Exercices du CH18 : Variables aléatoires discrètes

Exercices de la banque INP à étudier : ex 95 à 111 (en excluant les 101, 105, 107 et 112, déjà traités au chapitre précédent).

I Lois de probabilité, indépendance

Exercice 1 (*À la poste)
Dans une poste d'un petit village, on remarque qu'entre 10 heures et 11 heures, la probabilité pour que deux personnes entrent durant la même minute peut être considérée comme nulle et que l'arrivée des personnes est indépendante de la minute considérée. On a observé que la probabilité pour qu'une personne se présente entre la minute n et la minute $n+1$ est : $p = 0,1$. On veut calculer la probabilité pour que : 3,4,5,6,7,8... personnes se présentent au guichet entre 10h et 11h.

1. Définir une variable aléatoire adaptée, puis répondre au problème considéré.
2. Quelle est la probabilité pour qu'au moins 10 personnes se présentent au guichet entre 10h et 11h ? *Pour simplifier le calcul, on pourra utiliser un procédé d'approximation vu en cours !*

Exercice 2 (*Surbooking)
Des études effectuées par une compagnie aérienne montrent qu'il y a une probabilité $0,05$ qu'un passager ayant effectué une réservation ne se présente pas à l'embarquement. Dès lors elle vend toujours 94 billets pour ses avions à 90 places. On note A le nombre d'absents à l'embarquement.

1. Quelle est la loi suivie par A ?
   On pourra formuler une certaine hypothèse, même si elle n'est pas totalement justifiée.
2. Quelle est la probabilité pour qu'il y ait un problème à l'embarquement ?
3. Par quelle loi pourrait-il être assez raisonnable d'approcher la loi utilisée ici pour décrire A ? Comparer le résultat obtenu via cette approximation avec le résultat précédent.

Exercice 3 (*Minimum de deux lois géométriques)
Soient X et Y deux variables aléatoires discrètes indépendantes. On suppose qu'elles suivent des lois géométriques de paramètres respectifs $p \in]0,1[$ et $q \in]0,1[$.

1. Déterminer $\mathbb{P}(X > n)$ pour $n \in \mathbb{N}$.
2. En déduire la loi de $Z = \min(X, Y)$.
3. Calculer $\mathbb{P}(X = Y)$ et $\mathbb{P}(X > Y)$.

Exercice 4 (**)
Soient $n \in \mathbb{N}^*$ et p, q, r trois réels strictement positifs tels que $p + q + r = 1$.
Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{N}$. On suppose que la loi du couple (X, Y) est donnée par :

$$\forall (k,\ell) \in \mathbb{N}^2,\ \mathbb{P}\big((X,Y) = (k,\ell)\big) = \begin{cases} \dfrac{n!}{k!\ell![n-(k+\ell)]!} p^k q^\ell r^{n-(k+\ell)} & \text{si } k+\ell \leq n \\ 0 & \text{sinon} \end{cases}.$$

1. Déterminer les lois de X et de Y.
2. X et Y sont-elles indépendantes ?

Exercice 5 (Somme de deux lois de Poisson)**
Soient X et Y deux variables aléatoires discrètes indépendantes suivant des lois de Poisson de paramètres respectifs $\lambda > 0$ et $\mu > 0$.

1. Déterminer la loi de $X + Y$.
2. Reconnaître la loi de X sachant $X + Y = n$.

Exercice 6 (Colis détériorés)**
On suppose que le nombre N de colis expédiés chaque jour par une entreprise suit une loi de Poisson de paramètre $\lambda > 0$. Les envois sont supposés indépendants.
La probabilité que chaque colis a d'arriver détérioré est égale à $p \in]0\,;1[$.
Pour un jour donné, on note X le nombre de colis qui sont arrivés détériorés et Y le nombre de ceux arrivés en bon état.

