ÉCOLES NORMALES SUPÉRIEURES
ÉCOLE NATIONALE DES PONTS ET CHAUSSÉES

CONCOURS D'ADMISSION SESSION 2013

FILIÈRE BCPST

COMPOSITION DE MATHÉMATIQUES

Épreuve commune aux ENS de Cachan, Lyon, Paris et de l'ENPC

Durée : 4 heures

L'utilisation des calculatrices n'est pas autorisée pour cette épreuve.

⋆ ⋆ ⋆

Ce problème se compose de quatre parties. Les réponses à certaines questions de la partie I sont utiles dans les parties II et III. La partie IV peut être traitée indépendamment des autres.

L'objet du problème est d'étudier différentes questions liées à l'étude d'un modèle simplifié d'infection d'un organisme constitué d'une chaîne de cellules par un parasite. Le modèle précis d'infection change selon les parties du problème.

Il est recommandé de lire l'ensemble de l'énoncé attentivement. Il est demandé de veiller au soin de la présentation, ainsi qu'à la rigueur et à la concision des raisonnements.

1 Parasites, Fonctions génératrices et loi de Poisson

Soit Z une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$. On rappelle qu'une variable X de loi de Poisson de paramètre λ a pour distribution

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \dots$$

Étant donné une variable aléatoire X à valeurs dans les entiers naturels, on définit la *fonction génératrice* de X, notée g_X, par

$$g_X(s) = \mathbb{E}[s^X].$$

1. Exprimer $g_Z(s)$ la fonction génératrice de Z sous forme d'une série. Pour ces valeurs calculer la limite pour donner une expression simple de $g_Z(s)$.

On admettra que si $f(s) := \sum_{i=0}^{\infty} a_i s^i$ est bien définie pour tout $s \in I$ intervalle ouvert alors

$$\forall s \in I : f'(s) = \sum_{i=1}^{\infty} i a_i s^{i-1}$$

est également bien définie (i.e. on peut dériver terme à terme).

2. Soit Y une variable aléatoire à valeurs dans $\mathbb{N}$. Montrer que si $g_Y^{(n)}$ est la dérivée n-ième de g_Y on a

$$g_Y^{(n)}(1) = \mathbb{E}[Y(Y-1)\dots(Y-n+1)], n \geq 1.$$

   En déduire une expression de l'espérance et de la variance de Y à l'aide de g_Y et de ses dérivées et retrouver les expressions de l'espérance et de la variance de la loi de Poisson de paramètre λ.

3. Exprimer $\mathbb{P}(Y = k)$ pour $k \in \mathbb{N}$ à l'aide des dérivées de la fonction g_Y et en déduire que la fonction génératrice caractérise complètement la loi de Y.

Dans toute la suite du problème on considère la situation suivante. Un organisme est constitué d'une longue chaîne de cellules (ou de segments) reliés les uns aux autres comme sur la figure suivante.

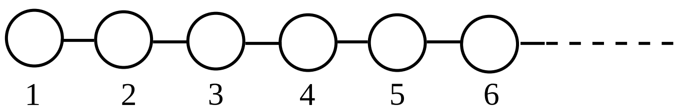

Figure 1: Une chaîne de cellules

La cellule -ou segment- numéro 1 est infectée par un parasite qui tente d'envahir l'organisme. Les parasites sont de deux types :

Actif : Seul le parasite initialement présent dans la cellule 1 est de ce type. Le parasite actif est le seul qui se déplace et qui peut se reproduire et infecter les cellules.

Passif : Les parasites passifs sont des descendants du parasite actif. Ils ne peuvent ni se reproduire ni se déplacer.

Lorsque le parasite actif est présent dans la cellule i il commence par produire Z_i descendants passifs qui s'installent dans la cellule i. Il tente ensuite d'infecter la cellule $i+1$ avec une probabilité de succès p. On suppose que les variables $Z_i, i = 1, 2, \ldots$ sont indépendantes et identiquement distribuées de loi commune une loi de Poisson de paramètre $\lambda > 0$. Lorsque le parasite actif échoue pour la première fois à envahir la cellule suivante il meurt et l'invasion s'arrête.

