Banque « Agro-Véto »
A BCPST - 0221

MODÉLISATION MATHÉMATIQUE ET INFORMATIQUE

Durée : 3 heures 30 minutes

L'usage d'une calculatrice est autorisé pour cette épreuve.

Chaque candidate ou candidat est responsable de la vérification de son sujet d'épreuve : pagination et impression de chaque page. Ce contrôle doit être fait en début d'épreuve. En cas de doute, il convient d'alerter au plus tôt l'équipe de surveillance qui vérifiera et, éventuellement, remplacera le sujet.

Ce sujet comporte 8 pages numérotées de 1 à 8.

Si, au cours de l'épreuve, une candidate ou un candidat repère ce qui lui semble être une erreur d'énoncé, elle ou il le signale sur sa copie et poursuit sa composition en expliquant les raisons des initiatives prises.

Le sujet se compose de 4 parties largement indépendantes hormis la question 5. de la partie 2 et la question 4.(d) de la partie 3. La partie 3 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.

1 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 :

$$\begin{cases} \dfrac{dz}{dt} = -\alpha z \\ z(0) = z_0 = 1300 \end{cases} \tag{1}$$

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

1. Donner la solution de l'équation différentielle (1) sur $\mathbb{R}^+$.
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 $\dfrac{z(t)}{z_0}$. Que représente la quantité $F(t)$ par rapport à la cohorte ?
   (b) Interpréter F en terme de lois de probabilités usuelles. En déduire l'espérance de vie $\mathcal{E}$ 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.
3. (a) Soit $\Delta_t \in \mathbb{R}^+$ un intervalle de temps. Que représente la quantité $z(t) - z(t + \Delta_t)$?
   (b) Pour tout $i \in \mathbb{N}$, on pose $t_i = i\Delta_t$. Donner une interprétation possible de la quantité $\mathcal{E}_{\Delta_t}$ suivante. On supposera que la somme est bien définie.

$$\mathcal{E}_{\Delta_t} = \frac{\sum_{i=0}^{+\infty} t_i\,(z(t_i) - z(t_{i+1}))}{z_0}. \tag{2}$$

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

$$\lim_{\Delta_t \to 0} \mathcal{E}_{\Delta_t} = \frac{\int_0^{+\infty} t \frac{dz}{dt} dt}{\int_0^{+\infty} \frac{dz}{dt} dt} \tag{3}$$

Calculer ce rapport et faire le lien avec la question 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.

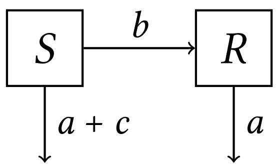
2 Modèle de Bernoulli

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 C^1 sur $\mathbb{R}^+$. Bernoulli pose le modèle suivant pour son problème :

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



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 $c = 1/64$.

1. (a) Expliquer pourquoi Bernoulli n'a pas fait l'approximation de a constant dans son modèle.

   (b) Interpréter $-\dfrac{dx}{dt}$ et cS. Expliquer alors pourquoi $a = -\dfrac{1}{x}\left(\dfrac{dx}{dt} + cS\right)$.

   (c) Justifier que :

$$\frac{dz}{dt} = \frac{1}{x}\left(\frac{dx}{dt} + cS\right) z \tag{4}$$

2. (a) Soit q la fonction définie sur $\mathbb{R}^+$ par $q = \dfrac{x}{S}$. Écrire l'équation différentielle vérifiée par q.

   (b) Résoudre l'équation différentielle et en déduire que :

$$q(t) = \frac{1}{8} + \frac{7}{8} e^{t/8} \tag{5}$$

   Ce résultat pourra être admis par la suite.

3. (a) Soit H la fonction définie sur $\mathbb{R}^+$ par $H = \ln \dfrac{z}{x}$, exprimer $\dfrac{dH}{dt}$ en fonction de q. En déduire que :

$$\frac{dH}{dt} = \frac{1}{8} \times \frac{1}{1 + 7e^{t/8}} \tag{6}$$

   Ce résultat pourra être admis par la suite.

