EXERCICE

Étant donné un entier n supérieur ou égal à 2, on considère un nuage de n points du plan, c'est-à-dire un n-uplet $((x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n))$ d'éléments de $\mathbb{R}^2$. On suppose que les réels $x_1,\ x_2\cdots$ (resp $y_1,\ y_2\cdots$) ne sont pas tous égaux.
On appelle moyenne arithmétique $\overline{x}$ et écart-type σ_x du n-uplet $x=(x_1,x_2,,x_n)$ les réels suivants :

$$\overline{x}=\frac{1}{n}\sum_{k=1}^{n}x_k \quad \text{et} \quad \sigma_x=\sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(x_k-\overline{x}\right)^2}$$

On définit de même la moyenne arithmétique $\overline{y}$ et l'écart-type σ_y du n-uplet $y=(y_1,y_2,\cdots,y_n)$
La covariance $\operatorname{cov}(x,y)$ et le coefficient de corrélation linéaire $r(x,y)$ du couple (x,y) sont donnés par :

$$\operatorname{cov}(x,y)=\frac{1}{n}\sum_{k=1}^{n}\left(x_k-\overline{x}\right)\left(y_k-\overline{y}\right) \quad \text{et} \quad r(x,y)=\frac{\operatorname{cov}(x,y)}{\sigma_x\times\sigma_y}$$

Soit f la fonction définie sur $\mathbb{R}^2$ à valeurs réelles qui, à tout couple (a,b) de $\mathbb{R}^2$, associe le réel $f(a,b)$ tel que :

$$f(a,b)=\sum_{k=1}^{n}\left(a\ x_k+b-y_k\right)^2$$

1. Les fonctions $(a,b)\to a$ et $(a,b)\to b$ sont C^2 sur $\mathbb{R}^2$ donc $(a,b)\to ax_k+b-y_k$ est somme de fonctions C^2 et f est C^2 comme composée et somme de fonctions C^2 sur $\mathbb{R}^2$
2. a) $\frac{\partial f}{\partial a}(a,b)=\sum_{k=1}^{n}2\left(a\ x_k+b-y_k\right)x_k$ et $\frac{\partial f}{\partial b}(a,b)=\sum_{k=1}^{n}2\left(a\ x_k+b-y_k\right)$
   Donc (a,b) est un point critique de f
   $\iff (S)\left\{\begin{array}{l} a\ \sum_{k=1}^{n}x_k^2+b\sum_{k=1}^{n}x_k-\sum_{k=1}^{n}y_kx_k=0 \\ a\ \sum_{k=1}^{n}x_k+nb-\sum_{k=1}^{n}y_k=0\end{array}\right.$

   b) par $L_1-\frac{1}{n}\sum_{k=1}^{n}x_kL_2$ on a :
   $(S)\iff\left\{\begin{array}{l} a\ \left[\sum_{k=1}^{n}x_k^2-\frac{1}{n}\left(\sum_{k=1}^{n}x_k\right)^2\right]-\sum_{k=1}^{n}y_kx_k+\frac{1}{n}\sum_{k=1}^{n}y_k\sum_{k=1}^{n}x_k=0 \\ a\ \sum_{k=1}^{n}x_k+nb-\sum_{k=1}^{n}y_k=0\end{array}\right.$
   Il faut faire apparaître les quantités $\overline{x}=\frac{1}{n}\sum_{k=1}^{n}x_k$ que l'on a par $\boxed{\sum_{k=1}^{n}x_k=n\ \overline{x}}$ et $\boxed{\sum_{k=1}^{n}y_k=n\ \overline{y}}$
   $\sigma_x^2=\frac{1}{n}\sum_{k=1}^{n}\left(x_k-\overline{x}\right)^2$ donc

$$\begin{aligned}\sum_{k=1}^{n}\left(x_k-\overline{x}\right)^2 &= \sum_{k=1}^{n}x_k^2-2\overline{x}\sum_{k=1}^{n}x_k+\sum_{k=1}^{n}\overline{x}^2 \\ &= \sum_{k=1}^{n}x_k^2-2n\overline{x}^2+n\overline{x}^2=\sum_{k=1}^{n}x_k^2-n\overline{x}^2\end{aligned}$$

   donc $\boxed{\sum_{k=1}^{n}x_k^2=n\sigma_x^2+n\overline{x}^2}$
   $\operatorname{cov}(x,y)=\frac{1}{n}$

$$\begin{aligned}\sum_{k=1}^{n}\left(x_k-\overline{x}\right)\left(y_k-\overline{y}\right) &= \sum_{k=1}^{n}x_k\ y_k-\overline{x}\sum_{k=1}^{n}y_k-\overline{y}\sum_{k=1}^{n}x_k+\sum_{k=1}^{n}\overline{x}\ \overline{y} \\ &= \sum_{k=1}^{n}x_k\ y_k-2n\overline{x}\ \overline{y}+n\overline{x}\ \overline{y}\end{aligned}$$

donc $\boxed{\sum_{k=1}^{n} x_k\ y_k = n\ \operatorname{cov}(x,y) + n\ \overline{x}\ \overline{y}}$

$(S) \iff \begin{cases} a\ \left[n\sigma_x^2 + n\overline{x}^2 - n\ \overline{x}^2\right] - n\ \operatorname{cov}(x,y) - n\ \overline{x}\ \overline{y} + n\ \overline{x}\ \overline{y} = 0 \\ a\ n\ \overline{x} + nb - n\ \overline{y} = 0 \end{cases}$

$\iff \begin{cases} a\ \sigma_x^2 - \operatorname{cov}(x,y) = 0 \\ a\ \overline{x} + b - \overline{y} = 0 \end{cases}$ or $\sum_{k=1}^{n}(x_k - \overline{x})^2 \neq 0$ car les x_k ne sont pas tous égaux (à $\overline{x}$) donc $\sigma_x^2 \neq 0$

$$\iff \begin{cases} a = \dfrac{\operatorname{cov}(x,y)}{\sigma_x^2} \\ b = -\dfrac{\operatorname{cov}(x,y)}{\sigma_x^2}\overline{x} + \overline{y} \end{cases}$$

