Il s'agit d'un très bel énoncé. Le reproche majeur qu'on peut lui faire est le morcellement des questions : le tout aurait été faisable en environ 25 questions moins élémentaires (et 15 aux mines !).

Partie I – Étude d'un cas simple

Q 1. Y est le rang du premier succès dans une suite d'expériences de Bernoulli **indépendantes** : cette variable aléatoire suit donc une loi géométrique de paramètre q (la probabilité de générer C).

$$\boxed{Y \hookrightarrow \mathcal{G}(q)\,;\ \mathbb{P}(Y=n)=p^{n-1}q \text{ pour tout } n\in\mathbb{N}^*}$$

Q 2. La série génératrice de Y est donc $\sum \frac{q}{p}(pt)^n$, de rayon de convergence $1/p$ (pour $r\geq 0$, la suite $\left(\frac{q}{p}(pr)^n\right)$ est bornée si et seulement si $pr\leq 1$, c'est-à-dire $r\leq \frac{1}{p}$). La somme de cette série vaut $G_Y(t)=\dfrac{q}{p}\displaystyle\sum_{n=1}^{+\infty}(pt)^n=\dfrac{q}{p}\dfrac{pt}{1-pt}$, et ainsi :

$$\boxed{R_Y=\frac{1}{p}>1 \text{ et } G_Y(t)=\frac{qt}{1-pt} \text{ pour tout } t\in]-1/p,1/p[.}$$

Puisque le rapport du jury en parle, précisons que $0<p<1$*, ce qui justifie* $\dfrac{1}{p}>1$*...*

Q 3. En tant que somme de série entière, on sait que G_Y est de classe $\mathcal{C}^\infty$ sur $]-R,R[$, les dérivées se calculant en dérivant terme à terme. Comme $R_Y>1$:

$$\boxed{G_Y \text{ est deux fois dérivable en 1.}}$$

Comme de plus pour tout $t\in]-1/p,1/p[$, $G_Y'(t)=\dfrac{q(1-pt)+qtp}{(1-pt)^2}=\dfrac{q}{(1-pt)^2}$ puis $G_Y''(t)=\dfrac{2pq}{(1-pt)^3}$, on a donc :

$$\boxed{G_Y'(1)=\frac{1}{q} \text{ et } G_Y''(1)=\frac{2p}{q^2}\cdot}$$

Q 4. Le lien entre l'espérance (respectivement la variance) et les dérivées de G_Y en 1 sont vaguement au programme... et surtout à savoir retrouver. Je pense qu'ils peuvent être donnés cash, après éventuellement un prudent calcul au brouillon, qui donnerait quelque chose comme :

$$G_Y'(1)=\sum_{n=1}^{+\infty}n\mathbb{P}(X=n)=\mathbb{E}(Y),$$

$$G''(Y)=\sum_{n=2}^{+\infty}n(n-1)\mathbb{P}(X=n)=\mathbb{E}(Y(Y-1))=\mathbb{E}(Y^2)-\mathbb{E}(Y)=\mathrm{Var}(Y)+(\mathbb{E}(Y))^2-\mathbb{E}(Y),$$

soit ici :

$$\boxed{\mathbb{E}(Y)=G_Y'(1)=\frac{1}{q} \text{ et } \mathrm{Var}(Y)=G_Y''(1)+G_Y'(1)-G_Y'(1)^2=\frac{p}{q^2}\cdot}$$

Et si on connaît son cours (ou qu'on ouvre le poly...), on retrouve bien l'expression connue des espérances et variance d'une géométrique (attention, ici q joue le rôle usuel de p...). Le rapport du jury laisse penser qu'on pouvait/devait parachuter les formules.

Partie II – Séries entières

Q 5. On est à nouveau face à une série entière géométrique de raison $1/a$; le cours ou l'argument vu plus haut nous donne son rayon de convergence :

$$\boxed{\text{Le rayon de convergence de la série entière } \sum \left(-\frac{1}{a^{n+1}}\right) z^n \text{ vaut } |a|.}$$

Q 6. Pour $|z| < |a|$, notre fine connaissance des séries géométriques nous donne :

$$\sum_{n=0}^{+\infty} \left(-\frac{1}{a^{n+1}}\right) z^n = -\frac{1}{a} \sum_{n=0}^{+\infty} \left(\frac{z}{a}\right)^n = -\frac{1}{a} \cdot \frac{1}{1 - z/a},$$

soit encore, comme annoncé :

$$\boxed{\text{Pour } |z| < |a|, \ \sum_{n=0}^{+\infty} \left(-\frac{1}{a^{n+1}}\right) z^n = \frac{1}{z-a}.}$$

Q 7. Il va essentiellement falloir sommer des termes d'une suite géométrique de raison $\dfrac{b}{a} \neq 1$:

$$v_n = \sum_{k=0}^{n} \frac{1}{a^{k+1}} \cdot \frac{1}{b^{n+1-k}} = \frac{1}{ab^{n+1}} \sum_{k=0}^{n} \left(\frac{b}{a}\right)^k = \frac{1}{ab^{n+1}} \cdot \frac{1 - \left(\frac{b}{a}\right)^{n+1}}{1 - \frac{b}{a}},$$

soit après une dernière agitation de termes :

$$\boxed{v_n = \frac{1}{b-a} \left(\frac{1}{a^{n+1}} - \frac{1}{b^{n+1}}\right).}$$

Q 8. On a $|a| < |b|$, donc $\dfrac{1}{|a|} > \dfrac{1}{|b|}$, donc $\left(\dfrac{1}{b}\right)^{n+1} = o\left(\left(\dfrac{1}{a}\right)^{n+1}\right)$, donc :

$$\boxed{v_n \sim \frac{1}{(b-a)a^{n+1}}.}$$

Q 9. Pour $r > 0$, on a $(v_n r^n)$ bornée si et seulement si $\left(\dfrac{1}{a(b-a)} \cdot \left(\dfrac{r}{a}\right)^n\right)$ l'est, c'est-à-dire si $r \leq |a|$:

$$\boxed{\text{Le rayon de convergence de } \sum v_n z^n \text{ vaut } |a|.}$$

Pour $|z| < |a|$, les séries $\sum u_n(a) z^n$ et $\sum v_n(a) z^n$ sont absolument convergentes, donc leur produit de Cauchy également, la somme du produit étant égal au produit des sommes des deux séries :

$$\boxed{\text{Pour } |z| < |a|, \ \sum_{n=0}^{+\infty} v_n z^n = \frac{1}{(z-a)(z-b)}.}$$

Q 10. Pour $t \in]-a, a[$, on a $f(t) = \sum_{n=0}^{+\infty} v_n t^{n+2} = \sum_{k=2}^{+\infty} v_{k-2} t^k$: il s'agit bien d'un développement en série entière... et le rayon de convergence de la série associée est le même que celui de $\sum v_n z^n$.

$$\boxed{\text{Ce qui est ce qu'on voulait montrer.}}$$

Q 11. On note que $g(t) = f(t) \cdot \dfrac{1}{t-c}$, et les arguments vus à la question 9 (produit de Cauchy de deux séries de rayon $\geq |a|$) nous permettraient de montrer de la même façon :

$\boxed{g \text{ est développable en série entière au voisinage de 0, avec un rayon de convergence } \geq |a|.}$

Les arguments à la Cauchy donnaient seulement une minoration du rayon, la valeur exacte n'étant acquise que par la forme particulière de la série entière.

