Devoir Surveillé 06

Le samedi 2 mars 2024
durée 3h00

Documents écrits, électroniques, calculatrices et téléphones portables interdits
La plus grande attention sera apportée à la qualité de la rédaction, syntaxe et orthographe comprise.
Numérotez les copies et questions et soulignez ou encadrez les résultats.
Vous pouvez utiliser les résultats des questions en amont d'une question donnée en la référençant clairement
Les calculs, même longs et fastidieux doivent figurer sur la copie

Si au cours de l'épreuve, vous repérez ce qui vous semble être une erreur d'énoncé, vous le signalerez sur votre copie et poursuivrez la composition.

Problème

Le problème se compose de 4 parties largement indépendantes hormis la question B.5 de la partie B et la question C.4.d de la partie C. La partie C est consacrée à l'informatique. Une annexe en dernière page du sujet rappelle quelques fonctions Python.

Les candidates et candidats pourront admettre le résultat d'une question pour répondre à une question postérieure à condition de le mentionner explicitement.

⋆⋆⋆

La variole est une maladie virale sévère, éradiquée grâce à la vaccination en 1979. Les premières formes de vaccination (par inoculation) présentant un risque important, de nombreux débats eurent lieu pour décider du bien-fondé de la pratique. Daniel BERNOULLI apporta en 1760 une résolution mathématique à ce problème, connue pour être un des premiers modèles biomathématiques.

L'idée de BERNOULLI était de comparer l'espérance de vie de son époque (26 ans et 7 mois) avec une espérance de vie estimée pour une population fictive systématiquement inoculée contre la variole.

Partie A
Calcul d'une espérance de vie

Dans cette partie on s'intéresse à une cohorte fictive non-soumise à la variole. Le temps sera noté t et décompté en années. On propose une première modélisation de l'évolution du nombre $z(t)$ d'individus de cette cohorte par l'équation :

$$\begin{cases} \dfrac{dz}{dt} = -\alpha.z \\ z(0) = z_0 = 1300 \end{cases} \tag{E}$$

où le taux de mortalité α est constant et la fonction z est de classe $\mathscr{C}^1$ sur $\mathbb{R}^+$.

A.1. Donner la solution de l'équation différentielle (E) sur $\mathbb{R}^+$.

A.2.a. On définit la fonction F sur $\mathbb{R}^+$ par $F(t) = 1 - e^{-\alpha t}$. Donner l'expression de $F(t)$ en fonction de $\frac{z(t)}{z_0}$. Que représente la quantité $F(t)$ par rapport à la cohorte ?

A.2.b. Interpréter F en terme de lois de probabilités usuelles. En déduire l'espérance de vie $\mathscr{E}$ de la cohorte fictive après avoir justifié que cette espérance existe. On pourra voir l'espérance de vie comme l'espérance de la loi associée à F.

A.3.a. Soit $\Delta_t \in \mathbb{R}_+^*$ un intervalle de temps. Que représente la quantité $z(t) - z(t+\Delta_t)$?
A.3.b. Pour tout $i \in \mathbb{N}$, on pose $t_i = i\Delta_t$. Donner une interprétation possible de la quantité $\mathscr{E}_{\Delta_t}$ suivante. On supposera que la somme est bien définie.

$$\mathscr{E}_{\Delta_t} = \frac{\sum_{i=0}^{+\infty} t_i.\left(z(t_i) - z(t_{i+1})\right)}{z_0}.$$

A.3.c. On admettra que quand Δ_t tend vers 0 la quantité $\mathscr{E}_{\Delta_t}$ tend vers le rapport d'intégrales suivant qu'on supposera bien défini :

$$\lim_{\Delta_t \to 0} \mathscr{E}_{\Delta_t} = \frac{\int_0^{+\infty} t \frac{dz}{dt}\, dt}{\int_0^{+\infty} \frac{dz}{dt}\, dt}$$

Calculer ce rapport et faire le lien avec la question A.2. Quelle est l'unité de α ? En déduire une interprétation de $1/\alpha$. Proposer une démarche pour calculer une approximation de l'espérance de vie d'une cohorte dont on ne connaîtrait que le nombre annuel de décès.

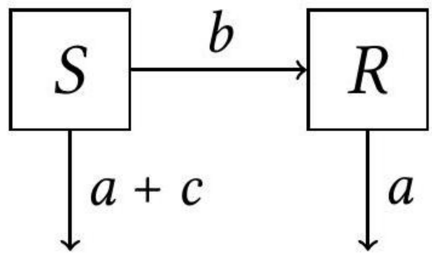
Partie B
Modèle de BERNOULLI

Dans cette partie on s'intéressera à la cohorte réellement observée par BERNOULLI. Il s'agissait d'une cohorte de 1300 bébés nés la même année et suivis durant 84 ans. Cette cohorte est considérée comme négligeable au sein d'une population stable dont on supposera que les taux d'infection par variole et les taux de décès ne varient pas. On établira dans cette partie la relation entre cette cohorte et une cohorte fictive qui ne subirait pas la variole. Les fonctions suivantes du temps $t \in \mathbb{R}^+$(toujours décompté en années) seront utilisées :

— S, R : nombres d'individus respectivement susceptibles (c'est-à-dire n'ayant jamais été atteints par la variole) et remis (donc immunisés) dans la cohorte observée
— x : nombre de survivants dans la cohorte observée
— z : nombre d'individus dans une cohorte fictive, initialement identique à la cohorte suivie, mais qui ne serait pas touchée par la variole

Ces fonctions seront supposées de classe $\mathscr{C}^1$ sur $\mathbb{R}^+$. BERNOULLI pose le modèle suivant pour son problème :

$$\begin{cases} \dfrac{dS}{dt} = -(a+b+c).S \\ \dfrac{dR}{dt} = b.S - a.R \\ \quad x = R + S \end{cases}.$$



où a est le taux de mortalité de cause indépendante de la variole par an, b est le taux d'individus susceptibles contractant la variole et guérissant par an et c est le taux d'individus susceptibles morts de la variole par an. Les taux b et c sont des constantes tandis que le taux a dépend du temps. On prendra $b = 7/64$ et $c = 1/64$.
B.1.a. Expliquer pourquoi BERNOULLI n'a pas fait l'approximation de a constant dans son modèle.
B.1.b. Interpréter $-\frac{dx}{dt}$ et $c.S$. Expliquer alors pourquoi $a = -\frac{1}{x}.\left(\frac{dx}{dt} + c.S\right)$.
B.1.c. Justifier que :

$$\frac{dz}{dt} = \frac{1}{x}.\left(\frac{dx}{dt} + c.S\right).z. \tag{F}$$

B.2.a. Soit q la fonction définie sur $\mathbb{R}^+$par $q = \frac{x}{S}$. Écrire l'équation différentielle vérifiée par q.
B.2.b. Résoudre cette équation différentielle et en déduire que :

$$q(t) = \frac{1}{8} + \frac{7}{8}.e^{t/8}$$

Ce résultat pourra être admis par la suite.
B.3.a. Soit H la fonction définie sur $\mathbb{R}^+$par $H = \ln \frac{z}{x}$, exprimer $\frac{dH}{dt}$ en fonction de q. En déduire que :

$$\frac{dH}{dt} = \frac{1}{8}.\frac{1}{1+7e^{t/8}}$$

Ce résultat pourra être admis par la suite.
B.3.b. Soit G la fonction définie sur $\mathbb{R}^+$ par $G(t) = \frac{e^{t/8}}{1+7e^{t/8}}$. On admettra que :

$$\frac{d\ln G}{dt} = \frac{dH}{dt} = \frac{d}{dt}\left(\ln \frac{z}{x}\right)$$

En déduire que :

$$z(t) = \frac{8e^{t/8}}{1+7e^{t/8}}.x(t)$$

B.4. Soit $y(t)$ le nombre d'individus dans une seconde cohorte fictive, initialement identique à la cohorte suivie, mais qui aurait été systématiquement inoculée juste avant l'instant initial. On donne $1/200$ la probabilité qu'un individu décède (immédiatement) des suites de l'inoculation. Justifier que $y = \gamma.z$ avec γ une constante à préciser.
B.5. Ainsi l'évolution du nombre d'individus $y(t)$ d'une cohorte fictive systématiquement inoculée est donnée par :

$$y(t) = \gamma.\frac{8e^{t/8}}{1+7e^{t/8}}.x(t)$$

Proposer une démarche pour calculer l'espérance de vie d'une population fictive systématiquement inoculée. On pourra utiliser la question A.3 de la première partie.

