

## PROBLÈMES DE MODÉLISATION : LE SIDA

Paul GIRAULT

Strasbourg, U.F.R. de Mathématiques et d'Informatique

L'observation et l'étude des épidémies ont une longue histoire. Il en est de même d'ailleurs des modèles et des explications qu'elles ont suscitées afin de justifier leur progrès et leurs causes. Rappelons que dans bien des cultures, et même en Europe avant l'invention des microbes ainsi que l'écrivait si joliment Marcel Pagnol, les épidémies comme les maladies étaient attribuées à l'action délétère d'esprits malfaisants ou de divinités mécontentes. Le SIDA, l'épidémie des années 1980 et probablement du vingtième siècle a même été présenté par beaucoup comme étant un châtement... envoyé par Dieu... C'est pourquoi on n'en admirera que plus Daniel Bernoulli (1760) [2] : en effet, dans un article dont le titre fait tout le charme des écrits anciens, il est le premier à avoir proposé un modèle mathématique en épidémiologie. L'objet de sa modélisation (qui avait pour principal ingrédient une équation différentielle ordinaire non linéaire) était de constater l'effet de l'inoculation de la vaccine sur la propagation de la variole. C'est certainement la première fois qu'un modèle mathématique fut utilisé pour imposer les avantages pratiques d'un programme contrôlé de vaccination! Néanmoins, ce sont les travaux théoriques de Kermack et MacKendrick (1927, 1932, 1933) utilisant les données de la peste de Bombay de 1905 qui eurent une influence décisive dans le développement des modèles mathématiques en épidémiologie. De nos jours, la littérature sur le sujet est particulièrement abondante... Dans la suite, nous allons exposer succinctement deux modèles simples de dynamiques de transmission du VIH inspirés des travaux d'Anderson et de ses collaborateurs (1986) afin d'allécher le lecteur aux problèmes de modélisation.

---

Le **virus de l'immunodéficience humaine** VIH conduit en général, mais pas toujours, au **syndrome de l'immunodéficience acquise** alias SIDA. Quand des anticorps au VIH sont détectés, le patient est infecté et dit être **séropositif**. Après la détection d'anticorps au VIH, il y a une période latente avant que le patient soit au dernier stade de la maladie, à savoir le SIDA. La durée de cette période dépend probablement des situations individuelles : elle n'est donc pas bien connue. Comme le montrent de nombreux documents, elle peut être de quelques mois à plusieurs années. En outre, la proportion de la population qui est séropositive est mal cernée. En fait, les paramètres épidémiologiques de base associés à la progression de la maladie sont encore mal connus. Ceci est dû en grande part aux nombreux problèmes sociaux créés par le rassemblement des données sur le nombre de personnes atteintes du VIH. Pourtant il est invraisemblable de croire que l'épidémie sera contenue si l'information n'est pas disponible dans un proche avenir. Rappelons aussi que, dans les pays développés, le SIDA est associé aux communautés homosexuelles alors que dans les autres contrées c'est l'extension hétérosexuelle qui prévaut. Toutefois, comme il appert de nombreux signes, le SIDA progresse nettement dans les communautés hétérosexuelles des pays développés. Pour finir sur cette présentation rapide du SIDA, il faut bien constater que la virulence et l'extension rapide de l'épidémie sont alarmantes.

Certains pronostiquent que ce sera une sévère épidémie mondiale, on parle alors de pandémie, dès la fin de ce siècle, comparable à la Peste Noire de 1347-1350 qui décima environ 25 % de la population européenne. Pour une plus ample information, on pourra consulter les ouvrages techniques [3], [4] et [5].

### I. Un premier modèle [1]

Comme on l'a vu précédemment, un problème majeur avec le SIDA est la durée variable de la période d'incubation entre le moment où le patient est diagnostiqué comme étant séropositif et le moment où il montre les symptômes du SIDA. Ceci a d'ailleurs d'importantes conséquences dans la propagation du virus. Par conséquent, le premier modèle que nous allons considérer est consacré à l'évolution de la maladie dans une population qui, infectée par VIH, voit tous ses membres atteints du SIDA.

