Analyse, Probabilité et Statistiques M1 MEEF **ML. Chabanol et C. Menini**

Eléments de correction du devoir surveillé du mardi 17 décembre 2013

Exercice 1

On suppose que par rapport à une certaine maladie, les individus peuvent être dans 3 états : soit ils ne sont pas malades mais ne sont pas immunisés (état J), soit ils sont malades (état M), soit ils sont immunisés (état I).

On suppose que pendant une certaine période de leur vie, chaque mois :

- *Les individus dans l'état J ont une probabilité de 0.02 de passer à l'état M, et ne peuvent pas passer directement à l'état I.*
- *Les individus dans l'état M ont une probabilité 0.1 de rester malade, et une probabilité 0.05 de guérir sans être immunisés*
- *Les individus immunisés le restent.*

On suppose qu'on peut modéliser la situation par une chaîne de Markov.

1. *Dessiner le graphe correspondant, et déterminer les composantes irréductibles. Quels sont les états transitoires ?*

Il faut faire attention de bien faire figurer une transition de M à I de poids 0.85. Attention aussi à ne pas oublier la boucle allant de I à I... Il y a deux composantes irréductibles : $\{J, M\}$ et $\{I\}$. Les deux états J et M sont transitoires.

2. *Donner la matrice A associée à cette chaîne de Markov.*

En numérotant les états dans l'ordre J, M, I, et en notant X_n l'état au mois numéro n, la matrice dont les éléments sont $a_{i,j} = P(X_{n+1} = i | X_n = j)$ est

$$A = \begin{pmatrix} 0.98 & 0.05 & 0 \\ 0.02 & 0.1 & 0 \\ 0 & 0.85 & 1 \end{pmatrix}$$

3.*On considère un individu dans l'état J à l'instant initial. Quelle est la loi de probabilité de son état de santé au bout de 2 mois ? Quelle est en fonction de A et de n, la loi de probabilité de son état de santé au bout de n mois ?*

La loi de X_0 est donc donnée par le vecteur colonne $\mu_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, et la loi de l'état de santé au bout de deux mois est $A^2.\mu_0 = \begin{pmatrix} 0.98^2 + 0.05 \times 0.02 \\ 0.98 \times 0.02 + 0.02 \times 0.1 \\ 0.02 \times 0.85 \end{pmatrix}$ La loi de l'état de santé au bout de n mois est donnée par $A^n.\mu_0$, c'est donc la première colonne de la matrice A^n.

Exercice 2 *On exprimera si nécessaire les réponses en termes de la fonction quantile q de la loi normale centrée réduite ou de sa fonction de répartition ϕ. On pourra le cas échéant utiliser $q_{0.975} \simeq 1.96$.*

1. *Soit X_n une suite de variables aléatoires suivant une loi binomiale de paramètres (n, p), avec $0 < p < 1$. Déterminer $p_0 = \lim_{n\to+\infty} P\left(p \in [\frac{X_n}{n} - \frac{1}{\sqrt{n}}, \frac{X_n}{n} + \frac{1}{\sqrt{n}}]\right)$.*

$$p \in [\frac{X_n}{n} - \frac{1}{\sqrt{n}}, \frac{X_n}{n} + \frac{1}{\sqrt{n}}] \Leftrightarrow \left|\frac{X_n}{n} - p\right| \leq \frac{1}{\sqrt{n}} \Leftrightarrow \left|\frac{X_n - np}{\sqrt{np(1-p)}}\right| \leq \frac{1}{\sqrt{p(1-p)}}$$

Or d'après le théorème de Moivre-Laplace,
$\frac{X_n - np}{\sqrt{np(1-p)}}$ converge en loi vers la loi normale centrée réduite, donc $p_0 = P(|Z| \leq \frac{1}{\sqrt{p(1-p)}})$ où Z suit une loi normale centrée réduite. Finalement, $p_0 = 2\phi(\frac{1}{\sqrt{p(1-p)}}) - 1$.

2. *Montrer que* $p_0 > 0.95$.

Pour tout $p \in]0,1[$, on vérifie aisément que $p(1-p) \leq \frac{1}{4}$, et donc que $\sqrt{p(1-p)}$ est inférieur ou égal à $1/2$ (maximum atteint pour $p = 1/2$, au sommet de la parabole d'équation $y = x(1-x)$). La fonction de répartition de la loi normale centrée réduite est une fonction strictement croissante donc

$$p_0 \geq 2\phi(2) - 1 > 2 \times 0.975 - 1 \quad \text{et} \quad 2 \times 0.975 - 1 = 0.95$$

($\phi(2) > 0.975$ car $2 > 1.96 \simeq q_{0.975}$).