Partie C
Estimation de la fonction a

Seul le résultat final de la question C.1 est utile pour répondre aux questions suivantes : si besoin, on pourra donc admettre dès le début de la question C.2 que la liste `Lnorm` est créée. Une annexe en dernière page du sujet rappelle quelques fonctions Python.
C.1. On dispose d'un fichier texte `data.txt`, composé de deux lignes, où figurent des données sous la forme suivante :
— la première ligne du fichier contient le nombre de morts par d'autres maladies, pour chaque année ;
— la seconde ligne contient le nombre de survivants au début de chaque année ;
— sur chaque ligne, les données sont séparées par un espace (caractère " ").

Par exemple, si l'on considère une version simplifiée avec seulement sept années, le fichier `data.txt` est constitué du texte suivant :
```
283 133 47 30 21 16 13
1000 855 798 760 732 710 692
```
L'objectif de cette question est d'extraire ces informations pour créer la liste `Lnorm` qui contient, pour chaque année, le nombre de morts par d'autres maladies divisé par le nombre de survivants :
```
Lnorm = [283 / 1000,133 / 855,47 / 798, ...]
```
C.1.a. Écrire une fonction d'entête `extraction_ch()` qui renvoie une liste composée de deux chaînes de caractères : une correspondant à la première ligne de `data.txt` et l'autre à la seconde. Cette fonction devra,

au moins, ouvrir le fichier `data.txt`, le lire puis le fermer. Pour l'ouverture, la lecture et la fermeture d'un fichier, on pourra utiliser les syntaxes décrites dans l'annexe (en dernière page du sujet).

Par exemple, sur le contenu simplifié ci-dessus, `extraction_ch()` renvoie :

```
['283 133 47 30 21 16 13\n', '1000 855 798 760 732 710 692\n']
```

C.1.b. Compléter la fonction suivante d'entête `ch_vers_list(ch)` prenant en entrée une chaîne de caractère (de même forme que celles renvoyées par `extraction_ch`) et qui renvoie la liste de nombres correspondant; ces nombres seront de type flottant. Si besoin, on rappelle que `float("6.5\n")` renvoie le flottant `6.5`.

Par exemple, `ch_vers_list ('1000 855 798 760 732 710 692')` renvoie :

```
[1000.0,855.0,798.0,760.0,732.0,710.0,692.0]
```

Les éléments complétés sont à justifier, en particulier le rôle de la variable `sh`.

```
def ch_vers_list(ch):
    L = _________________
    n = len(ch)
    sh = ""
    for k in range(0, n):
        if ch[k] != " ":
            sh = sh + ch[k]
        else:
            _______________
            _______________
    _______________
    return L
```

C.1.c. Écrire une fonction d'entête `division(L1, L2)` prenant en entrée deux listes de nombres flottants. Cette fonction renvoie le booléen `False` si les deux listes ne sont pas de la même longueur; sinon, elle renvoie la liste contenant les divisions terme à terme des éléments de `L1` par ceux de `L2` (on suppose qu'il n'y a pas de terme nul dans `L2`).

Par exemple, `division([2,10.5,6,4],[2,2,3,1])` renvoie : `[1.0,5.25,2.0,4.0]`

C.1.d. Proposer une suite d'instructions qui crée la liste `Lnorm` précédemment définie.

C.2. On se propose d'estimer la fonction a représentant le taux de morts hors variole au cours du temps. On note $(\ell_t)_{t\in\{0,\dots,83\}}$ les valeurs relevées sur la cohorte du nombre de morts hors variole pour les 84 années de l'étude, stockées dans la liste `Lnorm`. Le graphique des valeurs numériques de ℓ_t pour les 24 premières années est donné en Fig. 1. On admettra que le comportement en temps long reste stable. On définit la fonction $M : \mathbb{R}^3 \to \mathbb{R}$ par :

$$\forall(\lambda,\mu,\gamma)\in\mathbb{R}^3,\quad M(\lambda,\mu,\gamma)=\sum_{t=0}^{83}\left(\ell_t-\lambda\mathrm{e}^{-\mu t}-\gamma\right)^2$$

C.2.a. Un premier choix, non retenu pour la suite, aurait été de considérer, à la place de M, la fonction $\widetilde{M}$ suivante :

$$\forall(\alpha,\beta)\in\mathbb{R}^2,\quad \widetilde{M}(\alpha,\beta)=\sum_{t=0}^{83}\left(\ell_t-\alpha t-\beta\right)^2.$$

C.2.a.i. . Expliquer ce que l'on aurait cherché à faire en minimisant $\widetilde{M}$ par rapport à α et β.

C.2.a.ii. . Expliquer ce que l'on cherche à faire en minimisant M par rapport à λ, μ et γ; on pourra faire un dessin. Pourquoi choisir $t\mapsto\lambda\mathrm{e}^{-\mu t}+\gamma$ plutôt qu'une fonction polynomiale ?

C.2.b. À l'aide des données représentées sur la Fig. 1, indiquer avec justification quelle valeur numérique approximative peut être choisie pour γ. Faire de même pour λ. On note ces choix γ_0 et λ_0, stockés dans les

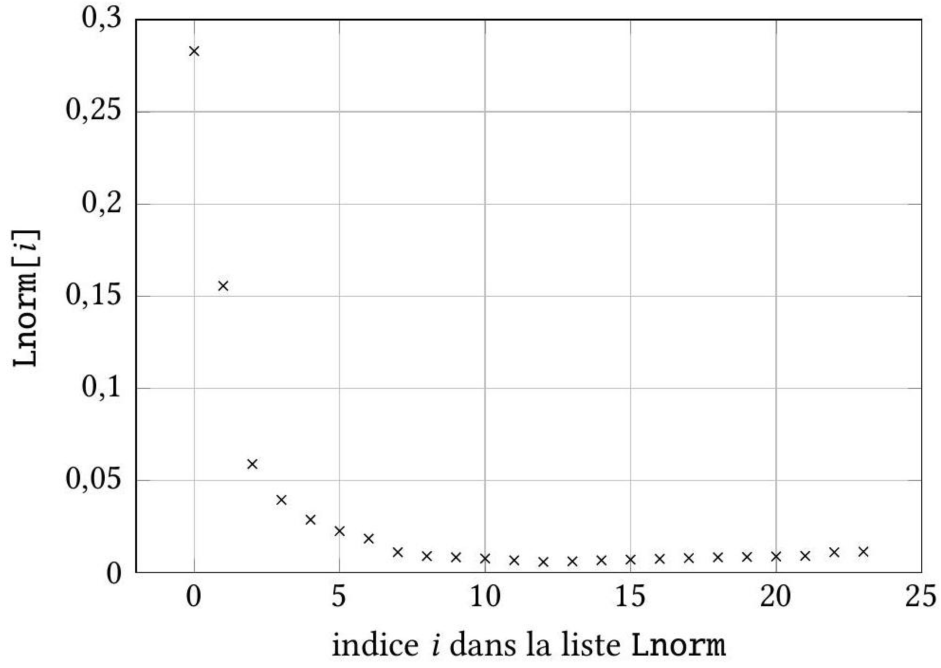


FIGURE 1 – Tracé des valeurs de ℓ_t.

variables globales gamma0 et lambda0. Dans tout le reste de cette question C.2, on considère que $\gamma = \gamma_0$ et $\lambda = \lambda_0$: seule μ reste à déterminer.

C.2.c. À quoi correspond la fonction suivante ? (La fonction Python exp est la fonction exponentielle.)

```
def M_mu(mu) :
    lamb = lambda0
    gam = gamma0
    s = 0
    for k in range(0, len(Lnorm)):
        s = s+(Lnorm[k]-lamb*exp(-mu*k)-gam)**2
    return s
```

C.2.d. Le tracé de la fonction M_mu est donné en Fig. 2. Expliquer si ce tracé est utile pour déterminer la valeur de μ cherchée. Si oui, proposer une valeur, notée μ_0 ; si non, que faudrait-il tracer ?

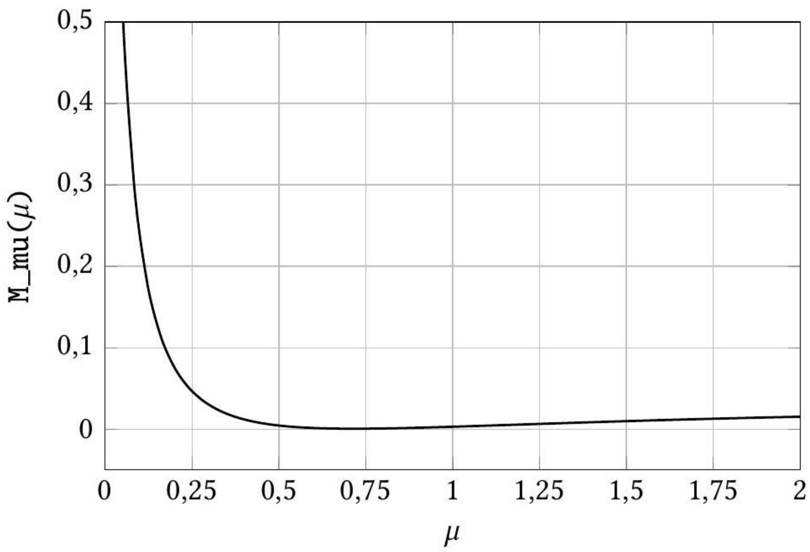


FIGURE 2 – Tracé de la fonction M_mu.

