d'où $\ln[S(t)/S(0)] = -\frac{a}{bN}R(t)$. En reportant dans la première équation, on trouve

$$\frac{dS}{dt} = -a\frac{S}{N}(N - S - R) = -a\frac{S}{N}\left(N - S + \frac{bN}{a}\ln[S(t)/S(0)]\right).$$

Le pic épidémique de la figure 1 correspond à $dI/dt = 0$, ce qui arrive d'après la deuxième équation quand $S(t) = Nb/a$. Notons que $S(t)$ est monotone décroissante puisque $dS/dt \leq 0$ d'après la première équation. On en déduit que si T est l'instant du pic épidémique, c'est-à-dire si

$$I(T) = \max_{t>0} I(t),$$

alors

$$T = \int_0^T dt = \int_{S(0)}^{bN/a} \frac{dS}{-a\frac{S}{N}\left(N - S + \frac{bN}{a}\ln[S/S(0)]\right)}.$$

Posons $s = S/N$. Alors

$$T = \frac{1}{a}\int_{b/a}^{1-1/N} \frac{ds}{s\left(1 - s + \frac{b}{a}\ln[s/(1-1/N)]\right)}. \tag{1}$$

Cherchons une majoration de T. Pour $b/a < s < 1 - 1/N$, on remarque que

$$\frac{1}{s\left(1 - s + \frac{b}{a}\ln[s/(1-1/N)]\right)} \leq \frac{1}{s\left(1 - s + \frac{b}{a}\ln s\right)}.$$

Or au voisinage de $s = 1$, on a

$$\frac{1}{s\left(1 - s + \frac{b}{a}\ln s\right)} \sim \frac{1}{(1-s)(1-b/a)}.$$

L'intégrale $\int_{b/a}^1 \frac{ds}{1-s}$ est divergente. Lorsque $N \to +\infty$, on a donc

$$T \leq \frac{1}{a}\int_{b/a}^{1-1/N} \frac{ds}{s\left(1 - s + \frac{b}{a}\ln s\right)} \sim \frac{1}{a}\int_{b/a}^{1-1/N} \frac{ds}{(1-s)(1-b/a)} \sim \frac{\ln N}{a-b}.$$

Cherchons maintenant une minoration de T. Pour tout $s > 0$, $\ln s \leq s - 1$. On a par ailleurs $\ln(1 - 1/N) \geq -2/N$ pour tout entier $N \geq 2$. On en déduit que pour tout $s \in]b/a, 1 - 1/N[$ et tout entier $N \geq 2$,

$$\begin{aligned} 1 - s + \frac{b}{a}\ln[s/(1-1/N)] &= 1 - s + \frac{b}{a}\ln s - \frac{b}{a}\ln(1-1/N) \\ &\leq 1 - s + \frac{b}{a}(s-1) + \frac{2b}{aN}. \end{aligned}$$

Donc

$$T \geq \frac{1}{a}\int_{b/a}^{1-1/N} \frac{ds}{s[1 - \frac{b}{a} + \frac{2b}{aN} - (1-b/a)s]}.$$

On tombe sur une fraction rationnelle de la forme $\frac{1}{s(u-vs)}$ avec $u = 1 - \frac{b}{a} + \frac{2b}{aN}$ et $v = 1 - b/a$. Elle se décompose en éléments simples comme $\frac{1}{u}(\frac{1}{s} + \frac{v}{u-vs})$. Une primitive est donc $\frac{1}{u}[\ln s - \ln(u - vs)] = \frac{1}{u} \ln \frac{s}{u-vs}$. Ainsi

$$\begin{aligned} T &\geq \frac{1}{a(1 - \frac{b}{a} + \frac{2b}{aN})} \ln \left\{ \frac{1 - 1/N}{b/a} \; \frac{1 - \frac{b}{a} + \frac{2b}{aN} - (1 - b/a)b/a}{1 - \frac{b}{a} + \frac{2b}{aN} - (1 - b/a)(1 - 1/N)} \right\} \\ &= \frac{1}{a - b + 2b/N} \ln \left\{ \frac{1 - 1/N}{b/a} \; \frac{(1 - b/a)^2 + \frac{2b}{aN}}{(1 + b/a)/N} \right\} \\ &= \frac{1}{a - b + 2b/N} \left[\ln N + \ln \left\{ \frac{1 - 1/N}{b/a} \; \frac{(1 - b/a)^2 + \frac{2b}{aN}}{(1 + b/a)} \right\} \right]. \end{aligned}$$

Le membre de droite est équivalent à $(\ln N)/(a - b)$ quand $N \to +\infty$. Comme c'est le même équivalent que celui du majorant obtenu ci-dessus, on en conclut que $T \sim (\ln N)/(a - b)$ quand $N \to +\infty$.