Considérons une population dont tous les individus sont infectés par le VIH au temps  $t = 0$ . On partitionne cette population en deux sous-ensembles et désignons par :

- .  $y(t)$  la fraction de la population totale qui a le SIDA au temps  $t$ ,
- .  $x(t)$  la fraction de la population totale qui est séropositive mais n'a pas encore le SIDA.

Par conséquent, on a  $x(t) = 1 - y(t)$ . Soit maintenant  $v(t)$  le taux de conversion de l'infection au SIDA. On obtient alors un modèle simple pour les dynamiques de ces deux populations, à savoir :

$$\frac{dx}{dt} = -v(t)x, \quad \frac{dy}{dt} = v(t)x, \quad x(0) = 1, \quad y(0) = 0$$

où  $x(t) + y(t) = 1$ .

On notera que ce modèle suppose que tous les individus infectés par le VIH auront le SIDA, ce qui n'est pas nécessairement le cas.

Il reste maintenant à caractériser la fonction  $v$ . Pour cela, si on suppose que le système immunitaire du patient, relativement aux maladies opportunes comme le cancer, est progressivement altéré depuis le début de l'infection, alors il est raisonnable de penser que  $v$  est une fonction croissante. Par conséquent, choisissons une fonction très simple, à savoir :

$$v(t) = at, \quad a > 0.$$

Après un calcul très simple, on obtient :

$$x(t) = \exp\left(-\frac{at^2}{2}\right), \quad y(t) = 1 - \exp\left(-\frac{at^2}{2}\right).$$

Peterman, Drotman et Curran (1985) ont présenté des données au sujet de 194 cas de SIDA associé à des transfusions sanguines. Les solutions précédentes furent

ensuite appliquées à ces données, le plus difficile étant de déterminer un ajustement correct du paramètre  $a$ . La valeur  $a = 0,237$  par an a été retenue. La courbe au tracé continu de la figure 1 montre le résultat que l'on obtient si on porte en ordonnée le taux de croissance  $dy/dt$  (en patients du SIDA) et en abscisse le temps (en années) écoulé depuis l'infection. La comparaison ainsi obtenue entre la théorie et les données peut être considérée comme assez bonne.