3. *Justifier l'existence d'un entier* n_0 *tel que pour tout* $n \geq n_0$, $P\left(p \in [\frac{X_n}{n} - \frac{1}{\sqrt{n}}, \frac{X_n}{n} + \frac{1}{\sqrt{n}}]\right) \geq 0.95$

Notons $p_n = P\left(p \in [\frac{X_n}{n} - \frac{1}{\sqrt{n}}, \frac{X_n}{n} + \frac{1}{\sqrt{n}}]\right)$. L'existence de n_0 résulte de la convergence de la suite (p_n) vers p_0. En prenant $\epsilon = \frac{p_0 - 0.95}{2} > 0$ et avec la définition de la limite

$$\exists n_0 \ \forall n \geq n_0 \quad 0.95 < p_0 - \epsilon \leq p_n \leq p_0 + \epsilon.$$

4. *On souhaite obtenir un intervalle de confiance pour une proportion p inconnue de personnes présentant un certain gène. On va pour cela effectuer une étude sur un sous-échantillon de n individus (qu'on supposera obtenu en choisissant ces n personnes au hasard indépendamment avec remise...).*
Quelle taille de l'échantillon pourrait-on prendre pour être sûr de pouvoir donner un intervalle de confiance asymptotique au niveau 0.95 de largeur inférieure à 0.02 ?
Les conditions dans lesquelles les personnes sont choisies conduisent à supposer que la variable aléatoire X_n qui à un échantillon de taille n associe le nombre de personnes présentant le gène étudié, suit une loi binomiale $\mathcal{B}(n,p)$.
Un intervalle de confiance asymptotique au seuil 0.95 est $[\frac{X_n}{n} - \frac{1}{\sqrt{n}}, \frac{X_n}{n} + \frac{1}{\sqrt{n}}]$, il est de largeur $\frac{2}{\sqrt{n}}$. Il est de largeur inférieure à 0.02 dès que $n \geq 10000$.

Problème 1
La définition donnée d'une fonction concave en Terminale ES, pour une fonction dérivable, est la suivante : Une fonction dérivable sur un intervalle I est dite concave sur cet intervalle si sa courbe représentative est entièrement située en-dessous de chacune de ses tangentes.

Dans la suite du problème f est toujours supposée dérivable.

Partie I. *Deux propriétés des fonctions concaves dérivables*
1. *Montrer que* f *est concave sur l'intervalle* I *si et seulement si*

$$\forall (x,y) \in I^2 \quad f(y) \leq f'(x)(y-x) + f(x) \tag{1}$$

Une équation de la tangente à la courbe représentative de f au point de coordonnées $(a, f(a))$ est $y = f'(a)(x-a) + f(a)$ et la courbe représentative de f est sous cette tangente si et seulement si

$$\forall b \in I \quad f(b) \leq f'(a)(b-a) + f(a).$$

Soit le résultat puisque ceci doit être vérifié pour tout a de I.

2. *Montrer que si l'inégalité (1) est vérifiée alors*

$$\forall (x,y) \in I^2, \quad x < y \Rightarrow f'(y) \leq \frac{f(y) - f(x)}{y - x} \leq f'(x) \tag{2}$$

Qu'en déduisez-vous sur la fonction dérivée f' *?*
De l'inégalité (1) on déduit pour $x < y$ (c'est-à-dire $y - x > 0$)

$$\frac{f(y) - f(x)}{y - x} \leq f'(x).$$

L'inégalité (1) étant vérifiée pour tout (x,y) de I^2, on en déduit en échangeant le rôle des variables muettes x et y

$$f(x) \leq f'(y)(x-y) + f(y)$$

puis toujours avec $x < y$ (c'est-à-dire $x - y < 0$)

$$\frac{f(x) - f(y)}{x - y} \geq f'(y).$$

Nous en déduisons que la dérivée f' est une fonction décroissante sur I.

3. *Soit f dérivable sur l'intervalle I. On suppose sa dérivée f' décroissante sur I. En étudiant pour x_0 fixé dans I la fonction auxiliaire*

$$\begin{array}{rcccl} \varphi & : & I & \to & \mathbb{R} \\ & & y & \mapsto & f'(x_0)(y - x_0) + f(x_0) - f(y) \end{array}$$

montrer que l'on a l'inégalité (1).

φ est dérivable sur I comme combinaison linéaire de fonctions dérivables et $\varphi'(y) = f'(x_0) - f'(y)$. $\varphi'(y) \geq 0$ équivaut à $y \geq x_0$ car f' est décroissante. La fonction φ est décroissante sur $I \cap]-\infty, x_0]$ et croissante sur $I \cap [x_0, +\infty[$, elle admet un minimum en x_0 et $\varphi(x_0) = 0$.

Nous avons donc pour tout y de I

$$f(y) \leq f'(x_0)(y - x_0) + f(x_0)$$

ceci étant de plus vrai pour tout x_0 de I, l'inégalité (1) est satisfaite.

