DEUXIEME EPREUVE DE MATHEMATIQUES
OPTION B

Durée : 4 heures

Nombre de sinistres en assurance non vie et hétérogénéité du portefeuille

Ce sujet aborde des questions de probabilités relatives à l'assurance non vie, néanmoins aucune connaissance en assurance n'est nécessaire, et toutes les questions se traitent avec les outils du programme.

1 Loi binomiale négative dont le premier paramètre est entier

Commençons par étudier la loi binomiale négative lorsque $n \in \mathbb{N} \setminus \{0\}$ et $p \in]0,1[$. Soit N une variable aléatoire suivant cette loi.
Cette loi discrète à support sur $\mathbb{N}$ est décrite par sa densité discrète $f_{n,p}$ donnée pour $k \in \mathbb{N}$ par

$$f_{n,p}(k) = P(N = k) = C_{k+n-1}^{k} \cdot p^n \cdot q^k,$$

où $q = 1 - p$ et où C_n^k désigne le coefficient binomial usuel

$$C_n^k = \frac{n}{k!(n-k)!}.$$

1. Montrer que la loi binomiale négative peut aussi s'écrire sous la forme suivante (qui explique l'origine de son nom) :

$$f_{n,p}(k) = C_{-n}^{k} p^n (-q)^k,$$

   où C_{-n}^k est le coefficient binomial généralisé, défini pour un entier négatif par

$$C_{-n}^k = \frac{(-n)(-n-1)\cdots(-n-k+1)}{k!}.$$

2. Montrer que pour tout x tel que $|x| < 1$,

$$(1-x)^{-1} = \frac{1}{1-x} = \sum_{k \in \mathbb{N}} x^k.$$

3. Montrer que $\forall r \in \mathbb{N}$, $\forall k \geq r+1$,

$$\frac{k-r}{r+1} C_k^r = C_k^{r+1}.$$

4. Montrer la formule du binome négatif :

$$(1-x)^{-r} = \sum_{k=0}^{+\infty} C_{-r}^{k} (-x)^k,$$

où C_{-r}^{k} est le coefficient binomial généralisé aux nombres négatifs défini plus haut. On pourra admettre (sans le redémontrer [1]) que pour $|x| < 1$, $\forall r \geq 1$,

$$(1-x)^{-(r+1)} = \sum_{k=r}^{+\infty} C_{k}^{r} x^{k-r}.$$

5. En utilisant cette formule, démontrer que la loi binomiale négative est bien une loi de probabilité pour $n \in \mathbb{N} \setminus \{0\}$ et $p \in]0,1[$.
6. Est-ce une loi de probabilité pour $n \in \mathbb{N} \setminus \{0\}$ et $p = 0$?
7. Est-ce une loi de probabilité pour $n \in \mathbb{N} \setminus \{0\}$ et $p = 1$?
8. Toujours en utilisant la formule du binôme négatif, montrer que l'espérance de N est

$$E(N) = \frac{nq}{p}.$$

9. Montrer que la variance de N est

$$Var(N) = \frac{nq}{p^2}.$$

10. La fonction Bêta incomplète, de paramètres $a > 0$ et $b > 0$, est définie pour $0 \leq x \leq 1$ par

$$B_{a,b}(x) = \int_0^x y^{a-1} (1-y)^{b-1} dy.$$

Pour $x = 1$, elle correspond à la fonction Bêta de paramètres $a > 0$ et $b > 0$:

$$B(a,b) = \int_0^1 y^{a-1} (1-y)^{b-1} dy.$$

1. Ceci se démontre par récurrence sur r à l'aide du théorème de dérivation des séries entières

La fonction Bêta incomplète régularisée $I_x(a,b)$, calculée en $x \in [0,1]$ et $a, b > 0$, s'obtient en divisant la fonction bêta incomplète par la fonction bêta (complète)

$$I_x(a,b) = \frac{B_{a,b}(x)}{B(a,b)}.$$

Dans le cas où a et b sont des entiers naturels, montrer en utilisant des intégrations par parties que $x \in]0,1[$ et $a, b \in \mathbb{N} \setminus \{0\}$,

$$I_x(a,b) = \sum_{j=a}^{a+b-1} \frac{(a+b-1)!}{j!(a+b-1-j)!} x^j (1-x)^{a+b-1-j}.$$

