

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 TD du mardi 10/03

Exercice 1 (*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 $\mathbb{P}(X > Y)$.

Corrigé de l'exercice 1

1. L'événement $(X > n)$ correspond à l'enchaînement de n échecs successifs indépendants (puisque X représente le rang du premier succès), donc $\mathbb{P}(X > n) = (1 - p)^n$ pour tout $n \in \mathbb{N}$.

Remarque

On peut aussi faire ce calcul de manière classique :

$$\forall n \in \mathbb{N}, \quad \mathbb{P}(X > n) = \sum_{k=n+1}^{+\infty} \mathbb{P}(X = k) = p \sum_{k=n+1}^{+\infty} (1-p)^{k-1} = \frac{p(1-p)^n}{1-(1-p)} = (1-p)^n.$$

2. On a $Z(\Omega) = \mathbb{N}^*$, et pour tout $n \in \mathbb{N}^*$:

$$\mathbb{P}(Z > n) = \mathbb{P}(\min(X, Y) > n) = \mathbb{P}((X > n) \cap (Y > n)).$$

Puisque X et Y sont indépendantes :

$$\mathbb{P}(Z > n) = \mathbb{P}(X > n) \times \mathbb{P}(Y > n) = (1-p)^n(1-q)^n = ((1-p)(1-q))^n$$

(et cette formule reste vraie pour $n = 0$ puisque $\mathbb{P}(Z > 0) = 1$). On en déduit :

$$\forall n \in \mathbb{N}^*, \quad \mathbb{P}(Z = n) = \mathbb{P}(Z > n-1) - \mathbb{P}(Z > n) = ((1-p)(1-q))^{n-1} - ((1-p)(1-q))^n,$$

c'est-à-dire

$$\forall n \in \mathbb{N}^*, \quad \mathbb{P}(Z = n) = (1 - (1-p)(1-q)) \times ((1-p)(1-q))^{n-1}.$$

On constate alors que $Z \sim \mathcal{G}(r)$ avec $r = 1 - (1-p)(1-q)$.

3. Puisque $[X = Y] = \bigcup_{n=1}^{+\infty} ([X = n] \cap [Y = n])$ (réunion disjointe), on obtient par indépendance de X et Y :

$$\mathbb{P}(X = Y) = \sum_{n=1}^{+\infty} \mathbb{P}(X = n)\mathbb{P}(Y = n) = pq \sum_{n=1}^{+\infty} (1-p)^{n-1}(1-q)^{n-1} = \frac{pq}{1-(1-p)(1-q)}.$$

De la même façon :

$$\mathbb{P}(X > Y) = \sum_{n=1}^{+\infty} \mathbb{P}([X > Y] \cap [Y = n]) = \sum_{n=1}^{+\infty} \mathbb{P}([X > n] \cap [Y = n]),$$

et toujours par indépendance de X et Y :

$$\mathbb{P}(X > Y) = \sum_{n=1}^{+\infty} \mathbb{P}(X > n)\mathbb{P}(Y = n) = q \sum_{n=1}^{+\infty} (1-p)^n(1-q)^{n-1} = \frac{q(1-p)}{1-(1-p)(1-q)}.$$

Exercice 2 (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 $\mu > 0$.

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

Corrigé de l'exercice 2

- Puisque $X(\Omega) = Y(\Omega) = \mathbb{N}$, on a $(X + Y)(\Omega) = \mathbb{N}$, et pour tout $n \in \mathbb{N}$:

$$[X + Y = n] = \bigcup_{0 \leq k \leq n} [X = k, Y = n - k],$$

donc

$$\mathbb{P}(X + Y = n) = \sum_{k=0}^n \mathbb{P}(X = k, Y = n - k),$$

et par indépendance de X et Y :

$$\mathbb{P}(X + Y = n) = \sum_{k=0}^n \mathbb{P}(X = k) \mathbb{P}(Y = n - k) = \sum_{k=0}^n e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^{n-k}}{(n-k)!},$$

c'est-à-dire

$$\mathbb{P}(X + Y = n) = \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} = \frac{e^{-(\lambda+\mu)}}{n!} (\lambda + \mu)^n.$$

On remarque que $X + Y \sim \mathcal{P}(\lambda + \mu)$.