C.3. Dans cette question, on dispose d'une fonction $f : x \mapsto f(x)$ définie sur un intervalle $[A,B]$ de $\mathbb{R}$ qui vérifie l'hypothèse suivante : il existe $x^* \in [A,B]$ tel que f est strictement décroissante sur $[A,x^*]$ et strictement croissante sur $[x^*,B]$; ainsi, f admet un minimum en x^*. L'objectif de la question est de déterminer x^* de façon approchée.

L'algorithme utilisé consiste à partir de $g_0 = A$ et $d_0 = B$ puis à construire une suite d'intervalles $[g_k,d_k]$ qui sont de plus en plus petits et qui contiennent x^* (ainsi, lorsqu'on aura atteint un intervalle «suffisamment petit», on aura localisé x^* avec une «précision suffisante»). Pour passer de $[g_k;d_k]$ à $[g_{k+1};d_{k+1}]$, on utilise le raisonnement suivant :

— on choisit deux nombres x^g et x^d qui vérifient $g_k < x^g < x^d < d_k$;
— si $f(x^g) < f\left(x^d\right)$ alors x^* se situe à gauche de x^d donc on pose $g_{k+1} = g_k$ et $d_{k+1} = x^d$;
— sinon si $f(x^g) > f\left(x^d\right)$ alors x^* se situe à droite de x^g donc on pose $g_{k+1} = x^g$ et $d_{k+1} = d_k$;
— sinon, c'est que $f(x^g) = f\left(x^d\right)$ donc x^* se situe dans $\left[x^g;x^d\right]$ et on pose $g_{k+1} = x^g$ et $d_{k+1} = x^d$.

On calcule cette suite d'intervalles et on s'arrête dès que la taille de l'intervalle $[g_k;d_k]$ est inférieure ou égale à ε (où $\varepsilon > 0$ est une valeur fixée) ; on renvoie alors, parmi les trois points g_k, d_k et m_k (où m_k est le milieu entre g_k et d_k), celui en lequel f est la plus petite.

C.3.a. On dispose de la fonction suivante, où f est une fonction et x, y deux réels.

```
def argmin2(f, x, y):
    if f(x) <= f(y):
        return x
    else:
        return y
```

Écrire une fonction `argmin3(f, x, y, z)` qui renvoie parmi les trois valeurs `x`, `y` et `z`, celle en laquelle `f` est minimale.

C.3.b. Écrire une fonction d'entête `minimum(f, A, B, eps)` qui programme l'algorithme déterminant x^*. Elle suivra la structure suivante (en utilisant autant de lignes que nécessaire dans le corps de boucle) ; par ailleurs, le choix de x^g et de x^d sera tel qu'ils découpent $[g_k;d_k]$ en trois parties égales. On prendra soin d'expliquer les choix effectués pour compléter la fonction.

```
def minimum(f, A, B, eps):
    g = A
    d = B
    while __________________ :
        #calcul de x^g et x^d :
        xg = __________________
        xd = __________________
        #calcul de g et d :
        .
        .
    return __________________
```

C.4. On revient dans cette question à la problématique de la question C.2 et on considère toujours γ et λ fixés aux valeurs γ_0 et λ_0 précédemment trouvées. Cependant, μ_0 est à déterminer à l'aide de la fonction `minimum` écrite en question C.3, plutôt que par lecture graphique.

C.4.a. Écrire une fonction d'entête `test_croissant(L)` prenant en entrée une liste de nombres et qui renvoie un booléen indiquant si les valeurs de `L` sont triées par ordre croissant.

Cette fonction devra parcourir `L` une seule fois ; en particulier, elle ne devra pas réaliser un quelconque tri.

C.4.b. À l'aide de `test_croissant`, proposer un raisonnement (sans écrire de code) pour tester si la fonction `M_mu` est bien croissante sur l'intervalle $[0,8;3]$.

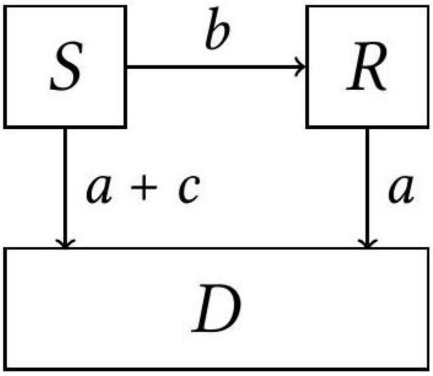
C.4.c. Proposer une instruction pour trouver la valeur μ_0 cherchée.

C.4.d. On trouve $\mu_0 \approx 0,7061$: commenter et conclure sur l'estimation de la fonction a. Que penser de l'hypothèse « a constant» du premier modèle ?

Partie D

Modèle individu-centré

Contrairement à l'étude de BERNOULLI on ne s'intéresse plus dans cette partie à toute la cohorte mais à un seul individu vivant dans la cohorte dont l'état (susceptible, remis ou mort) est modélisé par une variable aléatoire. Dans ce cas le modèle étudié peut être résumé par le schéma suivant dans lequel une classe d'état D a été rajoutée, correspondant à l'état mort :



Soit un individu donné dans la population. Le temps (en années) sera désormais noté n et ne prendra que des valeurs entières positives. On note X_n la variable aléatoire correspondant à l'état de l'individu au temps $n \in \mathbb{N}$: $X_n \in \{S, R, D\}$ où S, R et D désignent respectivement les états Susceptible, Remis et Mort. Les paramètres a, b et c représentent ici des probabilités conditionnelles et sont supposées constantes au cours du temps. Ainsi par exemple, quel que soit $n \in \mathbb{N}$, un individu Susceptible (S) au temps n aura une probabilité b d'être Remis (R) au temps $n+1$.

Pour tout instant $n \in \mathbb{N}$ on définit le vecteur Z_n des probabilités suivant :

$$Z_n = \begin{pmatrix} \mathbb{P}(X_n = S) \\ \mathbb{P}(X_n = R) \\ \mathbb{P}(X_n = D) \end{pmatrix}.$$

D.1. On suppose qu'au temps 0 l'individu est Susceptible. Donner alors les valeurs de Z_0 et Z_1.

D.2. On définit une matrice Q :

$$Q = \begin{pmatrix} 1-(a+b+c) & 0 & 0 \\ b & 1-a & 0 \\ a+c & a & 1 \end{pmatrix}.$$

D.2.a. Montrer que pour tout $n \in \mathbb{N}, Z_n = Q^n Z_0$ où Q^n désigne la puissance n de la matrice Q :

$$Q^n = \underbrace{Q.\cdots.Q}_{n \text{ fois}}.$$

D.2.b. Justifier qu'il existe deux constantes λ_1 et λ_2 telles que $0 < \lambda_1 < \lambda_2 < 1$ et telles que λ_1, λ_2 et 1 sont des valeurs propres de Q. Calculer des vecteurs propres associés à ces trois valeurs propres.

D.2.c. En déduire une matrice P telle que :

$$Q = P. \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 1 \end{pmatrix} .P^{-1}.$$

Calculer P^{-1}.

D.2.d. Montrer que pour tout $n \in \mathbb{N}$:

$$Z_n = P \begin{pmatrix} \lambda_1^n & 0 & 0 \\ 0 & \lambda_2^n & 0 \\ 0 & 0 & 1 \end{pmatrix} P^{-1} Z_0.$$

Dans la suite on admettra que :

$$Z_n = \begin{pmatrix} \lambda_1^n \\ \frac{-b\lambda_1^n + b\lambda_2^n}{b+c} \\ 1 - \frac{c\lambda_1^n + b\lambda_2^n}{b+c} \end{pmatrix}.$$

D.2.e. Calculer les limites $\lim_{n\to+\infty} \mathbb{P}(X_n = S)$, $\lim_{n\to+\infty} \mathbb{P}(X_n = R)$ et $\lim_{n\to+\infty} \mathbb{P}(X_n = D)$. Interpréter.