11. En déduire que la fonction de répartition de N au point $k \in \mathbb{N}$ est donnée par
$$F_N(k) = P(N \leq k) = I_x(a,b),$$
où $x \in]0,1[$ est à préciser, et où $a, b \in \mathbb{N} \setminus \{0\}$ sont à préciser également.
12. Montrer par récurrence sur k que pour $k \in \mathbb{N}$,
$$P(N \leq k) = 1 - q^{k+1} \sum_{i=0}^{n-1} C_{k+i}^{i} \cdot p^i.$$

2 Interprétation et liens avec d'autres lois

1. Considérons une suite d'épreuves de Bernoulli indépendantes dont la probabilité de succès (pour chaque essai) est $p \in]0,1[$ et celle d'échec $q = 1 - p$. Soit M la variable aléatoire correspondant au numéro de l'épreuve pour laquelle on observe le premier succès. Déterminer $P(M = k)$ pour $k \in \mathbb{N}$. En déduire que M suit la loi géométrique de paramètre $p \in]0,1[$.
2. Montrer que M suit une certaine loi binomiale négative pour des paramètres $n \in \mathbb{N} \setminus \{0\}$ et $p \in]0,1[$ à préciser. Cela montre que la loi géométrique est un cas particulier de la loi binomiale négative.
3. Par analogie avec l'interprétation précédente de loi géométrique, exprimer α et m en fonction de n et p dans cette phrase : la loi binomiale négative de paramètres $n \in \mathbb{N} \setminus \{0\}$ et $p \in]0,1[$ est la loi de probabilité de la variable aléatoire N qui comptabilise le nombre d'échecs

nécessaires avant obtention de m succès dans une suite d'épreuves de Bernoulli indépendantes dont la probabilité de succès (pour chaque essai) est $\alpha \in]0,1[$.

4. Si X suit une loi binomiale (et non binomiale négative) de paramètres $\nu \in \mathbb{N} \setminus \{0\}$ et $p \in]0,1[$, montrer que pour $\mu \in \mathbb{N}$ tel que $\mu \leq \nu$, $P(X \geq \mu)$ est la probabilité qu'après ν épreuves, il y ait au moins μ succès.
5. En déduire que si N suit une loi binomiale négative de paramètres $n \in \mathbb{N} \setminus \{0\}$ et $p \in]0,1[$, alors pour tout $k \in \mathbb{N}$,
$$P(N \leq k) = P(X \geq n),$$
lorsque X suit une loi binomiale (et non binomiale négative) de paramètres $\nu \in \mathbb{N} \setminus \{0\}$ et p, où ν est à préciser (l'exprimer en fonction de k et n).

3 Loi binomiale négative obtenue par mélange de lois de Poisson

Définition 1 Lois Poisson-mélange
Soit Θ une variable aléatoire admettant comme densité f_Θ telle que $P(\Theta \in]0,+\infty[) = 1$. On dit que N suit une loi Poisson-mélange de paramètres (λ, Θ) (avec Θ comme loi de mélange) si pour tout $n \in \mathbb{N}$, pour tout $\theta > 0$,
$$P(N = n \mid \Theta = \theta) = e^{-\lambda\theta}\frac{(\lambda\theta)^n}{n!}.$$

Pour tout $n \in \mathbb{N}$, on obtient alors $P(N = n)$ grâce à la formule suivante (qu'on pourra utiliser directement dans tout ce problème) :
$$P(N = n) = \int_0^{+\infty} P(N = n \mid \Theta = \theta) f_\Theta(\theta) d\theta = \int_0^{+\infty} e^{-\lambda\theta}\frac{(\lambda\theta)^n}{n!} f_\Theta(\theta) d\theta.$$

Rappelons que la densité d'une loi Gamma de paramètres $s > 0$ et $t > 0$ est donnée au point $x > 0$ par
$$g_{s,t}(x) = \frac{x^{s-1}e^{-x/t}}{\Gamma(s)t^s},$$

où $\Gamma(s)$ est la valeur de la fonction Gamma en $s > 0$:

$$\Gamma(s) = \int_0^{+\infty} t^{s-1}\, e^{-t}\, dt.$$

1. Montrer que pour $k \geq 0$,

$$P(N = k) = \int_0^{+\infty} \frac{\theta^{k+s-1} e^{-\theta \frac{t+1}{t}}}{k!\Gamma(s) t^s} d\theta.$$

2. En utilisant un changement de variable

$$u = \theta \frac{t+1}{t^\gamma},$$

   où $\gamma > 0$ est à préciser, montrer que :

$$P(N = k) = \left(\frac{t}{t+1}\right)^{s+k} \frac{1}{\Gamma(s) k! t^s} \int_0^{+\infty} \theta^{k+s-1} e^{-\theta} d\theta.$$

3. Montrer que pour $k \in \mathbb{N}$,

$$P(N = k) = \frac{\Gamma(k+s)}{\Gamma(s) k!} p^s q^k, \tag{1}$$

   où p et q sont à exprimer en fonction de t.