- Soit $n \in \mathbb{N}$ fixé. Puisque $0 \leq X \leq X + Y$, on a

$$\mathbb{P}_{(X+Y=n)}(X = k) = 0 \text{ si } k \notin [0, n].$$

Si $0 \leq k \leq n$, alors

$$\mathbb{P}_{(X+Y=n)}(X = k) = \frac{\mathbb{P}((X = k) \cap (X + Y = n))}{\mathbb{P}(X + Y = n)} = \frac{\mathbb{P}((X = k) \cap (Y = n - k))}{\mathbb{P}(X + Y = n)},$$

donc par indépendance de X et Y :

$$\mathbb{P}_{(X+Y=n)}(X = k) = \frac{\mathbb{P}(X = k) \mathbb{P}(Y = n - k)}{\mathbb{P}(X + Y = n)} = \frac{e^{-\lambda} \lambda^k}{k!} \times \frac{e^{-\mu} \mu^{n-k}}{(n-k)!} \times \frac{n!}{e^{-(\lambda+\mu)} (\lambda + \mu)^n},$$

c'est-à-dire

$$\mathbb{P}_{(X+Y=n)}(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

où $p = \frac{\lambda}{\lambda + \mu}$. On reconnaît là la loi binomiale $\mathcal{B}(n, \frac{\lambda}{\lambda + \mu})$.

Exercice 3 (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.

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

Corrigé de l'exercice 3

1. Puisque $0 \leq X \leq N$ et que $N(\Omega) = \mathbb{N}$, on a $X(\Omega) = \mathbb{N}$, et pour tout $k \in \mathbb{N}$:

$$\mathbb{P}(X = k) = \sum_{n=0}^{+\infty} \mathbb{P}(N = n) \mathbb{P}_{(N=n)}(X = k).$$

Or, la loi conditionnelle de X sachant $[N = n]$ est la loi binomiale $\mathcal{B}(n, p)$ (on compte le nombre de colis détériorés lors de n envois indépendants, avec probabilité de "succès" égale à p), donc

$$\mathbb{P}(X = k) = \sum_{n=0}^{+\infty} \mathbb{P}(N = n) \binom{n}{k} p^k (1-p)^{n-k} \mathbb{1}_{\{n \geq k\}} = \sum_{n=k}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{k} p^k (1-p)^{n-k},$$

c'est-à-dire

$$\mathbb{P}(X = k) = e^{-\lambda} p^k \sum_{j=0}^{+\infty} \frac{\lambda^{j+k}}{k! j!} (1-p)^j = e^{-\lambda} \frac{(\lambda p)^k}{k!} \sum_{j=0}^{+\infty} \frac{(\lambda(1-p))^j}{j!} = e^{-\lambda} \frac{(\lambda p)^k}{k!} e^{\lambda(1-p)} = e^{-\lambda p} \frac{(\lambda p)^k}{k!}.$$

Finalement on a $X \sim \mathcal{P}(\lambda p)$ et de même on montre que $Y \sim \mathcal{P}(\lambda(1-p))$.

2. Calculons la loi du couple (X, Y) . On a $X(\Omega) = Y(\Omega) = \mathbb{N}$ et pour tout couple $(k, l) \in \mathbb{N}^2$:

$$\mathbb{P}(X = k, Y = l) = \mathbb{P}(X = k, N = k + l) = \mathbb{P}(N = k + l) \times \mathbb{P}_{(N=k+l)}(X = k),$$

donc tout comme dans la question précédente :

$$\mathbb{P}(X = k, Y = l) = e^{-\lambda} \frac{\lambda^{k+l}}{(k+l)!} \binom{k+l}{k} p^k (1-p)^l = e^{-\lambda} \frac{1}{k! l!} p^k (1-p)^l,$$

qu'on peut réécrire :

$$\mathbb{P}(X = k, Y = l) = e^{-\lambda p} \frac{(\lambda p)^k}{k!} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^l}{l!} = \mathbb{P}(X = k) \mathbb{P}(Y = l).$$

Ainsi, les variables X et Y sont bien indépendantes.