Un exemple possible d'invasion peut donc être décrit comme suit :

- Le parasite actif est dans la cellule 1. Il produit $Z_1 = 3$ descendants passifs.
- Le parasite actif réussit à envahir la cellule 2. Il y produit $Z_2 = 0$ descendants passifs.
- Le parasite actif réussit à envahir la cellule 3. Il y produit $Z_3 = 1$ descendants passifs.
- Le parasite actif échoue à envahir la cellule 4 et meurt. L'invasion s'arrête.

Dans cet exemple 3 cellules ont été envahies et le nombre total de parasites à la fin de l'invasion est $3+0+1=4$.

4. Soit N le nombre total de cellules qui ont été infectés lorsque l'invasion s'arrête. Donner la loi de N (commencer par donner $\mathbb{P}(N=1), \mathbb{P}(N=2)$ et $\mathbb{P}(N=3)$).
5. Quelle est la probabilité pour qu'une cellule infectée soit saine (i.e. ne contiennent aucune copie du parasite) ?
6. Soit M la dernière (i.e. celle avec le plus haute numéro) cellule infectée qui contient des parasites. On a bien sûr $M \leq N$ (mais on peut avoir $M < N$). Calculer la loi de $N - M$ i.e ($\mathbb{P}(N - M = k), k = 0, 1, \ldots$). Donner la loi jointe de (M, N). Les variables M et N sont-elles indépendantes ?

On s'intéresse à présent à la loi du nombre total de parasites dans les cellules

$$T := \sum_{i=1}^{N} Z_i.$$

7. Calculer la fonction génératrice de N.

8. Soient X et Y deux variables aléatoires indépendantes et à valeurs dans $\mathbb{N}$, de fonctions génératrices respectives g_X et g_Y. Exprimer g_{X+Y} la fonction génératrice de $X+Y$ en fonction de s, g_X et g_Y.

9. En déduire la loi de $X+Y$ lorsque si X suit une loi de Poisson de paramètre λ, Y suit une loi de Poisson de paramètre μ et X et Y sont indépendantes.

10. Soient $(Y_i)_{i\in\mathbb{N}}$ des variables aléatoires indépendantes et identiquement distribuées de fonction génératrice g. Soit encore une variable X indépendante des Y_i de fonction génératrice h. En sommant sur les différentes valeur possible de X montrer que la fonction génératrice de $W = \sum_{i=1}^{X} Y_i$ (par convention $W = 0$ si $X = 0$) est donnée par
$$\mathbb{E}[s^W] = h(g(s)) = h \circ g(s).$$

11. En déduire une expression de la fonction génératrice de T
$$g_T(s) = \mathbb{E}[s^{\sum_{i=1}^{N} Z_i}].$$

2 Un modèle modifié

On modifie a présent le modèle. Plus une cellule contient de parasites et plus elle a de chances d'infecter sa voisine.

Si la cellule i est infectée elle contient Z_i copies du parasite. Chacun de ces parasites a une probabilité p d'infecter la cellule $i+1$. Si au moins un parasite parmi les Z_i de la cellule i réussit, la cellule $i+1$ est infectée. Si $Z_i = 0$ ou si tous les parasites de la cellule i échouent l'invasion s'arrête. Initialement, la cellule 1 est infectée. On suppose de nouveau que les variables Z_i sont indépendantes et identiquement distribuées de loi de Poisson de paramètre λ. Un exemple possible d'invasion peut donc être décrit comme suit :

- La cellule 1 est infectée. Elle contient $Z_1 = 4$ copies.
- Au moins l'un de ces 4 parasites réussit à infecter cellule 2. Elle contient $Z_2 = 7$ copies.
- Au moins l'un de ces 7 parasites réussit à infecter cellule 3. Elle contient $Z_3 = 2$ copies.
- Aucun de ces 2 parasites ne réussit à infecter cellule 4. L'invasion s'arrête.