Conclusion : $\boxed{f \text{ admet un unique point critique } \left(\hat{a},\hat{b}\right) = \left(\dfrac{\operatorname{cov}(x,y)}{\sigma_x^2}, -\dfrac{\operatorname{cov}(x,y)}{\sigma_x^2}\overline{x} + \overline{y}\right)}$

c) $r = \frac{\partial^2 f}{\partial a^2}(a,b) = 2\sum_{k=1}^{n} x_k^2 = 2n\left(\sigma_x^2 + \overline{x}^2\right)$
$s = \frac{\partial^2 f}{\partial a \partial b}(a,b) = 2\sum_{k=1}^{n}\ x_k = 2n\ \overline{x}$
$t = \frac{\partial^2 f}{\partial b^2}(a,b) = 2\sum_{k=1}^{n} 1 = 2n$
Donc en $\left(\hat{a},\hat{b}\right)$ on a $rt - s^2 = 4n^2\left(\sigma_x^2 + \overline{x}^2\right) - 4n^2\overline{x}^2 = 4n^2\sigma_x^2 > 0$
f présente donc un extremum local en $\left(\hat{a},\hat{b}\right)$ et comme $r > 0$, c'est un minimum local.

d) On calcule avec : $\left(\hat{a},\hat{b}\right) = \left(\dfrac{\operatorname{cov}(x,y)}{\sigma_x^2}, -\dfrac{\operatorname{cov}(x,y)}{\sigma_x^2}\overline{x} + \overline{y}\right)$
qui sont solutions de $\begin{cases} a\ \sum_{k=1}^{n} x_k^2 + b\sum_{k=1}^{n} x_k - \sum_{k=1}^{n} y_k x_k = 0 \\ a\ \sum_{k=1}^{n} x_k + nb - \sum_{k=1}^{n} y_k = 0 \end{cases}$ ... on va s'en resservir pour simplifier

$$\begin{aligned}
f\left(\hat{a},\hat{b}\right) &= \sum_{k=1}^{n}\left(\hat{a}\ x_k + \hat{b} - y_k\right)^2 \\
&= \sum_{k=1}^{n}\left(\hat{a}^2\ x_k^2 + \hat{b}^2 + y_k^2 + 2\hat{a}\ x_k\ \hat{b} - 2\hat{a}\ x_k y_k - 2\hat{b}\ y_k\right) \\
&= \hat{a}^2\sum_{k=1}^{n} x_k^2 + n\hat{b}^2 + \sum_{k=1}^{n} y_k^2 + 2\hat{a}\ \hat{b}\sum_{k=1}^{n} x_k - 2\hat{a}\sum_{k=1}^{n} x_k y_k - 2\hat{b}\sum_{k=1}^{n} y_k \\
&\quad \text{où on fait réapparaître les conditions de minima} \\
&= \hat{a}\left[\hat{a}\sum_{k=1}^{n} x_k^2 + \ \hat{b}\sum_{k=1}^{n} x_k - \sum_{k=1}^{n} x_k y_k\right] + \hat{b}\left[n\hat{b} + \hat{a}\sum_{k=1}^{n} x_k - \sum_{k=1}^{n} y_k\right] \\
&\quad + \sum_{k=1}^{n} y_k^2 - \hat{a}\sum_{k=1}^{n} x_k y_k - \hat{b}\sum_{k=1}^{n} y_k \\
&= \sum_{k=1}^{n} y_k^2 - \hat{a}\sum_{k=1}^{n} x_k y_k - \hat{b}\sum_{k=1}^{n} y_k \\
&= n\sigma_y^2 + n\overline{y}^2 - \frac{\operatorname{cov}(x,y)}{\sigma_x^2}\left(n\ \operatorname{cov}(x,y) + n\ \overline{x}\ \overline{y}\right) - \left(-\frac{\operatorname{cov}(x,y)}{\sigma_x^2}\overline{x} + \overline{y}\right)n\overline{y} \\
&= n\sigma_y^2 - \frac{\operatorname{cov}(x,y)}{\sigma_x^2}\left(n\ \operatorname{cov}(x,y) + n\ \overline{x}\ \overline{y}\right) + \frac{\operatorname{cov}(x,y)}{\sigma_x^2}\overline{x}n\overline{y} \\
&= n\sigma_y^2 - n\frac{\operatorname{cov}(x,y)^2}{\sigma_x^2}
\end{aligned}$$

d'autre part

$$\begin{aligned} n\sigma_y^2\left(1-r^2(x,y)\right) &= n\sigma_y^2\left(1-\frac{\operatorname{cov}(x,y)^2}{\sigma_x^2\times\sigma_y^2}\right) \\ &= n\sigma_y^2 - n\frac{\operatorname{cov}(x,y)^2}{\sigma_x^2} \end{aligned}$$

Conclusion : $\boxed{f\left(\hat{a},\hat{b}\right) = n\sigma_y^2\left(1-r^2(x,y)\right)}$

3. a) $f\left(\hat{a},\hat{b}\right)$ est une somme de termes positifs donc $f\left(\hat{a},\hat{b}\right)\geq 0$ et donc $1-r^2(x,y)\geq 0$.

   Conclusion : $\boxed{|r(x,y)|\leq 1.}$

   b) Lorsque $|r(x,y)|=1$ on a $f\left(\hat{a},\hat{b}\right)=0$ donc
   $\sum_{k=1}^n\left(\hat{a}\ x_k+\hat{b}-y_k\right)^2=0$ et (somme de termes positifs), $\left(\hat{a}\ x_k+\hat{b}-y_k\right)^2=0$ pour tout k

   Conclusion : $\boxed{\text{pour tout } k : y_k=\hat{a}\ x_k+\hat{b} \text{ les points sont alignés.}}$

PROBLÈME

Dans tout le problème, N désigne un entier naturel fixé supérieur ou égal à 2, et p un réel fixé de l'intervalle $]0,1[$.
On pose : $q=1-p$. Soit n un entier naturel quelconque.
Dans une population de N individus, on s'intéresse à la propagation d'un certain virus.
Chaque jour, on distingue dans cette population trois catégories d'individus : en premier lieu, les individus sains, c'est-à-dire ceux qui ne sont pas porteurs du virus, ensuite les individus qui viennent d'être contaminés et qui sont inoffensifs pour les autres, et enfin, les individus contaminés par le virus et qui sont contagieux.
Ces trois catégories évoluent jour après jour selon le modèle suivant :
- chaque jour n, chaque individu sain peut être contaminé par n'importe lequel des individus contagieux ce jour avec la même probabilité p, ces contaminations éventuelles étant indépendantes les unes des autres ;
- un individu contaminé le jour n devient contagieux le jour $n+1$;
- chaque individu contagieux le jour n redevient sain le jour $n+1$.