Partie III – Étude d'un cas intermédiaire

Q 12. Bien entendu, $p_1 = 0$. Ensuite, $p_2 = \mathbb{P}(C_1 \cap C_2) = \mathbb{P}(C_1)\mathbb{P}(C_2)$ (indépendance) donc $p_2 = q^2$. Enfin, la seule possibilité pour avoir $Z = 3$ est que les trois premières lettres soient PCC, donc toujours par indépendance : $p_3 = \mathbb{P}(P_1 \cap C_2 \cap C_3) = \mathbb{P}(P_1)\mathbb{P}(C_2)\mathbb{P}(C_3) = pq^2$.

$$\boxed{p_1 = 0,\ p_2 = q^2 \text{ et } p_3 = pq^2.}$$

Q 13. D'une part (P_1, C_1) est complet (deux événements disjoints, de réunion Ω), et d'autre part $C_1 = (C_1 \cap P_2) \cup (C_1 \cap C_2)$, l'union étant disjointe. Ainsi :

$$\boxed{(P_1, C_1 \cap P_2, C_1 \cap C_2) \text{ est un système complet d'événements.}}$$

Q 14. La formule des probabilités totales (pour le système complet d'événements précédent) nous donne :

$$p_n = \mathbb{P}(Z = n) = \mathbb{P}_{P_1}(Z = n)\mathbb{P}(P_1) + \mathbb{P}_{C_1P_2}(Z = n)\mathbb{P}(C_1P_2) + \mathbb{P}_{C_1C_2}(Z = n)\mathbb{P}(C_1C_2)$$

Si les deux premières lettres sont C_1C_2, alors $Z = 2$ donc $Z \neq n$: $\mathbb{P}_{C_1C_2}(Z = n) = 0$. Si maintenant la première lettre est P_1, alors l'automate est dans l'état zéro, donc la probabilité pour que $n - 1$ lettres plus tard il soit dans l'état 2 vaut $\mathbb{P}(Z = n - 1)$: $\mathbb{P}_{P_1}(Z = n) = \mathbb{P}(Z = n - 1) = p_{n-1}$. De même, $\mathbb{P}_{C_1P_2}(Z = n) = \mathbb{P}(Z = n - 2)$ (si après 2 lettres on est dans l'état 0, la probabilité de se retrouver dans l'état 2 après les $n - 2$ lettres suivantes est $\mathbb{P}(Z = n - 2) = p_{n-2}$). Enfin, $\mathbb{P}(P_1) = p$ et par indépendance, $\mathbb{P}(C_1P_2) = \mathbb{P}(C_1)\mathbb{P}(P_2) = pq$. Ainsi :

$$\boxed{\text{Pour tout } n \geq 3,\ p_n = p\,p_{n-1} + pq\,p_{n-2}.}$$

On se rassure en regardant les trois premiers termes de la suite, bien entendu...

Q 15. On fixe $t \in [-1, 1]$ et on multiplie la relation précédente par t^n, puis on somme pour $n \in [\![3, +\infty[\![$: les séries en jeu sont toutes absolument convergentes (les termes généraux sont positifs et majorés par $p_n = \mathbb{P}(Z = n)$, terme général d'une série absolument convergente). On obtient ainsi :

$$\sum_{n=3}^{+\infty} p_n t^n = pt \sum_{n=3}^{+\infty} p_{n-1} t^{n-1} + t^2 pq \sum_{n=3}^{+\infty} p_{n-2} t^{n-2}. \qquad (R)$$

Par décalages d'indices :

$$\sum_{n=3}^{+\infty} p_{n-1} t^{n-1} = \sum_{i=2}^{+\infty} p_i t^i = G_Z(t) - (p_0 + p_1 t) = G_Z(t)$$

et $\sum\limits_{n=3}^{+\infty} p_{n-2} t^{n-2} = G_Z(t)$. Puisque le membre de gauche de (R) vaut $G_Z(t) - p_2 t^2 = q^2 t^2$, on obtient $G_Z(t) - q^2t^2 = (pt + pqt^2)G_Z(t)$, ce qui n'est plus trop éloigné de l'objectif.

$$\boxed{\text{Pour tout } t \in [-1, 1],\ G_Z(t)(1 - pt - pqt^2) = q^2t^2.}$$

Q 16. Pfff... une de ces nombreuses questions élémentaires un peu usantes. Or donc :

$$Q(-1) = 1 + p - pq = 1 + p - p(1 - p) = 1 + p^2 \quad \text{et} \quad Q(1) = 1 - p - pq = q - pq = q(1 - p) = q^2$$

$$\boxed{Q(-1) = 1 + p^2 > 0 \text{ et } Q(1) = q^2 > 0}$$

Ces résultats seront utiles... deux questions plus loin, pour localiser les racines de Q...

