

M2 - Calcul Scientifique

Présentation du projet : Epidémiologie

Jean-Paul CHEHAB

LAMFA UMR CNRS 7352, Univ. Picardie Jules Verne

Plan

- 1 Premiers modèles EDO en épidémiologie avec R
 - Equation de KPP-Fisher
 - Lokta-Volterra

Plan

1 Premiers modèles EDO en épidémiologie avec R

- Equation de KPP-Fisher
- Lokta-Volterra

2 Modèles de type SIR

- Obtention du modèle
- Le modèle de base SIR pour les épidémies
- Estimation de paramètres et Inférence

Plan

1 Premiers modèles EDO en épidémiologie avec R

- Equation de KPP-Fisher
- Lokta-Volterra

2 Modèles de type SIR

- Obtention du modèle
- Le modèle de base SIR pour les épidémies
- Estimation de paramètres et Inférence

3 Applications

- Estimation de paramètres
- Prise en compte des décès (cas du COVID)

Plan

- 1 Premiers modèles EDO en épidémiologie avec R
 - Equation de KPP-Fisher
 - Lokta-Volterra
- 2 Modèles de type SIR
 - Obtention du modèle
 - Le modèle de base SIR pour les épidémies
 - Estimation de paramètres et Inférence
- 3 Applications
 - Estimation de paramètres
 - Prise en compte des décès (cas du COVID)
- 4 Extensions de SIR
 - Prise en compte du déspitage (cas du COVID)

Modèle à une espèce - Equation de KPP-Fisher

L'équation de Fisher de dynamique de populations

$$\begin{cases} \frac{d}{dt}x(t) = \theta x(1 - x/n), \\ x(0) = x_0. \end{cases}$$

Ici $x(t)$ est le nombre d'individus au temps t et n est une constante donnée

Modèle à deux espèces - Equations de Lokta-Volterra

On considère à présent un modèle à deux populations (équations de Lokta-Volterra), de type "proies-prédateurs" :

- $x(t)$ est le nombre d'individus proies au temps t
- $y(t)$ est le nombre d'individus prédateurs au temps t

Sans interaction entre proies et prédateurs

Le modèle est :

$$\begin{cases} \frac{d}{dt}x(t) = \alpha x, \\ \frac{d}{dt}y(t) = -\gamma y \\ x(0) = x_0, y(0) = y_0. \end{cases}$$

En l'absence de prédateurs, la population des proies croît exponentiellement tandis que celle des prédateurs, en l'absence des proies, elle, décroît exponentiellement.

Modèle à deux espèces - Equations de Lokta-Volterra

On considère à présent un modèle à deux populations (équations de Lokta-Volterra), de type "proies-prédateurs" :

- $x(t)$ est le nombre d'individus proies au temps t
- $y(t)$ est le nombre d'individus prédateurs au temps t

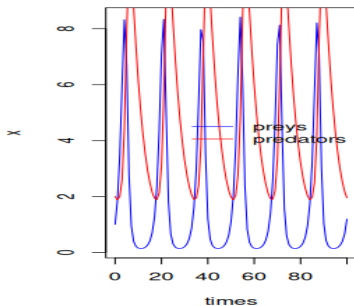
Le couplage

Il repose sur la correction suivante : au taux de croissance (positif) de la population de proies x diminue proportionnellement au nombre de prédateurs ; le taux de croissance (négatif) de la population de proies augmente proportionnellement au nombre de proies. On obtient :

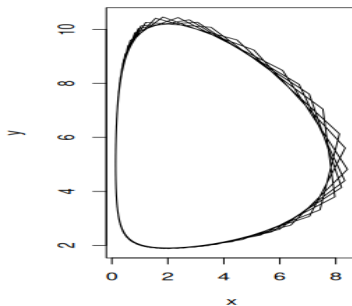
$$\begin{cases} \frac{d}{dt}x(t) = x(\alpha - \beta y), \\ \frac{d}{dt}y(t) = y(-\gamma + \delta x), \\ x(0) = x_0, y(0) = y_0. \end{cases}$$


```
install.packages("deSolve")
library(deSolve) # using the "ode" function
pars <- c(alpha = 1, beta = 0.2, delta = 0.5, gamma = 0.2)
# données initiales
init <- c(x = 1, y = 2)
# temps
times <- seq(0, 100, by = 1)
deriv <- function(t, state, pars) {
  with(as.list(c(state, pars)), {
    dx <- alpha * x - beta * x * y
    dy <- delta * beta * x * y - gamma * y
    return(list(c(x = dx, y = dy))) }) }
lv_results <- ode(init, times, deriv, pars)
# Graphique
x<-lv_results[,2]
y<-lv_results[,3]
# Partition en 1 lignes et 2 colonnes
par(mfrow = c(1,2))
plot(times,x,type="l",col="blue",main="Solutions en projection")
lines(times,y,col="red")
legende :
legend("right", c("preys", "predators"),
col = c("blue", "red"), lty = 1, bty = "n")
#portrait de phase
plot(x,y,type="l",col="black",main="Portrait de phase")
```

Solutions en projection



Portrait de phase



Compartiments

Definition (Compartiments)

On suit FBauer et Murray. Les compartiments sont les différents états en lesquels la population est divisée durant l'épidémie (sains, infectés, guéris ...)