On note alors X_n le nombre aléatoire d'individus contagieux le jour n.
On remarquera que si, pour un certain entier naturel i, on a $X_i=0$, alors on a aussi $X_{i+1}=0$.
Les variables aléatoires $X_0,X_1,\ldots,X_n$ sont supposées définies sur un même espace probabilisé $(\Omega,\mathcal{A},\mathrm{P})$, et$E(X_n)$ désigne, pour tout n de $\mathbb{N}$, l'espérance de X_n .

Partie I. Un cas particulier

Dans cette partie uniquement, on suppose que l'on a : $N=3$ et $p=1/3$. On considère les matrices S et R suivantes :

$$S=\begin{pmatrix}9&4&4&9\\0&4&5&0\\0&1&0&0\\0&0&0&0\end{pmatrix},\quad R=\begin{pmatrix}0&1&-6&1\\1&0&5&0\\-1&0&1&0\\0&-1&0&0\end{pmatrix}$$

L'ensemble $\mathcal{M}_{4,1}(\mathbb{R})$ des matrices colonnes à quatre lignes est confondu avec l'espace vectoriel $\mathbb{R}^4$.

1. Par la méthode de Gauss :

$$\begin{pmatrix} 0 & 1 & -6 & 1 \\ 1 & 0 & 5 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \quad \begin{matrix} L_1 \to L_2 \\ L_2 \to L_1 \\ L_3 + L_2 \to L_3 \end{matrix} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\Longleftrightarrow \begin{pmatrix} 1 & 0 & 5 & 0 \\ 0 & 1 & -6 & 1 \\ 0 & 0 & 6 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \quad L_4 + L_2 \to L_4 \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\Longleftrightarrow \begin{pmatrix} 1 & 0 & 5 & 0 \\ 0 & 1 & -6 & 1 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & -6 & 1 \end{pmatrix} \quad \begin{matrix} L_1 - \frac{5}{6} L_3 \\ L_2 + L_3 \\ L_3/6 \\ L_4 + L_3 \end{matrix} \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\Longleftrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad L_2 - L_4 \quad \begin{pmatrix} 0 & 1/6 & -5/6 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1/6 & 1/6 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\Longleftrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1/6 & -5/6 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 1/6 & 1/6 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \text{ Donc } R \text{ est inversible et } R^{-1} = \begin{pmatrix} 0 & 1/6 & -5/6 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 1/6 & 1/6 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

2. a) En s'inspirant de la question b) on teste si les colonnes de R sont propres pour S :

$$\begin{pmatrix} 9 & 4 & 4 & 9 \\ 0 & 4 & 5 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}$$ donc -1 est valeur propre de S associé à $(0, 1, -1, 0)$

$$\begin{pmatrix} 9 & 4 & 4 & 9 \\ 0 & 4 & 5 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$ donc 0 est valeur propre de S associé à $(1, 0, 0, -1)$

$$\begin{pmatrix} 9 & 4 & 4 & 9 \\ 0 & 4 & 5 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -6 \\ 5 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -30 \\ 25 \\ 5 \\ 0 \end{pmatrix} = 5 \begin{pmatrix} -6 \\ 5 \\ 1 \\ 0 \end{pmatrix}$$ donc 5 est valeur propre de S associé à $(-6, 5, 1, 0)$

$$\begin{pmatrix} 9 & 4 & 4 & 9 \\ 0 & 4 & 5 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 9 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$ donc 9 est valeur propre de S associé à $(1, 0, 0, 0)$

Conclusion : $\boxed{-1,\ 0,\ 5 \text{ et } 9 \text{ sont des valeurs propres de } S.}$

b) S étant d'ordre 4, et ayant 4 valeurs propres distinctes, la concaténations des vecteurs propres associés forme une base de $\mathbb{R}^4$.

Donc $R^{-1} S\ R = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 9 \end{pmatrix}$

c) et pour tout $n \in \mathbb{N} : S^n = RD^nR^{-1}$

$$= \begin{pmatrix} 0 & 1 & -6 & 1 \\ 1 & 0 & 5 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} (-1)^n & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 5^n & 0 \\ 0 & 0 & 0 & 9^n \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{6} & -\frac{5}{6} & 0 \\ 0 & 0 & 0 & -1 \\ 0 & \frac{1}{6} & \frac{1}{6} & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 9^n & 9^n - 5^n & 9^n - 5^n & 9^n \\ 0 & \frac{1}{6}(-1)^n + \frac{5}{6}5^n & \frac{5}{6}5^n - \frac{5}{6}(-1)^n & 0 \\ 0 & \frac{1}{6}5^n - \frac{1}{6}(-1)^n & \frac{5}{6}(-1)^n + \frac{1}{6}5^n & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