Q 17. Le polynôme $Q = 1 - pX - pqX^2$ (quelle idée de parler de l'application polynomiale Q, comme en prébac...) est de degré 2, coefficient dominant $-pq$ et possède deux racines distinctes a et b : on peut donc le factoriser sous la forme $Q = -pq(X-a)(X-b)$. Ainsi :

$$\boxed{\text{Pour tout } t \in \mathbb{R},\ Q(t) = -pq(t-a)(t-b)}$$

Q 18. *J'avais d'abord abordé cette question par un calcul absurde... avant de prendre 15cm de recul sur l'énoncé...*
L'application Q vérifie $Q(-1) = 1 > 0$, $Q(t) \underset{t\to-\infty}{\longrightarrow} -\infty$ (terme dominant) et est **continue**, donc le théorème des valeurs intermédiaires[1] nous assure qu'elle possède une racine dans $]-\infty, -1[$. De même, elle possède une racine dans $]1, +\infty[$. Par ailleurs, les deux racines de Q sont a et b avec $b < a$, donc : $b < -1 < 0 < 1 < a$. Ensuite :

$$|b| = -b = \frac{\sqrt{\Delta}+p}{2pq} > \frac{\sqrt{\Delta}-p}{2pq} = a = |a| > 1,$$

et ainsi :

$$\boxed{1 < |a| < |b|}$$

Q 19. D'après les questions précédentes, $f(t) = \dfrac{-\frac{q}{p}t^2}{(t-a)(t-b)}$, et la question 10 s'applique avec $\lambda = -\dfrac{q}{p}$ (la condition $|a| < |b|$ étant bien vérifiée).

$$\boxed{f \text{ est développable en série entière, de série entière associée } G_Z \text{ ayant un rayon de convergence } R_Z = |a|}$$

Q 20. Haaaaaa une question intéressante et fine (la façon dont le rapport du jury passe dessus négligemment me surprend...). Ce qu'on a montré dans ce qui précède, c'est que la relation $G_Z(t) = \dfrac{q^2t}{1-pt-pqt^2}$ est valable pour $t \in]1,1[$, et il s'agit donc de l'étendre à $]-|a|,|a|[$. Or on a ici :

$$\forall t \in]-1,1[, \qquad \sum_{n=0}^{+\infty} p_n t^n = \sum_{n=0}^{+\infty} v_n t^n$$

(notations de la deuxième partie), donc **par unicité du développement en série entière**, on a : pour tout $n \in \mathbb{N}$, $p_n = v_n$. La relation s'étend alors bien à $]-|a|,|a|[$.

$$\boxed{\text{Pour tout } t \in]-|a|,|a|[,\ G_Z(t) = \frac{q^2t^2}{1-pt-pqt^2}.}$$

Q 21. Une somme de série entière de rayon de convergence $R > 0$ est de classe $\mathcal{C}^\infty$ sur $]-R, R[$. Or ici, $R_Z = |a| > 1$, donc G_Z est deux fois dérivable en 1, donc :

$$\boxed{Z \text{ possède une espérance et une variance.}}$$

Il reste à calculer $\mathbb{E}(Z) = G_Z'(1)$. Ici :

$$\forall t \in]-R_Z, R_Z[, \qquad G_Z'(t) = \frac{2tq^2(1-pt-pqt^2) + q^2t^2(p+2pqt)}{(1-pt-pqt^2)^2}.$$

Pour évaluer en 1 on note que $1 - p - pq = q^2$ (question 16), donc :

$$\mathbb{E}(Z) = 2 + \frac{p}{q^2} + 2\frac{p}{q} = 2 + \frac{1-q}{q^2} + 2\frac{1-q}{q}.$$

Finalement, comme demandé :

1. Oui, bon... une extension.

$$\boxed{\mathbb{E}(Z) = \frac{1}{q^2} + \frac{1}{q}\cdot}$$

Q 22. Il s'agit d'établir : $\dfrac{1}{q} + \dfrac{1}{q^2} \geq 1 + \dfrac{p}{q^2}$, soit encore : $q+1 \geq q^2+p = q^2+1-q$, ou encore $q^2 \leq 2q$, c'est-à-dire $q \leq 2$. Cette dernière inégalité est vérifiée, et on a travaillé par équivalence (de l'importance du choix des mots quand on rédige...), ce qui prouve l'inégalité demandée.

$$\boxed{\mathbb{E}(Z) \geq \mathbb{E}(Y) + 1.}$$

Q 23. La première occurrence de CC est évidemment **strictement** précédée par la première occurrence de C, donc on a toujours $Z \geq Y+1$. Par croissance de l'espérance, on on déduit : $\mathbb{E}(Z) \geq 1+\mathbb{E}(Y)$.

$$\boxed{\text{Ben oui, c'était attendu !}}$$

Partie IV – Algèbre linéaire

Q 24. Si on a fini par comprendre cette histoire de « j'écris des choses en haut et à droite de la matrice pour comprendre ce que les colonnes représentent », alors on voit immédiatement qu'avec les notations qu'on imagine, on a $u(e_4) = 0$, ou encore en notant $X_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$, on a $AX_0 = 0 = 0.X_0$. Comme $X_0 \neq 0$, ceci prouve que :

$$\boxed{0 \text{ est valeur propre de } A, \text{ un vecteur propre associé étant } \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}}$$

Q 25. Soit $t \in \mathbb{R}$. On calcule $\chi_A(t)$ en développant par rapport à la première colonne (qui ne contient qu'un terme (potentiellement) non nul) puis les déterminants 3×3 par rapport à la première colonne (par exemple) :

$$\chi_A(t) = t\begin{vmatrix} t-p & 0 & -p \\ -q & t-q & 0 \\ 0 & -p & t \end{vmatrix} = t\left((t-p)(t-q)t + q(-(-p)(-p))\right) = t\left(t^3 - \underbrace{(p+q)}_{1}t^2 + pqt - qp^2\right),$$

soit donc :

$$\boxed{\text{Pour tout } t \in \mathbb{R},\ \chi_A(t) = t^4 - t^3 + \underbrace{pq}_{\alpha}\,t^2\,\underbrace{-qp^2}_{\beta}\,t + \underbrace{0}_{\gamma}}$$

Q 26. Question assez troublante... S est solution de (E_t) si et seulement si $A - tAS = L$, c'est-à-dire : $(I-tA)S = L$ (pas convaincu ? alors développez $(I-tA)S$!).

$$\boxed{\text{Ce qui prouve le résultat demandé !}}$$

Q 27. On rappelle que lorsque $\lambda \in \mathbb{K}$ et $M \in \mathcal{M}_n(\mathbb{K})$, la linéarité du déterminant **par rapport à chacune de ses colonnes** implique : $\det(\lambda M) = \lambda^n \det(M)$, donc ici pour $t \neq 0$:

$$\psi_A(t) = \det(I - tA) = \det\left(t\left(\frac{1}{t}I - A\right)\right) = t^4 \det\left(\frac{1}{t}I - A\right).$$

$$\boxed{\psi_A(t) = t^4\chi_A(1/t).}$$

Q 28. D'après ce qui précède et la question 25, on a :

$$\forall t \neq 0, \qquad \psi_A(t) = t^4\left(\frac{1}{t^4} - \frac{1}{t^3} + pq\frac{1}{t^2} - p^2q\frac{1}{t}\right) = 1 - t + pqt^2 - p^2qt^3.$$