D.3.a. Soit A_n l'évènement "l'individu vit exactement n années". Montrer que :

$$\mathbb{P}(A_n) = \frac{c(a+b+c)}{b+c}\lambda_1^n + \frac{ab}{b+c}\lambda_2^n.$$

Dans la suite on admettra que le premier terme est négligeable devant le deuxième et on prendra :

$$\mathbb{P}(A_n) \simeq \frac{ab}{b+c}\lambda_2^n.$$

D.3.b. Soit $\rho \in]0,1[$ et soit $\Sigma = \sum_{n=0}^{+\infty} n\rho^n$. Justifier que Σ est bien définie et montrer que :

$$\Sigma = \frac{\rho}{(\rho-1)^2}.$$

D.3.c. Montrer que l'espérance de vie d'un individu à la naissance est égale à :

$$\mathscr{E} = \frac{b(1-a)}{a(b+c)}.$$

On pourra voir l'espérance de vie comme l'espérance de la variable aléatoire

$$T = \max\{n \in \mathbb{N}, X_n \neq D\}.$$

D.3.d. Application numérique : dans le cas de la variole, calculer l'espérance de vie. On prendra $a = 2/64$, $b = 7/64$ et $c = 1/64$. Comparer avec l'espérance de vie réelle de 26 ans et 7 mois. Le choix de prendre a constant vous paraît-il justifié ?

Annexe - Rappels de syntaxes Python pour la partie C

- `f = open("monfichier.txt", "r")` ouvre le document intitulé `monfichier.txt` (en mode lecture) ; le fichier est alors désigné par le nom de variable `f`.
- `f.read()` lit l'intégralité du contenu du fichier `f` et le renvoie sous forme d'une chaine de caractères.
- `f.readline()` lit la première ligne du fichier `f` et la renvoie sous forme d'une chaine de caractères. Ré-exécuter `f.readline()` lit alors la deuxième ligne ; et ainsi de suite.
- `f.close()` ferme le fichier `f`.
- Le caractère `"\n"` (de longueur 1) désigne le passage à la ligne.

PYTHON
AGRO-VETO
2020

Listes

`[]`	Créer une liste vide
`[a]*n`	Créer une liste avec n fois l'élément `a`
`L.append(a)`	Ajoute l'élément `a` à la fin de la liste `L`
`L1 + L2`	Concatène les deux listes `L1` et `L2`
`len(L)`	Renvoie le nombre d'éléments de la liste `L`
`L.pop(k)`	Renvoie l'élément d'indice k de `L` et l'enlève de `L`
`L.remove(a)`	Enlève une fois la valeur `a` de la liste `L`
`max(L)`	Renvoie le plus grand élément de la liste `L`
`min(L)`	Renvoie le plus petit élément de la liste `L`
`sum(L)`	Renvoie la somme de tous les éléments de la liste `L`

Numpy

`import numpy as np`

`np.array()`	Transforme une liste en matrice `numpy`
`np.linspace(a,b,n)`	Crée une matrice ligne de n valeurs uniformément réparties entre a et b (inclus)
`np.zeros([n,m])`	Crée la matrice nulle de taille $n \times m$
`np.eye(n)`	Crée la matrice identité de taille n
`np.diag(L)`	Crée la matrice diagonale dont les termes diagonaux sont les éléments de la liste `L`
`np.transpose(M)`	Renvoie la transposée de M
`np.dot(M,P)`	Renvoie le produit matriciel MP
`np.sum(M)`	Renvoie la somme de tous les éléments de M
`np.prod(M)`	Renvoie le produit de tous les éléments de M
`np.max(M)`	Renvoie le plus grand élément de M
`np.min(M)`	Renvoie le plus petit élément de M
`np.shape(M)`	Renvoie dans un couple le format de la matrice M
`np.size(M)`	Renvoie le nombre d'éléments de M

Numpy.linalg

`import numpy.linalg as la`

`la.inv(M)`	Renvoie l'inverse de la matrice M si elle est inversible
`la.eigvals(M)`	Renvoie la liste des valeurs propres de M
`la.eig(M)`	Renvoie un couple `L,P` où `L` est la liste des valeurs propres de M et `P` la matrice de passage associée
`la.matrix_rank(M)`	Renvoie le rang de M

Random

`import numpy.random as rd`

`rd.rand()`	Simule une réalisation d'une variable $X \hookrightarrow \mathcal{U}([0,1])$
`rd.randint(a,b)`	Simule une réalisation d'une variable $X \hookrightarrow \mathcal{U}([\![a,b[\![)$
`rd.gauss(0,1)`	Simule une réalisation d'une variable $X \hookrightarrow \mathcal{N}(0,1)$
`rd.choice(L)`	Choisit aléatoirement un élément de la liste L

Math

`import numpy as np`

`np.atan(x)`	Renvoie $\arctan(x)$
`np.floor(x)`	Renvoie $\lfloor x \rfloor$
`np.factorial(n)`	Renvoie $n!$ si $n \in \mathbb{N}$
`np.sqrt(x)`	Renvoie $\sqrt{x}$ si $x \geq 0$
`np.log(x)`	Renvoie $\ln(x)$ si $x > 0$
`np.exp(x)`	Renvoie e^x

Logique

`a == b`	Teste l'égalité « $a = b$ »
`a != b`	Teste « $a \neq b$ »
`a < b`	Teste « $a < b$ »
`a <= b`	Teste « $a \leq b$ »
`a > b`	Teste « $a > b$ »
`a >= b`	Teste « $a \geq b$ »
`not A`	Renvoie la négation de A
`A and B`	Renvoie « A et B »
`A or B`	Renvoie « A ou B »
`True`	Constante booléenne « Vrai »
`False`	Constante booléenne « Faux »

Matplotlib.pyplot

`import matplotlib.pyplot as plt`

`plt.plot(X,Y,'+-r')`	Génère la courbe des points définis par les listes `X` et `Y` (abscisses et ordonnées) avec les options : • symbole : `'.'` point, `'o'` rond, `'h'` hexagone, `'+'` plus, `'x'` croix, `'*'` étoile, ... • ligne : `'-'` trait plein, `'--'` pointillé, `'-.'` alterné, ... • couleur : `'b'` **bleu** , `'r'` **rouge** , `'g'` **vert** , `'c'` **cyan** , `'m'` **magenta** , `'k'` **noir** , ...
`plt.bar(X,Y)`	Génère l'histogramme des points définis par les listes `X` et `Y` (abscisses et ordonnées)
`plt.axis('equal')`	Rend le repère orthonormé
`plt.xlim(xmin,xmax)`	Fixe les bornes de l'axe des abscisses
`plt.ylim(ymin,ymax)`	Fixe les bornes de l'axe des ordonnées
`plt.show()`	Affiche le graphique

Cette liste est non exhaustive. Les candidats sont libres d'utiliser les commandes de leur choix.

Correction DS 06

Correction Ex.–1

Partie A
Calcul d'une espérance de vie

Dans cette partie on s'intéresse à une cohorte fictive non-soumise à la variole. Le temps sera noté t et décompté en années. On propose une première modélisation de l'évolution du nombre $z(t)$ d'individus de cette cohorte par l'équation :

$$\begin{cases} \dfrac{dz}{dt} = -\alpha.z \\ z(0) = z_0 = 1300 \end{cases} \tag{E}$$

où le taux de mortalité α est constant et la fonction z est de classe $\mathscr{C}^1$ sur $\mathbb{R}^+$.

A.1. Par le cours, une fonction z, de classe $\mathscr{C}^1$ sur $\mathbb{R}^+$ vérifie l'équation différentielle du système (E) sur $\mathbb{R}^+$ si et seulement si il existe une constante $C \in \mathbb{R}$ telle que

$$\forall t \in \mathbb{R}^+, z(t) = C.e^{-\alpha.t}.$$

En utilisant la condition initiale du système (E), il vient que $C = z(0) = z_0 = 1300$ et donc la solution de (E) est z donnée par

$$\forall t \in \mathbb{R}^+, z(t) = z_0.e^{-\alpha.t} = 1300.e^{-\alpha.t}.$$

A.2.a. On a $F(t) = 1 - \frac{z(t)}{z_0}$. Vue [1] de la sorte $F(t)$ réprésente la proportion d'individus décédés de la cohorte à l'année t.

A.2.b. La fonction F (prolongée à 0 sur $\mathbb{R}^-$) est la fonction de répartition d'une v.a. T suivant une loi exponentielle de paramètre α. L'espérance de vie est $\mathscr{E} = \mathbb{E}(T) = \frac{1}{\alpha}$. Elle existe par le cours, l'intégrale généralisée $\int_0^{+\infty} t.\alpha.e^{-\alpha.t}\, dt$ étant convergente.

En fait –et ce point est hors programme– on peut voir T comme étant la variable aléatoire modèle, donnant l'année de décès d'un individu de la cohorte. *i.e.* les individus sont numérotés (par $k \in \mathbb{N}^*$), l'année du décès de l'individu k est donné par la v.a. T_k, la famille $(T_k)_{1 \leq k \leq ??}$ formant un échantillon de T.