3. Soit n un entier fixé de $\mathbb{N}$.

a) Sachant $(X_n = 0)$, aucun n'est contagieux donc $(X_{n+1} = 0)$ et donc $\mathrm{P}_{X_n=0}(X_{n+1} = 0) = 1$:
$\mathrm{P}_{X_n=0}(X_{n+1} = 1) = 0 : \mathrm{P}_{X_n=0}(X_{n+1} = 2) = 0 : \mathrm{P}_{X_n=0}(X_{n+1} = 3) = 0$

b) Sachant $(X_n = 3)$ on a alors $(X_{n+1} = 0)$ donc $\mathrm{P}_{X_n=3}(X_{n+1} = 0) = 1$:
$\mathrm{P}_{X_n=3}(X_{n+1} = 1) = 1 : \mathrm{P}_{X_n=3}(X_{n+1} = 2) = 0 : \mathrm{P}_{X_n=3}(X_{n+1} = 3) = 0$

c) Quand $[X_n = 1]$ il y a à l'instant n , 2 personnes pouvant être contaminés indépendamment par le seul contagieux, donc avec une probabilité de $p = \frac{1}{3}$
Donc $X_{n+1}/[X_n = 1] \hookrightarrow \mathcal{B}\left(2, \frac{1}{3}\right)$
Quand $[X_n = 2]$ il y a à l'instant n un seul individu pouvant être contaminé par l'un ou l'autre des contagieux (indépendamment).
La probabilité qu'il ne soit pas contaminé est donc $\frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$ et la probabilité qu'il le soit est de $\frac{5}{9}$
Donc $X_{n+1}/[X_n = 2] \hookrightarrow \mathcal{B}\left(1, \frac{4}{9}\right)$

d) $X_{n+1}/[X_n = 1] \hookrightarrow \mathcal{B}\left(2, \frac{1}{3}\right)$ donc $E(X_{n+1}/X_n = 1) = 2\frac{1}{3} = \frac{2}{3}$
$X_{n+1}/[X_n = 2] \hookrightarrow \mathcal{B}\left(1, \frac{4}{9}\right)$ donc $E(X_{n+1}/X_n = 2) = \frac{4}{9}$

4. On suppose, uniquement dans cette question, que X_0 suit la loi binomiale de paramètres $\left(3, \frac{1}{3}\right)$

a) $X_1(\Omega) = [[0, 3]]$ et avec le système complet d'événements $(X_0 = i)_{i \in [[0,3]]}$

$$\begin{aligned}
\mathrm{P}(X_1 = 0) &= \mathrm{P}_{X_0=0}(X_1 = 0)\,\mathrm{P}(X_0 = 0) + \mathrm{P}_{X_0=1}(X_1 = 0)\,\mathrm{P}(X_0 = 1) \\
&\quad + \mathrm{P}_{X_0=2}(X_1 = 0)\,\mathrm{P}(X_0 = 2) + \mathrm{P}_{X_0=3}(X_1 = 0)\,\mathrm{P}(X_0 = 3) \\
&= 1\binom{3}{0}\left(\frac{2}{3}\right)^3 + \binom{2}{0}\left(\frac{2}{3}\right)^2\binom{3}{1}\frac{1}{3}\left(\frac{2}{3}\right)^2 + \frac{4}{9}\binom{3}{2}\left(\frac{1}{3}\right)^2\frac{2}{3} + 1\left(\frac{1}{3}\right)^3 \\
&= \frac{8}{27} + \frac{4}{9}\frac{4}{9} + \frac{4}{9}\cdot\frac{2}{9} + \frac{1}{27} = \frac{17}{27} \\
\mathrm{P}(X_1 = 1) &= \mathrm{P}_{X_0=0}(X_1 = 1)\left(\frac{2}{3}\right)^3 + \mathrm{P}_{X_0=1}(X_1 = 1)\left(\frac{2}{3}\right)^2 \\
&\quad + \mathrm{P}_{X_0=2}(X_1 = 1)\frac{2}{9} + \mathrm{P}_{X_0=3}(X_1 = 1)\frac{1}{27} \\
&= 0 + 2\frac{2}{3}\frac{1}{3}\frac{4}{9} + \frac{5}{9}\frac{2}{9} + 0 \\
&= \frac{16}{81} + \frac{10}{81} = \frac{26}{81} \\
\mathrm{P}(X_1 = 2) &= 0 + \left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^2 + 0 + 0 = \frac{4}{81} \\
\mathrm{P}(X_1 = 3) &= 0
\end{aligned}$$

$(X_1 = 3)$ n'est possible que s'il y a au moins 3 individus non infectées et une de plus pour infecter ces trois là. Donc impossible ici.

$$E(X_1) = 0 + \frac{26}{81} + \frac{8}{81} = \frac{32}{81}$$

b) On a

$$\begin{aligned}\sum_{i=0}^{3} E(X_1/X_0 = i)\,\mathrm{P}([X_0 = i]) &= 0\mathrm{P}(X_0 = 0) + \frac{2}{3}\mathrm{P}(X_0 = 1) + \frac{4}{9}\mathrm{P}(X_0 = 2) + 0\mathrm{P}(X_0 = 3) \\ &= \frac{2}{3}3\frac{1}{3}\left(\frac{2}{3}\right)^2 + \frac{4}{9}3\left(\frac{1}{3}\right)^2\frac{2}{3} = \frac{8}{27} + \frac{8}{81} = \frac{32}{81} = E(X_1)\end{aligned}$$

Conclusion : $\boxed{E(X_1) = \sum_{i=0}^{3} E(X_1/X_0 = i)\,\mathrm{P}([X_0 = i])}$

5. Pour tout entier naturel n, on considère le vecteur U_n de $\mathbb{R}^4$ défini par :

$$U_n = \begin{pmatrix} u_n \\ v_n \\ w_n \\ t_n \end{pmatrix} = \begin{pmatrix} \mathrm{P}([X_n = 0]) \\ \mathrm{P}([X_n = 1]) \\ \mathrm{P}([X_n = 2]) \\ \mathrm{P}([X_n = 3]) \end{pmatrix}$$

a) On a $\mathrm{P}(X_n = 0) + \mathrm{P}(X_n = 1) + \mathrm{P}(X_n = 2) + \mathrm{P}(X_n = 3) = 1$ (système complet d'événements) donc
$u_n + v_n + w_n + t_n = 1$

b) Avec le système complet d'événements $([X_n = i])_{0 \le i \le 3}$ on a :

$$\begin{aligned}\mathrm{P}(X_{n+1} = 0) &= 1\mathrm{P}(X_n = 0) + \frac{4}{9}\mathrm{P}(X_0 = 1) + \frac{4}{9}\mathrm{P}(X_0 = 2) + 1\mathrm{P}(X_0 = 3) \\ \mathrm{P}(X_{n+1} = 1) &= 0\mathrm{P}(X_n = 0) + \frac{4}{9}\mathrm{P}(X_0 = 1) + \frac{4}{9}\mathrm{P}(X_0 = 2) + 0\mathrm{P}(X_0 = 3) \\ \mathrm{P}(X_{n+1} = 2) &= 0\mathrm{P}(X_n = 0) + \frac{1}{9}\mathrm{P}(X_0 = 1) + 0\mathrm{P}(X_0 = 2) + 0\mathrm{P}(X_0 = 3) \\ \mathrm{P}(X_{n+1} = 3) &= 0\mathrm{P}(X_n = 0) + 0\mathrm{P}(X_0 = 1) + 0\mathrm{P}(X_0 = 2) + 0\mathrm{P}(X_0 = 3)\end{aligned}$$