Il reste à voir que pour $t = 0$, le résultat (mais pas le calcul !) reste vrai puisque $\psi_A(0) = \det(I) = 1$.

$$\boxed{\text{Pour tout } t \in \mathbb{R},\ \psi_A(t) = -p^2qt^3 + pqt - t + 1.}$$

Q 29. Puisque ψ_A est continue et non nulle en 0 :

Il existe un voisinage de 0 *sur lequel* ψ_A *ne s'annule pas.*

Suer ce voisinage, $I_4 - tA$ est alors inversible, si bien que l'équation (E_t), qui est équivalente à $(I_4 - tA)S = L$, possède une unique solution (accessoirement, c'est $(I_4 - tA)^{-1}L$).

$$\boxed{\text{Pour } t \text{ au voisinage de } 0,\ (E_t) \text{ possède une unique solution.}}$$

Q 30. Le calcul par bloc de $(I_4 - tA)S$ donne (attention : les U_k sont des colonnes et les S_k des scalaires !) :

$$(I_4 - tA)S = \begin{pmatrix} U_1 \vdots U_2 \vdots U_3 \vdots U_4 \end{pmatrix}\begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = S_0U_1 + \cdots + S_3U_4.$$

Mais par ailleurs on a supposé $(I_4 - tA)S = L$. Ainsi :

$$\boxed{L = S_0U_1 + \cdots + S_3U_4}$$

Q 31. Le caractère multilinéaire puis alterné du déterminant nous assure que :

$$\det_{\mathcal{B}}(U_1, U_2, U_3, L) = S_0 \underbrace{\det_{\mathcal{B}}(\mathbf{U_1}, U_2, U_3, \mathbf{U_1})}_{0} + \cdots + S_2 \underbrace{\det_{\mathcal{B}}(U_1, U_2, \mathbf{U_3}, \mathbf{U_3})}_{0} + S_3 \det_{\mathcal{B}}(U_1, U_2, U_3, U_4)$$

Ainsi :

$$\boxed{\det_{\mathcal{B}}(U_1, U_2, U_3, L) = S_3 \det(U_1, U_2, U_3, U_4) = S_3 \det(I_4 - tA) = S_3\psi_A(t).}$$

Q 32. Au voisinage de 0, on a $\psi_A(t) \neq 0$, puis $S_3 = \dfrac{\det_{\mathcal{B}}(U_1, U_2, U_3, L)}{\psi_A(t)}\cdot$ Or le déterminant au numérateur se calcule très simplement en développant par rapport à la dernière colonne :

$$\det_{\mathcal{B}}(U_1, U_2, U_3, L) = \begin{vmatrix} 1-pt & 0 & -pt & 1 \\ -qt & 1-qt & 0 & 0 \\ 0 & -pt & 1 & 0 \\ 0 & 0 & -qt & 0 \end{vmatrix} = -\begin{vmatrix} -qt & 1-qt & 0 \\ 0 & -pt & 1 \\ 0 & 0 & -qt \end{vmatrix} = pq^2t^3$$

(oui, comptez bien le nombre de « moins » !). Comme on se souvient de l'expression de $\psi_A(t)$, on peut conclure !

$$\boxed{\text{Au voisinage de } 0,\ S_3 = \frac{pq^2t^3}{-p^2qt^3 + pqt^2 - t + 1}.}$$

Q 33. On se souvient que le déterminant d'une matrice est celui de sa transposée. Ou encore que le rang d'une matrice est également celui de sa transposée. Ainsi, $A - \alpha I$ est inversible si et seulement si ${}^tA - \alpha{}^tI = {}^tA - \alpha I$ l'est : A et ${}^tA$ ont le même spectre (j'imagine que ce résultat pouvait d'ailleurs être donné cash). Puisque λ est valeur propre de A, le résultat suit :

$$\boxed{\lambda \text{ est valeur propre de } {}^tA.}$$

Q 34. D'après la question précédente, il existe $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ **non nul** tel que ${}^tAX = \lambda X$, c'est-à-dire :

$$\begin{cases} px_1 & + & qx_2 & & & & & = & \lambda x_1 \\ & & qx_2 & + & px_3 & & & = & \lambda x_2 \\ px_1 & & & & & + & qx_4 & = & \lambda x_3 \\ & & & & & & 0 & = & \lambda x_4 \end{cases}$$

Puisque $\lambda = 0$, la dernière équation fournit $x_4 = 0$, donc x_1, x_2 et x_3 ne sont pas tous nuls, et par ailleurs la troisième équation devient $px_1 = \lambda x_3$.

$$\boxed{\text{Le système } (\mathcal{H}) \text{ possède une solution non nulle.}}$$

Q 35. Dans le cas où $M = |x_3| > 0$ (ben oui, la solution est non nulle, donc le coefficient de plus gros module est non nul), la troisième équation fournit $\lambda = p\dfrac{x_1}{x_3}$, donc $|\lambda| = p\dfrac{|x_1|}{|x_3|}\cdot$ Puisque $0 < p < 1$ et $\dfrac{|x_1|}{|x_3|} \leq 1$, on obtient bien $|\lambda| < 1$.
Dans le second cas, on regarde la deuxième équation. Par inégalité triangulaire :

$$|\lambda|\,|x_2| \leq q\,|x_2| + p\,|x_3|$$

donc $|\lambda| \leq q + p\dfrac{|x_3|}{|x_2|}$. Mais $\dfrac{|x_3|}{|x_2|} < 1$ et $p > 0$ (point crucial pour maintenir l'inégalité stricte) donc $p\dfrac{|x_3|}{|x_2|} < p$ puis $|\lambda| < p + q = 1$.
Le dernier cas se traite de la même façon, en s'intéressant à la première équation de $(\mathcal{H})$.

$$\boxed{|\lambda| < 1.}$$

Bien entendu, puisqu'exceptionnellement on traite d'inégalités STRICTES, il convient de faire bien attention aux arguments permettant de les obtenir, puisque ça y est enfin : vos mains écrivent naturellement, par défaut, des inégalités larges.