7 compartiments sont couramment utilisés : S , I , E , D , R , M et C :

- S compartiment des individus susceptibles (non encore infectés par la maladie)
- I compartiment des individus infectés
- E compartiment des individus exposés, compartiment intermédiaire qui permet de considérer des individus qui pourront devenir infectés après un temps de latence
- D compartiment des individus décédés
- R compartiment des individus guéris
- M compartiment des individus immunisés
- C compartiment des individus porteur de la maladie mais sans symptômes

Remarque : la classe I peut être à son tour subdivisée en sous-classes suivant la *charge virale* des individus, c'est à dire suivant la concentration en virus.

Règles

Definition (Règles)

Ce sont les lois qui spécifient la proportion des individus passant d'un compartiment à un autre.

Exemple : dans le cas de deux compartiments S et I , on doit spécifier la proportion d'individus susceptibles devenant infectés $p(S \rightarrow I)$ et une proportion d'individus infectés devenant guéris (à nouveau susceptibles ici) $p(I \rightarrow R)$.

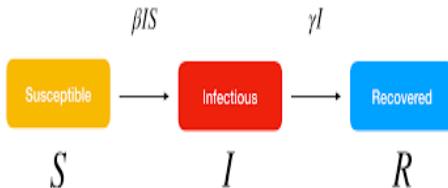


Figure – Exemple : règles du S-I-R

Ces règles sont incorporées dans les modèles mathématiques, tels ceux sous formes d'équations différentielles.

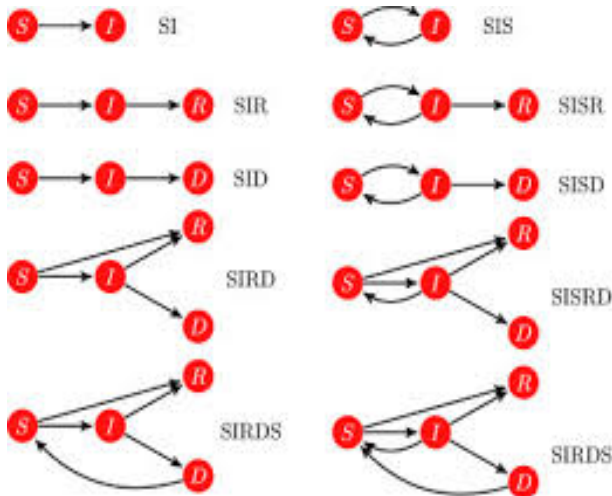


Figure – Les règles pour es différents modèles

Le modèle de base

On partage la population en 3 compartiments d'individus dont on cherche à déterminer le nombre en fonction du temps

- Le nombre d'individus susceptibles $S(t)$ au temps t (susceptible)
- Le nombre d'individus infectés $I(t)$ au temps t (infected)
- Le nombre d'individus guéris $R(t)$ au temps t (recovered)

Le modèle S-I-R correspond au système différentiel

$$\begin{cases} \frac{dS}{dt} = -\beta I(t) S(t), \\ \frac{dI}{dt} = \beta I(t) S(t) - \gamma I(t), \\ \frac{dR}{dt} = \gamma I(t). \end{cases} \quad (1)$$

Le paramètre β est le taux de contagion et γ est le taux de guérison.

La première équation correspond à la diminution de la population susceptible,

La seconde équation correspond à l'augmentation de la population infectée,

La troisième équation correspond à l'augmentation de la population guérie

Les modèles proposés sous forme d'EDO sont paramétriques et un enjeu crucial pour produire des outils de prédiction numériques est de pouvoir les estimer à l'aide des données. Il s'agit d'un problème inverse : connaissant les solutions, retrouver les paramètres du système de sorte à ce que le modèle numérique "colle au mieux" aux données. A titre d'exemple la connaissance du facteur R_0 de la Covid 19 a été un enjeu pour appréhender la tendance de l'épidémie.

Definition

L'inférence statistique est l'ensemble des techniques permettant d'induire les caractéristiques d'un groupe général (la population) à partir de celles d'un groupe particulier (l'échantillon), en fournissant une mesure de la certitude de la prédiction : la probabilité d'erreur.

Comme première démarche, nous allons estimer les paramètres du système d'EDO, à partir de simulations bruitées.