1. Déterminer les lois de X et de Y.
2. X et Y sont-elles indépendantes ?

Exercice 7 (*Loi hypergéométrique)**

1. Une urne contient N boules indiscernables : m blanches et $N - m$ noires.
   Étant fixé $n \in \{0, \ldots, N\}$, on prélève en une fois n boules de l'urne.
   On note X la variable aléatoire donnant le nombre de boules blanches prélevées.
   Déterminer $\mathbb{P}(X = i)$ pour $i \in \{0, \ldots, n\}$.
   On dit que X suit la loi hypergéométrique de paramètres n, N et m.
2. **Principe du maximum de vraisemblance** :
   (a) Soient m, n et $i \in \{0, \ldots, n\}$ fixés. Pour quelle valeur de N, la quantité $\mathbb{P}(X = i)$ est-elle maximale ? *On pourra calculer le rapport des $\mathbb{P}(X = i)$ pour deux valeurs consécutives de N...*
   (b) Une certaine quantité N d'animaux habitent une région. Des scientifiques procèdent à une capture de 50 d'entre eux, les marquent puis les relâchent. Plus tard, les scientifiques en capturent à nouveau 40 et constatent que 4 d'entre eux sont marqués. Selon la question précédente, quelle est l'estimation la plus vraisemblable de la population totale d'animaux ?
3. Soit X une variable aléatoire suivant une loi hypergéométrique de paramètres n, N et m. Lorsque m et $N - m$ sont grands par rapport à n, par quelle loi classique peut-on approcher la loi hypergéométrique ? Démontrer.

Exercice 8 (*Fonction de répartition d'une VAD réelle)**
Soit X une variable aléatoire discrète réelle. On définit sur $\mathbb{R}$ une fonction F_X, appelée **fonction de répartition** de X, par : $\forall x \in \mathbb{R}$, $F_X(x) = \mathbb{P}(X \leq x)$.

1. Montrer que F_X est une fonction croissante et déterminer ses limites en $\pm\infty$.
2. Montrer que F_X est continue à droite en tout point et qu'on a :
$$\forall x \in \mathbb{R},\ \mathbb{P}(X = x) = F_X(x) - \lim_{y \to x^-} F_X(y).$$
3. En déduire que F_X caractérise la loi de X.
   Préciser l'ensemble des points de discontinuité de F_X.
4. Montrer que l'ensemble des points de discontinuité de F_X est au plus dénombrable.

Exercice 9 (Taux de panne)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, X une variable aléatoire discrète, définie sur cet espace probabilisé, à valeurs dans $\mathbb{N}^*$ et vérifiant : $\forall n \in \mathbb{N}^*$, $\mathbb{P}(X \geq n) > 0$.
On appelle **taux de panne** associé à X la suite réelle $(x_n)_{n\in\mathbb{N}^*}$ de terme général
$$x_n = \mathbb{P}(X = n | X \geq n).$$

1. Déterminer une expression de $p_n = \mathbb{P}(X = n)$ à l'aide des x_k.
2. (a) Montrer que l'on a $0 \leq x_n < 1$ pour tout $n \in \mathbb{N}^*$ et que la série de terme général x_n diverge.
   (b) Réciproquement : soit $(x_n)_{n\in\mathbb{N}^*}$ une suite à valeurs dans $[0\,;1[$ telle que la série de terme général x_n diverge. Montrer qu'il existe une variable aléatoire dont le taux de panne est la suite (x_n).
3. Montrer que la variable X suit une loi géométrique ssi son taux de panne est constant.

II Espérance, variance, covariance

Exercice 10
Soit X une VAD à valeurs dans $\mathbb{N}$ telle que :

$$\forall\, n \in \mathbb{N}^*, \quad \mathbb{P}(X = n) = \frac{4}{n}\mathbb{P}(X = n-1).$$

Trouver la loi de X et son espérance.

Exercice 11 (Somme de n tirages avec remise)
Une urne contient N boules numérotées de 1 à N. On effectue n tirages avec remise et on appelle S_n la somme des numéros obtenus. Déterminer l'espérance et la variance de S_n.
Indication : comme cela arrive régulièrement, la bonne approche consiste à décrire S_n comme somme de variables aléatoires simples.

Exercice 12
Soit X une VAD suivant une loi géométrique de paramètre p. Calculer $E\left(\frac{1}{X}\right)$.

Exercice 13 (Une grenouille infatigable)
Une grenouille monte les marches d'un escalier (supposé infini...) en partant du sol et en sautant :
- ou bien une seule marche, avec probabilité $p \in]0\,;1[$;
- ou bien deux marches, avec probabilité $1-p$.

On suppose que ses sauts sont indépendants les uns des autres.