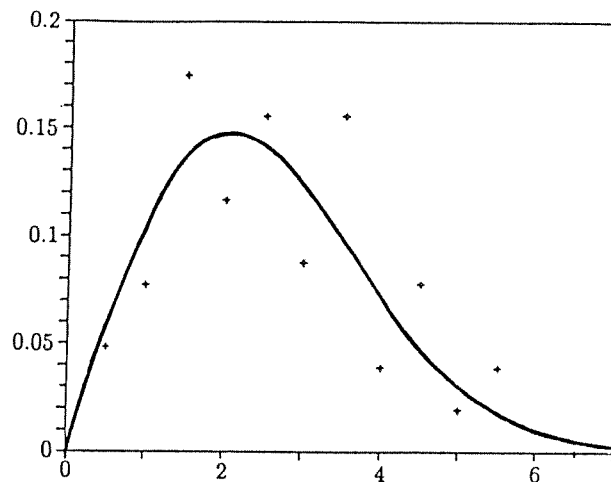


Figure 1

## II. Un modèle d'épidémie [1]

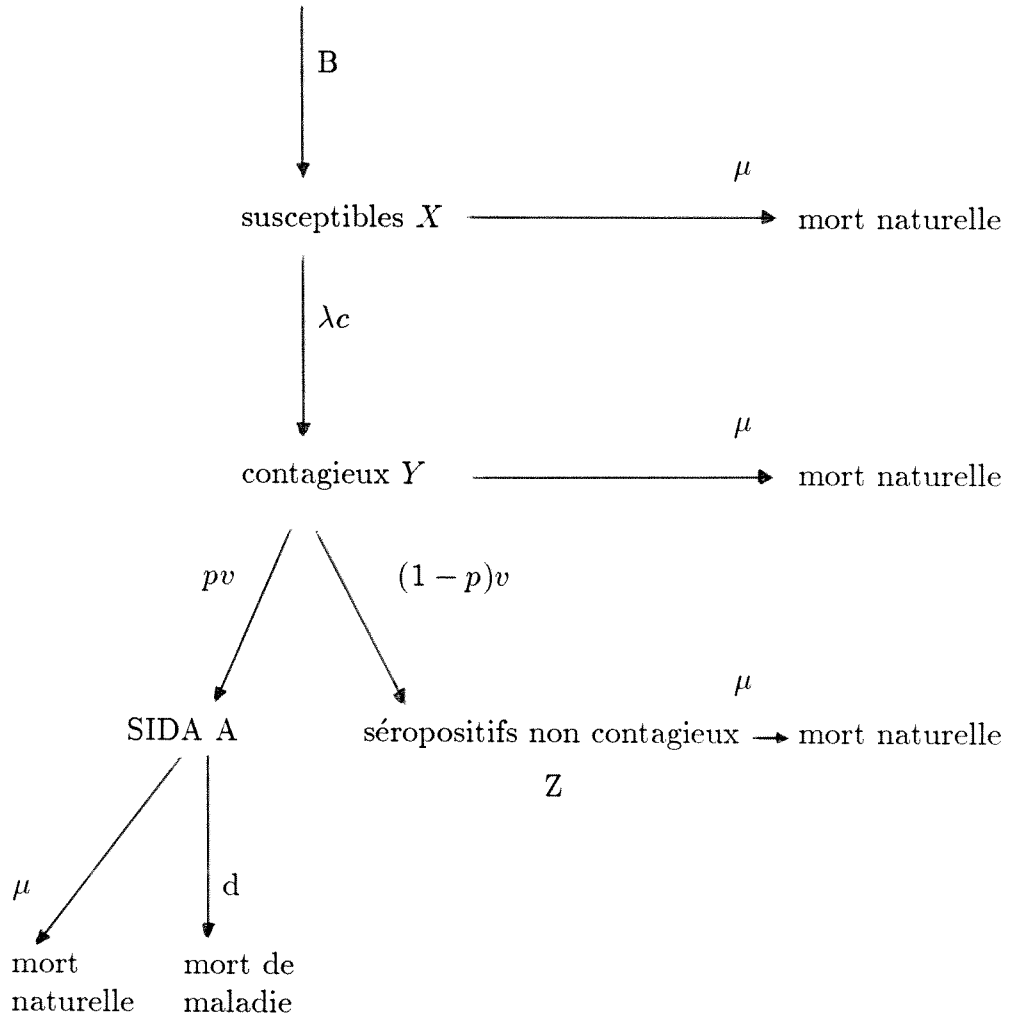
L'objet de nos préoccupations dans ce paragraphe est de modéliser une épidémie de SIDA dans une population d'homosexuels.

Considérons une population masculine d'homosexuels d'effectif  $N(t)$  au temps  $t$ . Désignons par

- .  $X(t)$  le nombre de susceptibles d'être atteints du SIDA,
- .  $Y(t)$  le nombre d'hommes contagieux (ou infectés),
- .  $Z(t)$  le nombre de séropositifs non contagieux,
- .  $A(t)$  le nombre de patients ayant le SIDA.

On suppose en outre que les susceptibles meurent naturellement au taux  $\mu$  et que les malades du SIDA meurent au taux  $d$ . Indiquons en passant qu'une valeur typique de  $1/d$  est de 9 à 12 mois. Evidemment les données précédentes ne sont pas suffisantes pour mettre en œuvre notre modélisation. Pour cela, aidons-nous du diagramme de la page suivante qui indique la "carrière" des individus.