Q 36. χ_A est un polynôme de degré 4, unitaire, et possédant 0 comme racine ; on sait alors qu'il est scindé sur $\mathbb{C}$ et que plus précisément il se factorise $X(X-\alpha_1)(X-\alpha_2)(X-\alpha_3)$, avec **a priori** les α_i pouvant être égaux entre eux, éventuellement nuls.
Mais ici, on a vu que χ_A possède 0 comme racine SIMPLE ($\chi_A'(0) = 1 \neq 0$), donc les α_i sont tous non nuls. Il reste à les réordonner pour que les modules soient croissants (et la question précédente nous assure que ces modules sont tous strictement plus petits que 1.

$$\boxed{\exists \lambda_1, \lambda_2, \lambda_3 \in \mathbb{C};\quad 0 < |\lambda_1| \leq |\lambda_2| \leq |\lambda_3| < 1 \ \text{ et } \ \forall t \in \mathbb{R},\ \chi_A(t) = t(t-\lambda_1)(t-\lambda_2)(t-\lambda_3).}$$

Q 37. On a, pour $t \neq 0$:

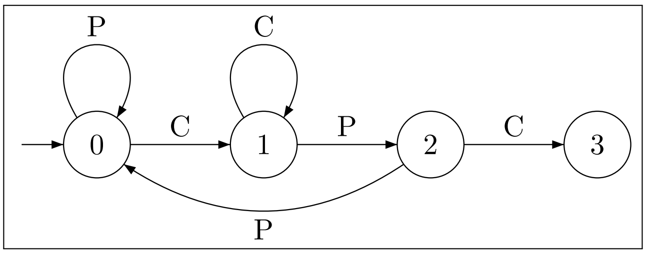
$$\begin{aligned} \psi_A(t) &= t^4 \chi_A(1/t) = t^4 \frac{1}{t}\left(\frac{1}{t} - \lambda_1\right)\left(\frac{1}{t} - \lambda_2\right)\left(\frac{1}{t} - \lambda_"\right) = t(1-\lambda_1 t)(1-\lambda_2 t)(1-\lambda_3 t) \\ &= -\lambda_1\lambda_2\lambda_3 t(t - 1/\lambda_1)(t-1/\lambda_2)(t-1/\lambda_3) \end{aligned}$$

Puisque $0 < |\lambda_1| \leq |\lambda_2| \leq |\lambda_3| < 1$, on a $1 < \left|\dfrac{1}{\lambda_3}\right| \leq \left|\dfrac{1}{\lambda_2}\right| \leq \left|\dfrac{1}{\lambda_1}\right|$, ce qui nous incite à poser $a = \dfrac{1}{\lambda_3}$, $b = \dfrac{1}{\lambda_2}$, $c = \dfrac{1}{\lambda_1}$ et $\mu = -\lambda_1\lambda_2\lambda_3 (= -p^2 q) \neq 0$ Il reste à noter que la relation prouvée pour $t \neq 0$ s'étend à $t = 0$ (les deux membres sont égaux à 1).

$$\boxed{\text{Et c'est gagné.}}$$

Partie V – Étude d'un dernier cas

Q 38. J'utilise la notation des informaticiens et non des zessistes pour ce type d'automates : les flèches arrivent sur des états et pas au mileu d'autres flèches !



L'idée (qui reste intuitive...) est de s'assurer de passer à l'état 3 au premier moment où on aura rencontré le mot CPC. ***À chaque instant, la question est donc : quel est le plus gros préfixe (début) de CPC que je viens de rencontrer ?***

Q 39. Bien entendu :

$$\boxed{S_0(0) = \begin{pmatrix} p_{0,0} \\ p_{0,1} \\ p_{0,2} \\ p_{0,3} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}}$$

Q 40. On ne détaille que la première équation, les autres étant de même nature. Le principe est le même qu'à la question 14 : on conditionne l'événement $E_{n,0}$ selon le système complet d'événements $(E_{n-1,0}, E_{n-1,1}, E_{n-1,2}, E_{n-1,3})$, sachant que la modélisation du problème nous donne les probabilités conditionnelles $\mathbb{P}_{E_{n-1,i}}(E_{n,0})$: elles sont données par les flèches entrantes dans l'état 0 :

$$\mathbb{E}_{n,0} = \underbrace{\mathbb{P}(E_{n-1,0})}_{p_{n-1,0}} \underbrace{\mathbb{P}_{E_{n-1,0}}(E_{n,0})}_{p} + \underbrace{\mathbb{P}(E_{n-1,1})}_{p_{n-1,1}} \underbrace{\mathbb{P}_{E_{n-1,0}}(E_{n,0})}_{0} + \underbrace{\mathbb{P}(E_{n-1,2})}_{p_{n-1,2}} \underbrace{\mathbb{P}_{E_{n-1,2}}(E_{n,0})}_{p} + \underbrace{\mathbb{P}(E_{n-1,3})}_{p_{n-1,3}} \underbrace{\mathbb{P}_{E_{n-1,3}}(E_{n,0})}_{0},$$

ce qui nous donne exactement la première équation demandée.
Une lecture attentive du graphe nous donne les autres équations, basées sur le même principe.

$\boxed{\text{Et c'est gagné.}}$

Q 41. Même principe qu'à la question 15 ! On prend la première relation du système précédent. On multiplie par t^n, en laissant de coté tp dans les deux termes du membre de droite. Toutes les séries étant absolument convergentes (les coefficients de ces séries entières sont positifs et de somme 1), on peut joyeusement sommer **pour** n **allant de** 1 **à** $+\infty$. Le membre de gauche est alors $S_0(t) - p_{0,0} = S_{0,t} - 1$, alors que celui de droite est $tpS_1(t) + tpS_2(t)$, et c'est gagné. Le même principe fournit les autres équations.

$\boxed{\text{Encore gagné !}}$

Q 42. Hum, que vaut $tAS(t) + L$? Mais mais mais, c'est $S(t)$!

$\boxed{\text{Bon, ben oui !}}$

Q 43. La fonction génératrice $G_T = S_3$ vaut, d'après la question 32 : $S_3(t) = \dfrac{pq^2t^3}{-p^2qt^3 + pqt^2 - t + 1}$, pour t au voisinage de 0. Mais le raisonnement fait en fin de partie II (questions 19 et 20) nous assure qu'en fait, S_3 est de rayon de convergence égal au module de la plus petite racine du dénominateur, et que la relation s'étend du voisinage de 0 à $]-R, R[$. La question 37 nous assure que ces racines sont de module strictement plus grand que 1, ce qui prouve : $R_T > 1$.

$$\boxed{R_T > 1 \text{ et pour tout } t \in]-R_T, R_T[,\ G_T(t) = \frac{pq^2t^3}{-p^2qt^3 + pqt^2 - t + 1}.}$$