1. Dans cette question, on observe n sauts de la grenouille et on note X_n le nombre de fois où elle a sauté 1 marche, et Y_n le nombre marches franchies.
   Déterminer la loi de X_n puis exprimer Y_n en fonction de X_n et en déduire $E(Y_n)$ et $V(Y_n)$.
2. Pour $k \geq 1$, on note p_k la probabilité que la grenouille passe par la marche k.
   Déterminer une relation de récurrence satisfaite par p_k puis l'expression de p_k en fonction de k.

Exercice 14 (Un premier exemple de loi binomiale négative)
Une bactérie a la probabilité p d'être touchée par un laser. On envoie un rayon laser par seconde. La bactérie ne meurt que lorsqu'elle est touchée r fois ($r \in \mathbb{N}^*$).
Déterminer la loi de la variable aléatoire X égale à la durée de vie de la bactérie, ainsi que son espérance (de vie).
On pourra utiliser la relation $\sum\limits_{k=r}^{+\infty} \binom{k}{r} x^{k-r} = \frac{1}{(1-x)^{r+1}}$ *pour* $x \in]-1,1[$*, tout en expliquant d'où vient cette formule.*

Exercice 15 (Loi binomiale négative : une autre approche)
On dit qu'une variable aléatoire X suit une **loi binomiale négative** de paramètres n et p lorsque $X(\Omega) = [\![n, +\infty[\![$ et $\mathbb{P}(X = k) = \begin{pmatrix} k-1 \\ n-1 \end{pmatrix} p^n (1-p)^{k-n}$.

1. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes suivant toutes une loi géométrique de paramètre p.
   Montrer que $X_1 + \cdots + X_n$ suit une loi binomiale négative de paramètres n et p.
   On pourra raisonner par récurrence sur n.
2. En déduire espérance et variance d'une loi binomiale négative de paramètres n et p.

Exercice 16
Soit (A_n) une suite croissante d'événements. Vérifier que l'indicatrice de leur réunion est la limite simple des fonctions indicatrices $\mathbb{1}_{A_n}$ et interpréter la propriété de continuité croissante à l'aide de l'espérance.

Exercice 17 (Problème de la prise de sang)
On considère un groupe de N personnes ayant la probabilité p (petite) de souffrir d'une maladie. On dispose d'un test fiable par prise de sang pour déceler cette maladie. On souhaite trouver une stratégie pour minimiser le coût total. On note X la variable aléatoire égale au nombre de tests effectués.
La première stratégie consiste à tester tout le monde. Dans ce cas, X est constante, égale à N.
La seconde stratégie est la suivante : on fixe un entier k (divisant N pour simplifier). On groupe les N personnes en paquets de k personnes dont on mélange le sang. On teste chacun des paquets puis, le cas échéant, on teste chaque personne dans les paquets malsains.

1. Montrer qu'avec la seconde stratégie, on a : $E(X) = \dfrac{N}{k} + N\big(1-(1-p)^k\big)$.

   Remarque
   Ce type de calcul se fait plus naturellement avec la notion (hors programme) d'espérance conditionnelle, cf. l'exercice ci-après.

2. Montrer que pour p très petit, cette espérance est minimale pour k proche de $\sqrt{\frac{1}{p}}$ et qu'elle vaut alors environ $2N\sqrt{p}$.
3. Conclusion ?

Exercice 18 (Utilisation d'une fonction génératrice)

1. *Préliminaire* : À l'aide d'une série entière, déterminer une expression simple, pour tout $x \in]-1\,;1[$ et tout $n \geq 1$, de $\sum\limits_{k=0}^{+\infty}(k+n)(k+n-1)\cdots(k+1)x^k$.
2. Soient $n \in \mathbb{N}$, $a > 0$, $p \in [0\,;1[$ et X une variable aléatoire à valeurs dans $\mathbb{N}$ dont la loi est donnée par :
$$\mathbb{P}(X=k) = a\binom{n+k}{k}p^k.$$
   Calculer la fonction génératrice de X et en déduire l'espérance et la variance de X.
   En cours de calcul, on trouvera la valeur que possède nécessairement a...

