Correction du devoir surveillé n° 6

Problème 1

Partie A : Une matrice diagonalisable

1. La famille d'événements (A_{2n}, B_{2n}) est un système complet d'événements. D'après la formule des probabilités totales :
$$u_{n+1} = P(A_{2n+2}) = P(A_{2n}) \times P_{A_{2n}}(A_{2n+2}) + P(B_{2n}) \times P_{B_{2n}}(A_{2n+2}).$$
D'après l'énoncé $P_{A_{2n}}(A_{2n+2}) = p$ et $P_{B_{2n}}(A_{2n+2}) = q$, donc $u_{n+1} = pu_n + qv_n$.

2. a) $C_0 = \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ car initialement le virus est sur le serveur A.
De même que dans la question précédente, on peut montrer que $v_{n+1} = qu_n + pv_n$. On a donc $C_{n+1} = MC_n$ avec $M = \begin{pmatrix} p & q \\ q & p \end{pmatrix}$.

   b) La matrice M est symétrique à coefficients réels donc elle est diagonalisable.
Déterminons les valeurs propres de M. On cherche donc les réels λ tels que $M - \lambda I_2$ n'est pas inversible, c'est-à-dire tels que $\det(M - \lambda I_2) = 0$. Or on a :
$$\det(M - \lambda I_2) = (p - \lambda)^2 - q^2 = (p + q - \lambda)(p - q - \lambda) = (1 - \lambda)(p - q - \lambda).$$
Les valeurs propres de M sont donc 1 et $p - q = 2p - 1$.
Comme M est une matrice de $\mathscr{M}_2(\mathbb{R})$ admettant deux valeurs propres distinctes, ses sous-espaces propres sont des sous-espaces vectoriels de $\mathscr{M}_{2,1}(\mathbb{R})$ chacun de dimension 1. Il suffit donc de trouver un vecteur propres non nul pour chaque valeur propre pour avoir une base de chaque sous-espace propre.
   - On peut voir que $M\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, donc $E_1(M) = \text{vect}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$.
   - On a aussi $M\begin{pmatrix} 1 \\ -1 \end{pmatrix} = (p - q)\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, donc $E_{p-q}(M) = \text{vect}\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right)$.

   $\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$ est une base de vecteurs propres de M, donc on sait que l'on a
$$M = PDP^{-1}, \text{ avec } D = \begin{pmatrix} p - q & 0 \\ 0 & 1 \end{pmatrix} \text{ et } P = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

   c) $P \times P^T = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = 2I_2$.
On a donc $P \times \frac{1}{2}P^T = \frac{1}{2}P^T \times P = I_2$ et ainsi $P^{-1} = \frac{1}{2}P^T = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.

3. a) Montrons par récurrence que la propriété $\mathscr{P}(n)$: « $C_n = PD^nP^{-1}C_0$ » est vraie pour tout $n \in \mathbb{N}$.
   - Comme $D^0 = I_2$, on a bien $PD^0P^{-1}C_0 = C_0$.
   - Soit $n \in \mathbb{N}$ fixé. Supposons que $\mathscr{P}(n)$ est vraie.
D'après la question 2.a :
$$C_{n+1} = MC_n = PDP^{-1} \times PD^nP^{-1}C_0 = PD^{n+1}P^{-1}C_0.$$
Donc $\mathscr{P}(n+1)$ est vérifiée.

   Grâce au principe de récurrence on a montré que pour tout $n \in \mathbb{N}$, $C_n = PD^nP^{-1}C_0$.
En calculant le produit matriciel on obtient que $\begin{cases} u_n = \frac{1}{2}(1 + (p - q)^n) \\ v_n = \frac{1}{2}(1 - (p - q)^n) \end{cases}$.

   b) Comme $p \in]0; 1[$ on a $-q \in] - 1; 0[$ et donc $p - q \in] - 1; 1[$. Ainsi, $\lim\limits_{n \to +\infty}(p - q)^n = 0$.
Les suites (u_n) et (v_n) sont donc convergente et $\lim\limits_{n \to +\infty} u_n = \lim\limits_{n \to +\infty} v_n = \frac{1}{2}$.

4. Si le virus est initialement en B, on a $C_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ et donc $\begin{cases} u_n = \dfrac{1}{2}(1-(p-q)^n) \\ v_n = \dfrac{1}{2}(1+(p-q)^n) \end{cases}$.