4. En déduire que lorsque s est entier, N suit une loi binomiale négative de paramètres à préciser.

Les masses de probabilité décrites par l'équation (1) fournissent donc une généralisation de la loi binomiale négative pour un premier paramètre réel strictement positif. Les formules donnant la moyenne, la variance et la fonction de répartition (sous la forme de la fonction Bêta généralisée régularisée) s'étendent sans aucun problème à ce cas plus général.

4 Applications en assurance non vie

Le mélange de loi de Poisson peut servir à prendre en compte l'hétérogénéité d'un portefeuille d'assurances : différents groupes d'assurés peuvent avoir des probabilités très différentes de déclarer des sinistres, et leur nombre de sinistres déclarés dans une année peuvent souvent être modélisé par des variables aléatoires suivant des lois de Poisson de paramètres différents. En prenant des hypothèses raisonnables, le nombre de sinistres pour l'ensemble du portefeuille peut donc souvent être modélisé par une variable aléatoire suivant une loi Poisson mélange.

Considérons quatre assureurs automobile : les Assurances de la Jeunesse Australienne (AJA), nombre de sinistres en 2010 noté N_1 ;
les Assurances des Services Sanitaires des Entreprises (ASSE), nombre de sinistres en 2010 noté N_2 ;
l'assureur des Obligataires Libres (OL), nombre de sinistres en 2010 noté N_3, et les Entreprises et Sociétés Transaméricaines d'Assurance Civile (ESTAC), nombre de sinistres en 2010 noté N_4.
On supposera dans toute la suite du problème que les variables aléatoires N_1, N_2, N_3 et N_4 sont mutuellement indépendantes. On supposera également que chaque assureur dispose d'estimations des deux premiers moments du nombre de sinistres en 2010 et choisit la loi du nombre de sinistres de manière à respecter ces deux premiers moments. Les lois disponibles sont les lois de Poisson et les lois binomiales négatives.

1. En supposant que $E(N_1) = 100$ et $Var(N_1) = 150$, l'assureur AJA doit-il choisir une loi de Poisson ou une loi binomiale négative ? Justifier la réponse.
2. Quel(s) paramètre(s) convient-il de prendre pour la loi de N_1 ?
3. En supposant que $E(N_2) = Var(N_2) = 2$, l'assureur ASSE (à qui il ne reste plus qu'un seul assuré, ne conduisant d'ailleurs pas très bien) doit-il choisir une loi de Poisson ou une loi binomiale négative ? De quel(s) paramètre(s) ?
4. Montrer que si M_1 et M_2 sont deux variables aléatoires indépendantes suivant des lois binomiales négatives de paramètres respectifs n_1 et p_1, et n_2 et p_2 avec $p_1 = p_2$, alors $M_1 + M_2$ suit aussi une loi binomiale négative de paramètres n et p à préciser.
5. Justifier ce résultat dans le cas où r_1 et r_2 sont entiers à la lumière de la question 3 de la partie 2.
6. On suppose que N_3 suit une loi binomiale négative de paramètres $n_3 = 100$ et $p_3 = 2/3$, et que N_1 suit la loi déterminée aux deux premières questions de cette partie. Supposons que les assureurs OL et AJA décident de se regrouper sous le nom COnglomérat des Jeunes

Obligataires Libres (COJOL). Quelle est la loi du nombre $N_1 + N_3$ de sinistres en 2010 pour l'assureur COJOL ?

7. On suppose que N_4 suit une loi de Poisson de paramètre 23. Les assureurs ASSE et ESTAC, risquant de s'adresser au même public en 2010, souhaitent fusionner. En supposant que N_2 suit la loi obtenue à la troisième question de cette partie, quelle est la loi du nombre de sinistres $N_2 + N_4$ en 2010 de ce nouveau conglomérat [2] ?

2. Toute ressemblance entre les noms de ces assureurs et des associations ou entreprises est (presque) purement fortuite et ne saurait surtout pas être prise au premier degré...