Remarque

Ceci est assez contre-intuitif car X et Y sont liées par une relation ($X + Y = N$), mais la somme N étant elle-même aléatoire, cela n'empêche visiblement pas l'indépendance de X et Y !

Bien sûr, si le nombre d'envois de colis est un entier $n \in \mathbb{N}^$ fixé à l'avance, X et Y (qui suivent alors les lois binomiales respectives $\mathcal{B}(n, p)$ et $\mathcal{B}(n, 1-p)$), ne seront alors plus indépendantes.*

Exercice 4 (***)Loi hypergéométrique)

- Une urne contient N boules indiscernables : m blanches et $N - m$ noires. Étant fixé $n \in \{0, \dots, N\}$, 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 $i \in \{0, \dots, n\}$. On dit que X suit la loi hypergéométrique de paramètres n , N et m .
- Principe du maximum de vraisemblance :**
 - Soient m , n et $i \in \{0, \dots, n\}$ 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 ...
 - 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 ?
- 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.

Corrigé de l'exercice 4

1. On peut modéliser l'expérience par l'univers Ω formé des parties à n éléments d'un ensemble à N éléments, muni de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$ et de l'équiprobabilité. Pour tout $i \in \{0, \dots, n\}$, l'événement $[X = i]$ correspond alors aux parties composées de i boules choisies parmi les m blanches et $n - i$ boules choisies parmi les $N - m$ noires. On en déduit :

$$\forall i \in \{0, \dots, n\}, \quad \mathbb{P}([X = i]) = \frac{|[X = i]|}{|\Omega|} = \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}}$$

(avec comme d'habitude la convention : pour tout $(a, b) \in \mathbb{N}^*$, $\binom{b}{a} = 0$ si $a > b$).
On note alors $X \sim \mathcal{H}(n, N, m)$ (loi hypergéométrique de paramètres n, N, m).

Remarque

Cette loi importante est la "loi des tirages simultanés", ou "loi des tirages sans remise" (cela revient au même). On va voir plus loin que dans certaines conditions, on peut l'approcher par une loi binomiale classique.

2. (a) Soient m, n et $i \in \{0, \dots, n\}$ fixés. Posons pour tout $N \geq n$:

$$u_{N,i} = \mathbb{P}(X = i) = \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}}.$$

On résout

$$\begin{aligned} u_{N+1,i} \geq u_{N,i} &\iff \frac{\binom{N+1-m}{n-i}}{\binom{N+1}{n}} \geq \frac{\binom{N-m}{n-i}}{\binom{N}{n}} \iff \binom{N+1-m}{n-i} \binom{N}{n} \geq \binom{N-m}{n-i} \binom{N+1}{n} \\ &\iff \frac{(N+1-m)!N!}{(n-i)!(N+1-m-n+i)!n!(N-n)!} \geq \frac{(N-m)!(N+1)!}{(n-i)!(N-m-n+i)!n!(N+1-n)!} \\ &\iff \frac{(N+1)-m}{(N+1-n)+(i-m)} \geq \frac{N+1}{N+1-n} \iff -m(N+1-n) \geq (N+1)(i-m) \\ &\iff mn \geq (N+1)i \iff N+1 \leq \frac{mn}{i}. \end{aligned}$$

Le terme maximal de la suite $(u_{N,i})_{N \geq n}$ est donc $u_{N_0,i}$ avec $N_0 = \lfloor \frac{nm}{i} \rfloor$ (le plus grand entier inférieur ou égal à mn/i).