   Les limites sont alors toujours de $\dfrac{1}{2}$.

Partie B : Moyenne et variance empiriques

5. X_i est une variable finie donc elle admet une espérance et une variance. De plus

$$E(X_i) = -P(X_i = -1) + P(X_i = 1) = -v_i + u_i = (p-q)^i.$$

   On a aussi $E(X_i^2) = P(X_i = -1) + P(X_i = 1) = 1$, donc, d'après la formule de Kœnig-Huygens :

$$V(X_i) = E(X_i^2) - E(X_i)^2 = 1 - (p-q)^{2i}.$$

6. a) Sachant que le virus est sur le serveur A au bout de $2i$ semaines, calculer la probabilité qu'il se retrouve sur le serveur A au bout de $2j$ semaines revient au même que de calculer la probabilité que le virus soit sur le serveur A initialement puis qu'il se retrouve sur le serveur A au bout de $2(j-i)$ semaines car dans le premier cas, les mouvements effectués avant la semaine $2i$ n'ont pas d'influence sur les mouvements suivants.
   On a donc bien $P_{[X_i=1]}(X_j = 1) = P(X_{j-i} = 1)$.
   b) $P([X_i = 1] \cap [X_j = 1]) = P(X_i = 1) \times P_{[X_i=1]}(X_j = 1) = u_i \times P(X_{j-i} = 1) = u_i \times u_{j-i}$.
   Donc $P([X_i = 1] \cap [X_j = 1]) = \dfrac{(1+(p-q)^i)(1+(p-q)^{j-i})}{4}$.
   c) $P([X_i = 1] \cap [X_j = -1]) = P(X_i = 1) \times P_{[X_i=1]}(X_j = -1) = u_i \times P(X_{j-i} = -1) = u_i \times v_{j-i}$.
   Donc $P([X_i = 1] \cap [X_j = -1]) = \dfrac{(1+(p-q)^i)(1-(p-q)^{j-i})}{4}$.
   Pour $P_{[X_i=-1]}(X_j = ...)$ on utilise les résultats de la question A.4. car c'est comme si le virus était initialement en B.
   Donc $P([X_i = -1] \cap [X_j = 1]) = \dfrac{(1-(p-q)^i)(1-(p-q)^{j-i})}{4}$ et $P([X_i = -1] \cap [X_j = -1]) = \dfrac{(1-(p-q)^i)(1+(p-q)^{j-i})}{4}$.
   d)

$$\begin{aligned} P(X_iX_j = 1) &= P([X_i = 1] \cap [X_j = 1]) + P([X_i = -1] \cap [X_j = -1]) \\ &= \frac{1+(p-q)^{j-i}}{2} = P(X_{j-i} = 1) \end{aligned}$$

   X_iX_j suit donc la même loi que X_{j-i}. On en déduit donc que $E(X_iX_j) = E(X_{j-i}) = (p-q)^{j-i}$.

7. a) Par linéarité de l'espérance :

$$E(M_n) = \frac{1}{n+1}\sum_{i=0}^{n} E(X_i) = \frac{1}{n+1}\sum_{i=0}^{n}(p-q)^i = \frac{1-(p-q)^{n+1}}{2q(n+1)},$$

   car $p - q = 2p - 1 \neq 1$ car $p \neq 1$.
   b) *La récurrence n'est pas indispensable ici mais comme l'énoncé le demande, on s'y colle !*
   Montrons par récurrence que la propriété $\mathscr{P}(n)$: « $E(M_n^2) = \dfrac{1}{n+1} + \dfrac{2}{(n+1)^2}\displaystyle\sum_{0\leqslant i<j\leqslant n} E(X_iX_j)$ »est vraie pour tout $n \in \mathbb{N}$.
   — Pour $n = 0$, on a $M_0 = X_0$ donc

$$E(M_0^2) = E(X_0^2) = V(X_0) + E(X_0)^2 = 1 - (p-q)^0 + \big((p-q)^0\big)^2 = 1.$$

   De plus $\dfrac{1}{0+1} + \dfrac{2}{(0+1)^2}\displaystyle\sum_{0\leqslant i<j\leqslant 0} E(X_iX_j) = 1$ (car la somme est vide).
   Donc $\mathscr{P}(0)$ est bien vérifiée.