A.3.a. Soit $\Delta_t > 0$ un intervalle de temps. La quantité $z(t) - z(t + \Delta_t)$ est la variation de population entre les instants t et $t + \Delta_t$, c'est à dire le nombre de décès pendant l'intervalle [2] de temps $[t, t + \Delta_t]$.

A.3.b. Pour $i \in \mathbb{N}$, on pose $t_i = i\Delta_t$. On a$t_0 = 0$, $t_{i+1} = t_i + \Delta_t$ et (par téléscopie, vu que $z_{t_i} \to 0$ lorsque $i \to +\infty$),

$$z_0 = \sum_{i=0}^{+\infty} (z(t_i) - z(t_{i+1}))$$

et donc

$$\mathscr{E}_{\Delta_t} = \frac{1}{z_0} \sum_{i=0}^{+\infty} t_i.(z(t_i) - z(t_{i+1}))$$

1. J'estime cette question tendancieuse car elle nécéssite de disposer de l'écriture précédente pour F qui est peut-être naturelle mais n'est certainement pas la seule. De toute façon, on va voir que cet énoncé est de qualité très médiocre avec des questions vagues, une terminologie approximative, etc... mais bon, hein, c'est ce que vous devez subir...

2. Ca, c'est un intervalle, Δ_t est en fait une durée

est la moyenne (pondérée par la proportion de décès pendant l'intervalle de temps $[t_i, t_i + \Delta_t]$) des instants t_i.

Il s'agit donc de l'âge moyen de décès des individus de la cohorte, âge moyen mesuré avec une précision de Δ_t.

A.3.c. On admet que quand Δ_t tend vers 0 la quantité $\mathscr{E}_{\Delta_t}$ tend vers le rapport d'intégrales suivant :

$$\lim_{\Delta_t \to 0} \mathscr{E}_{\Delta_t} = \frac{\int_0^{+\infty} t \frac{dz}{dt} dt}{\int_0^{+\infty} \frac{dz}{dt} dt}$$

Vu que $\frac{dz}{dt} \leq 0$, ce rapport d'intégrales est en fait l'espérance d'une variable aléatoire (à valeurs ≥ 0) dont une densité est

$$f(t) = -C.\frac{dz}{dt}.\mathbb{1}_{\{t \geq 0\}}$$

où (pourvu que $z(t) \to 0$ lorsque $t \to +\infty$) :

$$\frac{1}{C} = -\int_0^{+\infty} \frac{dz}{dt}.dt = z_0$$

en l'espèce, il s'agit d'une v.a. T exponentielle de paramètre α et donc

$$\lim_{\Delta_t \to 0} \mathscr{E}_{\Delta_t} = \mathbb{E}(T) = \frac{1}{\alpha}$$

La quantité $\frac{1}{\alpha}$ est en années, α est en années^{-1}.

Pour calculer une approximation de l'espérance de vie d'une cohorte dont on ne connaîtrait que le nombre annuel de décès :

— Si on connait chaque N_i à l'année t_i, on calcule

$$\frac{\sum_i N_i.t_i}{\sum_i N_i}$$

— Si, sur une année, on connait l'âge de chaque décédé de l'année, on peut comptabiliser le nombre de morts N_a pour chaque âge a et calculer [3]

$$\frac{\sum_a N_a.a}{\sum_a N_a}.$$

Partie B
Modèle de BERNOULLI

B.1.a. Qui [4] le sait ?

Si a est constant, on peut complètement résoudre le modèle :

$$S(t) = S(0)e^{-(a+b+c)t}, R(t) = C.e^{-at} + D.e^{-(a+b+c)t}$$

où $C + D = R(0)$ et D est choisi pour que R vérifie son équation différentielle :

$$\frac{dR}{dt} - b.S + a.R = (-(a+b+c)D - bS(0))e^{-(a+b+c).t}$$

On prend donc $D = -\frac{b}{a+b+c}S(0)$.

3. C'est de cette manière que l'espérance de vie est calculée effectivement par l'INED
4. Expliquer pourquoi BERNOULLI n'a pas fait l'approximation de a constant dans son modèle.

On dispose donc de formules explicites permettant d'évaluer d'une part a et $a+b+c$

D'un autre côté, les temps de l'époque étaient durs ! Guerre, famines, épidémies, l'hypothèse d'un taux de mortalité constant serait déraisonnable alors que ce qui est cherché c'est le sur-effet du à la variole, ceci, indépendamment des conditions extérieures [5].

Enfin, voir l'explication donnée en C.2.a.ii concernant l'évolution par rapport à l'âge, du taux de mortalité infantile entre 0 et 10 ans.

B.1.b. La quantité $-\frac{dx}{dt}$ est la variation (diminution) de la population totale de la cohorte, $-\frac{1}{x}.\frac{dx}{dt}$ est le taux de mortalité. $c.S$ est la quantité de personnes décédant du fait de la variole, donc $\frac{dx}{dt}+c.S$ est la quantité de personnes décédant pour une auter raison que la variole et donc $=-\frac{1}{x}.\left(\frac{dx}{dt}+c.S\right)$ devrait être le taux de mortalité indépendante de la variole dans la population totale, c'est à dire a.

On vérifie ceci par le calcul :

$$\frac{dx}{dt}+c.S=\frac{dR}{dt}+\frac{dS}{dt}+cS=b.S-a.R-(a+b+c).S+c.S=-a.(R+S)=-a.x$$

B.1.c. De ce qu'on vient de dire, une population non touchée par la variole doit avoir un taux de mortalité de a, c'est à dire :

$$\frac{dz}{dt}=-a.z=\frac{1}{x}.\left(\frac{dx}{dt}+c.S\right).z. \tag{F}$$

B.2.a. Soit q la fonction définie sur $\mathbb{R}^+$par $q=\frac{x}{S}$.

On a $q=1+\frac{R}{S}$ et :

$$\frac{dq}{dt}=\frac{1}{S^2}.\left(S.\frac{dR}{dt}-R.\frac{dS}{dt}\right)=\frac{1}{S^2}.\left(b.S^2-a.S.R+(a+b+c)R.S\right)=b+(b+c)\frac{R}{S}=b+(b+c).(q-1)$$

et donc

$$\frac{dq}{dt}=-c+(b+c)q.$$

B.2.b. Il s'agit d'une équation différentielle linéaire, du premier ordre, à coefficients constants de second membre constant (comme l'équation de décharge d'un condensateur). La solution générale de l'équation diférentielle homogène associée est $t\mapsto\lambda.e^{(b+c)t}$, une solution particvulière en est la constante $t\mapsto\frac{c}{c+b}$ et donc, il existe $\lambda\in\mathbb{R}$ telle que

$$\forall t\geq 0,\, q(t)=\frac{c}{c+b}+\lambda.e^{(b+c)t}=\frac{1}{8}+\lambda.e^{\frac{1}{8}.t}$$

A l'instant 0, il n'y a pas encore eu d'infectés et donc tout le monde est susceptible, *i.e.* $q(0)=x(0)/S(0)=1$ et donc $\lambda=\frac{7}{8}$ et finalement :

$$\forall t\geq 0,\, q(t)=\frac{1}{8}+\frac{7}{8}.e^{t/8}$$

B.3.a. Soit H la fonction définie sur $\mathbb{R}^+$par $H=\ln\frac{z}{x}$, on a donc

$$\frac{dH}{dt}=\frac{1}{z}.\frac{dz}{dt}-\frac{1}{x}.\frac{dx}{dt}=c.S.\frac{1}{x}=c.\frac{1}{q}$$

En substituant à q l'expression $\frac{1}{8}+\frac{7}{8}.e^{t/8}$ trouvée à la question précédente et $c=\frac{1}{64}$, il vient :

$$\frac{dH}{dt}=\frac{1}{8}.\frac{1}{1+7e^{t/8}}$$

5. ce n'est pas totalement raisonnable, dans une population mal nourrie, il est probable que cette malnutrition amplifie la mortalité due à la variole.