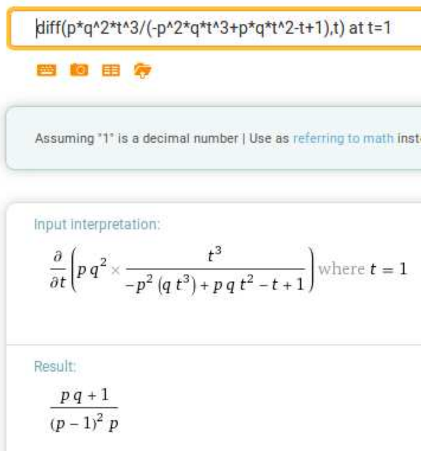
Q 44. Comme pour la question 21 : G_T est de classe $\mathcal{C}^\infty$ sur $]-R_T, R_T[$ qui contient $[-1, 1]$, donc G_T est deux fois dérivables en 1, donc :

$$\boxed{T \text{ possède une espérance et une variance.}}$$

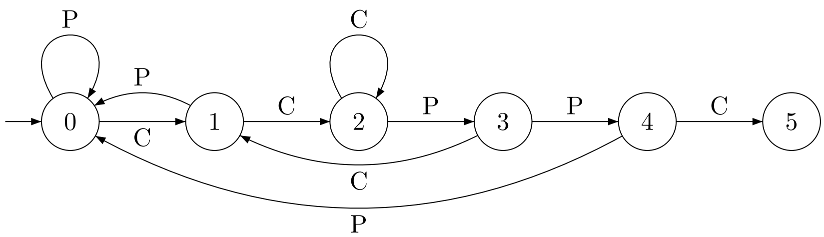
Q 45. Il reste à calculer courageusement $\mathbb{E}(T) = G_T'(1)$. J'ai obtenu (promis, sans tricher) :

$$\boxed{\mathbb{E}(T) = \frac{1 + q - q^2}{q^2(1-q)}.}$$

Jugeant trop compliqué de refaire le calcul, j'ai préféré programmer une petite simulation Python... dont les résultats sont spectaculaires (voir l'annexe Python) !
Bon, j'ai ensuite lancé Wolfram Alpha... et c'est spectaculaire à nouveau !



Q 46. Commençons par construire l'automate associé à la recherche de ce motif. C'est la même chose que pour la question 18 : quel est le plus gros préfixe de CCPPC que je viens de rencontrer ? Si c'est par exemple CCP, j'aimerais lire P (le plus gros préfixe devient CCPC, je passe à l'état 4) ; mais si c'est C, alors je viens de lire CCPC : le plus gros préfixe de CCPPC en cours est C et je passe dans l'état 1.

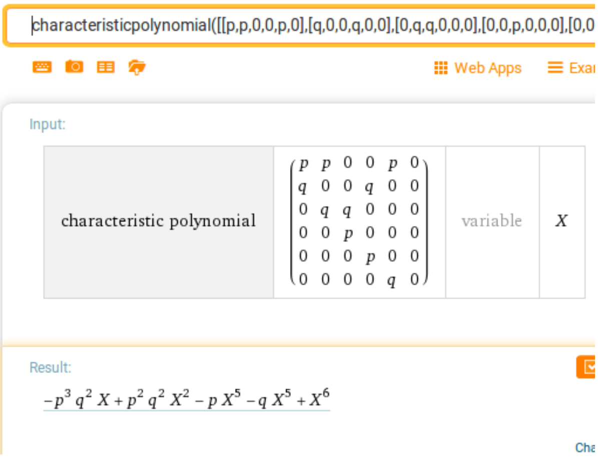


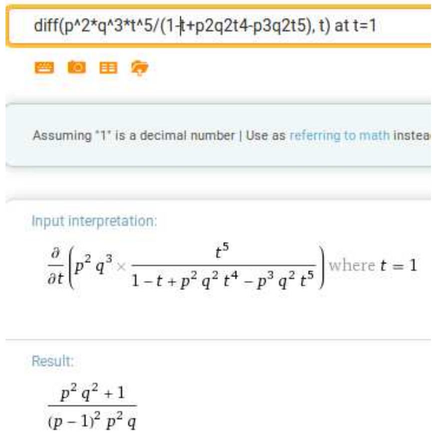
Maintenant, la matrice associée à cet automate se construit en plaçant, en position (i, j), la probabilité de passer de l'état j à l'état i (oui, dans ce sens ; et les numérotations partent de 0) :

$$A = \begin{pmatrix} p & p & 0 & 0 & p & 0 \\ q & 0 & 0 & q & 0 & 0 \\ 0 & q & q & 0 & 0 & 0 \\ 0 & 0 & p & 0 & 0 & 0 \\ 0 & 0 & 0 & p & 0 & 0 \\ 0 & 0 & 0 & 0 & q & 0 \end{pmatrix}$$

On cherche l'espérance de S_5, qui est une fraction avec $p^2q^3t^5$ au numérateur (« il se lit sous la diagonale »), et $\psi_A(t)$ au dénominateur, qui est aussi $t^6\chi_A(1/t)$. Pas envie de calculer tout ça ? ? ? Moi non plus...

Notez la souplesse quasiment scandaleuse de la syntaxe pour le deuxième appel (puissances oubliées, écrites « comme on les prononce »).





$$\boxed{\mathbb{E}(T) = \frac{1+q^2(1-q)^2}{q^3(1-q)^2}\cdot}$$

Un brin d'ADN peut être vu comme un mot (sur un alphabet à 4 lettres) de longueur de l'ordre de quelques centaines de millions (ou moins, ou plus !). Les recherches de propriétés de ce mot sont très consommatrices « d'algorithmes du texte » (un peu plus élaborés que la simple recherche d'un motif, mais qui restent souvent de nature proche).

L'automate associé à un motif donné m permet de parcourir un grand texte (disons de taille n) et repérer les occurrences de m sans revenir en arrière (ce qu'on ferait avec un algorithme naïf). Le temps de recherche du motif m passe alors de $n\,|m|$ à n, ce qui est une amélioration considérable, surtout si m a une longueur de l'ordre de n (ce qui arrive dans la vraie vie).

Mais tout ceci nécessite d'avoir construit l'automate préalablement. L'algorithme naïf (chercher le plus gros suffixe qui soit un préfixe...) est de complexité $|m|^3$, ce qui est rédhibitoire si m contient de l'ordre du milliard (ou même seulement du million) de caractères. En 1970, Knuth, Morris et Pratt ont conçu un algorithme très intelligent pour construire cet automate en temps linéaire en $|m|$. Le très belle idée consiste (mais ça constitue un sujet complet d'informatique de l'option !) grosso-modo à construire l'automate en utilisant à la volée ce qui a déjà été construit jusque là.

