

Devoir surveillé - mardi 17 décembre 2013

*Durée 4h - Documents non autorisés. Calculatrice autorisée.***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 ?
2. Donner la matrice A associée à cette chaîne de Markov.
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 ?

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 $p_0 = \lim_{n \rightarrow +\infty} P\left(p \in \left[\frac{X_n}{n} - \frac{1}{\sqrt{n}}, \frac{X_n}{n} + \frac{1}{\sqrt{n}}\right]\right)$.
 2. Montrer que $p_0 > 0.95$.
 3. Justifier l'existence d'un entier n_0 tel que pour tout $n \geq n_0$, $P\left(p \in \left[\frac{X_n}{n} - \frac{1}{\sqrt{n}}, \frac{X_n}{n} + \frac{1}{\sqrt{n}}\right]\right) \geq 0.95$
 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 ?

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

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

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

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

Qu'en déduisez-vous sur la fonction dérivée f' ?

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

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

montrer que l'on a l'inégalité (1).

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

5. Montrer que la fonction $\ln$ est concave.

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

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

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

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

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

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

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

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

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

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

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

9. Montrer que

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

10. (bonus) Si l'on suppose de plus f positive, donner une interprétation géométrique de l'inégalité (3).

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 $n \geq 2$

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

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

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

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

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

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

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 $\mathcal{B}(n, p)$.

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 $k \in \mathbb{N}^*$, $P(X = k) = q^{k-1}p$).

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 , $P(X = k) = \exp(-\lambda) \frac{\lambda^k}{k!}$).

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 $g_X(1)$.

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 ?

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 $p \in]0, 1[$.

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 g_X caractérise la loi de la variable aléatoire X .

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 $t \in [-1, 1]$, $g_{X+Y}(t) = g_X(t)g_Y(t)$.

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 $t \in [-1, 1]$, $g_{S_n}(t) = (f(t))^n$.

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

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

9. En déduire que pour tout $t \in [-1, 1]$

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

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

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

11. Dans le cas particulier où X suit une loi de Bernoulli $\mathcal{Ber}(p)$ de paramètre $p \in]0, 1[$ et N une loi de Poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda > 0$, déterminer la loi de $Y = S_N$.

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 $E(N)$.