4. *Enoncer la propriété des fonctions concaves dérivables, qui vient d'être montrée à ce stade du problème.*

On a montré qu'une fonction f dérivable sur un intervalle I est concave sur I si et seulement si sa dérivée f' est une fonction décroissante sur I.

5. *Montrer que la fonction* ln *est concave.*

La fonction ln est définie et dérivable sur $I =]0, +\infty[$, de dérivée la fonction $x \mapsto \frac{1}{x}$. La fonction dérivée de ln est décroissante sur I donc ln est concave sur I.

6. *Le but de cette question est de démontrer que la courbe représentative d'une fonction dérivable et concave est située au-dessus de chacune de ses cordes, c'est-à-dire que l'on a*

$$\forall (a, b) \in I^2 \ \forall t \in [0, 1] \quad (1 - t)f(a) + tf(b) \leq f\left((1 - t)a + tb\right).$$

On suppose $a < b$. Étudier sur $[0, 1]$ les variations de la fonction

$$t \mapsto f\left((1 - t)a + tb\right) - (1 - t)f(a) - tf(b)$$

et en déduire l'inégalité attendue. (Ind. : on pourra penser à utiliser le théorème des accroissements finis)

Notons h cette fonction, elle est dérivable sur $[0, 1]$ comme composée et combinaison linéaire de fonctions dérivables,

$$\forall t \in [0, 1] \quad h'(t) = (b - a)f'\left((1 - t)a + tb\right) - (f(b) - f(a)) = (b - a)\left[f'\left((1 - t)a + tb\right) - \frac{f(b) - f(a)}{b - a}\right].$$

f est continue sur $[a, b]$ et dérivable sur $]a, b[$, avec le théorème des accroissements finis

$$\exists c \in]a, b[\ : \ \frac{f(b) - f(a)}{b - a} = f'(c).$$

Puisque c est dans $]a, b[$, il existe $t_0 = \frac{c-a}{b-a}$ dans $]0, 1[$ tel que $c = (1 - t_0)a + t_0 b$. Ainsi

$$\begin{array}{l} t \leq t_0 \Rightarrow (1 - t)a + tb \leq c \Rightarrow f'\left((1 - t)a + tb\right) \geq f'(c) \\ t \geq t_0 \Rightarrow (1 - t)a + tb \geq c \Rightarrow f'\left((1 - t)a + tb\right) \leq f'(c) \end{array}$$

h est croissante sur $[0, t_0]$ et décroissante sur $[t_0, 1]$. $h(0) = h(1) = 0$ donc h est positive sur $[0, 1]$ et la courbe représentative de f est au dessus de chacune de ses cordes.

Partie II. *Encadrement de l'intégrale d'une fonction concave*

Le but de cette partie est de montrer le résultat suivant

Soit une fonction f concave et dérivable sur un intervalle I, soient $x_0 \in I$ et $h > 0$ tels que $[x_0-h, x_0+h] \subset I$, alors

$$\frac{h}{2}\left[f(x_0-h)+2f(x_0)+f(x_0+h)\right] \leq \int_{x_0-h}^{x_0+h} f(u)\,du \leq 2hf(x_0) \tag{3}$$

7. *Soit une fonction f concave et dérivable sur un intervalle I, soient $x_0 \in I$ et $h > 0$ tels que $[x_0-h, x_0+h] \subset I$.*
En utilisant l'inégalité (1) pour $x = x_0$, montrer que

$$\int_{x_0-h}^{x_0+h} f(u)\,du \leq 2hf(x_0)$$

En prenant $x = x_0$ et $y = u$ dans l'inégalité (1) et avec la positivité de l'intégrale

$$\begin{aligned}\int_{x_0-h}^{x_0+h} f(u)\,du &\leq \int_{x_0-h}^{x_0+h} \big(f'(x_0)(u-x_0)+f(x_0)\big)\,du \\ &\leq \left[f'(x_0)\tfrac{(u-x_0)^2}{2}\right]_{x_0-h}^{x_0+h} + f(x_0)\times 2h \\ &\leq 0+2hf(x_0)\end{aligned}$$

8. *Montrer que*

$$\int_{x_0-h}^{x_0} f(u)\,du = h\int_0^1 f((1-t)(x_0-h)+tx_0)\,dt$$

puis que

$$\int_{x_0-h}^{x_0} f(u)\,du \geq \frac{h}{2}\left(f(x_0-h)+f(x_0)\right)$$

L'égalité est obtenue en faisant le changement de variables $u = (1-t)(x_0-h)+tx_0$ (l'application affine $t \mapsto (1-t)(x_0-h)+tx_0$ est bien une bijection de $[0,1]$ sur $[x_0-h, x_0]$). En utilisant que la courbe représentative de f est au dessus de chacune de ses cordes (question 6) et la positivité de l'intégrale, on obtient

$$\begin{aligned}h\int_0^1 f((1-t)(x_0-h)+tx_0)\,dt &\geq h\int_0^1 \left((1-t)f(x_0-h)+tf(x_0)\right)\,dt \\ &\geq hf(x_0-h)\left[-\tfrac{(1-t)^2}{2}\right]_0^1 + hf(x_0)\left[\tfrac{t^2}{2}\right]_0^1 \\ &\geq \tfrac{h}{2}\left(f(x_0-h)+f(x_0)\right)\end{aligned}$$