B.3.b. Soit G la fonction définie sur $\mathbb{R}^+$ par $G(t) = \frac{e^{t/8}}{1+7e^{t/8}}$. Comme on admet que :

$$\frac{d\ln G}{dt} = \frac{dH}{dt} = \frac{d}{dt}\left(\ln\frac{z}{x}\right)$$

il vient qu'il existe une constante C telle que

$$\forall t \geq 0, \ln G(t) - C = \left(\ln\frac{z(t)}{x(t)}\right)$$

Comme à $t = 0$, $z(0) = x(0)$ (c'est comme ça qu'a été choisi z, « initialement identique » à x), on en déduit que $\ln\frac{1}{8} = C$ et donc, en passant à l'exponentielle, il vient

$$\forall t \geq 0, \frac{z(t)}{x(t)} = 8.G(t) = \frac{8e^{t/8}}{1+7e^{t/8}}$$

c'est à dire :

$$\forall t \geq 0, z(t) = \frac{8e^{t/8}}{1+7e^{t/8}}.x(t)$$

B.4. L'hypothèse signifie que cette cohorte à la même dynamique que celle décrite par z (y satisfait la même équation différentielle que z) mais avec un effectif initial amputé du fait des décès dus à l'inoculation. On a donc (après résolution de l'equation différentielle), l'existence d'une constante C telle que :

$$\forall t \geq 0, \ln G(t) - C = \left(\ln\frac{y(t)}{x(t)}\right)$$

Comme à $t = 0$, $y(0) = \gamma.x(0) = \gamma.z(0)$ (avec $\gamma = \frac{199}{200}$ c'est comme ça qu'on exprime que $\frac{1}{200}$ des personnes inoculées à décédé tout de suite), on en déduit que $\ln\frac{1}{8} = C + \ln\gamma$ et donc, en passant à l'exponentielle, il vient

$$\forall t \geq 0, \frac{y(t)}{x(t)} = 8.\gamma.G(t) = \gamma.\frac{8e^{t/8}}{1+7e^{t/8}}$$

c'est à dire :

$$\forall t \geq 0, y(t) = \gamma.\frac{8e^{t/8}}{1+7e^{t/8}}.x(t) = \gamma.z(t)$$

NB : Alternativement, on aurait pu définir $Q = \frac{y}{z}$ qui vérifie $Q' = 0$ (du fait que z et y satisfont la même EDO présentée en B.1.c) et $Q(0) = \gamma$ pour conclure que Q vaut constamment $\gamma = \frac{199}{200}$.

B.5. Ainsi l'évolution du nombre d'individus $y(t)$ d'une cohorte fictive systématiquement inoculée est donnée par :

$$y(t) = \gamma.\frac{8e^{t/8}}{1+7e^{t/8}}.x(t)$$

En utilisant la formule de la question A.3.c de la première partie, on pourrait poser l'espérance de vie d'une population fictive systématiquement inoculée :

$$\mathscr{E}_{in} = \frac{\int_0^{+\infty} t.\frac{dy}{dt}\,dt}{\int_0^{+\infty}\frac{dy}{dt}\,dt}$$

— Le dénominateur de cette expression vaut $-y(0)$ pourvu que $y(t) \to 0$ lorsque $t \to +\infty$,

— Le numérateur de cette expression vaut, par ipp :

$$\int_0^{+\infty} t.\frac{dy}{dt}\,dt = [t.y(t)]_0^{+\infty} - \int_0^{+\infty} y(t)\,dt = -\int_0^{+\infty} y(t)\,dt$$

pourvu que $t.y(t) \to 0$ lorsque $t \to +\infty$. On aurait donc :

$$\mathscr{E}_{in} = \int_0^{+\infty}\frac{y(t)}{y(0)}\,dt.$$

Partie C
Estimation de la fonction a

C.1.a.

```
def extraction_ch():
    f = open("data.txt", "r")
    ligne_1 = f.readline()
    ligne_2 = f.readline()
    return [ligne_1, ligne_2]
```

C.1.b.

```
def ch_vers_list(ch):
    L = [] #init liste L vide
    n = len(ch)
    sh = ""
    for k in range(0, n):
        if ch[k] != " ":
            sh = sh + ch[k] #Accumule dans sh des chiffres formant un nombre
        else:
            L.append(float(sh)) #Quand tt est accumulé,
            #on le transforme en flottant et ajout dans L
            sh = "" #On va recommencer l'accumulation dans sh remis à zero
    L.append(float(sh))
    return L
```

C.1.c.

```
def division(L1, L2):
    """
    retourne la liste des elts de L1 divisés terme à terme par ceux de L2
    """
    n = len(L1)
    if n != len(L2):
        return False
    return [L1[i]/L2[i]  for i in range(n)]
```

C.1.d.

```
L = extraction_ch()
mort = ch_vers_list(L[0])
survivant = ch_vers_list(L[1])
Lnom = division(mort, survivant)
```

C.2.
C.2.a.i. . Si on considère la fonction $\widetilde{M}$:

$$\forall(\alpha,\beta)\in\mathbb{R}^2,\quad \widetilde{M}(\alpha,\beta)=\sum_{t=0}^{83}(\ell_t-\alpha t-\beta)^2,$$

et que l'on cherche α et β minimisant $\tilde{M}$, on cherche à approximer la fonction $t\mapsto\ell_t$ par une fonction affine (approximation par la droite des moindres carrés)
C.2.a.ii. . Ici, en minimisant M par rapport à λ, μ et γ; on cherche à approximer – aus ens des moindres carrés– $t\mapsto\ell_t$ par une fonction du type $t\mapsto\lambda e^{-\mu t}+\gamma$.

On peut rappeler que ℓ_t est, à l'année t, le nombre de morts par d'autres maladies (taux de mortalité a) divisé par le nombre de survivants x, c'est donc, dans les notations de la partie B, la quantité $a.(R+S)/(R+S)$, c'est à dire a.

Vu le graphique obtenu en Fig. 1, la forme du graphe de $t \mapsto \ell_t$, il est légitime que l'on cherche à estimer a par une fonction de la forme $t \mapsto \lambda e^{-\mu t} + \gamma$ et pas par une fonction polynomiale qui n'hexhiberait pas se type de comportement de décroissance avec asymptote en $+\infty$.

On peut revenir sur la question de la constance ou pas de a au vu de ce graphioque, une piste, qui me semble s'éloigner du sujet peut être la suivante : $a(t)$ est le taux de mortalité du à d'autres maladies que la variole à l'instant t et on suit une cohorte qui comporte toujours les mêmes individus, t est donc lié à l'âge de ces individus si on les prend tous au départ du même âge (disons qu'on les prend à la naissance, $t = 0$ correspond à l'âge 0.

Que le taux de mortalité dépende de l'âge est chose naturelle, indépendamment des contrinates extérieures qu'on pouvait évoquer en B.1.a, surtout à cette époque où la mortalité infantile est énorme. Ce qui ressort de la figure 1, c'est qu'à cette époque, le taux de mortalité hors variole des enfants décroit exponetiellement entre 0 et 10 ans.

C.2.b. La valeur γ est la limite de $\lambda.e^{\mu.t} + \gamma$ lorsque $t \to +\infty$. On pourrait prendre pour γ la limite en l'infini que l'on peut voir[6] sur la figure, et pour λ, remarquer que la valeur en 0 vaut 0.28 et donc, par rapport au modèle proposé $\lambda + \gamma = 0.28$? On prend donc $\gamma_0 = 0.01$ et $\lambda_0 = 0.27$.

Rq : la phrase : « Dans tout le reste de cette question C.2, on considère que $\gamma = \gamma_0$ et $\lambda = \lambda_0$: seule μ reste à déterminer. » est en incohérence avec le fait que l'on minimise M sur les trois variables. Ici, on « devine » deux paramètres et on minimise une fonction d'une variable pour trouver le troisième paramètre.

C.2.c. La fonction décrite dans le script

```
def M_mu(mu) :
    lamb = lambda0
    gam = gamma0
    s = 0
    for k in range(0, len(Lnorm)):
        s = s+(Lnorm[k]-lamb*exp(-mu*k)-gam)**2
    return s
```

est donc le codage de la fonction $\mu \mapsto M(\lambda_0, \mu, \gamma_0)$.

C.2.d. Le tracé de la fonction `M_mu` montre que le minimum de celle-ci est atteint du côté de $\mu_0 \simeq 0,7$.

C.3.a.

```
def argmin3(f, x, y, z):
    meill_xy = argmin2(f, x, y) # choisit entre x et y celui qui a la plus petite
    return argmin2(f, meill_xy, z)  # choisit ensuite entre ce dernier et z
```

C.3.b.

```
def minimum(f, A, B, eps):
    g = A
    d = B
    while d - g > eps:
        # calcul de xg xd
        xg = g + (d-g)/3
        xd = d - (d-g)/3
        #calcul de g et d
        if f(xg) < f(xd):
            d = xd
        elif f(xg) > f(xd):
            g = xg
        else:
            g = xg
            d = xd
    return argmin3(f, g, d, (g+d)/2)
```

6. Noter qu'on ne voit rien du tout : poser ce genre de questions sur la base de graphiques d'aussi piètre qualité ?

C.4.a.

```
def test_croissant(L):
    '''
    retourne vrai/faux selon que L est triee par ordre croissant
    '''
    for i in range(len(L)-1):
        if L[i+1] < L[i]:
            return False
    return True
```

C.4.b. On forme une liste des valeurs de `M_mu` sur un échantillonnage croissant $[0,8;3]$ (obtenu par un `mu = np.linspace(0.8,3,N)` puis `L = [M_mu(mu_i) for mu_i in mu]`) et on teste cette liste à l'aide de `test_croissant(L)`.
C.4.c. On peut utiliser `minimum(M_mu, 0,2,0.01)` pour évaluer μ_0 rendant minimum M_μ.
C.4.d. On trouve $\mu_0 \approx 0,7061$: pour commenter et conclure sur l'estimation de la fonction a, il faudrait tracer le graphe de de $t \mapsto \tilde{a}(t) = \gamma_0 + \lambda_0.e^{-\mu_0.t}$ par dessus celui de $t \mapsto \ell_t$, sans ça, on ne voit pas trop ?

