CONCOURS D'ENTREE

DEUXIEME EPREUVE DE MATHEMATIQUES

Durée : 4 heures

Calculatrice non autorisée

OPTION B

Risque d'ouragan en Floride : saisonnalité et réchauffement climatique.

Ce sujet aborde de manière très simplifiée des questions de probabilités inspirées de problèmes rencontrés en assurance catastrophes naturelles. Néanmoins, aucune connaissance préalable en actuariat n'est nécessaire. Il est conseillé de bien lire la fin de la partie 1 avant d'aborder la partie 2. Rappelons que d'une façon générale, une condition sur des paramètres a, b et c pour qu'une certaine propriété soit vérifiée peut éventuellement ne faire intervenir qu'un sous-ensemble strict de $\{a, b, c\}$.

1 Préliminaires : processus de Poisson homogènes et inhomogènes

On commence par définir les notions de processus de Poisson homogène et inhomogène et établir ou poser comme vérifiées des propriétés qui nous seront utiles par la suite.
Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées, de loi exponentielle de paramètre $\lambda > 0$:

$$\forall x \geq 0, \quad P(Y_1 > x) = e^{-\lambda x}.$$

Soit $T_0 = 0$ et, pour tout $n \in \mathbb{N} \setminus \{0\}$,

$$T_n = \sum_{k=1}^{n} Y_k.$$

On utilisera les notations $E(X)$ et $Var(X)$ respectivement pour l'espérance et la variance d'une variable aléatoire. Pour $A \subset \mathbb{R}$, 1_A désigne la fonction indicatrice de A, définie pour $x \in \mathbb{R}$ par $1_A(x) = 1$ si $x \in A$ et $1_A(x) = 0$ sinon.

1. Déterminer l'espérance et la variance de T_n pour $n \geq 1$.
2. Montrer que pour $n \geq 1$, T_n admet pour densité f_{T_n}, définie pour $s \in \mathbb{R}$ par

$$f_{T_n}(s) = e^{-\lambda s} \lambda^n \frac{s^{n-1}}{(n-1)!} 1_{\mathbb{R}^+}(s).$$

3. Reconnaître la loi de T_n pour $n \geq 1$.
4. Montrer par des intégrations par parties successives que la fonction de répartition F_{T_n} de T_n vérifie pour $n \geq 1$ et $x \in \mathbb{R}$:

$$F_{T_n}(x) = 1 - \sum_{k=0}^{n-1} e^{-\lambda x} \frac{(\lambda x)^k}{k!} 1_{\mathbb{R}^+}(x). \tag{1}$$

5. Pour $t \geq 0$, soit

$$N_t = \sum_{n \geq 1} 1_{\{T_n \leq t\}} = \sup\{n, T_n \leq t\}.$$

   Montrer que pour $t > 0$, la variable aléatoire N_t vérifie

$$P(N_t = 0) = P(T_1 > t).$$

6. Ecrire l'événement $\{N_t = n\}$ à l'aide des événements $\{N_t \geq k\}$, $k \geq 1$.
7. Ecrire l'événement $\{N_t \geq n\}$ à l'aide d'un événement faisant intervenir certain(s) T_k, $k \geq 1$.
8. Déduire des deux questions précédentes et de l'équation (1) que pour $t > 0$, N_t suit une loi de Poisson de paramètre λt.

On supposera connus dans la suite les résultats suivants (qu'on peut donc utiliser directement, sans démonstration) :

- On appelle dans ce cas $(N_t)_{t \geq 0}$ un processus de Poisson homogène d'intensité λ. On a de plus : $\forall\, 0 < s < t$, $N_t - N_s$ suit une loi de Poisson de paramètre $(t-s)\lambda$ et est indépendant de N_s.
- Dans la suite, on parlera aussi de processus de Poisson inhomogène de fonction d'intensité (positive) $t \rightarrow \lambda(t)$: dans ce cas, on ne pourra plus utiliser les résultats précédents. Si $(N_t)_{t \geq 0}$ est un processus de Poisson inhomogène de fonction d'intensité $t \rightarrow \lambda(t)$, alors $\forall\, 0 < s < t$, $N_t - N_s$ suit une loi de Poisson de paramètre

$$\int_s^t \lambda(u) du$$

  et est indépendant de N_s.

– Lorsqu'on modélise le nombre d'ouragans en Floride entre la date 0 et la date t par un processus de Poisson $(N_t)_{t\geq 0}$ homogène d'intensité λ (respectivement inhomogène de fonction d'intensité $t \to \lambda(t)$), cela signifie que le nombre d'ouragans entre la date s et la date $t > s$ est représenté par $N_t - N_s$, qui suit une loi de Poisson de paramètre $(t-s)\lambda$ (respectivement de paramètre $\int_s^t \lambda(u)du$), et indépendante de N_s. Comme $N_0 = 0$ presque sûrement, le nombre d'ouragans entre la date 0 et la date t est représenté par $N_t = N_t - N_0$.