9. *Montrer que*

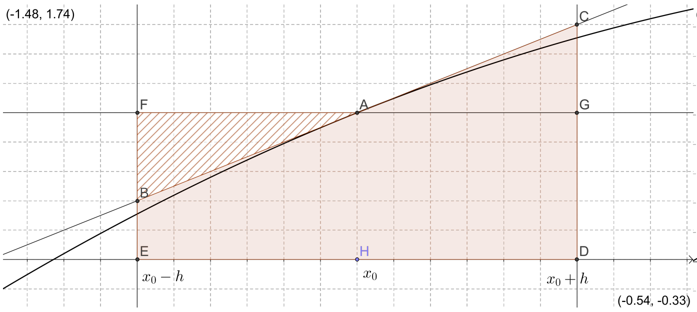
$$\int_{x_0}^{x_0+h} f(u)\,du \geq \frac{h}{2}\left(f(x_0)+f(x_0+h)\right)$$

Découle de l'inégalité précédente en remplaçant x_0-h par x_0 et x_0 par x_0+h puisque la fonction f est encore concave sur $[x_0, x_0+h]$ (on peut aussi faire le changement de variable $u = (1-t)x_0+t(x_0+h)$ et calquer la preuve faite à la question précédente).
10. *(bonus)Si l'on suppose de plus f positive, donner une interprétation géométrique de l'inégalité (3).*
f étant positive $A = \int_{x_0-h}^{x_0+h} f(u)\,du$ est l'aire (en unité d'aire) du domaine compris entre l'axe des abscisses, la courbe représentative de la fonction f, les droites d'équations $x = x_0-h$ et $x = x_0+h$.
La courbe étant sous ses tangentes, cette aire A est inférieure ou égale à l'aire du trapèze $BEDC$ qui est aussi celle du rectangle $FEDG$ car les triangles AFB et AGC ont même aire (cf. A milieu de $[FG]$). Ceci est illustré sur la figure (a) et nous donne l'inégalité

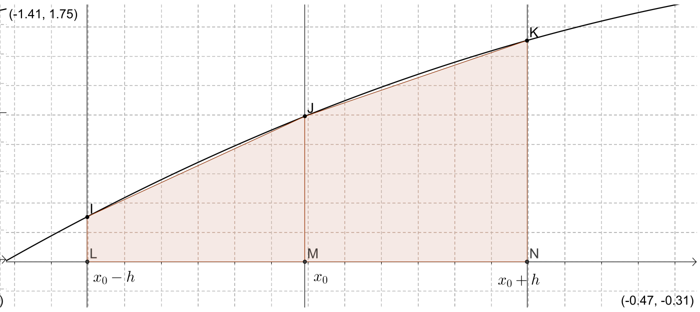
$$\int_{x_0-h}^{x_0+h} f(u)\,du \leq 2hf(x_0).$$

La courbe étant au dessus de ses cordes, cette aire A est supérieure ou égale à la sommes des aires des deux trapèzes $ILMJ$ et $JMNK$. Ceci est illustré sur la figure (b) et nous donne l'inégalité

$$h\left[\frac{f(x_0-h)+f(x_0)}{2}+\frac{f(x_0)+f(x_0+h)}{2}\right] \leq \int_{x_0-h}^{x_0+h} f(u)\,du.$$



(a) tangente



(b) cordes

Partie III. *Encadrement de n!*
Avec les résultats des parties I et II, nous avons que pour tout entier $k \geq 2$

$$\frac{1}{4}\left[\ln(k-\frac{1}{2}) + 2\ln(k) + \ln(k+\frac{1}{2})\right] \leq \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \ln(x)\,dx \leq \ln(k)$$

ceci pourra être utilisé sans démonstration.
11. *Montrer que pour tout entier $n \geq 2$*

$$\sum_{k=2}^{n} \ln(k) + \frac{1}{4}\sum_{k=2}^{n} \ln\left(\frac{4k^2-1}{4k^2}\right) \leq \int_{\frac{3}{2}}^{n+\frac{1}{2}} \ln(x)\,dx \leq \sum_{k=2}^{n} \ln(k)$$