Il ne reste plus qu'à supposer la population uniformément mêlée. Alors un modèle fondé sur le diagramme est donné par la formule qui lui succède :



$$\left\{ \begin{array}{l} \frac{dX}{dt} = B - \mu X - \lambda c X, \quad (1) \\ \frac{dY}{dt} = \lambda c X - (v + \mu) Y, \quad (2) \\ \frac{dZ}{dt} = (1-p)v Y - \mu Z, \quad (3) \\ \frac{dA}{dt} = pv Y - (d + \mu) A, \quad (4) \\ N(t) = X(t) + Y(t) + Z(t) + A(t). \quad (5) \end{array} \right.$$

Précisons maintenant la signification des différents paramètres introduits dans le diagramme (ou le système différentiel) précédent :

- .  $B$  est le taux d'immigration d'hommes susceptibles dans la population,
- .  $\mu$  est le taux de décès naturel,
- .  $\lambda$  est la probabilité d'acquérir l'infection à partir d'un partenaire choisi au hasard.

PROBLÈMES DE MODÉLISATION : LE SIDA

On a  $\lambda = \frac{\beta Y}{N} \simeq \frac{\beta Y}{X+Y+Z}$  où  $\beta$  est la probabilité de transmission.

- .  $c$  est le nombre de partenaires sexuels,
  - .  $d$  est le taux de décès consécutif au SIDA,
  - .  $p$  est la proportion de séropositifs contagieux,
  - .  $v$  est le taux de conversion de l'infection vers le SIDA : on le supposera constant.
- Avec  $v$  constant,  $1/v$  est le temps d'incubation moyen de la maladie.

On a donc obtenu un système différentiel d'ordre 4 non linéaire qu'il n'est pas question de résoudre puisque c'est une chose que l'on ne sait pas faire. On va plutôt en faire une étude qualitative qui sera tout aussi suggestive.

Après avoir remarqué que  $N(t) = X(t) + Y(t) + Z(t) + A(t)$  n'est pas constant, si on additionne (1), (2), (3), (4), on obtient :

$$\frac{dN}{dt} = B - \mu N - dA. \quad (6)$$

On constate que s'il n'y avait pas de SIDA (i.e.  $A = 0$ ), l'état stationnaire de la population serait  $N^* = \frac{B}{\mu}$ .

A la vue du système, la question qui se pose est de savoir quand une épidémie se développe. La traduction mathématique se fait alors de la façon suivante : une épidémie apparaît si la fonction  $Y$  qui dénombre le nombre d'individus infectés est croissante dans un petit voisinage de  $t = 0$ . Par conséquent, si dans (2), au temps  $t = 0$ , un individu infecté est introduit dans une population de susceptibles sans autre infection, on a initialement  $X \simeq N$  et pour un temps  $t$  proche de 0

$$\frac{dY}{dt} \simeq (\beta c - v - \mu)Y \simeq v(R_0 - 1)Y. \quad (7)$$

On notera que la dernière égalité provient du fait que le temps d'incubation moyen  $1/v$  est beaucoup plus court que l'espérance de vie  $1/\mu$  d'un susceptible, i. e.  $\mu \ll v$ . Eh bien, on en déduit le seuil approximatif pour qu'une épidémie se déclenche, à savoir :

$$R_0 \simeq \frac{\beta c}{v} > 1. \quad (8)$$

On constate que le nombre  $R_0$  appelé le taux basique de reproduction de la maladie ne dépend que du nombre de partenaires sexuels  $c$ , de la probabilité de transmission  $\beta$  et du temps d'incubation moyen de la maladie  $1/v$ .