(b) Soit G la fonction définie sur $\mathbb{R}^+$ par $G(t) = \dfrac{e^{t/8}}{1+7e^{t/8}}$. On admettra que :

$$\frac{d\ln G}{dt} = \frac{dH}{dt} = \frac{d}{dt}\left(\ln\frac{z}{x}\right) \tag{7}$$

En déduire que :

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

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. On donne 1/200 la probabilité qu'un individu décède des suites de l'inoculation. Justifier que $y = \gamma z$ avec γ une constante à préciser.
5. Ainsi l'évolution du nombre d'individus $y(t)$ d'une cohorte fictive systématiquement inoculée est donnée par :

$$y(t) = \gamma \times \frac{8e^{t/8}}{1+7e^{t/8}}x(t) \tag{9}$$

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

3 Estimation de la fonction a

Seul le résultat final de la question 1 est utile pour répondre aux questions suivantes : si besoin, on pourra donc admettre dès le début de la question 2 que la liste `Lnorm` est créée. Une annexe en dernière page du sujet rappelle quelques fonctions Python.

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, ...]`.

   (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']
```

   (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.

   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`.

```python
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) 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]
```

(d) Proposer une suite d'instructions qui crée la liste `Lnorm` précédemment définie.

Si besoin, on rappelle que `float("6.5\n")` *renvoie* `6.5`.

2. On se propose d'estimer la fonction a représentant le taux de morts hors variole au cours du temps. On note $(\ell_t)_{t\in\{0,\ldots,83\}}$ 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 figure 1. On admettra que le comportement en temps long reste stable. On définit la fonction $M : \mathbb{R}^3 \longrightarrow \mathbb{R}$ par :

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

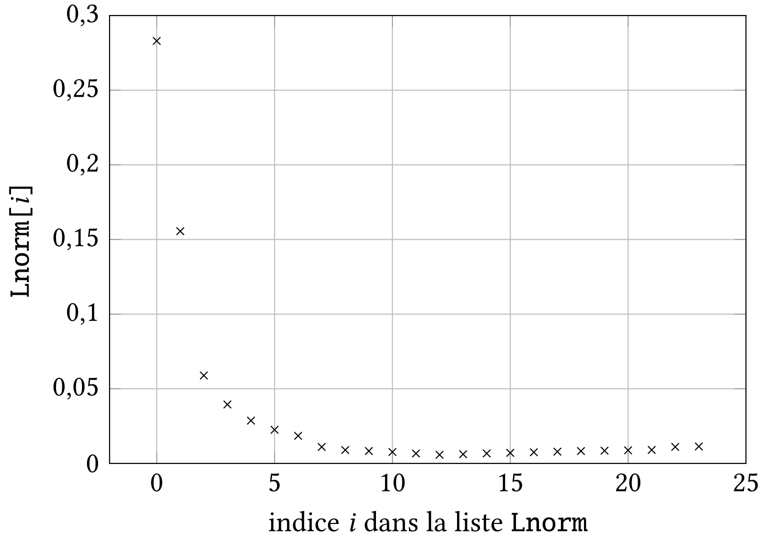


Figure 1 – Tracé des valeurs de ℓ_t

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

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

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

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 $t \mapsto \lambda e^{-\mu t} + \gamma$ plutôt qu'une fonction polynomiale ?

(b) À l'aide des données représentées sur la figure 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 variables globales `gamma0` et `lambda0`. Dans tout le reste de cette question 2, on considère que $\gamma = \gamma_0$ et $\lambda = \lambda_0$: seule μ reste à déterminer.

(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
```

(d) Le tracé de la fonction `M_mu` est donné en figure 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 ?