Sans ces algorithmes efficaces, le séquençage de l'ADN aurait probablement été sans objet : un informaticien donne du travail à des milliers de biologistes :-)

```
# -*- coding: utf-8 -*-
"""
Created on Thu Jan 18 11:23:24 2018

@author: stephane
"""

from random import random

"""
Je code les automates d'états 0, ..., n avec une liste de taille n, auto[k] étant la liste des deux transitions depuis l'état k :
auto[k][0] et auto[k][1] sont les états auxquels on accède en lisant respectivement les lettres P et C
"""

autoC = [[0, 1]]
autoCC = [[0, 1], [0, 2]]
autoCPC = [[0, 1], [2, 1], [0, 3]]
autoCCPPC = [[0, 1], [0, 2], [3, 2], [4, 1], [0, 5]]


def experience(automate, p):
    n = len(automate)
    q, etapes = 0, 0  # q est l'état courant : on s'arrête quand il atteint n
    while q < n:
        etapes += 1
        if random() < p:  # la lettre est P
            q = automate[q][0]
        else:
            q = automate[q][1]
    return etapes


def theo1(p):
    return 1/(1-p)


for p in [1/4, 1/2, 3/4]:
    print(p, theo1(p), sum(experience(autoC, p) for _ in range(10**5)) / 10**5)

"""
0.25 1.3333333333333333 1.33282
0.5 2.0 2.00177
0.75 4.0 3.99561
"""


def theo2(p):
    return 1/(1-p)**2+1/(1-p)

for p in [1/4, 1/2, 3/4]:
    print(p, theo2(p), sum(experience(autoCC, p) for _ in range(10**5)) / 10**5)

"""
0.25 3.1111111111111107 3.10865
0.5 6.0 6.00052
0.75 20.0 20.02166
"""


def theo3(p):
    q = 1-p
    return p**2*q**3*(p**2*q**2-4*p+5)/(q**2*(p**2*q**2+1)**2)
```



```
for p in [1/4, 1/2, 3/4]:
    print(p, theo3(p), sum(experience(autoCCPPC, p) for _ in range(10**5)) / 10*
*5)

"""
0.25 8.444444444444445 8.43946
0.5 10.0 9.98587
0.75 25.333333333333332 25.39219
"""


def theo4(p):
    q = 1-p
    return (p**2*q**2+1)/(p**2*q**3)

for p in [1/4, 1/2, 3/4]:
    print(p, theo4(p), sum(experience(autoCCPPC, p) for _ in range(10**5)) / 10*
*5)

"""
0.25 39.25925925925926 39.36461
0.5 34.0 33.94907
0.75 117.77777777777777 117.76952
"""
```

1/ CONSIGNES GENERALES :

Présentation du sujet :

Le sujet proposait l'étude d'un automate qui génère à des instants successifs les lettres C ou P. On s'intéressait alors particulièrement au temps d'attente Y d'une certaine séquence de lettres et à sa fonction génératrice. Le problème comportait 5 parties largement indépendantes les unes des autres. La partie I consistait à retrouver les résultats du cours sur la loi géométrique en probabilité. La partie II établissait des résultats préliminaires d'analyse sur les séries entières. En partie III, un cas intermédiaire était abordé : son étude nécessitait d'établir une formule de récurrence linéaire d'ordre 2 et les propriétés de Y en découlaient assez simplement. Des résultats particuliers d'algèbre linéaire étaient ensuite établis en partie IV afin d'étudier un cas plus complet en partie V.
La problématique du sujet était largement inspirée par l'étude des automates telle qu'elle est abordée par Graham, Knuth et Patashnik dans le livre « Mathématiques concrètes ».
Dans son ensemble, le sujet cherchait à évaluer les candidats sur les connaissances d'une partie la plus large possible du programme. La difficulté des questions se voulait progressive et de nombreuses questions correspondaient à des applications directes du cours ; les questions ouvertes n'étaient pas bloquantes si bien qu'un candidat n'ayant pas trouvé le bon résultat à une question ouverte pouvait continuer sa composition sans difficulté. Le sujet se donnait également pour but d'évaluer chez les candidats les six compétences explicitement détaillées dans le programme de PCSI-PC. La longueur relative du sujet était un point essentiel d'évaluation de l'efficacité des candidats.

Problèmes constatés par les correcteurs :

L'épreuve de cette année s'est avérée très discriminante. Un nombre plus conséquent de candidats a su mobiliser ses connaissances et ses compétences afin de réussir pleinement les deux tiers du problème. En revanche, le reste des candidats s'est retrouvé assez rapidement en difficulté et n'a pu réussir correctement que les questions du problème faisant intervenir des compétences de manipulation élémentaire des outils mathématiques ou de représentation.

La présentation écrite est très bonne, en règle générale. La mise en œuvre de stratégies pour répondre correctement à un maximum de questions a encore progressé si bien que les copies traitant sérieusement moins de 15 questions sont quasiment inexistantes. Les candidats ayant traité la quasi-totalité du sujet sont moins rares cette année.

La tonalité « probabiliste » du sujet a pu faire peur à certains candidats mais la variété des parties du programme présentées a toujours permis d'évaluer ces candidats sur leurs compétences en analyse et en algèbre.

Tout cela ne doit pas pour autant occulter les graves lacunes qu'ont rencontrées les correcteurs cette année encore. On constate notamment :

- une maîtrise très insuffisante du socle des connaissances ; les définitions moins courantes sont souvent oubliées ;
- les candidats s'engagent souvent dans des calculs sans prendre de recul alors que le résultat peut être obtenu avec beaucoup plus de concision ;
- les liens logiques entre les différentes parties d'un même raisonnement ne sont pas toujours très clairs ;
- beaucoup de candidats essayent de « bluffer » le correcteur en faisant croire qu'ils ont obtenu honnêtement le résultat ; cette attitude n'a pas manqué d'être préjudiciable à l'évaluation globale de la copie.

Sur quelques points du programme de PCSI-PC, nous ajouterons les remarques détaillées suivantes :

- en probabilités, l'indépendance des expériences dans la loi géométrique, la définition précise d'un système complet d'évènements, la formule des probabilités totales et la croissance de l'espérance doivent être mieux connues ;
- en algèbre, la définition de vecteur propre et le lien avec sa valeur propre associée, la multilinéarité du déterminant, les propriétés de la transposée d'une matrice, le théorème de d'Alembert-Gauss et la multiplicité d'une racine doivent être mieux compris ;
- en analyse, la manipulation de la règle de d'Alembert pour les séries entières, les équivalents simples, la manipulation d'inégalités, les propriétés de la valeur absolue, le produit de Cauchy, le développement en série entière et la dérivabilité d'un quotient de fonctions doivent être mieux assimilés.