Je ne comprends pas ce qu'est l'hypothèse « a constant» du premier[7] modèle.

Partie D
Modèle individu-centré

$$Z_n = \begin{pmatrix} \mathbb{P}(X_n = S) \\ \mathbb{P}(X_n = R) \\ \mathbb{P}(X_n = D) \end{pmatrix}.$$

D.1. On suppose qu'au temps 0 l'individu est susceptible. On a alors

$$Z_0 = \begin{pmatrix} \mathbb{P}(X_0 = S) \\ \mathbb{P}(X_0 = R) \\ \mathbb{P}(X_0 = D) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \text{ et } Z_1 = \begin{pmatrix} \mathbb{P}(X_1 = S) \\ \mathbb{P}(X_1 = R) \\ \mathbb{P}(X_1 = D) \end{pmatrix} = \begin{pmatrix} 1-(a+b+c) \\ b \\ a+c \end{pmatrix}.$$

On a ces dernières valeurs par interprétation des probabilités conditionnelles et application de la formule des probabilités totales pour calculer chacune des probabilités $\mathbb{P}(X_1 = S)$, $\mathbb{P}(X_1 = R)$, $\mathbb{P}(X_1 = D)$ à l'aide du SCEI $(\{X_1 = S\}, \{X_1 = R\}, \{X_0 = D\})$:

$$\begin{aligned}
\mathbb{P}(X_1 = S) &= \mathbb{P}(X_1 = S|X_0 = S).\mathbb{P}(X_0 = S) + \mathbb{P}(X_1 = S|X_0 = R).\mathbb{P}(X_0 = R) + \mathbb{P}(X_1 = S|X_0 = D).\mathbb{P}(X_0 = D) \\
\mathbb{P}(X_1 = R) &= \mathbb{P}(X_1 = R|X_0 = S).\mathbb{P}(X_0 = S) + \mathbb{P}(X_1 = R|X_0 = R).\mathbb{P}(X_0 = R) + \mathbb{P}(X_1 = R|X_0 = D).\mathbb{P}(X_0 = D) \\
\mathbb{P}(X_1 = D) &= \mathbb{P}(X_1 = D|X_0 = S).\mathbb{P}(X_0 = S) + \mathbb{P}(X_1 = D|X_0 = R).\mathbb{P}(X_0 = R) + \mathbb{P}(X_1 = D|X_0 = D).\mathbb{P}(X_0 = D)
\end{aligned}$$

qui se réécrit (schéma des transferts) en

$$\begin{aligned}
\mathbb{P}(X_1 = S) = {} & \mathbb{P}(X_1 = S|X_0 = S).\mathbb{P}(X_0 = S) + && 0.\mathbb{P}(X_0 = R) + 0.\mathbb{P}(X_0 = D) \\
\mathbb{P}(X_1 = R) = {} & b.\mathbb{P}(X_0 = S) + && \mathbb{P}(X_1 = R|X_0 = R).\mathbb{P}(X_0 = R) + \mathbb{P}(X_1 = R|X_0 = D).\mathbb{P}(X_0 = D) \\
\mathbb{P}(X_1 = D) = {} & (a+c).\mathbb{P}(X_0 = S) + && a.\mathbb{P}(X_0 = R) + \underbrace{1}_{\text{Eih bennek, eih blavek}} .\mathbb{P}(X_0 = D)
\end{aligned}$$

et finalement en (les trois probabilités de passage manquantes se calculant par complémentation à 1) :

7. partie A ? ce n'est pas un modèle sérieux, B ? a n'y est pas constant...

$$\begin{array}{rrr}
\mathbb{P}(X_1=S)= & (1-(a+b+c)).\mathbb{P}(X_0=S)+ & 0.\mathbb{P}(X_0=R)+0.\mathbb{P}(X_0=D)\\
\mathbb{P}(X_1=R)= & b.\mathbb{P}(X_0=S)+ & (1-a).\mathbb{P}(X_0=R)+0.\mathbb{P}(X_0=D)\\
\mathbb{P}(X_1=D)= & (a+c).\mathbb{P}(X_0=S)+ & a.\mathbb{P}(X_0=R)+1.\mathbb{P}(X_0=D)
\end{array}$$

D.2. On définit la matrice Q :

$$Q=\begin{pmatrix} 1-(a+b+c) & 0 & 0\\ b & 1-a & 0\\ a+c & a & 1\end{pmatrix}.$$

D.2.a. Le raisonnement (de conditionnement sur les valeurs possibles de X_0)) fait dans la question précédente pour calculer Z_1 à partir de Z_0 montre qu'indépendamment de la valeur de Z_0 donnée par l'hypothèse initiale, on a $Z_1=Q.Z_0$

Pour $n\in\mathbb{N}$, on passe de Z_n à Z_{n+1} de la façon dont on passe de Z_0 à Z_1 et on obtient :

$$\forall n\in\mathbb{N},\, Z_{n+1}=Q.Z_n$$

Une récurrence de type géométrique, **à rédiger proprement** montrer que

$$\forall n\in\mathbb{N},\, Z_n=Q^n.Z_0.$$

D.2.b. Comme la matrice Q est triangulaire inférieure, ses valeurs sont les valeurs sur la diagonale :

$$\mathrm{Spec}(Q)=\{\underbrace{1-(a+b+c),1-a,1}_{\lambda_1<\lambda_2<1}\}$$

Plus précisément $\lambda\in\mathrm{Spec}(Q)\Leftrightarrow Q-\lambda.I_3$ non inversible et donc

$$\lambda\in\mathrm{Spec}(Q)\Leftrightarrow\begin{pmatrix}1-(a+b+c)-\lambda & 0 & 0\\ * & 1-a-\lambda & 0\\ * & * & 1-\lambda\end{pmatrix}\text{ non inversible}$$

Le critère de non-inversibilité portant sur la nullité de l'un des éléments diagonaux d'une matrice triangulaire nous permet de conclure.

Calculons les espaces propres associés à chacune de ces valeurs propres (en donnant à chaque fois un vecteur propre générateur de cet espace) :

— $\lambda=1$, calcul de $\mathrm{Ker}\,(Q-I_3)$. L'équation $(Q-I_3).X=0$ d'inconnue $X=\begin{pmatrix}x\\y\\z\end{pmatrix}\in\mathbb{R}^3$ se réécrit sous forme de système :

$$\left\{\begin{array}{rrcl} (-(a+b+c)).x & & = & 0\\ bx & -a.y & = & 0\\ (a+c).x & +a.y & = & 0\end{array}\right.$$

ceci équivaut à $x=0$, $y=0$ et z quelconque, *i.e.*

$$\mathrm{Ker}\,(Q-I_3)=\mathrm{Vect}\left\langle\begin{pmatrix}0\\0\\1\end{pmatrix}\right\rangle$$

— $\lambda = 1-a$, calcul de Ker $(Q-(1-a)I_3)$. L'équation $(Q-I_3).X = 0$ d'inconnue $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ se réécrit sous forme de système :

$$\left\{ \begin{array}{rrrcl} (-(b+c)).x & & & = & 0 \\ bx & & & = & 0 \\ (a+c).x & +a.y & +az & = & 0 \end{array} \right.$$

ceci équivaut à $x = 0$, $y+z = 0$ (car $a \neq 0$, *i.e.*

$$\text{Ker}\,(Q-(1-a)I_3) = \text{Vect}\left\langle \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \right\rangle$$

— $\lambda = 1-(a+b+c)$, calcul de Ker $(Q-(1-(a+b+c))I_3)$. L'équation $(Q-I_3).X = 0$ d'inconnue $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ se réécrit sous forme de système :

$$\left\{ \begin{array}{rrrcl} 0.x & & & = & 0 \\ bx & +(b+c).y & & = & 0 \\ (a+c).x & +a.y & +(a+b+c)z & = & 0 \end{array} \right.$$

ceci équivaut à (système échelonné quite à échanger les places des inconnues : (z,x,y), d'inconnue secondaire y) : $\exists \mu \in \mathbb{R}$

$$x = -\frac{b+c}{b}.\mu, y = \mu, z = \frac{1}{a+b+c}\left(+\frac{(b+c)(a+c)}{b}\mu - a.\mu\right) = \frac{\mu}{(a+b+c).b}((b+c)(a+c)-ab) = -\frac{c}{b}\mu$$

i.e.

$$\text{Ker}\,(Q-(1-(a+b+c))I_3) = \text{Vect}\left\langle \begin{pmatrix} -\frac{b+c}{b} \\ 1 \\ \frac{c}{b} \end{pmatrix} \right\rangle = \text{Vect}\left\langle \begin{pmatrix} -(b+c) \\ b \\ c \end{pmatrix} \right\rangle$$

D.2.c. En posant

$$P = \begin{pmatrix} -(b+c) & 0 & 0 \\ b & -1 & 0 \\ c & 1 & 1 \end{pmatrix}$$

on a, P inversible car triangulaire sans élémetns nuls sur la diagonale et

$$P^{-1}.Q.P = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ où } \lambda_1 = 1-(a+b+c),\ \lambda_2 = 1-a$$

Pour calculer P^{-1}, résolvons le système $P.X = Y$ d'inconnue $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, de second membre $Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$.