Dans cet exemple, trois cellules ont été infectées et le nombre total de parasites lorsque l'invasion s'arrête est $4 + 7 + 2 = 13$.

12. Quelle est la probabilité pour que la cellule 2 ne soit pas infectée. Quelle est la probabilité que la cellule $n+1$ soit infectée sachant que la cellule n est infectée. Donner la loi de N.

Soit l'évènement $I_k := \{$ la cellule k est envahie $\} = \{N \geq k\}$ où N est toujours le nombre de cellules infectées. Pour un évènement A on note $\mathbf{1}_A$ la fonction caractéristique de A, i.e. la variable qui vaut 1 si A se réalise et 0 sinon. La notation $A^{\complement}$ désigne l'évènement complémentaire de A.

13. Donner la loi jointe du couple de variables $(Z_1, \mathbf{1}_{I_2})$. Calculer les $\tilde{p}_k := \mathbb{P}(Z_1 = k|I_2)$ pour $k = 1, 2, \ldots$ Donner l'expression de $\tilde{g}_Z$ la fonction génératrice qui correspond à la distribution $(\tilde{p}_k)_{k\geq 1}$.
14. Calculer les $\bar{p}_k := \mathbb{P}(Z_1 = k|I_2^{\complement}), k = 0, 1, \ldots$ Donner l'expression de $\bar{g}_Z$ la fonction génératrice qui correspond à la distribution $(\bar{p}_k)_{k\geq 1}$.
15. montrer que pour tout $i \geq 1, j \geq 0$ et $n \geq 1$ on a
$$\mathbb{P}(Z_n = i|\{Z_{n+1} = j\} \cap I_{n+1}) = \mathbb{P}(Z_n = i|I_{n+1}).$$

On vient de montrer que conditionnellement à I_{n+1} les variables Z_n et Z_{n+1} sont indépandantes. Plus généralement, on peut en déduire que conditionnellement à $N = n$ les variables $Z_1, Z_2, \ldots, Z_n$ forment une famille ind´pandante de lois respectives $(\tilde{p}_i)_{i\in\mathbb{N}}$ pour les variables $Z_1, \ldots, Z_{n-1}$ et $(\bar{p}_i)_{i\in\mathbb{N}}$ pour Z_n.

16. Soit T le nombre total de copies du parasites présentes dans les cellules, $T = \sum_i^N Z_i$. Donner une expression de
$$g_T(s) = \mathbb{E}[s^T]$$
à l'aide de $g_N, \tilde{g}_Z$ et $\bar{g}_Z$.
17. On cherche à distinguer lequel des deux modèle est le meilleur. Pour cela on réalise l'expérience suivante : on introduit la parasite dans la cellule 1 de l'organisme et on laisse l'infection se développer. Lorsque l'invasion s'arrête on compte le nombre de parasites présents dans chaque cellule. Par contre on ne sait pas distinguer une cellule infectée mais saine (sans parasites) d'une cellule non-infectée. On suppose que l'on est capable de répéter cette expérience un grand nombre de fois. Comment feriez-vous pour déterminer lequel des deux modèles est le plus réaliste ?

3 Un test statistique

On se place de nouveau dans le cadre du modèle de la section 1. Comme dans la question ci-dessus, on suppose que l'on répète un grand nombre de fois l'expérience suivante : on infecte la cellule 1 d'une chaîne de cellules puis l'on observe le nombre total de cellules infectées à la fin de l'invasion. On note N_k ce nombre pour l'expérience numéro k.

18. La suite
$$(\theta_n)_{n\in\mathbb{N}} := \left(\frac{\sum_{i=1}^n N_i}{n}\right)_{n\in\mathbb{N}}$$
converge-t-elle ? Si oui vers quelle valeur ?