En sommant k de 2 à n dans l'inégalité rappelée en début de partie III, nous avons avec la relation de Chasles

$$\frac{1}{4}\sum_{k=2}^{n} 4\ln(k) + \frac{1}{4}\sum_{k=2}^{n}\left[\ln(k-\frac{1}{2}) - 2\ln(k) + \ln(k+\frac{1}{2})\right] \leq \int_{\frac{3}{2}}^{n+\frac{1}{2}} \ln(x)\,dx \leq \sum_{k=2}^{n} \ln(k)$$

puis le résultat attendu en remarquant que

$$\ln(k-\frac{1}{2}) - 2\ln(k) + \ln(k+\frac{1}{2}) = \ln\left(\frac{(k-\frac{1}{2})(k+\frac{1}{2})}{k^2}\right) = \ln\left(\frac{4k^2-1}{4k^2}\right)$$

12. *Montrer que pour tout $u > 0$, $\ln(u) \leq u - 1$; en déduire que pour tout entier $n \geq 2$*

$$\sum_{k=2}^{n} \ln\left(\frac{4k^2}{4k^2-1}\right) \leq \frac{1}{2}\sum_{k=2}^{n}\left(\frac{1}{2k-1} - \frac{1}{2k+1}\right) \leq \frac{1}{6}$$

Nous avons vu dans la partie I que la fonction ln est concave sur $]0,+\infty[$, la courbe représentative de ln est sous sa tangente au point de coordonnées $(1,0)$, c'est-à-dire

$$\forall u > 0 \quad \ln(u) \leq u - 1.$$

Nous en déduisons que pour tout entier k compris entre 2 et n

$$\ln\left(\frac{4k^2}{4k^2-1}\right) \leq \frac{4k^2}{4k^2-1} - 1 \leq \frac{1}{4k^2-1} \leq \frac{1}{2}\left(\frac{1}{2k-1} - \frac{1}{2k+1}\right).$$

Nous avons une somme télescopique,

$$\sum_{k=2}^{n}\left(\frac{1}{2k-1} - \frac{1}{2k+1}\right) = \frac{1}{3} - \frac{1}{2n+1} \leq \frac{1}{3}$$

13. *Calculer $\int_{\frac{3}{2}}^{n+\frac{1}{2}} \ln(x)\,dx$*

En faisant une intégration par parties

$$\int_{\frac{3}{2}}^{n+\frac{1}{2}} \ln(x)\,dx = [x\ln(x) - x]_{\frac{3}{2}}^{n+\frac{1}{2}} = (n+\frac{1}{2})\ln(n+\frac{1}{2}) - n - \frac{3}{2}\ln(\frac{3}{2}) + 1.$$

14. *Déduire de ce qui précède qu'il existe existe des constantes $C_1 > 0$ et $C_2 > 0$ telles que pour tout entier $n \geq 2$,*

$$C_1 e^{-n}\left(n+\frac{1}{2}\right)^{n+\frac{1}{2}} \leq n! \leq C_2 e^{-n}\left(n+\frac{1}{2}\right)^{n+\frac{1}{2}}$$

Avec les questions 11, 12 et 13 et en remarquant que pour $n \geq 2$, $\sum\limits_{k=2}^{n} \ln(k) = \ln(n!)$, nous avons que

$$(n+\frac{1}{2})\ln(n+\frac{1}{2}) - n - \frac{3}{2}\ln(\frac{3}{2}) + 1 \leq \ln(n!) \leq (n+\frac{1}{2})\ln(n+\frac{1}{2}) - n - \frac{3}{2}\ln(\frac{3}{2}) + 1 + \frac{1}{6}$$

on conclue en utilisant la croissance de la fonction exponentielle, les constantes $C_1 = \frac{2\sqrt{6}}{9}e$ et $C_2 = e^{1/6}C_1$ conviennent.

Problème 2.
Partie I. *Fonction génératrice d'une variable aléatoire à valeurs dans $\mathbb{N}$*
Pour une variable aléatoire réelle X à valeurs dans $\mathbb{N}$, on pose, pour tout réel t pour lequel cela a un sens (c'est-à-dire pour lequel la série converge),

$$g_X(t) = E(t^X) = \sum_{k=0}^{+\infty} P(X=k)t^k$$

où E désigne l'espérance. La fonction g_X est appelée fonction génératrice *de la variable aléatoire X. Dans la suite du problème on considèrera que $0^0 = 1$, ce qui permet ici par exemple d'affirmer ici que $g_X(0) = P(X=0)$.*
1. *Pour chacun des cas suivants, donner l'ensemble de définition de g_X et la déterminer :*
1.a. *On suppose que X suit une loi binomiale $\mathcal{B}(n,p)$.*
Pour $k \geq n+1$ $P(X=k) = 0$, nous avons une somme finie donc g_X est définie sur $\mathbb{R}$ et

$$g_X(t) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} t^k = (pt+1-p)^n$$

avec la formule du binôme.
1.b. *On suppose que X suit une loi géométrique $\mathcal{G}(p)$ de paramètre $p \in]0,1[$ (on pourra noter $q = 1-p$). (On rappelle qu'on a alors pour tout $k \in \mathbb{N}^*$, $P(X=k) = q^{k-1}p$).*
Lorsque cela a un sens

$$g_X(t) = \sum_{k=1}^{+\infty} q^{k-1}pt^k = pt\sum_{k=1}^{+\infty}(qt)^{k-1}.$$