Donc $U_{n+1} = \frac{1}{9}S\ U_n$

Conclusion : $\boxed{U_{n+1} = \frac{1}{9}S\ U_n}$

c) *Conclusion :* $\boxed{M = \frac{1}{9}S = \frac{1}{9}RDR^{-1}}$ donc les valeurs propres de M sont celles de $\frac{1}{9}D$ soit $-\frac{1}{9}, 0, \frac{5}{9}$ et 1.

d) $$M^n = \left(\frac{1}{9}\right)^n \begin{pmatrix} 9^n & 9^n - 5^n & 9^n - 5^n & 9^n \\ 0 & \frac{1}{6}(-1)^n + \frac{5}{6}5^n & \frac{5}{6}5^n - \frac{5}{6}(-1)^n & 0 \\ 0 & \frac{1}{6}5^n - \frac{1}{6}(-1)^n & \frac{5}{6}(-1)^n + \frac{1}{6}5^n & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Donc $u_n = u_0 + \left(1 - \left(\frac{5}{9}\right)^n\right) v_0 + \left(1 - \left(\frac{5}{9}\right)^n\right) w_0 + t_0$ avec $t_0 + u_0 = 1 - v_0 - w_0$

Conclusion : $\boxed{u_n = 1 - \left(\frac{5}{9}\right)^n (v_0 + w_0)}$

Conclusion : $\boxed{v_n = \left(\frac{1}{6}\left(\frac{-1}{9}\right)^n + \frac{5}{6}\left(\frac{5}{9}\right)^n\right) v_0 + \left(\frac{5}{6}\left(\frac{5}{9}\right)^n - \frac{1}{6}\left(\frac{-1}{9}\right)^n\right) w_0}$

6. On pose : $F = \bigcup_{n=0}^{+\infty} [X_n = 0]$

   a) F signifie que $X_n = 0$ au moins une fois (il existe une date à laquelle aucun individus n'est plus contagieux.) donc que le virus a disparu.

   b) F n'est pas une réunion d'incompatibles....
   Pour tout entier n : $\mathrm{P}(F) \geq \mathrm{P}(X_n = 0) = u_n \to 1$ donc le virus finit par disparaître presque sûrement, quelle que soit la loi de la variable aléatoire initiale $X_{0.}$

Partie II. Le cas général

On suppose que pour tout entier naturel n et pour tout entier i de $[[0, N]]$, on a : $\mathrm{P}(X_n = i) > 0$. On suppose également que pour tout couple (i, j) de $[[0, N]]^2$, le réel $q_{i,j}$ défini par : $q_{i,j} = \mathrm{P}_{[X_n=i]}([X_{n+1} = j])$ est indépendant de n.
Soit Q la matrice de $\mathcal{M}_{N+1}(\mathbb{R})$ définie par : $Q = (q_{i,j})_{0 \leq i,j \leq N}$

1. a) Pour tout j de $[[0, N]]$: $q_{0,j} = \mathrm{P}_{[X_n=0]}[X_{n+1} = j]$
   Quand $X_n = 0$, il n'y a plus de virus et $X_{n+1} = 0$ donc
   Conclusion : $\boxed{q_{0,j} = 0 \text{ si } j \neq 0 \text{ et } q_{0,0} = 1}$
   $q_{N,j} = \mathrm{P}_{[X_n=N]}[X_{n+1} = j]$
   Quand $[X_n = N]$ tous les individus sont guéris le jour $n + 1$.
   Conclusion : $\boxed{q_{N,j} = 0 \text{ si } j \neq 0 \text{ et } q_{N,0} = 1}$.
   Si $[X_n = i]$, il y a au moins i individus guéris à $n + 1$ donc $X_{n+1} \leq N - i < N$ si $i > 0$ et si $X_n = 0$ alors $X_{n+1} = 0$.
   Dans tous les cas, $X_{n+1} = N$ est impossible.
   Conclusion : $\boxed{q_{i,N} = 0 \text{ pour tout } i \text{ de } [[0, N]]}$

   b) Si $[X_n = i]$, il y a au moins i individus guéris à $n + 1$ donc $X_{n+1} \leq N - i$ et $[X_{n+1} = j]$ est impossible si $j > N - i$
   Conclusion : $\boxed{\text{si } j > N - i, \text{ alors } q_{i,j} = 0.}$

   c) Pour tout i de $[[1, N - 1]]$, quand $[X_n = i]$, pour chaque individu, la probabilité d'être infecté par chacun des malade est p.
   la probabilité de n'être infecté par aucun des malades est donc $(1 - p)^i = q^i$ et la probabilité d'être infecté est donc $(1 - q^i)$
   les $N - i$ individus sains étant infectés indépendamment
   Conclusion : $\boxed{X_{n+1}/[X_n = i] \hookrightarrow \mathcal{B}(N - i\ ,\ 1 - q^i)}$

2. a) En s'inspirant de la partie I, la première colonne de Q est $(q_{0,j})_{0 \leq j \leq N} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ donc

   $$Q \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$ et 1 est valeur propre de Q associé à $(1, 0, \ldots, 0)$

   b) Soit λ une valeur propre de Q, et $V = \begin{pmatrix} V(0) \\ V(1) \\ \vdots \\ V(N) \end{pmatrix}$ un vecteur propre associé à λ.

   On pose : $|V(i)| = \max\limits_{0 \leq j \leq N} |V(j)|$.