— Soit $n \in \mathbb{N}$ fixé tel que $\mathscr{P}(n)$ est vérifiée.

On a alors $M_{n+1} = \dfrac{1}{n+2}(X_0 + \ldots + X_{n+1}) = \dfrac{n+1}{n+2} M_n + \dfrac{1}{n+2} X_{n+1}$.

On en déduit que :

$$\begin{aligned}
E(M_{n+1}^2) &= E\left(\frac{(n+1)^2}{(n+2)^2} M_n^2 + \frac{2(n+1)}{(n+2)^2} M_n X_{n+1} + \frac{1}{(n+2)^2} X_{n+1}^2\right) \\
&= \frac{(n+1)^2}{(n+2)^2} E(M_n^2) + \frac{2(n+1)}{(n+2)^2} E(M_n X_{n+1}) + \frac{1}{(n+2)^2} E(X_{n+1}^2) \quad \text{linéarité de l'espérance} \\
&= \frac{n+1}{(n+2)^2} + \frac{2}{(n+2)^2} \sum_{0 \leqslant i < j \leqslant n} E(X_i X_j) + \frac{2}{(n+2)^2} \sum_{i=0}^{n} E(X_i X_{n+1}) + \frac{1}{(n+2)^2} \\
&= \frac{1}{n+2} + \frac{2}{(n+2)^2} \sum_{0 \leqslant i < j \leqslant n+1} E(X_i X_j).
\end{aligned}$$

$\mathscr{P}(n+1)$ est donc bien vérifiée.

Grâce au principe de récurrence, on a montré que pour tout $n \in \mathbb{N}$, $E(M_n^2) = \dfrac{1}{n+1} + \dfrac{2}{(n+1)^2} \displaystyle\sum_{0 \leqslant i < j \leqslant n} E(X_i X_j)$

Voici ce que ça donne sans récurrence !

Pour tout $n \in \mathbb{N}$, $M_n^2 = \dfrac{1}{(n+1)^2} \displaystyle\sum_{i=0}^{n} X_i^2 + \dfrac{2}{(n+1)^2} \sum_{0 \leqslant i < j \leqslant n+1} X_i X_j$.

Par linéarité de l'espérance et d'après les résultats précédents, on a donc :

$$\begin{aligned}
E(M_n^2) &= \frac{1}{(n+1)^2} \sum_{i=0}^{n} E(X_i^2) + \frac{2}{(n+1)^2} \sum_{0 \leqslant i < j \leqslant n+1} E(X_i X_j) \\
&= \frac{1}{n+1} + \frac{2}{(n+1)^2} \sum_{0 \leqslant i < j \leqslant n} E(X_i X_j).
\end{aligned}$$

c) En reprenant la définition de la variance empirique du cours, on a $S_n^2 = \dfrac{1}{n+1} \displaystyle\sum_{i=0}^{n} X_i^2 - M_n^2$.

Par linéarité de l'espérance et d'après les questions précédentes :

$$\begin{aligned}
E(S_n^2) &= \frac{1}{n+1} \sum_{i=0}^{n} E(X_i^2) - E(M_n^2) \\
&= 1 - \frac{1}{n+1} - \frac{2}{(n+1)^2} \sum_{0 \leqslant i < j \leqslant n} E(X_i X_j) \\
&= 1 - \frac{1}{n+1} - \frac{2}{(n+1)^2} \sum_{0 \leqslant i < j \leqslant n} (p-q)^{j-i} \\
&= \frac{n}{n+1} - \frac{2}{(n+1)^2} \sum_{j=1}^{n} \sum_{i=0}^{j-1} (p-q)^{j-i} \\
&= \frac{n}{n+1} - \frac{2}{(n+1)^2} \sum_{j=1}^{n} (p-q)^j \frac{1 - \frac{1}{(p-q)^j}}{1 - \frac{1}{p-q}} \text{ car } p - q \neq 1 \\
&= \frac{n}{n+1} - \frac{2(p-q)}{(n+1)^2(p-q-1)} \sum_{j=1}^{n} (p-q)^j + \frac{2(p-q)n}{(n+1)^2(p-q-1)} \\
&= \frac{n}{n+1} - \frac{2(p-q)^2}{(n+1)^2(p-q-1)} \times \frac{1-(p-q)^n}{1-(p-q)} + \frac{2n(p-q)}{(n+1)^2(p-q-1)}
\end{aligned}$$