19. Rappeler l'énoncé du théorème de la limite centrale.

20. Démontrer que la variance de N est donnée par la formule

$$\text{Var}(N) = \frac{p}{(1-p)^2}.$$

21. On suppose $n = 100$ et $p = 1/2$. Calculer, à l'aide de la table ci-dessous et de la valeur $1/\sqrt{2} \simeq 0,7071$ un encadrement de la probabilité que $\theta_n > 2,3$.

x	1	1.2	1.4	1.6	1.8	2	2.4	2.8	3	3.4
$\Phi(x)$	0.841	0.885	0.919	0.945	0.964	0.977	0.992	0.9974	0.9987	0.9997

Dans la table ci-dessus $\phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx$ est la fonction de répartition d'une loi normale centrée réduite.

On veut tester l'hypothèse $p \leq 0,5$. Pour cela on commence par supposer que $p = 0,5$. On se donne un *niveau de confiance*, $\eta = 95\%$ et on rejette l'hypothèse à ce niveau de confiance si le θ_n observé avait une probabilité inférieure à $1 - \eta$ sous l'hypothèse.

22. D'après la question précédente, rejette-t-on l'hypothèse $p \leq 0,5$ au niveau de confiance 95% si l'on observe un $\theta_n \geq 2,3$?

4 Un processus de branchement

On cherche à présent à décrire l'évolution de l'infection au cours du temps. Pour cela on va modéliser la situation en oubliant dans un premier temps qu'il y a plusieurs cellules dans l'organisme. On se contente de suivre le nombre total de parasites dans l'organisme au cours du temps. Le temps est discret et indexé par les entiers $n = 0, 1, 2, \ldots$ On note $T(n)$ le nombre total de parasites à l'instant n. À l'instant initial $n = 0$ il y a un unique parasite $T(0) = 1$. On appelle génération n les parasites en vie à l'instant n.

Soit μ une distribution de probabilité sur $\mathbb{N}$ et soit Z une variable de loi μ, i.e. $\mathbb{P}(Z = i) = \mu(i)$. On note $m := \sum_{i=0}^{\infty} i\mu(i) = \mathbb{E}(Z)$ son espérance. On passe de la génération n à la génération $n+1$ en remplaçant chaque individu de la génération n par un nombre aléatoire, indépendant de descendants de loi μ. Autrement dit, si $Z_{i,n}$ désigne le nombre de descendants de l'individu i dans la génération n on a

$$T(n+1) = \sum_{i=1}^{T(n)} Z_{i,n}.$$

23. Calculer $\mathbb{E}(T(1)), \mathbb{E}(T(2))$. Donner une expression pour $\mathbb{E}(T(n))$.

On s'intéresse ici à la probabilité d'extinction de l'infection, i.e. est-ce qu'il existe un temps n fini aléatoire tel que $T(n) = 0$? Posons $\tau := \inf\{n > 0; T(n) = 0\}$. On note g la fonction génératrice de μ, i.e. $g(s) := \mathbb{E}(s^Z)$ définie pour $s \in [0, 1]$.

24. Exprimer $\mathbb{P}(Z = 0)$ en fonction de g.

25. En utilisant la question 8 donner une expression de $\mathbb{P}(T(2) = 0)$ et plus généralement de $u_n := \mathbb{P}(T(n) = 0)$ à l'aide de g.

26. Représenter graphiquement la fonction g sur $[0, 1]$. Interpréter sur la figure les quantités $m = \mathbb{E}(Z)$ et $\mathbb{P}(Z = 0)$. On veillera à ce que la représentation respecte les aspects qualitatifs de la fonction g (sens da variation, signe de la dérivée seconde). On distinguera les situations $m < 1, m = 1$ et $m > 1$.

27. On se place dans le cas $m > 1$. Montrer que l'équation $g(x) = x$ a une unique solution q sur $]0, 1[$. Montrer que si $u_n > q$ alors $u_{n+1} \in]q, u_n[$. De même, montrer que si $u_n < q$ alors $u_{n+1} \in]u_n, q[$

28. Représenter sur le graphe de la fonction g (dans le cas $m > 1$) les trois premiers termes de la suite (u_n).

29. On se place toujours dans le cas $m > 1$. Montrer que u_n converge vers la limite $q < 1$.

30. On suppose à présent que $m \leq 1$. Prouver que $u_n := \mathbb{P}(T(n))$ converge vers 1. Placer les premiers termes de la suite u_n sur le graphe de g dans le cas $m \leq 1$.