- (b) Dans cette expérience, l'effectif total N est inconnu, l'effectif "marqué" est $m = 50$ (qui correspond au nombre de boules blanches), l'effectif "de pioche" est $n = 40$, et le nombre d'animaux marqués obtenus dans l'échantillon est $i = 4$, donc l'effectif total le plus probable est $N = \lfloor \frac{nm}{i} \rfloor = 500$.
3. Quand m et $N - m$ sont grands par rapport à n , cela signifie que le nombre de boules tirées est négligeable devant le nombre de boules de chaque couleur. On peut donc "oublier" le fait que l'on tire simultanément les n boules, et considérer que cela revient à peu près au même d'effectuer n tirages successifs indépendants (donc avec remise).

Il est donc raisonnable d'espérer que la loi hypergéométrique $\mathcal{H}(n, N, m)$ va être proche de la loi binomiale $\mathcal{B}(n, p)$, où $p = \frac{m}{N}$ est la probabilité de piocher une boule blanche lors de chacun des n tirages indépendants. Montrons cette approximation. Pour $0 \leq i \leq n$ fixés, on a

$$\begin{aligned} \mathbb{P}(X = i) &= \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}} \\ &= \frac{m(m-1) \cdots (m-i+1)}{i!} \times \frac{(N-m)(N-m-1) \cdots (N-m-n+i+1)}{(n-i)!} \times \frac{n!}{N(N-1) \cdots (N-n+1)} \\ &= \binom{n}{i} \frac{Q_1(m) Q_2(N-m)}{Q_3(N)}, \end{aligned}$$

où Q_1, Q_2, Q_3 sont des polynômes unitaires de degrés respectifs $i, n - i, n$. Par hypothèse, les variables m et $N - m$ (et donc N) sont beaucoup plus grandes que n , donc que $(i, n - i, n)$, on peut donc approcher les valeurs des trois polynômes par leurs équivalents en $+\infty$, c'est-à-dire leur terme de plus haut degré :

$$\mathbb{P}(X = i) \simeq \binom{n}{i} \frac{m^i (N - m)^{n-i}}{N^n} = \binom{n}{i} \left(\frac{m}{N}\right)^i \left(1 - \frac{m}{N}\right)^{n-i},$$

et on retrouve bien la loi binomiale $\mathcal{B}(n, \frac{m}{N})$.

Exercice 5 (***) 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 :

$$\forall x \in \mathbb{R}, \mathbb{P}(X = x) = F_X(x) - \lim_{y \rightarrow x^-} F_X(y).$$

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.

Corrigé de l'exercice 5

1. Pour $x \leq y$ réels, on a l'inclusion d'événements $[X \leq x] \subset [X \leq y]$, donc par croissance d'une probabilité : $F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y)$, ce qui prouve la croissance de F_X sur \mathbb{R} . De plus F_X est majorée par 1 et minorée par 0, donc elle possède des limites finies en $\pm\infty$. Vu que $\lim_{x \rightarrow +\infty} F_X(x)$ existe, on a en particulier l'atteindre "par les entiers" :

$$\lim_{x \rightarrow +\infty} F_X(x) = \lim_{n \rightarrow +\infty} F_X(n) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq n),$$

et par continuité croissante d'une probabilité, on obtient

$$\lim_{x \rightarrow +\infty} F_X(x) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} [X \leq n]\right) = \mathbb{P}(\Omega) = 1$$

(en effet, pour tout $\omega \in \Omega$, il existe $n \in \mathbb{N}$ tel que $X(\omega) \leq n$, donc $\Omega \subset \bigcup_{n \in \mathbb{N}} [X \leq n]$ et l'inclusion réciproque est évidente).

On obtient de même en utilisant la continuité décroissante que :

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq -n) = \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} [X \leq -n]\right) = \mathbb{P}(\emptyset) = 0.$$

2. Soit $x \in \mathbb{R}$. Puisque F_X est monotone, elle possède des limites finies à gauche et à droite en x (pas nécessairement égales!). On peut atteindre ces limites par n'importe quelle suite, par exemple :

$$\lim_{y \rightarrow x^+} F_X(y) = \lim_{n \rightarrow +\infty} F_X\left(x + \frac{1}{n}\right) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(X \leq x + \frac{1}{n}\right)$$

Par continuité décroissante, on obtient donc :

$$\lim_{y \rightarrow x^+} F_X(y) = \mathbb{P}\left(\bigcap_{n \in \mathbb{N}^*} [X \leq x + \frac{1}{n}]\right) = \mathbb{P}([X \leq x]) = F_X(x),$$

et donc F_X est continue à droite en x .