On reconnait une série géométrique de raison qt, elle converge si et seulement si $|qt| < 1$. g_X est définie sur $]-\frac{1}{q}, \frac{1}{q}[$ et

$$g_X(t) = \frac{pt}{1-qt}$$

1.c. *On suppose que X suit une loi de Poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda > 0$ (on rappelle qu'on a alors pour tout entier k, $P(X=k) = \exp(-\lambda)\frac{\lambda^k}{k!}$).*
Lorsque cela a un sens

$$g_X(t) = \sum_{k=0}^{+\infty} \exp(-\lambda)\frac{\lambda^k}{k!}t^k.$$

$\frac{\lambda^{k+1}}{(k+1)!}\frac{k!}{\lambda^k} = \frac{\lambda}{k+1}$ d'où $\lim\limits_{k\to+\infty} \frac{\lambda^{k+1}}{(k+1)!}\frac{k!}{\lambda^k} = 0$, la série entière a un rayon de convergence égal à $+\infty$. g_X est définie sur $\mathbb{R}$ et

$$g_X(t) = \exp(-\lambda)\sum_{k=0}^{+\infty}\frac{(t\lambda)^k}{k!} = \exp(\lambda(t-1))$$

Remarque : on pouvait aussi directement reconnaitre le développement en série entière de la fonction exponentielle au point λt et en déduire que g_X est définie sur $\mathbb{R}$.

On revient au cas général.

2. *Montrer que $g_X(t)$ est bien définie pour tout $t \in [-1,1]$ et donner la valeur de $g_X(1)$.*

$$\forall t \in [-1,1]\ \forall k \in \mathbb{N} \quad |P(X=k)t^k| \leq P(X=k)$$

et $\sum\limits_{k=0}^{+\infty} P(X=k) = 1$. La série est absolument convergente, donc convergente pour $t \in [-1,1]$. Ainsi $g_X(t)$ est bien définie pour tout $t \in [-1,1]$ et on a vu de plus que $g_X(1) = \sum\limits_{k=0}^{+\infty} P(X=k) = 1$.

On rappelle qu'une des définitions du rayon de convergence d'une série entière $\sum_{n\geq 0} a_n t^n$ est

$$R = \sup\{r \geq 0 \mid \sum_{n\geq 0} |a_n| r^n < +\infty\}.$$

On rappelle également qu'une série entière est $\mathcal{C}^\infty$ sur $]-R,R[$.

3. *Expliquer pourquoi le rayon de convergence R de la série entière $\sum_{k=0}^{+\infty} P(X=k)t^k$ est supérieur ou égal à 1. Que vaut-il lorsque X est une variable aléatoire prenant un nombre fini de valeurs ?*

Ici $a_k = P(X=k)$ est positif pour tout entier k. On vient de voir à la question 3 que $g_X(1) = 1 < +\infty$ donc 1 est élément de $\{r \geq 0 \mid \sum_{n\geq 0}|a_n|r^n < +\infty\}$, R la borne supérieure de cet ensemble est donc supérieure ou égale à 1.

Lorsque X ne prend qu'un nombre fini de valeurs alors la somme est finie (comme dans le cas de la binomiale vue à la question 1.a) et la fonction génératrice est définie sur $\mathbb{R}$, dans ce cas $R = +\infty$.

4. *On suppose $R > 1$. Justifier que X admet une espérance et donner une relation entre $g'_X(1)$ et $E[X]$. Retrouver à l'aide de la question 1 la valeur de l'espérance d'une variable aléatoire suivant une loi géométrique $\mathcal{G}(p)$ de paramètre $p \in]0,1[$.*

La série entière, c'est-à-dire g_X, est une fonction $\mathcal{C}^\infty$ sur $]-R,R[$. Si $R > 1$ alors elle est en particulier dérivable au point 1 et sa dérivée est la somme des dérivées termes à termes.

$$\forall t \in]-R,R[\quad g'_X(t) = \sum_{k=1}^{+\infty} kP(X=k)t^{k-1}$$

soit pour $t=1$, $g'_X(1) = \sum_{k=1}^{+\infty} kP(X=k)$ est fini. On en déduit que X admet une espérance et que $E(X) = g'_X(1)$.