En remarquant que $p - q - 1 = -2q$, on obtient

$$E(S_n^2) = \frac{n}{n+1} + \frac{p-q}{(n+1)^2 q} \left[\frac{(p-q)(1-(p-q)^n)}{2q} - n\right].$$

Partie C : Équation matricielle

8. On a $D_{2n} = C_n$.
9. a) On sait que $C_{n+1} = MC_n$. On a donc $D_{2n+2} = MD_{2n}$.
   Mais, par hypothèse de l'énoncé, on a aussi $D_{2n+2} = N^2 D_{2n}$.
   On en déduit donc que, pour tout entier n, $MD_{2n} = N^2 D_{2n}$, ce qui donne$(N^2 - M)C_n = 0$.

   b) D'après la question précédente on a $(N^2 - M)C_0 = 0$ et $(N^2 - M)C_1 = 0$.
   Or $(C_0, C_1) = \left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}p\\q\end{pmatrix}\right)$ est une famille libre car $q \neq 0$ qui contient deux vecteurs donc c'est une base de $\mathscr{M}_{2,1}(\mathbb{R})$.
   On peut donc en déduire que, pour tout $X \in \mathscr{M}_{2,1}(\mathbb{R})$, $(N^2 - M)X = 0$ et donc $N^2 - M = 0$.

   c) On a $\Delta^2 = P^{-1}NPP^{-1}NP = P^{-1}N^2P = P^{-1}MP = D$.
   On écrit alors $\Delta = \begin{pmatrix}a & b\\ c & d\end{pmatrix}$, et l'égalité $\Delta^2 = D$ s'écrit alors :
$$\begin{cases} a^2 + bc = p - q \\ ab + bd = 0 \\ ac + dc = 0 \\ bc + d^2 = 1 \end{cases} \Leftrightarrow \begin{cases} a^2 + bc = p - q \\ b = 0 \text{ ou } a + d = 0 \\ c = 0 \text{ ou } a + d = 0 \\ bc + d^2 = 1 \end{cases}$$
   On remarque alors que si $a + d = 0$ alors $a^2 - d^2 = 0$ et donc en soustrayant la ligne 1 à la ligne 4, on a $a^2 - d^2 = p - q - 1 = 0$, d'où $2q = 0$ et ainsi $q = 0$ ce qui est contraire à notre hypothèse.
   On en déduit donc que $\Delta^2 = D$ implique $\begin{cases} a^2 = p - q \\ b = 0 \\ c = 0 \\ d^2 = 1 \end{cases}$. On obtient bien dans ces conditions une matrice Δ qui est diagonale.
   Comme l'énoncé impose l'existence de N et donc de Δ, cela impose que $p - q = a^2 \geqslant 0$.
   On en déduit alors les matrices Δ solutions de $\Delta^2 = D$:
$$\begin{pmatrix}\sqrt{p-q} & 0\\ 0 & 1\end{pmatrix}, \begin{pmatrix}-\sqrt{p-q} & 0\\ 0 & 1\end{pmatrix}, \begin{pmatrix}\sqrt{p-q} & 0\\ 0 & -1\end{pmatrix}, \begin{pmatrix}-\sqrt{p-q} & 0\\ 0 & -1\end{pmatrix}.$$

   d) Comme $N = P\Delta P^{-1}$, on en déduit 4 matrices possibles pour N :
$$\frac{1}{2}\begin{pmatrix}\sqrt{p-q}+1 & 1-\sqrt{p-q}\\ 1-\sqrt{p-q} & \sqrt{p-q}+1\end{pmatrix}, \frac{1}{2}\begin{pmatrix}-\sqrt{p-q}+1 & 1+\sqrt{p-q}\\ 1+\sqrt{p-q} & -\sqrt{p-q}+1\end{pmatrix},$$
$$\frac{1}{2}\begin{pmatrix}\sqrt{p-q}-1 & -1-\sqrt{p-q}\\ -1-\sqrt{p-q} & \sqrt{p-q}-1\end{pmatrix}, \frac{1}{2}\begin{pmatrix}-\sqrt{p-q}-1 & -1+\sqrt{p-q}\\ -1+\sqrt{p-q} & -\sqrt{p-q}-1\end{pmatrix}$$
   Comme on cherche à ce que N soit à coefficients positifs ou nuls, les deux seules matrices répondant au conditions demandées sont :
$$\frac{1}{2}\begin{pmatrix}\sqrt{p-q}+1 & 1-\sqrt{p-q}\\ 1-\sqrt{p-q} & \sqrt{p-q}+1\end{pmatrix}, \frac{1}{2}\begin{pmatrix}-\sqrt{p-q}+1 & 1+\sqrt{p-q}\\ 1+\sqrt{p-q} & -\sqrt{p-q}+1\end{pmatrix}.$$