Par le pivot de GAUSS, les sytème suivants sont équivalents à $PX = Y$:

$$\left\{ \begin{array}{rrrcl} -(b+c)x_1 & & & = & y_1 \\ bx_1 & -x_2 & & = & y_2 \\ cx_1 & +x_2 & +x_3 & = & y_3 \end{array} \right. \xrightarrow{L_3 \leftarrow L_3+L_1+L_2} \left\{ \begin{array}{rrrcl} -(b+c)x_1 & & & = & y_1 \\ bx_1 & -x_2 & & = & y_2 \\ & & +x_3 & = & y_1+y_2+y_3 \end{array} \right.$$

$$\xrightarrow{L_1 \leftarrow -\frac{1}{b+c}L_1} \left\{ \begin{array}{rrrcl} x_1 & & & = & -\frac{1}{b+c}y_1 \\ bx_1 & -x_2 & & = & y_2 \\ & & x_3 & = & y_1+y_2+y_3 \end{array} \right. \xrightarrow{L_2 \leftarrow -L_2+bL_1} \left\{ \begin{array}{rrrcl} x_1 & & & = & -\frac{1}{b+c}y_1 \\ & x_2 & & = & -\frac{b}{b+c}y_1 \quad -y_2 \\ & & x_3 & = & y_1 \quad +y_2 \quad +y_3 \end{array} \right.$$

Ceci montre que

$$P^{-1} = \begin{pmatrix} -\frac{1}{b+c} & 0 & 0 \\ -\frac{b}{b+c} & -1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

D.2.d. Montrons par récurrence (sur n) que pour tout $n \in \mathbb{N}$:

$$Z_n = P \begin{pmatrix} \lambda_1^n & 0 & 0 \\ 0 & \lambda_2^n & 0 \\ 0 & 0 & 1 \end{pmatrix} P^{-1} Z_0.$$

— Pour $n = 0$, il s'agit de vérifier que $Z_n = P.I_3.P^{-1}.Z_0 = Z_0$: c'est vrai !
— Supposons la formule vraie à un certain rang n. On a alors par D.2.a

$$\begin{array}{rcl} Z_{n+1} & = & Q.Z_n \\ \textbf{[diago } Q\textbf{]} & = & P^{-1}.\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.P.Z_n \\ \textbf{[formule pour } Z_n\textbf{]} & = & P^{-1}.\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.\underbrace{P.P^{-1}}_{I_3}.\begin{pmatrix} \lambda_1^n & 0 & 0 \\ 0 & \lambda_2^n & 0 \\ 0 & 0 & 1 \end{pmatrix}.P.Z_0 \\ \textbf{[Calcul]} & = & P^{-1}.\begin{pmatrix} \lambda_1^{n+1} & 0 & 0 \\ 0 & \lambda_2^{n+1} & 0 \\ 0 & 0 & 1 \end{pmatrix}.P.Z_0 \end{array}$$

Ce qui est la formule pour $n+1$.

Dans la suite on admettra que (issu du calcul explicite du quadruple produit) :

$$\forall n \in \mathbb{N}, Z_n = \begin{pmatrix} \lambda_1^n \\ \frac{-b\lambda_1^n + b\lambda_2^n}{b+c} \\ 1 - \frac{c\lambda_1^n + b\lambda_2^n}{b+c} \end{pmatrix}.$$

D.2.e. On a donc, en observant les composantes de Z_n que :

$$\forall n \in \mathbb{N}, \mathbb{P}(X_n = S) = \lambda_1^n, \mathbb{P}(X_n = R) = \frac{-b\lambda_1^n + b\lambda_2^n}{b+c} \text{ et } \mathbb{P}(X_n = D) = 1 - \frac{c\lambda_1^n + b\lambda_2^n}{b+c}.$$

Comme $0 \leq \lambda_1, \lambda_2 < 1$, lorsque $n \to +\infty$, $\lambda_1^n \to 0$, $\lambda_2^n \to 0$ et par combinaisons linéaires de limites :

$$\begin{array}{rcl} \lim\limits_{n \to +\infty} \mathbb{P}(X_n = S) & = & 0 \\ \lim\limits_{n \to +\infty} \mathbb{P}(X_n = R) & = & 0 \\ \lim\limits_{n \to +\infty} \mathbb{P}(X_n = D) & = & 1 \end{array}$$

On peut interpréter ceci en disant que la probabilité de mourir un jour est de 1.

D.3.a. Soit A_n l'évènement "l'individu vit exactement n années". Considérons l'événement D_n, de probabilité $\mathbb{P}(X_n = D)$: à l'année n, l'individu est mort. On a

$$A_n = D_{n+1} \setminus D_n$$

et donc

$$\begin{array}{rcl}
\mathbb{P}(A_n) & = & \mathbb{P}(D_{n+1}) - \mathbb{P}(D_n) \\
 & = & 1 - \dfrac{c\lambda_1^{n+1} + b\lambda_2^{n+1}}{b+c} - \left(1 - \dfrac{c\lambda_1^n + b\lambda_2^n}{b+c}\right) \\
 & = & \left(\dfrac{c}{c+b} - \lambda_1\right)\lambda_1^n + \left(\dfrac{b}{c+b} - \lambda_2\right)\lambda_2^n \\
\textbf{[valeurs } \lambda_1, \lambda_2\textbf{, calcul]} & = & \left(\dfrac{c(a+b+c)}{b+c}\right)\lambda_1^n + \dfrac{ab}{b+c}\lambda_2^n.
\end{array}$$

Dans la suite on admettra que le premier terme est négligeable devant le deuxième et on prendra :

$$\mathbb{P}(A_n) \simeq \frac{ab}{b+c}\lambda_2^n.$$

D.3.b. Soit $\rho \in]0,1[$ et soit $\Sigma = \sum_{n=0}^{+\infty} n\rho^n$. C'est un élément de cours à savoir refaire (valeur de la série géométrique dérivée première) que :

$$\Sigma = \frac{\rho}{(\rho-1)^2}.$$

D.3.c. L'espérance de vie d'un individu à la naissance est égale à $\mathbb{E}(T)$ où T est le nombre d'année que vit l'individu. En d'autres termes $\{T = n\} = A_n$ et donc (avec la simplification faite), $\mathbb{P}(T = n) = \frac{ab}{b+c}\lambda_2^n$

On a alors

$$\mathscr{E} = \mathbb{E}(T) = \frac{ab}{b+c}\frac{\lambda_2}{(1-\lambda_2)^2} = \frac{ab}{b+c}\frac{1-a}{(1-(1-a))^2} = \frac{b(1-a)}{a(b+c)}$$

D.3.d. Application numérique : dans le cas de la variole (en prenant) $a = 2/64$, $b = 7/64$ et $c = 1/64$, on obtient

$$\mathscr{E} = \frac{7 \times 62}{2 \times (7+1)} = \frac{7 \times 31}{8} = 28 - \frac{7}{8} = 27 + \frac{1}{8}$$

Pour comparer avec l'espérance de vie réelle de 26 ans et 7 mois, ben même si ce n'est pas éloigné, on n'a pas trop d'éléments vu que les valeurs données sont un peu arbitraires.....

Le choix de prendre a constant n'est pas vraiment justifié : comme on l'a vu, le taux de mortalité dépend probablement de l'âge de l'invidu, il faudrait donc– a minima– travailler avec une suite (a_n) qui donne des probabilités de transition dépendant du temps avec une matrice Q_n copmme Q mais avec a remplacé par a_n, a_n obtenu par exemple comme dans la partie C. Les calculs sont tout à fait faisables, notamment car la matrice diagonalisant Q, diagonalise aussi Q_n et il faudra remplacer dans D.3 λ_2^n par $\prod_{k=1}^{n}(1-a_n)$.