Nous avons vu à la question 1.b. que la fonction génératrice d'une variable aléatoire X suivant une loi géométrique $\mathcal{G}(p)$ est définie sur $]-\frac{1}{q},\frac{1}{q}[$, le rayon de convergence est donc dans ce cas $R = \frac{1}{q} > 1$. On retrouve que X admet une espérance et

$$\forall t \in]-\frac{1}{q},\frac{1}{q}[\quad g'_X(t) = \frac{p}{(1-qt)^2}$$

soit $E(X) = g'_X(1) = \frac{1}{p}$.

5. *On rappelle qu'il découle de la question 3 que g_X est $\mathcal{C}^\infty$ sur $]-1,1[$. Exprimer pour tout $n \in \mathbb{N}$ (en fonction des $P(X=k)$, $k \in \mathbb{N}$), $g_X^{(n)}(0)$ la dérivée nième de g_X en 0.*

En déduire que si X et Y sont deux variables aléatoires telles que $g_X = g_Y$ alors X et Y suivent la même loi.

Pour tout entier n, g_X est n fois dérivable sur $]-1,1[$ et sa dérivée n-ième est la sommes des dérivées n-ième terme à terme (propriété des séries entières). Ainsi

$$\forall n \in \mathbb{N}\ \forall t \in]-1,1[\quad g_X^{(n)}(t) = \sum_{k=n}^{+\infty} \frac{k!}{(k-n)!} P(X=k) t^{k-n}$$

et $g_X^{(n)}(0) = n! P(X=n)$.
Ainsi si $g_X = g_Y$, on a que pour tout entier n, $g_X^{(n)}(0) = n!P(X=n) = g_Y^{(n)}(0) = n!P(Y=n)$. Pour tout entier n, $P(X=n) = P(Y=n)$, les variables aléatoires X et Y suivent la même loi. On connait la loi d'une variable aléatoire discrète dès que l'on connait sa fonction génératrice.

Partie II *Somme d'un nombre aléatoire de variables aléatoires*

6. *Montrer que si X et Y sont deux variables aléatoires* indépendantes *à valeurs dans* $\mathbb{N}$ *alors pour tout* $t \in [-1,1]$, $g_{X+Y}(t) = g_X(t)g_Y(t)$.
Pour tout réel t (pour $t \neq 0$ c'est clair, on peut vérifier que c'est aussi vrai pour $t=0$), $g_{X+Y}(t) = E(t^{X+Y}) = E(t^X t^Y)$; les variables aléatoires X et Y étant indépendantes il en est de même pour t^X et t^Y et $E(t^X t^Y) = E(t^X)E(t^Y) = g_X(t)g_Y(t)$.

Dans la suite de cette partie, $(X_k)_{k\geq 1}$ désigne une suite de variables réelles à valeurs dans $\mathbb{N}$, indépendantes et de même loi. On pose $S_0 = 0$ et, pour tout $n \in \mathbb{N}^*$, $S_n = \sum_{k=1}^n X_k$. On note f la fonction génératrice commune à toutes les variables aléatoires X_k.

7. *Montrer que, pour tout $n \in \mathbb{N}$ et tout $t \in [-1,1]$, $g_{S_n}(t) = (f(t))^n$.*
Montrons par récurrence

$$(H_n)\ :\quad \forall t \in [-1,1] \quad g_{S_n}(t) = (f(t))^n$$

Pour $n=0$, $P(S_0=0)=1$ et $g_{S_0}(t) = 1 = (f(t))^0$ pour tout $t \in [-1,1]$. (H_0) est vérifiée.
Montrons que pour tout entier n, (H_n) implique (H_{n+1}).
Commençons par remarquer que S_n et X_{n+1} sont indépendantes. En effet :
Si $n=0$, S_0 est constante et donc indépendante de X_1.
Si $n \neq 0$, les variables aléatoires X_k, $k \in \mathbb{N}^*$ sont indépendantes, donc S_n et X_{n+1} sont indépendantes.
Avec la question 6 et (H_n)

$$\forall t \in [-1,1] \quad g_{S_{n+1}}(t) = g_{S_n+X_{n+1}}(t) = g_{S_n}(t)g_{X_{n+1}}(t) = (f(t))^n f(t) = (f(t))^{n+1}$$

(H_{n+1}) est vérifiée.
On a montré que (H_0) est vérifiée, que pour tout entier n, (H_n) implique (H_{n+1}), alors, (H_n) est vérifiée pour tout entier n.

On considère maintenant une variable aléatoire réelle N à valeurs dans $\mathbb{N}$, indépendante des variables X_k, $k \in \mathbb{N}$, et on définit $Y = S_N$ (il s'agit de la somme d'un nombre aléatoire de variables aléatoires). On admettra que Y est bien une variable aléatoire réelle à valeurs dans $\mathbb{N}$.
On notera, h la fonction génératrice de N, g la fonction génératrice de Y.
On cherche maintenant à déterminer g en fonction de f et h.