Quand on contemple un système différentiel, une des premières recherches à effectuer, ce sont ses états stationnaires (1). Dans le cas qui nous préoccupe, on

---

(1) Soit  $U \subset \mathbb{R}^n$  un domaine de  $\mathbb{R}^n$  et  $v$  un champ de vecteurs sur  $U$ . Les états stationnaires du système différentiel

$$\dot{x} = v(x),$$

sont les vecteurs réels  $x^*$  tels que  $v(x^*) = 0$ .

trouve un unique état stationnaire défini par :

$$\begin{aligned} X^* &= \frac{(v + \mu)N^*}{c\beta}, & Y^* &= \frac{(d + \mu)(\beta - \mu N^*)}{pvd} \\ Z^* &= \frac{(1 - p)(d + \mu)(\beta - \mu N^*)}{pd\mu}, & A^* &= \frac{\beta - \mu N^*}{d} \\ N^* &= \frac{B\beta[\mu(v + d + \mu) + vd(1 - p)]}{[v + \mu][\beta(d + \mu) - pv]}. \end{aligned} \quad (9)$$

Ainsi, si une épidémie se développe, le système (1) évolue vers l'état stationnaire donné par  $(X^*, Y^*, Z^*, A^*)$ . Plus précisément, on peut montrer que si on linéarise le système différentiel au voisinage de son état stationnaire,  $(X, Y, Z, A)$  tend vers  $(X^*, Y^*, Z^*, A^*)$  à la manière d'oscillations amorties avec une période d'oscillation donnée en fonction des paramètres du modèle. Avec des valeurs très courantes des paramètres, la période des éruptions épidémiques est de l'ordre de 30 à 40 ans.

On peut obtenir quelques informations intéressantes à partir d'une analyse du système concernant le début de l'épidémie. En effet, comme la population consiste essentiellement en susceptibles  $X \simeq N$ , l'équation donnant la croissance de la population de séropositifs  $Y$  est alors facile à obtenir. De (7), on déduit :

$$Y(t) = Y(0) \exp[v(R_0 - 1)t], \quad (10)$$

où  $Y(0)$  est le nombre d'individus infectés introduits dans la population susceptible. Rappelons que si  $R_0 > 1$ , on est en présence d'une épidémie. Il est alors particulièrement intéressant de déterminer le **temps de doublement** de la population infectée, c'est-à-dire le temps  $t_d$  tel que  $Y(t_d) = 2Y(0)$ . On trouve facilement :

$$t_d = \frac{\ln 2}{r} = \frac{\ln 2}{v(R_0 - 1)},$$

où  $r = v(R_0 - 1)$  est appelé le taux de croissance intrinsèque. On voit alors que plus  $R_0$  est grand, plus le temps de doublement est court.

Si on substitue (10) dans (3), on obtient :

$$\frac{dA}{dt} = pvY(0) \exp(rt) - (d + \mu)A.$$

Comme on peut supposer qu'au début de l'épidémie aucun patient n'a le SIDA, donc que  $A(0) = 0$ , alors

$$A(t) = pvY(0) \frac{\exp(rt) - \exp[(d + \mu)t]}{r + d + \mu}. \quad (12)$$

Il reste à fournir une estimation des différents paramètres. Par exemple, Anderson et May ont obtenu  $r = 0,88$  par an : pour cela ils ont utilisé des données

## PROBLÈMES DE MODÉLISATION : LE SIDA

concernant une population masculine composée d'homosexuels et de bisexuels de 6875 individus qui fréquentaient une clinique de San Francisco de 1978 à 1985. Plus généralement une estimation grossière des différents paramètres donnée à titre indicatif par les mêmes auteurs est :

$$\begin{aligned}
 R_0 &= 3 \text{ à } 4 \\
 d + \mu &\simeq d = 1 \text{ à } 1,33 \text{ par an} \\
 p &= 10\% \text{ à } 30\% \text{ (probablement plus)} \\
 v &\simeq 0,22 \text{ par an} \\
 c &= 2 \text{ à } 6 \text{ partenaires par mois.}
 \end{aligned}$$

Avec ces données, le temps  $t_d$  de doublement pour la classe  $Y$  est d'environ 9 mois. Des simulations numériques du système différentiel (1) - (4) donnent un bon aperçu du développement de l'épidémie après l'introduction du VIH dans une population masculine d'homosexuels susceptibles. La figure 2 montre une telle simulation : le modèle prédit que le nombre de séropositifs atteint un maximum 12 à 15 ans après introduction du VIH dans la population.