2 Modélisation du nombre d'ouragans

1. On commence par modéliser le nombre d'ouragans en Floride entre la date 0 (correspondant au 1er janvier 2007) et la date t (l'unité de temps est l'année) par un processus de Poisson homogène d'intensité λ. On suppose qu'en moyenne, il y a deux ouragans par an en Floride. Comment doit-on choisir λ pour respecter cette contrainte ?
2. A cause du réchauffement climatique, certains experts prétendent que la fréquence moyenne de ces ouragans augmentera dans le futur. Afin de tenir compte de l'avis de ces experts, votre cabinet de conseil modélise maintenant l'arrivée des ouragans par un processus de Poisson inhomogène de fonction d'intensité affine
$$\lambda(t) = \alpha t + \beta.$$
Ces experts prétendent qu'il y aura en moyenne lors de l'année 2027 deux fois plus d'ouragans par an que lors de l'année 2007. Comment calibrer α et β pour que le nombre moyen d'ouragans en 2007 reste inchangé (égal à 2) et que le nombre moyen d'ouragans en 2027 soit égal à 4 ?
3. Finalement votre actuaire conseil qui est venu à la réunion en 4×4 décide que le réchauffement climatique n'a rien à voir là-dedans et néglige ce phénomène. Néanmoins, il souhaite modéliser la saisonnalité de ces événements. La saison des ouragans en Floride étant de début juin (correspondant aux dates $5/12 + k,\ k \in \mathbb{N}$) à fin novembre (correspondant aux dates $11/12 + k,\ k \in \mathbb{N}$) avec un pic fin août-début septembre (correspondant aux dates $8/12 + k,\ k \in \mathbb{N}$), votre actuaire conseil décide de considérer une fonction d'intensité du type
$$\lambda(t) = a + bf\left(t - \lfloor t \rfloor\right),$$

où $\lfloor x \rfloor$ désigne la partie entière de x, $\gamma \in [1/4, 3/4]$, et f est la fonction définie pour $x \in [0,1]$ par

$$\begin{aligned} f(x) &= 0 \quad \text{pour } 0 \le x \le \gamma - \frac{1}{4}, & (2)\\ &= x - \left(\gamma - \frac{1}{4}\right) \quad \text{pour } \gamma - \frac{1}{4} \le x \le \gamma, & (3)\\ &= \gamma + \frac{1}{4} - x \quad \text{pour } \gamma \le x \le \gamma + \frac{1}{4}, & (4)\\ &= 0 \quad \text{pour } \gamma + \frac{1}{4} \le x \le 1. & (5) \end{aligned}$$

(a) A quelle condition sur a, b et γ le nombre moyen d'ouragans par an est-il inchangé par rapport au modèle à intensité constante (pour lequel le nombre moyen d'ouragans par an est égal à 2) ?

(b) A quelle condition sur a, b et γ la variance du nombre d'ouragans par an est-elle inchangée par rapport au modèle à intensité constante (pour lequel le nombre moyen d'ouragans par an est égal à 2) ?

(c) Quelle valeur de γ est la plus cohérente ?

(d) On suppose que γ prend la valeur déterminée à la question précédente. Votre actuaire conseil, décidément très calé en ouragans, estime que lorsqu'au moins un ouragan est survenu dans l'année civile, la probabilité conditionnelle qu'il n'y ait pas eu d'ouragan depuis le début de cette année civile jusqu'à début juin (correspondant aux dates $5/12 + k$, $k \in \mathbb{N}$) est de 99%. A quelle condition sur a et b cela correspond-il ?

(e) En déduire les valeurs de a, b et γ qui calibrent le modèle (c'est-à-dire qui permettent de respecter les contraintes fixées). On pourra utiliser que

$$K = \frac{12}{5} \ln\left(e^{-2} + 99\% \left(1 - e^{-2}\right)\right) \simeq -0,0208$$

et que $16K \simeq 0,333$.

3 Pertes cumulées sur une année

Le but est maintenant d'étudier la loi des pertes cumulées sur l'année 2007 (entre $t = 0$ et $t = 1$). Le nombre d'ouragans sur cette période est représenté par une variable aléatoire N_1 à valeurs dans $\mathbb{N}$. Soit $(W_k)_{k\geq 1}$ une suite de variables aléatoires i.i.d. et indépendantes de N_1, également à valeurs dans $\mathbb{N}$. Pour $k \geq 1$, W_k représente le coût du k-ème sinistre (accident) en euros. La perte cumulée sur l'année 2007 est donc représentée par la variable aléatoire S à valeurs dans $\mathbb{N}$, définie par $S = 0$ si $N_1 = 0$ et

$$S = \sum_{k=1}^{N_1} W_k \quad \text{sinon.}$$

Pour $m \in \mathbb{N}$, notons $\pi_W(m) = P(W_1 = m)$.