Problème 2

Partie A : Autour de la loi géométrique

1. a) La première alerte peut survenir au plus tôt le premier jour, dans ce cas $X_1 = 0$, et il n'y a pas de valeur maximale que peut prendre X_1. X_1 prend forcément une valeur entière.
   En résumé, $\boxed{X_1(\Omega) = \mathbb{N}}$.
   Notons, pour $i \in \mathbb{N}^*$, A_i l'événement « une alerte survient le i-ème jour après l'installation du sismomètre ».
   On peut alors écrire, $[X_1 = 0] = A_1$ et pour $k \neq 0$ $[X_1 = k] = \overline{A_1} \cap \ldots \cap \overline{A_k} \cap A_{k+1}$.
   On en déduit que $P(X_1 = 0) = p$ et, comme les événements $(A_i)_{i\in\mathbb{N}^*}$ sont mutuellement indépendants, on a pour $k \in \mathbb{N}^*$:
$$P(X_1 = k) = P(\overline{A_1}) \times \ldots \times P(\overline{A_k}) \times P(A_{k+1}) = p(1-p)^k.$$
   En résumé, $\boxed{\text{pour tout } k \in \mathbb{N},\ P(X_1 = k) = p(1-p)^k.}$

b) D'après la question précédent $(1+X_1)(\Omega)=\mathbb{N}^*$ et, pour tout $k\in\mathbb{N}^*$,

$$P(X_1+1=k)=P(X_1=k-1)=p(1-p)^{k-1}.$$

$\boxed{X_1+1 \text{ suit bien la loi géométrique de paramètre } p.}$

2. a) On peut tout d'abord remarquer que si $k<j$ alors $P_{[X_1=j]}(X_2=k)=0$.
Soit maintenant $(j,k)\in\mathbb{N}^2$ tel que $j\leqslant k$. On reprend les notations de la question 1.a).
On a alors $P_{[X_1=j]}(X_2=j)=P(A_{j+2})=p$ et pour $k>j$, on peut alors écrire :

$$\begin{aligned}
P_{[X_1=j]}(X_2=k)&=P(\overline{A_{j+2}}\cap\ldots\cap\overline{A_{k+1}}\cap A_{k+2})\\
&=P(\overline{A_{j+2}})\times\ldots\times P(\overline{A_{k+1}})\times P(A_{k+2}) && \text{evt indépendants}\\
&=(1-p)^{k-j}\times p
\end{aligned}$$

En résumé $\boxed{P_{[X_1=j]}(X_2=k)=\begin{cases}0 & \text{si } j>k\\ (1-p)^{k-j}p & \text{si } j\leqslant k\end{cases}.}$

b) D'après la formule des probabilités totales appliquée avec le système complet d'événements $([X_1=j])_{j\in\mathbb{N}}$, on a, pour tout $k\in\mathbb{N}$:

$$\begin{aligned}
P(X_2=k)&=\sum_{j=0}^{+\infty}P(X_1=j)P_{[X_1=j]}(X_2=k)\\
&=\sum_{j=0}^{k}p(1-p)^j\times p(1-p)^{k-j}+\sum_{j=k+1}^{+\infty}p(1-p)^j\times 0\\
&=p^2\sum_{j=0}^{k}(1-p)^k\\
&=(k+1)p^2(1-p)^k=(k+1)q^kp^2.
\end{aligned}$$

On a bien montré que $\boxed{\forall k\in\mathbb{N},\ P(X_2=k)=(k+1)p^2q^k.}$

Partie B : Généralisation de la situation précédente

3. a) Montrons par récurrence que la propriété $\mathscr{P}(k)$: « $\binom{n+k}{k}=\sum_{j=0}^{k}\binom{n+j-1}{j}$ » est vraie pour tout $k\in\mathbb{N}$.