Exercice 19 (Espérance conditionnelle)
Soit X une VAD réelle et admettant une espérance. Soit aussi A un événement de probabilité non nulle. On définit $E(X|A) = \sum\limits_{x\in X(\Omega)} x\mathbb{P}(X=x|A)$.
Soit $(B_n)_{n\geq 1}$ un système complet d'événements de probabilités non nulles. Montrer que :
$$E(X) = \sum_{n=1}^{+\infty}\mathbb{P}(B_n)E(X|B_n).$$

Exercice 20
Soit (X_1, X_2, Y_1, Y_2) une famille de 4 variables aléatoires réelles et discrètes admettant toutes un moment d'ordre 2. Montrer que si (X_1, X_2) est indépendant de (Y_1, Y_2), alors
$$Cov(X_1+Y_1, X_2+Y_2) = Cov(X_1, X_2) + Cov(Y_1, Y_2).$$

Exercice 21
On note Y_1 et Y_2 les variables aléatoires donnant respectivement le nombre de 1 et le nombre de 2 obtenus au cours de n lancers indépendants de dés équilibrés.
Calculer $Cov(Y_1, Y_2)$ puis le *coefficient de corrélation* ρ de ces deux variables, défini par
$$\rho = \frac{Cov(Y_1, Y_2)}{\sigma_{Y_1}\sigma_{Y_2}}.$$

III Inégalités, Loi faible des grands nombres

Exercice 22 (Production)
On suppose qu'une usine fabrique en moyenne 50 objets par semaine.

1. Quelle est la probabilité pour que l'usine fabrique au moins 75 objets la semaine prochaine ?
2. Si l'on sait que la variance de la production hebdomadaire est de 25, quelle est la probabilité pour que le nombre d'objets fabriqués la semaine prochaine soit compris entre 40 et 60 ?

Exercice 23 (Convergence en probabilité)
Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires définies sur un même espace probabilisé.
On suppose que chaque X_n suit une loi de Bernoulli de paramètre p_n et que les X_n sont deux à deux indépendantes. Pour $n \geq 1$, on pose $S_n = \dfrac{X_1 + \cdots + X_n}{n}$ et $m_n = \dfrac{p_1 + \cdots + p_n}{n}$.
Montrer que $\forall \varepsilon > 0$, $\mathbb{P}(|S_n - m_n| \geq \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0$.

Exercice 24 (Nombre moyen d'urnes vides)
On dispose de n urnes et de $N = na$ boules, avec $(n, a) \in (\mathbb{N}^*)^2$. Ces boules sont réparties dans les urnes de façon indépendante et équiprobable.
On note Y_n la variable aléatoire donnant le nombre d'urnes vides et $S_n = \dfrac{Y_n}{n}$.

1. Calculer $E(S_n)$ et $V(S_n)$.
   On évitera un calcul direct et on commencera plutôt par interpréter de manière simple Y_n.
2. Montrer : $\forall \varepsilon > 0$, $\lim\limits_{n\to+\infty} \mathbb{P}(|S_n - e^{-a}| \geq \varepsilon) = 0$.

IV Fonctions génératrices

Exercice 25 (Loi binomiale négative, deuxième approche)
On souhaite étudier le temps d'attente du k-ième succès dans une série d'expériences aléatoires indépendantes ayant une même probabilité de réussite égale à $p \in [0\,;1]$.
On note $(X_n)_{n\geq 1}$ la suite de variables de Bernoulli modélisant la succession des expériences. Le temps d'attente du k-ième succès définit une variable aléatoire T_k déterminée par :

$$T_k = n \iff (X_1 + \cdots + X_n = k \text{ et } X_1 + \cdots + X_{n-1} < k)\,.$$

1. Déterminer la loi ainsi que la fonction génératrice de T_1.
2. Même question avec $T_k - T_{k-1}$ pour $k \geq 2$.
3. En déduire la fonction génératrice de T_k.
4. Exprimer la loi de T_k.

Exercice 26 (*Points fixes de permutations)**
On considère la variable X_n égale au nombre de cycles disjoints d'une permutation de l'ensemble $\{1, \cdots, n\}$.

1. Établir : $\forall i \in \{2, \cdots, n+1\}$, $\mathbb{P}(X_{n+1} = i) = \dfrac{\mathbb{P}(X_n = i-1)}{n+1} + \dfrac{n\mathbb{P}(X_n = i)}{n+1}$.
2. En déduire, si G_n désigne la fonction génératrice de la loi X_n, qu'on a : $G_{n+1}(t) = \dfrac{n+t}{n+1} G_n(t)$.
3. En déduire l'expression factorisée du polynôme G_n.
4. Que vaut $E(X_n)$?