1. Pour $k \geq 1$, considérons la somme des k premiers montants

   $$S_k = \sum_{i=1}^{k} W_i.$$

   Pour $k \geq 1$ et $m \in \mathbb{N}$, soit

   $$\pi_W^{(k)}(m) = P(S_k = m).$$

   On définit $\pi_W^{(0)}(0) = 1$ et $\pi_W^{(0)}(m) = 0$ pour $m \geq 1$. Montrer pour $k \geq 1$ que

   $$\pi_W^{(k)}(m) = \sum_{p=0}^{m} \pi_W^{(k-1)}(m-p)\pi_W(p).$$

2. Pour $m \in \mathbb{N}$, exprimer $P(S = m)$ à l'aide des probabilités conditionnelles

   $$P\left(S = m \mid N_1 = k\right), \quad k \geq 1,$$

   et en déduire l'égalité classique suivante pour $m \in \mathbb{N}$:

   $$P(S = m) = \sum_{k\in\mathbb{N}} P(N_1 = k)\pi_W^{(k)}(m).$$

3. La fonction génératrice d'une variable aléatoire N à valeurs dans $\mathbb{N}$ est définie au point u par

   $$G_N(u) = E\left[u^N\right] = \sum_{n\in\mathbb{N}} P(N = n)u^n.$$

   On admettra que cette série converge pour $u \in [0, R[$, où $R \in [1, +\infty[$. On admettra également dans toute la suite que les dérivées successives de G_N peuvent être obtenues en dérivant sous le signe Σ les termes de la série. Montrer que $P(N = 0) = G_N(0)$ et que la fonction génératrice de N détermine la loi de N.

4. On suppose désormais que le plus petit des rayons de convergence (qu'on appellera toujours R) des fonctions génératrices considérées dans la suite vérifie $R > 1$. Pour $k \geq 1$, on considère le moment factoriel d'ordre k de N défini par

$$\mu_N^{(k)} = E\left[N(N-1)\dots(N-k+1)\right].$$

   Exprimer $E(N)$ et $Var(N)$ en fonction de moments factoriels de N.

5. Exprimer pour $k \geq 1$ les moments factoriels $\mu_N^{(k)}$ d'ordre k de N à l'aide de dérivées successives de G_N.

6. Déterminer la fonction génératrice et les moments factoriels d'ordre k (pour $k \geq 1$) de N lorsque N suit une loi de Poisson de paramètre λ.

7. Lorsque N et N' sont des variables aléatoires indépendantes à valeurs dans $\mathbb{N}$, exprimer $G_{N+N'}(u)$ pour $u \in]0, R[$ à partir de $G_N(u)$ et de $G_{N'}(u)$.

8. Montrer à l'aide des questions précédentes que pour $u \in]0, R[$,

$$G_S(u) = G_{N_1}\left[G_{W_1}(u)\right].$$

9. En déduire l'expression de la perte cumulée moyenne $E(S)$ en fonction de $E(N_1)$ et de $E(W_1)$.

10. En déduire également l'expression de la variance de la perte cumulée $Var(S)$ en fonction de $E(N_1)$, $E(W_1)$, $Var(N_1)$ et $Var(W_1)$. Cette formule est un cas particulier du théorème de décomposition de la variance.

11. On souhaite comparer la variance de S suivant deux types de modèles : dans le premier modèle, on considère donc que le nombre N_1 d'ouragans touchant la Floride sur une année suit une loi de Poisson de paramètre 2 ; dans le deuxième modèle, on considère qu'exactement 20 ouragans se forment chaque année au large, et que chaque ouragan a une probabilité p d'atteindre la Floride. On suppose également que les événements "l'ouragan numéro k atteint la Floride", $1 \leq k \leq 20$ sont mutuellement indépendants. Quelle est la loi du nombre d'ouragans qui touchent la Floride sur une année dans le deuxième modèle ?

12. Comment choisir p dans le deuxième modèle afin que le nombre moyen d'ouragans touchant la Floride sur une année soit le même pour les deux modèles ?
13. Comparer la variance de la perte cumulée sur une année $Var(S)$ dans les deux modèles.
14. Proposer une amélioration du deuxième modèle qui tient compte du réchauffement climatique (on ne demande pas de refaire les calculs, mais d'expliquer en quoi le modèle proposé prend en compte le réchauffement climatique ou la possibilité d'un réchauffement climatique).