— Pour $k=0$, on a d'une part $\binom{n+0}{0}=1$ et d'autre part $\sum_{j=0}^{0}\binom{n+j-1}{j}=\binom{n-1}{0}=1$.

Donc $\mathscr{P}(0)$ est vraie.

— Soit $k\in\mathbb{N}$ fixé et tel que $\mathscr{P}(k)$ est vraie. On a alors :

$$\begin{aligned}
\binom{n+k+1}{k+1}&=\binom{n+k}{k}+\binom{n+k}{k+1} && \text{formule de Pascal}\\
&=\sum_{j=0}^{k}\binom{n+j-1}{j}+\binom{n+k}{k+1} && \text{d'après l'hypothèse de récurrence}\\
&=\sum_{j=0}^{k+1}\binom{n+j-1}{j}.
\end{aligned}$$

On a alors montré que $\mathscr{P}(k+1)$ est vérifiée.

Grâce au principe de récurrence, on a montré que $\boxed{\text{pour tout } k\in\mathbb{N},\ \binom{n+k}{k}=\sum_{j=0}^{k}\binom{n+j-1}{j}.}$

b) Montrons par récurrence que la propriété $\mathscr{P}(n)$: « $\forall k \in \mathbb{N},\ P(X_n = k) = \binom{n+k-1}{k} q^k p^n$ » est vraie pour tout $n \in \mathbb{N}^*$.

— Pour $n = 1$, on a d'une part, pour tout $k \in \mathbb{N}$, $P(X_1 = k) = pq^k$ et d'autre part $\binom{1+k-1}{k} q^k p^1 = pq^k$.
Donc $\mathscr{P}(1)$ est vraie.

— Soit $n \in \mathbb{N}^*$ fixé et tel que $\mathscr{P}(n)$ est vraie.
À l'aide du même raisonnement qu'à la question A.2.a) on peut remarquer que :

$$\forall (j,k) \in \mathbb{N}^2, \quad P_{[X_n=j]}(X_{n+1} = k) = \begin{cases} 0 & \text{si } j > k \\ (1-p)^{k-j} p & \text{si } j \leqslant k. \end{cases}$$

Ainsi, en appliquant la formule des probabilités totales avec le système complet d'événements $([X_n = j])_{j\in\mathbb{N}}$, on a

$$\begin{aligned}
\forall k \in \mathbb{N}, \quad P(X_{n+1} = k) &= \sum_{j=0}^{+\infty} P(X_n = j) P_{[X_n=j]}(X_{n+1} = k) \\
&= \sum_{j=0}^{k} \binom{n+j-1}{j} q^j p^n \times q^{k-j} p + \sum_{j=k+1}^{+\infty} \binom{n+j-1}{j} q^j p^n \times 0 \\
&\qquad \text{d'après } \mathscr{P}(n) \text{ et la remarque ci-dessus} \\
&= p^{n+1} q^k \sum_{j=0}^{k} \binom{n+j-1}{j} \\
&= \binom{n+k}{k} p^{n+1} q^k.
\end{aligned}$$

On a alors montré que $\mathscr{P}(n+1)$ est vérifiée.

Grâce au principe de récurrence, on a montré que $\boxed{\text{pour tout } n \in \mathbb{N}^* \text{ et } k \in \mathbb{N},\ P(X_n = k) = \binom{n+k-1}{k} q^k p^n.}$

4. a) Pour $n = 1$, on a $\sum_{i=1}^{1} X_1^{(i)} = X_1^{(1)}$ qui suit la loi $\mathscr{G}^{-1}(p)$.

Or, pour $n = 1$ et pour tout $k \in \mathbb{N}$, $\binom{1+k-1}{k} q^k p^1 = q^k p = P(X_1^{(1)} = k)$.