Comme les $|V(j)|$ sont positifs ou nuls et non tous nul (un vecteur propre n'est pas nul) alors la composante $V(i)$ n'est pas nulle.
La ligne i du système $QV = \lambda V$ est $\sum_{j=1}^{N} q_{i,j} V(j) = \lambda V(i)$
Donc $|\lambda V(i)| = \left|\sum_{j=1}^{N} q_{i,j} V(j)\right| \leq \sum_{j=1}^{N} |q_{i,j} V(j)|$ les $q_{i,j}$ étant positifs :

$$|\lambda| |V(i)| \leq \sum_{j=1}^{N} q_{i,j} |V(j)| \leq \sum_{j=1}^{N} q_{i,j} |V(i)| = |V(i)| \sum_{j=1}^{N} q_{i,j}$$

et $\sum_{j=1}^{N} q_{i,j} = 1$ ($(X_{n+1} = j)$ est un SCE avec la probabilité conditionnelle sachant $(X_n = i)$).
Conclusion : $\boxed{|\lambda| \leq 1}$

3. On pose pour tout entier naturel n : $U_n = \begin{pmatrix} \mathrm{P}([X_n = 0]) \\ \mathrm{P}([X_n = 1]) \\ \vdots \\ \mathrm{P}([X_n = N]) \end{pmatrix}$

   $(X_n = j)_{j \in [0,N]}$ est un système complet d'événements donc

$$\begin{aligned} \mathrm{P}(X_{n+1} = i) &= \sum_{j=0}^{N} \mathrm{P}_{[X_n = i]}(X_{n+1} = j)\, \mathrm{P}(X_n = j) \\ &= \sum_{j=0}^{N} q_{i,j} \mathrm{P}(X_n = j) \end{aligned}$$

   qui est la $i + 1^{ème}$ ligne de QU_n (en partant de la ligne 1 pour $i = 0$)
   Conclusion : $\boxed{U_{n+1} = QU_n \text{ et } M = Q}$
   On suppose jusqu'à la fin de la partie II que la matrice M est diagonalisable, et que $\mathcal{B} = (V_0, \ldots, V_N)$ est une base de vecteurs propres de M telle que, pour tout k de $[[0, N]]$ le vecteur propre V_k est associé à une valeur propre λ_k

   De plus, on suppose que : $\lambda_0 = 1$, $V_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ (vecteur propre que l'on avait donné.)

   et que pour tout k de $[[0, N]]$, on a : $|\lambda_k| < 1$ (on avait déjà $|\lambda_k| \leq 1$)

4. On décompose alors le vecteur U_0 sur la base $\mathcal{B}$: $U_0 = \displaystyle\sum_{k=0}^{N} \alpha_k V_k$

   a) On a $U_n = M^n U_0 = \sum_{k=0}^{N} \alpha_k M^n V_k$
   Et comme $MV_k = \lambda_k V_k$ on a alors $M^n V_k = (\lambda_k)^n V_k$ et
   Conclusion : $\boxed{U_n = \sum_{k=0}^{N} \alpha_k (\lambda_k)^n V_k}$

   b) On note, pour tout couple (k, i) de $[[0, N]]^2$ $V_k(i)$ la $(i + 1)$-ième composante du vecteur V_k.
   $\mathrm{P}[X_n = i]$ est la $i^{ème}$ composante de $U_n = \sum_{k=0}^{N} \alpha_k (\lambda_k)^n V_k$ donc somme des $i^{èmes}$ composantes de V_k
   Conclusion : $\boxed{\mathrm{P}[X_n = i] = \sum_{k=0}^{N} \alpha_k (\lambda_k)^n V_k(i)}$

c) Or, $(\lambda_k)^n \to 0$ si $k \neq 0$ car $\lambda_k < 1$, et $(\lambda_0)^n = 1$ donc $\mathrm{P}[X_n = i] = \sum_{k=0}^{N} \alpha_k (\lambda_k)^n V_k(i) \to \alpha_0 V_0(i)$.

Or $V_0(i) = 0$ si $i \neq 0$ donc

Conclusion : $\boxed{\text{pour tout } i \text{ de } [[1,N]], \text{ on a : } \lim_{n\to+\infty} \mathrm{P}([X_n = i]) = 0.}$

d) Donc $\mathrm{P}(X_n > 0) \to 0$ et la probabilité que le virus disparaisse ($(X_n = 0) = \overline{(X_n > 0)}$) tend vers 1.

Et comme $\mathrm{P}(F) \geq \mathrm{P}(X_n = 0)$ alors$\mathrm{P}(F) = 1$

Conclusion : $\boxed{\begin{array}{l}\text{le virus finit par disparaître presque sûrement,} \\ \text{quelle que soit la loi de la variable aléatoire initiale } X_0\end{array}}$.

Partie III. Estimations ponctuelle et par intervalle de confiance de p

On suppose que le paramètre p, qui exprime la probabilité qu'un individu contagieux transmette le virus à un individu sain, est inconnu, et on cherche à l'estimer.

On rappelle que : $q = 1 - p$.

Pour m entier supérieur ou égal à m, on considère un m-échantillon $(Y_1, Y_2, \ldots, Y_m)$ de variables aléatoires indépendantes, de même loi de Bernoulli de paramètre p, définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathrm{P})$

On pose : $\overline{Y_m} = \frac{1}{m}\sum_{i=1}^{m} Y_i$.

Dans toute la suite de cette partie, on note ε un réel strictement positif quelconque.