En revanche, pour la limite à gauche, on obtient par continuité croissante :

$$\lim_{y \rightarrow x^-} F_X(y) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(X \leq x - \frac{1}{n}\right) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} [X \leq x - \frac{1}{n}]\right) = \mathbb{P}([X < x]) = F_X(x) - \mathbb{P}(X = x),$$

d'où le résultat voulu.

3. Si deux VAD X et Y vérifient $F_X = F_Y$, alors la formule précédente montre que pour tout réel x , on a

$$\mathbb{P}(X = x) = \mathbb{P}(Y = x),$$

donc X et Y ont même loi.

Les points de discontinuités sont exactement les $x \in \mathbb{R}$ tels que $\lim_{x^-} F_X < F_X(x)$, c'est-à-dire les x tels que $\mathbb{P}(X = x) > 0$. Il s'agit des points de l'image $X(\Omega)$ qui ont une probabilité non nulle.

Remarque

Pour une VAD $X : \Omega \rightarrow \mathbb{R}$, on remarque que la fonction de répartition F_X est constante par morceaux (escalier), et les "sauts" sont les points non négligeables de $X(\Omega)$.

4. D'après ce qui précède, les points de discontinuités de F_X font partie de $X(\Omega)$, qui est supposé au plus dénombrable, dont ces points forment eux aussi un ensemble au plus dénombrable.

Remarque

Plus généralement (et hors contexte probabiliste), on peut montrer que l'ensemble des points de discontinuité d'une fonction monotone est un ensemble au plus dénombrable.

Exercice 6 (***)Taux de panne

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probablisé, X une variable aléatoire discrète, définie sur cet espace probablisé, 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 $(x_n)_{n \in \mathbb{N}^*}$ dont le terme général est la probabilité conditionnelle :

$$x_n = \mathbb{P}_{(X \geq n)}(X = n).$$

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.

Corrigé de l'exercice 6

1. On a

$$x_n = P_{(X \geq n)}(X = n) = \frac{P((X = n) \cap (X \geq n))}{P(X \geq n)} = \frac{P(X = n)}{P(X \geq n)}.$$

D'autre part :

$$p_n = P(X = n) = P(X \geq n) - P(X \geq n + 1), \quad P(X = n) = x_n P(X \geq n),$$

donc $P(X \geq n + 1) = (1 - x_n)P(X \geq n)$.

Puis, en itérant, $P(X \geq n) = \left(\prod_{k=1}^{n-1} (1 - x_k) \right) P(X \geq 1) = \prod_{k=1}^{n-1} (1 - x_k)$, vu que $P(X \geq 1) = 1$.

Donc $p_n = \prod_{k=1}^{n-1} (1 - x_k) - \prod_{k=1}^n (1 - x_k) = x_n \prod_{k=1}^{n-1} (1 - x_k)$.

2. (a) On a $x_n \in [0, 1]$ en tant que probabilité.
De plus, pour tout $n \geq 1$, $x_n \neq 1$ car sinon on aurait pour un certain n_0 , $x_{n_0} = 1$ et donc :

$$P(X \geq n_0 + 1) = \prod_{k=1}^{n_0} (1 - x_k) = 0,$$

ce qui contredit les hypothèses. D'où $\forall n \in \mathbb{N}^*, 0 \leq x_n < 1$.

Puis, d'après la question précédente, $\prod_{k=1}^n (1 - x_k) = P(X \geq n + 1)$ donc

$$\sum_{k=1}^n \ln(1 - x_k) = \ln(P(X \geq n + 1)).$$

Or, $P(X \geq n + 1) = \sum_{k=n+1}^{+\infty} P(X = k) \xrightarrow{n \rightarrow +\infty} 0$ comme reste d'une série convergente.