Donc $\boxed{\text{la propriété est bien vérifiée pour } n = 1.}$

On suppose maintenant que pour un $n \in \mathbb{N}^*$ fixé, l'énoncé suivant est vrai :

$\mathscr{E}(n)$: « si $(X_1^{(i)})_{i\in[\![1,n]\!]}$ est une famille de n variables aléatoires mutuellement indépendantes qui suivent toutes la loi $\mathscr{G}^{-1}(p)$ alors $\sum_{i=1}^{n} X_1^{(i)}$ suit la loi $\mathscr{B}^{-1}(n,p)$. »

Pour vérifier que cette propriété est vraie au rang $n+1$, on se donne $(X_1^{(i)})_{i\in[\![1,n+1]\!]}$ une famille de $n+1$ variables aléatoires mutuellement indépendantes qui suivent toutes la loi $\mathscr{G}^{-1}(p)$.

D'après le lemme des coalitions on peut alors affirmer que $\boxed{\sum_{i=1}^{n} X_1^{(i)} \text{ et } X_1^{(n+1)} \text{ sont indépendantes.}}$

b) Soit $k \in \mathbb{N}$. On peut remarquer que $\left[\sum_{i=1}^{n+1} X_1^{(i)} = k\right] = \bigcup_{j=0}^{k} \left[\sum_{i=1}^{n} X_1^{(i)} = j\right] \cap [X_1^{(n+1)} = k-j]$ donc

$$\begin{aligned}
P\left(\sum_{i=1}^{n+1} X_1^{(i)} = k\right) &= \sum_{j=0}^{k} P\left(\left[\sum_{i=1}^{n} X_1^{(i)} = j\right] \cap [X_1^{(n+1)} = k-j]\right) && \text{union d'evt disjoints}\\
&= \sum_{j=0}^{k} P\left(\sum_{i=1}^{n} X_1^{(i)} = j\right) \times P(X_1^{(n+1)} = k-j) && \text{evt indépendants}\\
&= \sum_{j=0}^{k} \binom{n+j-1}{j} q^j p^n \times pq^{k-j} && \text{hyp de récurrence et def de } \mathscr{G}^{-1}(p)\\
&= \binom{n+k}{k} p^{n+1} q^k && \text{même calcul que question 3.b)}
\end{aligned}$$

On a donc bien montré que $\sum_{i=1}^{n+1} X_1^{(i)}$ suit la loi $\mathscr{B}^{-1}(n+1,p)$, c'est-à-que que notre énoncé est vrai au rang $n+1$.

Grâce au principe de récurrence on a montré que $\boxed{\text{l'énoncé } \mathscr{E}(n) \text{ est vrai pour tout } n \in \mathbb{N}^*.}$

c) En écrivant X_n et X_m respectivement comme des sommes de variables aléatoires indépendantes suivant toutes la loi $\mathscr{G}^{-1}(p)$ on peut voir $X_n + X_m$ comme une somme de $n+m$ variables aléatoires indépendantes suivant toutes la loi $\mathscr{G}^{-1}(p)$ et donc on peut penser que $\boxed{X_n + X_m \text{ suit la loi } \mathscr{B}^{-1}(n+m,p).}$

Partie C : Une loi binomiale et une loi normale

5. a) Il y a ici une succession de $n+k$ épreuves de Bernoulli identiques et indépendantes : chaque jour on note s'il y a eu une alerte.

   Chaque jour, la probabilité d'avoir une alerte est égale à p et Y_{n+k} compte le nombre de réalisations de l'événement « il y a une alerte ».

   Ainsi, $\boxed{Y_{n+k} \text{ suit une loi binomiale de paramètres } (n+k,p).}$

   b) L'événement $[Y_{n+k} < n]$ signifie « lors des $n+k$ premiers jours il y a eu strictement moins de n alertes ». Ceci est équivalent à dire que « la n-ième alerte a lieu à un jour portant un numéro strictement supérieur à $n+k$ ».

   C'est encore équivalent à « au moment de la n-ième alerte strictement plus que $n+k$ jours se sont passés »ce qui est encore équivalent à dire « au moment de la n-ième alerte il y a eu strictement plus que k jours sans alerte »ce qui est exactement l'événement $[X_n > k]$.

   On a donc bien $\boxed{[Y_{n+k} < n] = [X_n > k].}$

6. a) Il nous faut, dans cette question, déterminer l'espérance et la variance de X_n.

   Pour éviter de faire de gros calculs, il est astucieux d'utiliser la question B.4.