Exercice 1 : Equation de KPP-Fisher

- ❶ Ecrire un code en R de simulation de l'équation de KPP- Fisher. On pourra s'inspirer de celui sur Lokta-Volterra
- ❷ Faire tourner le code avec les spécificités

```
# definir les parametres du modele
model.par <- c(theta = c(0.1))
# definir la donnee initiale
state <- c(X = 0.1)
times <- seq(0,100,length.out=101)
mod <- ode(y=state,times=times,func=ode.model,parms = model.par)
nobs <- length(times)
scale <- 0.5
noise <- scale * rnorm(n = nobs, mean = 0 , sd = 1)
observ <- mod[,2] + noise
observ contiendra les données bruitées.
```
- ❸ Représenter la solution en fonction du temps

Estimation optimale de paramètres

On suppose connu le relevé de valeurs expérimentales. On cherche à déterminer les paramètres de l'équation de sorte à minimiser l'écart quadratique entre les valeurs expérimentales et celles générées par le modèle mathématique. On se donne

- un vecteur initial `init`
- une liste de temps discrets `times`
- une fonction ODE `ODESYS`
- une liste de noms de paramètres, par exemple `names(parameters) <- c("beta", "gamma")`

```
init <- c(S = N-Infected[1], I = Infected[1], R = 0)
RSS <- function(parameters) {
  names(parameters) <- c("beta", "gamma")
  out <- ode(y = init, times = times, func = ODESYS, parms = parameters)
  # fit contiendra la sortie
  fit <- out[, 3]
  # fit contiendra la sortie
  sum((ValeurExperimentale - fit)^2)
}
```

L'appel de cette fonction

```
Opt <- optim(c(0.5, 0.5), RSS, method = "L-BFGS-B", lower = c(0, 0),
upper = c(1, 1))
Optpar <- setNames(Opt$par, c("beta", "gamma"))
Optpar # les paramètres optimaux
```

Exercice 2 : Inférence de KPP-Fisher

Reprendre l'exercice 1

- 1 Calculer le paramètre θ à partir de la solution bruitée et le comparer à celle initiale.

- 2 Tracé de la courbe avec les écart quadratiques. On suivra le modèle en R :

```
# prévalences observées
with(flu, plot(times, cases, pch = 19, col = "red", ylim = c(0,
600)))
# modèles te prévalences prédites
with(predictions, lines(time, I, col = "red", type = "o"))
# les erreurs
segments(flu$times, flu$cases, predictions$time, predictions$x)
```

Exercice 3 : Système Lokta Volterra

1 Mêmes questions qu'aux exercices 1 et 2. Conclusions ?

Modèle SIR : taux de reproduction de base R_0

Une des données fondamentales à estimer est le fameux facteur R_0 , c'est à dire le taux de reproduction de base : le nombre moyen d'individus qu'une personne contagieuse pourra infecter ; il est de 3 pour le Covid-19.

Le taux R_0 (l'indice 0 se rapporte au temps initial) est défini dans une population totalement saine à l'exception d'un individu. Une condition pour que le nombre d'infectés décroît est $\frac{dI}{dt}|_{t=0} = I_0(rS_0 - a) < 0$, soit $S_0 < a/r$ avec r le taux de d'infection, $1/a$ est le temps moyen d'infection. On définit alors

$$R_0 = \frac{rS_0}{a}$$

- si $R_0 < 1$, alors l'individu infecté contamine en moyenne moins d'une personne et la maladie disparaîtra
- si $R_0 > 1$, la maladie peut se propager et se transformer en épidémie

La connaissance de R_0 permet donc de caler les modèles et de prévoir la propagation de la maladie.

Un enjeu important est d'arriver à un temps t^* tel que $R^* = \frac{rS^*}{a} < 1$ pour que la maladie puisse disparaître.

Exercice 4 : Système SIR

Mêmes questions.

- 1 Il s'agit ici surtout d'estimer le facteur R_0 et d'apprécier la précision de l'estimation.

Le covid-19 provoque des décès qu'il faut prendre en compte. On introduit alors une nouvelles inconnue $D(t)$ reliée au nombre d'infecté par $D'(t) = dI(t)$. On prendra $d = 0.0113$. Le nouveau système est

$$\left\{ \begin{array}{l} \frac{dS}{dt} = -b I(t) S(t), \\ \frac{dI}{dt} = b I(t) S(t) - g I(t) - d I(t), \\ \frac{dR}{dt} = g I(t), \\ \frac{dD}{dt} = d I(t). \end{array} \right. \quad (2)$$

Les effets du confinements sont très relatifs : à 90%, la moitié de la population française sera infecté en 8 mois.

Il faut incorporer au modèle la détection des sujets contaminés et leur mise en quarantaine : on introduit la nouvelle variable $Q(t)$ dont la variation temporelle est proportionnelle au nombre d'infectés : $Q'(t) = f I(t)$ avec $0 < f < 1$. Le nouveau système est

$$\left\{ \begin{array}{l} \frac{dS}{dt} = -b I(t) S(t), \\ \frac{dI}{dt} = b I(t) S(t) - g I(t) - f I(t) - d I(t), \\ \frac{dR}{dt} = g I(t), \\ \frac{dD}{dt} = d I(t), \\ \frac{dQ}{dt} = f I(t). \end{array} \right. \quad (3)$$

Exercice 5

Exercice de recherche libre. Proposez d'autres méthodes de traitement des données pour l'estimation de paramètres