Les deux dernières questions permettent de conclure que l'infection s'éteint en temps finit presque sùrement (i.e. avec probabilité un) si et seulement si $m \leq 1$. Si $m > 1$ alors l'extinction s'éteint en temps fini avec probabilité q et survit pour toujours avec probabilité $1 - q$.

On s'intéresse à présent à la position des parasites au cours du temps. Le parasite initial de l'infection est dans la cellule 1. Chaque parasite choisit au moment de sa naissance soit de rester dans la même cellule que son parent (avec probabilit $1 - p$) soit de sauter de 1 vers la droite (avec probabilité p). À l'instant n on note $(X_1(n), X_2(n), \ldots)$ la suite du nombre de parasites dans les cellules $1, 2, \ldots$ et on a

$$T(n) := \sum_{i=1}^{\infty} X_i(n).$$

De ce fait, la trajectoire d'un parasite *typique* au cours du temps est donnée par le processus

$$S_n = \sum_{i=1}^{n} U_i$$

où les U_i sont des variables de Bernoulli de paramètre p indépendantes entre elles.

31. Étudier la convergence de S_n/n.

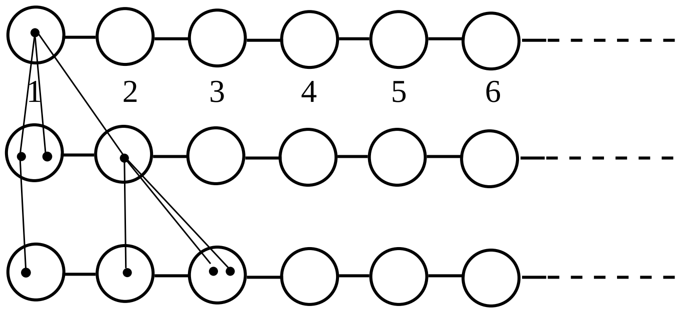

Figure 2: Une réalisation possible du processus de branchement : les trois premièrres générations.

Pour $a > 0$ on cherche à présent à estimer le nombre de parasite à droite de la position an quand n devient grand.

32. Montrer, que pour $k \in \mathbb{N}$ on a
$$\mathbb{E}(\sum_{i \geq \lceil an \rceil} X_i(n) | T(n) = k) = k\mathbb{P}(S_n > an).$$
En déduire que
$$\mathbb{E}(\sum_{i \geq \lceil an \rceil} X_i(n)) = m^n \mathbb{P}(S_n > an).$$

33. Donner une approximation de $P(S_n > an)$ pour n grand à l'aide de la loi normale.

Afin de déterminer le plus grand a tel qu'il y a des parasites au dessus de an on va utiliser des bornes sur la queue de distribution de la loi normale $\bar{\phi}(x)$
$$\bar{\phi}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$$

34. En utilisant que pour $x > 0$ et $t > x$ on a $1 \leq t/x$ montrer que
$$\bar{\phi}(x) \leq \frac{e^{-x^2/2}}{x\sqrt{2\pi}}.$$

35. À l'aide d'une intégration par partie bien choisie, montrer que
$$\bar{\phi}(x) \geq \frac{e^{-x^2/2}}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3} \right).$$

36. En déduire
$$a^* := \sup\{a > 0 : \lim_{n \to \infty} \mathbb{E}(\sum_{i \geq an} X_i(n)) = \infty\}.$$

On peut montrer que a^* est vraiment la vitesse de propagation de l'infection, c'est-à-dire que si M_n est la position du parasite le plus à droite à l'instant n, alors $M_N/N \to a^*$ avec probabilité 1.