2/ REMARQUES SPECIFIQUES :

Partie I (Etude d'un cas simple) :

Q1. La loi géométrique a été reconnue dans la très grande majorité des cas mais l'indépendance des expériences de Bernoulli donnant lieu à un temps d'attente est très rarement évoquée.
Q2. La maîtrise des arguments permettant de conclure à la valeur du rayon de convergence n'est pas au rendez-vous. Beaucoup utilisent la règle de d'Alembert pour les séries (souvent sans précaution avec le cas $x = 0$) et affirment que ce critère donne une condition nécessaire et suffisante de convergence de la série entière. L'oubli des valeurs absolues est fréquent. L'inégalité $1/p>1$ n'est pas souvent traitée.
Q3. La justification de la dérivabilité est souvent laborieuse.
Q4. La formule donnant la variance de Y à l'aide des dérivées de G en 1 est mal connue.

Partie II (Séries entières) :

Q5. Même genre de problème qu'en **Q2**.
Q6 et Q7. Bien réussies.
Q8. Réussie seulement dans une bonne moitié des cas et rarement bien justifiée. Confusions entre suites et constantes, entre inégalité et négligeabilité, disparition des constantes multiplicatives dans l'équivalent simple.
Q9. La question pouvait se faire avec **Q8** pour le rayon de convergence et avec le produit de Cauchy pour l'égalité.
Pour l'égalité, les candidats ont souvent opté pour la version calculatoire à l'aide de **Q7** sans reconnaître de produit de Cauchy.
Q10 et Q11. De graves confusions entre les notions « développable en série entière au voisinage de 0 » et « fonction prolongeable par continuité en 0 » ou encore « fonction de classe C^∞ ».

Partie III (Etude d'un cas intermédiaire) :

Q12. Les résultats sont souvent justes mais la décomposition des évènements en réunions et intersections est rarement faite.
Q13. La définition d'un système complet d'évènements est très rarement donnée correctement. Beaucoup de confusions avec une conséquence de la définition : vérifier que la somme des probabilités vaut 1 n'est pas suffisant.
Q14. La formule des probabilités totales (dont l'emploi était largement suggéré par la question **Q13**) a été assez rarement utilisée. Les candidats ont donc souvent paraphrasé le résultat sans la moindre justification mathématique convaincante.
Q15. Ce calcul d'analyse un peu technique a posé des problèmes aux candidats faibles.
Q16. La question la mieux réussie de tout le problème.
Q17. Une question simple du niveau du secondaire qui pouvait se faire avec concision grâce aux éléments indiqués pour faire gagner du temps. Malgré cela, certains candidats ont effectué le développement de $(t - a)(t - b)$ en se trompant dans le calcul.
Q18. Souvent traitée mais très mal réussie. Les candidats ont de gros problèmes de manipulation des valeurs absolues. La fonction valeur absolue est assimilée à une fonction croissante ou décroissante sur R suivant les besoins. L'inégalité $a>1$ est souvent parachutée après des calculs obscurs et sans issue.
Q19 et Q20. Le lien avec les questions précédentes est souvent fait.
Q21. Le raisonnement pour obtenir l'existence de l'espérance et de la variance est souvent confus. Le résultat final n'est pas toujours atteint par manque de maîtrise dans le calcul de la dérivée et dans les simplifications tenant compte du fait que $p = 1 - q$.
Q22. Bien réussie.
Q23. Grave confusion entre $E(Z) \geq E(Y)$ et $E(Z) \geq E(Y)+1$. Beaucoup d'arguments trop vagues sur le résultat en moyenne. Quasiment personne ne justifie que $Z \geq Y+1$ pour utiliser ensuite la croissance de l'espérance.

Partie IV (Algèbre linéaire) :

Q24. Que de perte de temps sur cette question. La plupart des candidats commence cette question en calculant le polynôme caractéristique qui est demandé en question suivante. La question faisait un tout et déterminer un vecteur propre associé à 0 permettait directement de vérifier que 0 était valeur propre.
Q25. Bien réussie.
Q26. Très bien réussie à part quelques candidats qui ne font pas attention au sens dans lequel les produits matriciels sont exécutés.
Q27. La propriété de multilinéarité n'étant connue que par une moitié des candidats, les autres ont utilisé la question **Q28** pour répondre à la question.

Q28. Beaucoup de candidats trouvent ce polynôme en refaisant un calcul direct et long de déterminant. En utilisant **Q27**, c'était immédiat dans le cas où t est différent de 0. Pour t = 0, c'est aussi immédiat par propriété de la matrice identité.
Q29. Cette question a souvent posé des problèmes. Le lien entre la non-nullité du déterminant et l'existence et unicité d'une solution du système n'est pas clair pour beaucoup.
Q30, Q31 et Q32. Bien réussies.
Q33. Pour réussir cette question, la plupart des candidats refait un calcul long de déterminant !
Q34. Le résultat était souvent parachuté. Avec quelques détails, on comprenait parfois que le candidat choisissait de prendre x_4 = 0.
Q35. Cette question a été mal réussie à cause de la confusion fréquente entre inégalité stricte et inégalité large. Dans le cas i), il est souvent utilisé que |x_1|<|x_3| pour aboutir au résultat.
Q36. Le théorème de d'Alembert-Gauss est parfois évoqué à travers le fait que tout polynôme complexe est scindé. Toutefois, il est très rare d'avoir l'argument sur la multiplicité de 0 dans le polynôme (confusion fréquente avec la dimension du sous-espace propre associé à 0).
Q37. La question a été rarement traitée. Le lien avec **Q27** et **Q36** a été vu mais il ne fallait pas oublier le cas t = 0.

Partie V (Etude d'un dernier cas) :

Q38 et Q39. Bien réussies.
Q40. La rédaction n'évoque quasiment jamais de système complet d'évènements ni de formule des probabilités totales.
Q41, Q42, Q43 et Q44. Bien réussies par ceux qui ont traité ces questions.
Q45. Le calcul est rarement bien maîtrisé jusqu'à une formule juste.
Q46. La représentation du schéma a été réussie par ceux qui ont traité cette question. En revanche, la matrice associée était rarement correcte. La méthode générale a été bien comprise par les rares candidats à avoir traité cette dernière question.