   En effet, grâce à cette question, on peut affirmer que X_n admet la même espérance et variance que $\sum_{i=1}^{n} X_1^{(i)}$ où les $X_1^{(i)}$ vérifient les hypothèses de la question 4.

   D'après la linéarité de l'espérance on a donc $E(X_n) = \sum_{i=1}^{n} E(X_1^{(i)}) = nE(X_1^{(1)})$, car les $X_1^{(i)}$ suivent toutes la même loi.

   D'après la question A.1.b) et par linéarité de l'espérance,

$$E(X_1^{(1)}) = E(\underbrace{X_1^{(1)} + 1}_{\text{suit une loi géom}}) - 1 = \frac{1}{p} - 1 = \frac{q}{p}.$$

   On a donc $E(X_n) = n\dfrac{q}{p}$.

   Les variables $X_1^{(i)}$ étant mutuellement indépendantes, on a aussi $V(X_n) = \sum_{i=1}^{n} V(X_1^{(i)}) = nV(X_1^{(1)})$.

Et, par propriété de la variance :

$$V(X_1^{(1)}) = V(\underbrace{X_1^{(1)}+1}_{\text{suit une loi géom}}) = \frac{1-p}{p^2} = \frac{q}{p^2}.$$

On a donc $V(X_n) = n\dfrac{q}{p^2}$. On a bien montré que $\boxed{Z_n \hookrightarrow \mathscr{N}\left(\dfrac{nq}{p}, \dfrac{nq}{p^2}\right).}$

b) Le théorème central limite :
Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles définies sur un même espace probabilisé et indépendantes.
On suppose que les X_n suivent toutes la même loi et qu'elles admettent une variance non nulle.
On note μ leur espérance commune et σ leur écart-type commun.
On rappelle que :

$$\forall n \in \mathbb{N}^*, \quad M_n = \frac{1}{n}\sum_{k=1}^{n} X_k \quad \text{et} \quad M_n^* = \frac{M_n - E(M_n)}{\sqrt{V(M_n)}} = \frac{M_n - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Alors la suite $(M_n^*)_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire suivant la loi $\mathscr{N}(0,1)$.
Autrement dit, pour tout $x \in \mathbb{R}$, on a :

$$\lim_{n\to+\infty} P(M_n^* \leqslant x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t = \Phi(x).$$

On rappelle que Φ désigne la fonction de répartition de la loi $\mathscr{N}(0,1)$.

Application :
Soit $(Y_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes et telle que pour tout $i \in \mathbb{N}^*$, Y_i suit la loi $\mathscr{G}^{-1}(p)$.

D'après la question B.4., X_n et $\displaystyle\sum_{i=1}^{n} Y_i$ suivent la même loi. Donc :

$$P\left(a < \frac{\frac{X_n}{n} - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}} < b\right) = P\left(a < \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}} < b\right).$$

Or, d'après le théorème central limite appliqué à la suite de variables aléatoires $(Y_i)_{i\in\mathbb{N}^*}$, $\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} Y_i\right)^*$ converge en loi vers une variable aléatoire suivant la loi normale centrée réduite.

De plus, $\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} Y_i\right)^* = \dfrac{\frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}}$, donc

$$\lim_{n\to+\infty} P\left(a < \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}} < b\right) = \int_a^b \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-t^2/2}\mathrm{d}t.$$

D'autre part, on a vu que Z_n suit une loi normale donc Z_n^* suit la loi normale centrée réduite. Or :

$$Z_n^* = \frac{Z_n - E(Z_n)}{\sqrt{V(Z_n)}} = \frac{Z_n - \frac{nq}{p}}{\sqrt{\frac{nq}{p^2}}} = \frac{\frac{Z_n}{n} - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}}.$$

On a donc $P\left(a < \dfrac{\frac{Z_n}{n} - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}} < b\right) = \displaystyle\int_a^b \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-t^2/2}\mathrm{d}t$. (quantité qui ne dépend pas de n.)

En conclusion, $\boxed{\displaystyle\lim_{n\to+\infty} P\left(a < \frac{\frac{X_n}{n} - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}} < b\right) = \lim_{n\to+\infty} P\left(a < \frac{\frac{Z_n}{n} - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}} < b\right) = \int_a^b \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-t^2/2}\mathrm{d}t.}$