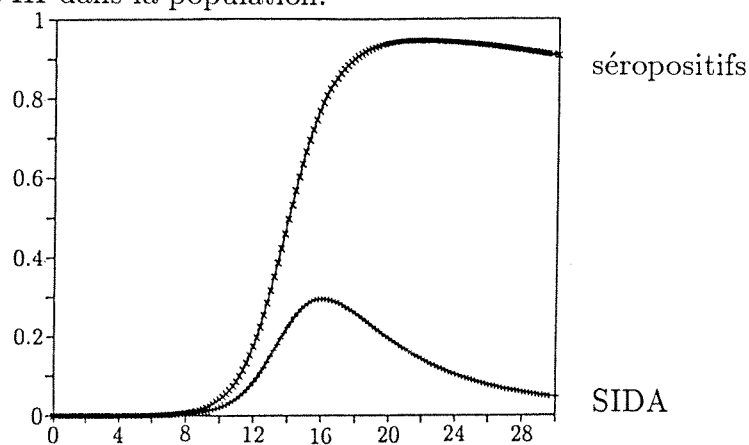


Figure 2

Les données numériques qui ont été employées pour cette simulation numérique sont toujours tirées des travaux d'Anderson et de ses collaborateurs [1]. Nous conseillons vivement le lecteur intéressé de se reporter à leur lecture pour le détail des calculs. Voici ces données :

$$\begin{aligned}
 A(0) &= Z(0) = 0 \\
 X(0) + Y(0) &= N(0) = 100.000 \\
 B &= 13333,3 \text{ par an} \\
 v &= 0,2 \text{ par an} \\
 \mu &= \frac{1}{32} \text{ par an} \\
 d &= 1 \text{ par an} \\
 p &= 0,3 \\
 R_0 &\simeq \frac{\beta c}{v} = 5,15.
 \end{aligned}$$

Le graphique montre la proportion des séropositifs et la proportion de ceux qui ont le SIDA reportées en ordonnée; en abscisse le temps est compté en années. On comparera la courbe de la population atteinte du SIDA avec celle donnée figure 1.

### Conclusion finale

Malgré la simplicité des modèles, les résultats sont conformes aux observations concernant les populations masculines d'homosexuels. Evidemment on en a proposés bien d'autres. D'ailleurs, on trouvera une revue des modèles mathématiques de dynamiques de transmission du VIH les plus usuels répertoriés par Isham (1988) [6]. Avec l'accumulation de plus de données et de plus d'informations sur l'épidémie, des modèles plus sophistiqués seront nécessaires. Comme l'épidémie se déplace vers la communauté hétérosexuelle, de nouveaux modèles ont été batis qui tiennent compte de ce nouvel état des faits. Pour finir, des estimations mêmes grossières du temps de doublement sont en eux-mêmes du plus haut intérêt.

### Bibliographie sommaire

- [1] Anderson R.M., Medley G.F, May R.M., Johnson A.M.– *A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causitive agent of AIDS*. IMA J. Maths. Appl. in Medecine and Biol. 3, 229-263 (1986).
- [2] Bernoulli D.– *Essai d'une nouvelle analyse causée par la petite vérole et des avantages de l'inoculation pour la prévenir*. Histoire de l'Acad. Roy. Sci. (Paris) avec Mém. des Math. et Phys., Mém., 1-45 (1760).
- [3] Cassuto J.P., Pesce A., Quaranta J.F. : *Le SIDA*, coll. "Que sais-je?" n° 2332, P.U.F.
- [4] Cassuto J.P., Pesce A., Quaranta J.F. : *SIDA et Infection à VIH*, Masson, Paris (1992).
- [5] Hirschel B. : *Le SIDA. Guide du praticien, diagnostic, traitement, prise en charge..*– Ed. Médecine et Hygiène, Genève (1993).
- [6] Isham V. : *Mathematical modelling of the transmission dynamics of HIV infection and AIDS : a review.*– J. Roy. Statist. Soc., A 151, 5-30 (1988).