Dans l'exemple de la figure 1, on a $T \simeq 71{,}4$ jours tandis que $(\ln N)/(a-b) \simeq 72{,}0$ jours. L'approximation semble assez bonne. Dans la figure 2, on fait varier la taille N de la population ainsi que le paramètre a, avec b fixé. On remarque que $(\ln N)/(a - b)$ tantôt surestime, tantôt sous-estime T.

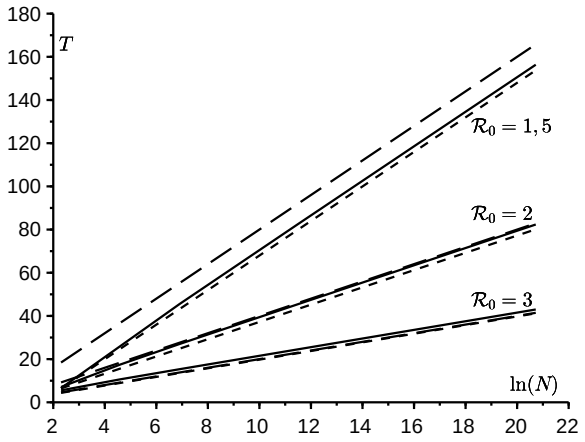


FIGURE 2 – Comparaison entre d'une part la date T du pic de l'épidémie (ligne continue) calculée avec l'intégrale de la formule (1), d'autre part l'équivalent $(\ln N)/(a - b)$ (grands pointillés) et l'approximation $[\ln(2N(1 - b/a)^2)]/(a - b)$ (petits pointillés), en fonction de $\ln(N)$ lorsque N varie de 10 à 10^8 pour trois valeurs du paramètre a telles que $\mathcal{R}_0 = a/b \in \{1{,}5 \; ; \; 2 \; ; \; 3\}$.

Notons que $a-b$ est le taux de croissance de l'épidémie à ses débuts puisqu'on a alors, pour N grand, $S(t) \simeq N$ et $dI/dt \simeq (a-b)I$. Ce taux peut être estimé à partir des données épidémiologiques. Connaissant la population totale N, on peut donc prévoir la date du pic de l'épidémie. En pratique, l'hypothèse sous-jacente au modèle d'un mélange homogène de la population reste néanmoins douteuse à l'échelle d'un pays, un peu moins à l'échelle d'une ville.

Historiquement, Kermack et McKendrick avaient résolu de manière approchée le système S-I-R en supposant que la reproductivité $\mathcal{R}_0$ était proche de 1, ce qui conduit à une petite épidémie. En utilisant un développement limité du type $e^x \simeq 1 + x + x^2/2$, ils étaient parvenus à une équation de Riccati résoluble explicitement, de sorte que

$$I(t) \simeq \frac{X}{\mathrm{ch}^2(Yt - Z)}, \tag{2}$$

où X, Y et Z s'expriment de manière particulièrement compliquée en fonction des paramètres N, a et b (voir par exemple [3]). Le pic épidémique a lieu en $T \simeq Z/Y$. Notons au passage qu'à notre connaissance l'approximation (2) n'a jamais été démontrée rigoureusement, ce pourquoi on utilise ci-dessus et ci-dessous le signe informel $\simeq$. Supposons maintenant en plus que $1/N$ soit petit devant $(\mathcal{R}_0 - 1)^2$. Dans ce cas,