1. a) $E\left(\overline{Y_m}\right) = \frac{1}{m}\sum_{i=1}^{m} E(Y_i) = \frac{mp}{m} = p$

Donc le biais de $\overline{Y_m}$ comme estimateur de p est $E\left(\overline{Y_m}\right) - p = 0$

Sa variance est : $V\left(\frac{1}{m}\sum_{i=1}^{m} Y_i\right) = \frac{1}{m^2}\sum_{i=1}^{m} V(Y_i)$ car les Y_i sont indépendantes. et $V\left(\overline{Y_m}\right) = \frac{mpq}{m^2} = \frac{pq}{m}$

Conclusion : $\boxed{\text{Donc son risque quadratique est : } \frac{pq}{m} + 0^2 = \frac{pq}{m}}$

b) $p \in \left[\overline{Y_m} - \sqrt{\frac{5}{m}}, \overline{Y_m} + \sqrt{\frac{5}{m}}\right] \iff \left|\overline{Y_m} - p\right| \leq \sqrt{\frac{5}{m}}$

Donc (Bienaymé-Tchebycheff)

$$\mathrm{P}\left(p \in \left[\overline{Y_m} - \sqrt{\frac{5}{m}}, \overline{Y_m} + \sqrt{\frac{5}{m}}\right]\right) \geq 1 - \frac{V\left(\overline{Y_m}\right)}{\sqrt{\frac{5}{m}}^2} = \frac{pq}{5}$$

Or $p \to p(1-p)$ est maximale en $p = \frac{1}{2}$ et vaut $\frac{1}{4}$ en ce point donc

$$\mathrm{P}\left(p \in \left[\overline{Y_m} - \sqrt{\frac{5}{m}}, \overline{Y_m} + \sqrt{\frac{5}{m}}\right]\right) \geq 1 - \frac{1}{20} = 0,95$$

Conclusion : $\boxed{\begin{array}{l}\left[\overline{Y_m} - \sqrt{\frac{5}{m}}, \overline{Y_m} + \sqrt{\frac{5}{m}}\right] \text{ est un intervalle de confiance de } p \text{ .} \\ \text{au niveau de confiance } 0,95\end{array}}$.

2. Soit θ un réel positif.

a) $\left[\overline{Y_m} - p \geq \varepsilon\right] = \left[\overline{Y_m} \geq \varepsilon + p\right] = \left[m\theta\overline{Y_m} \geq m\theta(\varepsilon + p)\right] = \left[\exp\left(m\theta\overline{Y_m}\right) \geq \exp(m\theta(p + \varepsilon))\right]$

Conclusion : $\boxed{\mathrm{P}\left(\left[\overline{Y_m} - p \geq \varepsilon\right]\right) = \mathrm{P}\left(\left[\exp\left(m\theta\overline{Y_m}\right) \geq \exp(m\theta(p+\varepsilon))\right]\right)}$

b) L'espérance est alors

$$\begin{aligned} E(T) &= \sum_{k\in T(\Omega)} k\mathrm{P}(T=k) \\ &= \sum_{k\in T(\Omega):k<a} k\mathrm{P}(T=k) + \sum_{k\in T(\Omega):k\geq a} k\mathrm{P}(T=k) \text{ avec} \\ \sum_{k\in T(\Omega):k\geq a} k\mathrm{P}(T=k) &\geq \sum_{k\in T(\Omega):k\geq a} a\mathrm{P}(T=k) = a\mathrm{P}(T\geq a) \end{aligned}$$

Conclusion : $\boxed{\mathrm{P}([T\geq a]) \leq \dfrac{E(T)}{a} \text{ pour } a>0}$

c) Soit g la fonction définie sur $\mathbb{R}^+$ par : $g(x) = \ln(p\ e^x + q)$.
De la question précédente on déduit :

$$\begin{aligned} \mathrm{P}\left(\left[\overline{Y_m} - p \geq \varepsilon\right]\right) &= \mathrm{P}\left(\left[\exp\left(m\theta\overline{Y_m}\right) \geq \exp\left(m\theta(p+\varepsilon)\right)\right]\right) \\ &\leq \frac{E\left(\exp\left(m\theta\overline{Y_m}\right)\right)}{\exp\left(m\theta(p+\varepsilon)\right)} \text{ car } \exp\left(m\theta(p+\varepsilon)\right) > 0 \end{aligned}$$

Reste à calculer

$$\begin{aligned} \exp\left(m\theta\overline{Y_m}\right) &= \exp\left(\theta\sum_{k=1}^m Y_k\right) = \prod_{k=1}^m \exp(\theta Y_k) \text{ indépendantes donc} \\ E\left(\exp\left(m\theta\overline{Y_m}\right)\right) &= \prod_{k=1}^m E(\exp(\theta Y_k)) \end{aligned}$$

et $E(\exp(\theta Y_k)) = e^0 q + e^\theta p = q + e^\theta p$ donc

$$\begin{aligned} \prod_{k=1}^m E(\exp(\theta Y_k)) &= \left(q+e^\theta p\right)^m \\ &= \exp\left(m\ln\left(q+e^\theta p\right)\right) \\ &= \exp(mq(\theta)) \end{aligned}$$

Finalement,

$$\mathrm{P}\left(\left[\overline{Y_m} - p \geq \varepsilon\right]\right) \leq \frac{\exp(mq(\theta))}{\exp(m\theta(p+\varepsilon))} = \exp\left(m\left[g(\theta) - \theta(p+\varepsilon)\right]\right)$$

Conclusion : $\boxed{\mathrm{P}\left(\left[\overline{Y_m} - p \geq \varepsilon\right]\right) \leq \exp\left(m\left[g(\theta) - \theta(p+\varepsilon)\right]\right)}$

d) Sur $\mathbb{R}^+$ $p\ e^x + q > 0$ donc g est de classe C^2 sur $\mathbb{R}^+$ comme composée de fonctions C^2

$$\begin{aligned}
g'(x) &= \frac{1}{pe^x+q}pe^x \\
g''(x) &= p\frac{e^x(pe^x+q)-pe^xe^x}{(pe^x+q)^2} \\
&= \frac{pqe^x}{(pe^x+q)^2} \\
g'''(x) &= pq\frac{e^x(pe^x+q)-2pe^xe^x}{(pe^x+q)^3} \\
&= pq\frac{e^xq-pe^xe^x}{(pe^x+q)^3} \\
&= pqe^x\frac{q-pe^x}{(pe^x+q)^3}
\end{aligned}$$

nulle en $e^x = q/p$ où

$$\begin{aligned}
g''(\ln(q/p)) &= \frac{pq\frac{q}{p}}{\left(p\frac{q}{p}+q\right)^2} \\
&= \frac{p^2q^2}{(pq+qp)^2} = \frac{1}{4}
\end{aligned}$$

x	0	$\ln(q/p)$	$+\infty$
$q-pe^x$	$\searrow +$	0	$\searrow -$
$g'''(x)$	$+$		$-$
$g''(x)$	$\nearrow$	$\frac{1}{4}$	$\searrow$