Donc $\prod_{k=1}^n \ln(1 - x_k) \xrightarrow{n \rightarrow +\infty} -\infty$.

Or, soit on a $x_n \rightarrow 0$, d'où $\ln(1 - x_n) \sim -x_n$ et donc on a prouvé que $\sum x_n$ diverge par comparaison de séries à termes négatifs. Soit on a $x_n \not\rightarrow 0$ et on a aussi $\sum x_n$ qui diverge (grossièrement).

Donc, dans tous les cas, $\sum x_n$ diverge.

- (b) Soit $(x_n)_{n \in \mathbb{N}^*}$ une suite à valeurs dans $[0; 1[$ telle que la série de terme général x_n diverge. Conformément aux calculs précédents, on définit :

$$p_1 = x_1, \quad \forall n \geq 2, \quad p_n = x_n \prod_{k=1}^{n-1} (1 - x_k).$$

On note qu'on a : $\forall n \geq 2, p_n = u_{n-1} - u_n$, où $u_n = \prod_{k=1}^n (1 - x_k)$ donc

$$P_n = \sum_{k=1}^n p_k = p_1 + u_1 - u_n = x_1 + 1 - x_1 - u_n = 1 - u_n.$$

Or $\ln(u_n) = \sum_{k=1}^n \ln(1 - x_k) \leq \sum_{k=1}^n (-x_k)$ par concavité de $x \mapsto \ln(1+x)$, et donc, vu que $\sum x_n$ diverge et est à termes positifs, $\ln(u_n) \rightarrow -\infty$, c'est-à-dire $u_n \rightarrow 0$ et donc $P_n \rightarrow 1$.

Ainsi : $\forall n \geq 1, p_n \in [0, 1[$ et $(p_n)_{n \geq 1}$ est sommable de somme égale à 1 donc on définit une probabilité sur \mathbb{N}^* en posant : $\forall n \in \mathbb{N}^*, P(\{n\}) = p_n = x_n \prod_{k=1}^{n-1} (1 - x_k)$.

Et toute variable aléatoire X à valeurs dans \mathbb{N}^* qui admet comme loi de probabilité P a pour taux de panne la suite $(x_n)_{n \geq 1}$ car :

$$P(X \geq n) = 1 - P(X \leq n - 1) = 1 - \sum_{k=1}^{n-1} p_k = \prod_{k=1}^{n-1} (1 - x_k) > 0, \text{ et}$$

$$\frac{P(X = n)}{P(X \geq n)} = \frac{x_n \prod_{k=1}^{n-1} (1 - x_k)}{\prod_{k=1}^{n-1} (1 - x_k)} = x_n.$$

3. \Rightarrow : On suppose $X \sim \mathcal{G}(p)$ avec $p \in]0, 1[$.

Alors : $\forall n \geq 1, P(X = n) = p(1 - p)^{n-1}$ et $P(X \geq n) = (1 - p)^{n-1} > 0$ donc $x_n = \frac{P(X=n)}{P(X \geq n)} = p$ et ainsi le taux de panne est bien constant.

\Leftarrow : On suppose le taux de panne constant. Ainsi : $\exists p \in [0, 1]$ tq $\forall n \geq 1, \frac{P(X=n)}{P(X \geq n)} = p$.
Et $p \neq 0$ car sinon $P(X = n) = 0$ pour tout $n \geq 1$ et donc $P(\Omega) = 0$: impossible.

Les calculs précédents donnent alors : $P(X = n) = p_n = x_n \prod_{k=1}^{n-1} (1 - x_k) = p(1 - p)^{n-1}$, d'où $X \sim \mathcal{G}(p)$.

II TD du mercredi 11/03

Exercice 7 (*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.

Corrigé de l'exercice 7

Exercice 8 (*)

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

Corrigé de l'exercice 8

Exercice 9 (**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 $\sum_{k=r}^{+\infty} \binom{k}{r} x^{k-r} = \frac{1}{(1-x)^{r+1}}$ pour $x \in]-1, 1[$, tout en expliquant d'où vient cette formule.