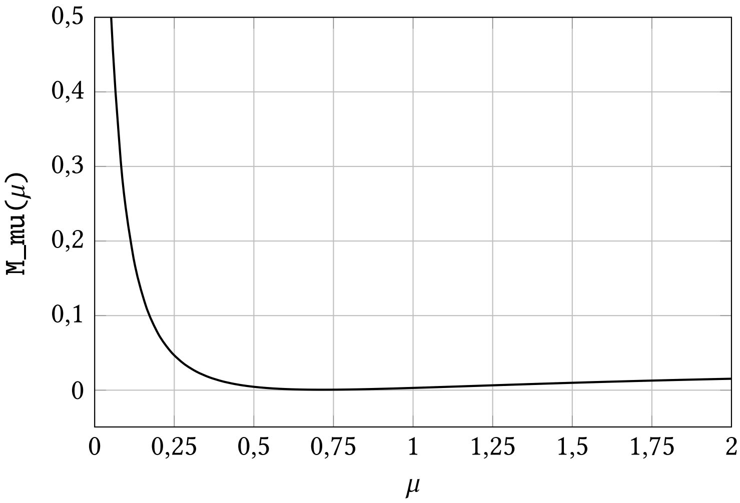


FIGURE 2 – Tracé de la fonction `M_mu`

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 $f(x^g) < f(x^d)$ 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 $f(x^g) > f(x^d)$ 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 $f(x^g) = f(x^d)$ donc x^* se situe dans $[x^g\,;x^d]$ 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.

(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.

(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 __________
```

4. On revient dans cette question à la problématique de la question 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 3, plutôt que par lecture graphique.

   (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.*

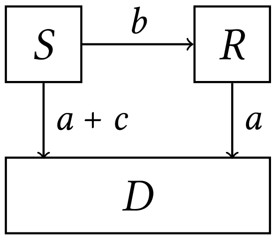
   (b) À l'aide de `test_croissant`, proposer un raisonnement (sans écrire de code) pour tester si la fonction `M_mu` est bien croissante sur l'intervalle $[0{,}8\ ; 3]$.

   (c) Proposer une instruction pour trouver la valeur μ_0 cherchée.

   (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 ?

4 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 susceptibles, remis et morts. 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 au temps n aura une probabilité b d'être remis au temps $n+1$.

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

$$Z_n = \begin{pmatrix} \mathbb{P}(X_n = S) \\ \mathbb{P}(X_n = R) \\ \mathbb{P}(X_n = D) \end{pmatrix} \tag{10}$$

1. On suppose qu'au temps 0 l'individu est susceptible. Donner alors les valeurs de Z_0 et Z_1.
2. On définit une matrice Q :

$$Q = \begin{pmatrix} 1-(a+b+c) & 0 & 0 \\ b & 1-a & 0 \\ a+c & a & 1 \end{pmatrix} \tag{11}$$

   (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 :

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

   (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.

   (c) En déduire une matrice P telle que :

$$Q = P \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 1 \end{pmatrix} P^{-1} \tag{12}$$

   Calculer P^{-1}.

   (d) Montrer que pour tout $n \in \mathbb{N}$:

$$Z_n = P \begin{pmatrix} \lambda_1^n & 0 & 0 \\ 0 & \lambda_2^n & 0 \\ 0 & 0 & 1 \end{pmatrix} P^{-1} Z_0 \tag{13}$$

Dans la suite on admettra que :

$$Z_n = \begin{pmatrix} \lambda_1^n \\ \dfrac{-b\lambda_1^n + b\lambda_2^n}{b+c} \\ 1 - \dfrac{c\lambda_1^n + b\lambda_2^n}{b+c} \end{pmatrix} \tag{14}$$

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

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 :

$$\mathbb{P}(A_n) \simeq \frac{ab}{b+c}\lambda_2^n \tag{15}$$

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

$$\Sigma = \frac{\rho}{(\rho-1)^2} \tag{16}$$

(c) Montrer que l'espérance de vie d'un individu à la naissance est égale à :

$$\mathcal{E} = \frac{b(1-a)}{a(b+c)} \tag{17}$$

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) 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 3

- `f = open("monfichier.txt", "r")` ouvre le document intitulé `monfichier.txt` (en mode lecture) ; le fichier est alors désigné par la 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.

FIN DU SUJET