Conclusion : $\boxed{\text{pour tout } x \text{ de } \mathbb{R}^+, : |g''(x)| \leq \frac{1}{4}}$

e) On étudie la différence :$h(\theta) = g(\theta) - \theta p - \dfrac{\theta^2}{8}$ qui est C^2 sur $\mathbb{R}^+$

$$\begin{aligned}
h'(\theta) &= g'(\theta) - p - \frac{\theta}{4} \\
h''(\theta) &= g''(\theta) - \frac{\theta}{4} \leq 0
\end{aligned}$$

avec $g'(0) = \frac{p}{p+q} = p$ donc $h'(0) = 0$ et h' décroissante donc $h' \leq 0$
$h(0) = g(0) = \ln(p+q) = 0$ et h décroissante donc

Conclusion : $\boxed{g(\theta) \leq \theta p + \frac{\theta^2}{8}}$

f) $h(x) = \frac{x^2}{8} - \varepsilon x$.
h est dérivable sur $\mathbb{R}^+$ et $h'(x) = \frac{x}{4} - \varepsilon$
En $+\infty$: $h(x) = \frac{x^2}{8} - \varepsilon x = x^2\left(\frac{1}{8} - \frac{\varepsilon}{x}\right) \to +\infty$

x	0	4ε	$+\infty$
$h'(x)$	$-$	$+$	affine
$h(x)$	$0 \searrow$	$-2\varepsilon^2 \nearrow$	$+\infty$

Conclusion : $\boxed{\text{Et on a donc } h(x) = \frac{x^2}{8} - \varepsilon x \geq -2\varepsilon^2 \text{ pour tout } x \geq 0.}$
On met les inégalités bout à bout :

$$
\begin{aligned}
\mathrm{P}\left(\left[\overline{Y_m}-p\geq\varepsilon\right]\right) &\leq \exp\left(m\left[g\left(\theta\right)-\theta\left(p+\varepsilon\right)\right]\right)\\
&\leq \exp\left(m\left[\theta p+\frac{1}{8}\theta^2-\theta\left(p+\varepsilon\right)\right]\right)=\exp\left(m\left[\frac{1}{8}\theta^2-\theta\varepsilon\right]\right)
\end{aligned}
$$

L'inégalité $h\left(x\right)\geq-2\varepsilon^2$. n'est pas ici dans le bon sens.

Mais l'inégalité précédente est vraie pour tout $\theta\geq0$, donc en particulier pour $\theta=4\varepsilon$ où l'on a $\exp\left(m\left[\frac{1}{8}\theta^2-\theta\varepsilon\right]\right)=\exp\left(-2m\varepsilon^2\right)$ et donc

Conclusion : $\boxed{\mathrm{P}\left(\left[\overline{Y_m}-p\geq\varepsilon\right]\right)\leq\exp\left(-2m\varepsilon^2\right)}$

3. On pose : $\overline{W_m}=\frac{1}{m}\sum_{i=1}^m\left(1-Y_i\right)$
   $1-Y_i$ est une variable de Bernouilli de paramètre q
   Donc on retrouve la propriété précédente en substituant p à q et q à p
   Conclusion : $\boxed{\mathrm{P}\left(\left[\overline{W_m}-q\geq\varepsilon\right]\right)\leq\exp\left(-2m\varepsilon^2\right)}$

4. a) On retravaille $\overline{W_m}=1-\dfrac{1}{m}\displaystyle\sum_{i=1}^m Y_i=1-\overline{Y_m}$

Donc

$$
\begin{aligned}
\left[\overline{W_m}-q\geq\varepsilon\right] &= \left[1-\overline{Y_m}-q\geq\varepsilon\right]\\
&= \left[p-\overline{Y_m}\geq\varepsilon\right]\\
&= \left[\overline{Y_m}-p\leq\varepsilon\right]
\end{aligned}
$$

et comme

$$\left[\left|\overline{Y_m}-p\right|\geq\varepsilon\right]=\left[\overline{Y_m-p}\leq\varepsilon\right]\cup\left[\overline{Y_m}-p\geq\varepsilon\right]$$

réunion d'incompatibles

$$
\begin{aligned}
\mathrm{P}\left[\left|\overline{Y_m}-p\right|\geq\varepsilon\right] &= \mathrm{P}\left[\overline{Y_m}-p\leq\varepsilon\right]+\mathrm{P}\left[\overline{Y_m}-p\geq\varepsilon\right]\\
&= \mathrm{P}\left[\overline{W_m}-q\geq\varepsilon\right]+\mathrm{P}\left[\overline{Y_m}-p\geq\varepsilon\right]
\end{aligned}
$$

Conclusion : $\boxed{\mathrm{P}\left(\left|\overline{Y_m}-p\right|\geq\varepsilon\right)\leq 2\ \exp\left(-2m\varepsilon^2\right)}$

b) Sachant que $\ln\left(0.025\right)\approx-3.688$, pour $\varepsilon=\sqrt{\frac{1.844}{m}}$ on a

$$
\begin{aligned}
2e^{-2m\varepsilon^2} &= 2e^{-2m\frac{1.844}{m}}\\
&= 2e^{-3.688}\\
&\approx 2\cdot0.025=0.05
\end{aligned}
$$

Donc avec $\varepsilon=\sqrt{\dfrac{1.844}{m}}$, $\left[\overline{Y_m}-\sqrt{\dfrac{1.844}{m}},\overline{Y_m}+\sqrt{\dfrac{1.844}{m}}\right]$ est un intervalle de confiance au niveau de risque 0.05 ou de confiance $0,95$

Précédemment on avait $\left[\overline{Y_m}-\sqrt{\dfrac{5}{m}},\overline{Y_m}+\sqrt{\dfrac{5}{m}}\right]$ pour la même confiance.

un intervalle $\sqrt{5/1.844}=1.6$ fois plus petit. (un petit mieux pour un gros travail !)