Corrigé de l'exercice 9

Exercice 10 (**Loi binomiale négative : une autre approche)

On dit qu'une variable aléatoire X suit une **loi binomiale négative** de paramètres n et p lorsque

$$X(\Omega) = [n, +\infty[\text{ et } \mathbb{P}(X = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}.$$

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

Corrigé de l'exercice 10

Exercice 11 (**Espérance conditionnelle)

Soit X une VAD réelle et admettant une espérance. Soit aussi A un événement de probabilité non nulle. On définit $E(X|A) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x|A)$.

Soit $(B_n)_{n \geq 1}$ un système complet d'événements de probabilités non nulles. Montrer que :

$$E(X) = \sum_{n=1}^{+\infty} \mathbb{P}(B_n) E(X|B_n).$$

Corrigé de l'exercice 11

Exercice 12 ()**

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

$$\text{Cov}(X_1 + Y_1, X_2 + Y_2) = \text{Cov}(X_1, X_2) + \text{Cov}(Y_1, Y_2).$$

Corrigé de l'exercice 12**Exercice 13 (**)**

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 $\text{Cov}(Y_1, Y_2)$ puis le *coefficient de corrélation* ρ de ces deux variables, défini par

$$\rho = \frac{\text{Cov}(Y_1, Y_2)}{\sigma_{Y_1} \sigma_{Y_2}}.$$

Corrigé de l'exercice 13**III TD du lundi 16/03****Exercice 14 (*Production)**

On suppose qu'une usine fabrique en moyenne 50 objets par semaine.

1. Majorer 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, que dire de la probabilité pour que le nombre d'objets fabriqués la semaine prochaine soit compris entre 40 et 60 ?

Corrigé de l'exercice 14**Exercice 15 (*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 $S_n = \frac{X_1 + \dots + X_n}{n}$ et $m_n = \frac{p_1 + \dots + p_n}{n}$.

Montrer que $\forall \varepsilon > 0, \mathbb{P}(|S_n - m_n| \geq \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0$.

Corrigé de l'exercice 15**Exercice 16 (**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 $S_n = \frac{Y_n}{n}$.

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 : $\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|S_n - e^{-a}| \geq \varepsilon) = 0$.

Corrigé de l'exercice 16

Exercice 17 (*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 $\sum_{k=0}^{+\infty} (k+n)(k+n-1) \cdots (k+1)x^k$.
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 ...

Corrigé de l'exercice 17**Exercice 18 (**Trucage impossible)**

À l'aide de fonctions génératrices, prouver qu'il est impossible de truquer deux dés de sorte que la somme d'un lancer des deux dés suive une loi uniforme sur $\{2, \dots, 12\}$.

On pourra s'intéresser aux racines des fonctions formées.

Corrigé de l'exercice 18**Exercice 19 (***)Points fixes de permutations)**

On considère la variable X_n égale au nombre de cycles disjoints d'une permutation de l'ensemble $\{1, \dots, n\}$.

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

Corrigé de l'exercice 19**IV Exercices supplémentaires****Exercice 20 (*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.

Corrigé de l'exercice 20

Exercice 21 ()**

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 :

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

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

Corrigé de l'exercice 21**Exercice 22 (**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 .

Corrigé de l'exercice 22**Exercice 23 (*)**

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.

Corrigé de l'exercice 23

Exercice 24 (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 : $E(X) = \frac{N}{k} + N(1 - (1 - p)^k)$.

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 ?

Corrigé de l'exercice 24**Exercice 25 (*Utilisation d'une fonction génératrice)**

Soit X une variable aléatoire telle que : $\forall k \in \mathbb{N}^*, P(X = k) = \frac{k-1}{2^k}$.

1. Vérifier par le calcul que $\sum_{k=1}^{+\infty} P(X = k) = 1$.
2. Donner la fonction génératrice de X . Quel est son rayon de convergence ?
3. La variable X admet-elle une espérance finie ? Si oui, que vaut-elle ?

Corrigé de l'exercice 25**Exercice 26 (**)**

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.

Corrigé de l'exercice 26