8. Justifier que pour tout entier naturel k

$$P(Y=k) = \sum_{n=0}^{+\infty} P[(Y=k)\cap(N=n)] = \sum_{n=0}^{+\infty} P[(S_n=k)\cap(N=n)]$$

N est à valeurs dans $\mathbb{N}$, donc $\{(N=n) \mid n \in \mathbb{N}\}$ réalise une partition de l'univers sur lequel est défini N, d'où

$$(Y=k) = \bigcup_{n\in\mathbb{N}} [(Y=k)\cap(N=n)]$$

avec $[(Y=k)\cap(N=n)]$, $n \in \mathbb{N}$, 2 à 2 disjoints. Par σ-additivité d'une probabilité

$$P(Y=k) = \sum_{n=0}^{+\infty} P[(Y=k)\cap(N=n)]$$

et l'on conclut en remarquant que $Y(\omega) = S_N(\omega) = S_n(\omega)$ pour tout $\omega \in (N = n)$.
9. *En déduire que pour tout $t \in [-1, 1]$*

$$g(t) = \sum_{k=0}^{+\infty} \left(\sum_{n=0}^{+\infty} P(S_n = k)\, P(N = n) \right) t^k.$$

Par définition

$$g(t) = \sum_{k=0}^{+\infty} P(Y = k) t^k = \sum_{k=0}^{+\infty} \left(\sum_{n=0}^{+\infty} P[(S_n = k) \cap (N = n)] \right) t^k$$

avec la question 8.
On conclue en utilisant l'indépendance de N avec les variables aléatoires X_k, $k \in \mathbb{N}$, et donc l'indépendance de N avec S_n pour tout entier n. Ainsi pour tous entiers k et n

$$P[(S_n = k) \cap (N = n)] = P(S_n = k)\, P(N = n).$$

10. *En déduire que, pour tout $t \in [-1, 1]$, $g(t) = (h \circ f)(t)$. (Ind. On pourra utiliser sans démonstration que si $\sum_{k=0}^{+\infty} \left(\sum_{n=0}^{+\infty} |a_{k,n}| \right) < +\infty$ alors*

$$\sum_{k=0}^{+\infty} \left(\sum_{n=0}^{+\infty} a_{k,n} \right) = \sum_{n=0}^{+\infty} \left(\sum_{k=0}^{+\infty} a_{k,n} \right)$$

Nous savons que pour tout $t \in [-1, 1]$

$$\sum_{k=0}^{+\infty} \left(\sum_{n=0}^{+\infty} P(S_n = k)\, P(N = n) |t|^k \right) \leq \sum_{k=0}^{+\infty} \left(P(S_n = k) \sum_{n=0}^{+\infty} P(N = n) \right) \leq 1.$$

Nous pouvons comme dans le rappel permuter les deux sommes et avec les questions 7 et 9, pour tout $t \in [-1, 1]$

$$g(t) = \sum_{n=0}^{+\infty} \left(P(N = n) \sum_{k=0}^{+\infty} P(S_n = k)\, t^k \right) = \sum_{n=0}^{+\infty} P(N = n) g_{S_n}(t) = \sum_{n=0}^{+\infty} P(N = n)(f(t)^n = (h \circ f)(t).$$

11. *Dans le cas particulier où X suit une loi de Bernoulli $\mathcal{B}er(p)$ de paramètre $p \in]0, 1[$ et N une loi de Poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda > 0$, déterminer la loi de $Y = S_N$.*
Avec les notations de la question 10, avec les fonctions génératrices trouvées aux questions 1.a (pour $n = 1$) et 1.c

$$f(t) = pt + 1 - p \quad h(t) = \exp(\lambda(t - 1)) \quad g(t) = (h \circ f)(t) = \exp(\lambda(pt - p)) = \exp(p\lambda(t - 1)).$$

Nous avons vu à la question 5 que la fonction génératrice d'une variable aléatoire caractérise sa loi. On reconnait ici la fonction génératrice d'une variable aléatoire qui suit une loi de Poisson de paramètre $p\lambda$, $Y = S_N$ suit une loi $\mathcal{P}(p\lambda)$.
12. *Dans le cas particulier où les variables aléatoires X et N prennent un nombre fini de valeurs, justifier que X, N et $Y = S_N$ admettent des espérances puis calculer $E(Y)$ en fonction de $E(X)$ et $E(N)$.*
X, N et $Y = S_N$ prennent un nombre fini de valeurs, nous avons vu à la question 3 que le rayon de convergence de leurs séries génératrices est $+\infty$ et qu'alors elles admettent une espérance (question 4).
Avec la composition et en remarquant que $f(1) = 1$

$$E(Y) = g'(1) = h'(f(1))f'(1) = h'(1)f'(1) = E(N)E(X).$$