$$Y \simeq \frac{a-b}{2}, \quad Z \simeq \frac{1}{2}\ln[2N(1-b/a)^2]$$

(voir par exemple [4]), de sorte que

$$T \simeq \frac{\ln[2N(1-b/a)^2]}{a-b}\,.$$

Cette approximation est également représentée dans la figure 2 ; elle semble particulièrement proche de la vraie valeur de T lorsque $\mathcal{R}_0$ est le plus proche de 1, comme on pourrait s'y attendre. Avec N qui tend vers l'infini, on a néanmoins $T \simeq \frac{\ln N}{a-b}$ et on retrouve le même comportement asymptotique. Notre analyse basée sur la formule (1) est valable cependant pour tout $\mathcal{R}_0 > 1$ et pas seulement pour $\mathcal{R}_0$ proche de 1 ; de plus, elle est rigoureuse.

3 Conclusion

On a trouvé que le pic de l'épidémie avait lieu à un temps T tel que $T \sim (\ln N)/(a-b)$ lorsque N tend vers l'infini. Ce résultat semble nouveau malgré sa simplicité. Dans la littérature scientifique, on n'a trouvé que des informations sur la hauteur du pic $I(T)$, qui se détermine facilement grâce à l'intégrale première du système. En effet, Avec

$$S(T) = \frac{N\,b}{a}, \quad \ln\frac{S(T)}{S(0)} = -\frac{a}{bN}R(T), \quad N = S(T) + I(T) + R(T),$$

on obtient

$$\frac{I(T)}{N} = 1 - \frac{b}{a} + \frac{b}{a}\log\left(\frac{N}{N-1}\,\frac{b}{a}\right) \simeq 1 - \frac{b}{a} + \frac{b}{a}\log\frac{b}{a}$$

(voir par exemple [1, équation 7.26]). On remarque bien dans certaines références comme [7, p. 238] des inversions où l'on considère le temps t comme fonction des variables monotones R ou S, ce qui conduit à des intégrales comme (1) à un changement de variable près. Mais comme le système différentiel S-I-R y était écrit sous la forme $dS/dt = -\alpha SI$, l'idée n'était probablement pas venue de prendre $\alpha = a/N$, ce qui est d'une certaine manière plus réaliste et nécessaire pour que la limite $N \to +\infty$ donne des résultats intéressants.

Pour une épidémie comme celle de coronavirus, le modèle S-I-R manque un peu de réalisme. Cela conduit à considérer des modèles avec plus de compartiments, par exemple avec un compartiment supplémentaire pour les personnes dans la phase latente qui sont infectées mais pas encore contagieuses. On pourra consulter à ce sujet [5], qui essaie aussi d'estimer les paramètres à partir de la courbe épidémique. L'étude analytique du pic épidémique dans ce cadre-là est sans doute plus difficile.

Remerciements

On remercie Hisashi Inaba pour ses commentaires sur le texte.

Références

[1] B. Anselme, *Biomathématiques*, Dunod, Paris, 2015.

[2] N. Bacaër, *Histoires de mathématiques et de populations*, Cassini, Paris, 2009.

[3] N. Bacaër, *Le modèle de Kermack et McKendrick pour la peste à Bombay et la reproductivité nette d'un type avec saisonnalité*. J. Math. Biol. 64 (2012) 403-422, hal.archives-ouvertes.fr/hal-01340008.

[4] N. Bacaër, *Sur la taille finale des épidémies dans un environnement périodique*, C. R. Biol. 342 (2019) 119-123.

[5] N. Bacaër, *Un modèle mathématique des débuts de l'épidémie de coronavirus en France* (2020) hal.archives-ouvertes.fr/hal-02509142

[6] M. Corlosquet-Habart, J. Janssen, R. Manca, *Modélisation stochastique du risque de pandémie : stratégies de couverture et d'assurance*, Lavoisier, Cachan, 2012.

[7] J. Gani, *Processus stochastiques de population*, P.-L. Hennequin (éd.), École d'été de probabilités de Saint-Flour IV, Springer, Berlin, 1975, p. 188-293.

[8] A. Hillion, *Les Théories mathématiques des populations*, Presses Universitaires de France, Paris, 1986.

[9] A.J. Lotka, *Théorie analytique des associations biologiques, 2e partie*, Hermann, Paris, 1939.