# Variables aléatoires discrètes

S6 Polytech

1. **Autour des expériences de Bernoulli.** On considère une urne contenant $N$ boules, dont $M$ boules blanches et $N-M$ boules rouges. Donner dans chacun des cas suivants la loi, l'espérance et la variance de la variable aléatoire $X$.
   (a) On tire au hasard une boule de l'urne ; $X$ est la variable qui vaut 1 si la boule est blanche et $-1$ si la boule est rouge.
   (b) On tire au hasard des boules de l'urne, en les remettant à chaque fois, jusqu'à ce qu'on obtienne un boule blanche ; $X$ est le nombre d'essais nécessaires.
   (c) On tire au hasard des boules de l'urne, en les remettant à chaque fois ; $X$ est le nombre de boules blanches obtenues après $n$ essais.
   (d) On tire au hasard des boules de l'urne, en les retirant à chaque fois, jusqu'à ce qu'on obtienne un boule blanche ; $X$ est le nombre d'essais nécessaires.
   (e) On tire au hasard des boules de l'urne, en les retirant de l'urne à chaque fois ; $X$ est le nombre de boules blanches obtenues après $n \leq \min(M, N-M)$ essais.

2. **Une formule pour l'espérance d'une variable positive.** Soit $X$ une variable aléatoire discrète à valeurs positives, dans $\mathbb{N}$.
   (a) Montrer que $\mathbb{E}[X] = \sum_{k=0}^{\infty} \mathbb{P}[X > k] = \sum_{k=0}^{\infty} \mathbb{P}_X[\{k+1, k+2, \dots\}]$.
   (b) Application : dans une urne contenant $B$ boules blanches et $N$ boules noires, on retire des boules une à une jusqu'à obtenir une boule blanche. Montrer que si $X$ est le nombre de tirages nécessaires, alors

   $$\mathbb{E}[X] = \frac{N+B+1}{B+1}$$

   (indication : on pourra remarquer que si $M = B+N$, alors la somme à calculer vérifie l'équation de récurrence $S(N, M) = 1 + \frac{N\, S(N-1, M-1)}{M}$.)

3. **Fonctions indicatrices.** Soit $(\Omega, \mathbb{P})$ un espace de probabilité, $A \subset \Omega$ et $1_A$ la fonction sur $\Omega$ définie par :

   $$1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon.} \end{cases}$$

   On dit que $1_A$ est la fonction indicatrice de $A$, et elle fournit une variable aléatoire à valeurs dans $\{0, 1\}$. Donner la loi de cette variable discrète $1_A$. Calculer $\mathbb{E}[1_A]$.

4. **L'inégalité de Markov.** Soit $X$ une variable aléatoire discrète, à valeurs positives. Montrer que

   $$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

   pour tout $t > 0$. On pourra comparer $X$ et $X\, 1_{X \geq t}$, et utiliser l'exercice précédent.

5. **L'inégalité de Bienaymé-Chebyshev.** Soit $X$ une variable aléatoire discrète, telle que $\mathrm{var}(X)$ soit bien définie. Montrer que pour tout $t > 0$,

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{var}(X)}{t^2}.$$

Commenter. Si $X \sim \mathcal{B}(n,p)$, en déduire que pour tout $\varepsilon > 0$,

$$\mathbb{P}[|X - np| \geq n\varepsilon] \to_{n\to\infty} 0.$$

6. **L'inégalité de Chernov.** On considère encore une variable aléatoire $X$ discrète.
   (a) Montrer que pour tout $t \in \mathbb{R}$,

   $$\mathbb{P}[X \geq t] \leq \mathrm{e}^{\inf_{\lambda \in \mathbb{R}_+} \left(\log \mathbb{E}[\mathrm{e}^{\lambda X}] - \lambda t\right)}.$$

   (b) Calculer $\log \mathbb{E}[\mathrm{e}^{\lambda X}]$ lorsque $X \sim \mathcal{B}(n,p)$ (indication : écrire $X$ comme une somme de $n$ variables de Bernoulli indépendantes, et factoriser l'espérance $\mathbb{E}[\mathrm{e}^{\lambda X}]$).
   (c) En déduire que si $\varepsilon > 0$, alors il existe $0 < C(\varepsilon) < 1$ telle que

   $$\mathbb{P}[X \geq n(p+\varepsilon)] \leq (C(\varepsilon))^n.$$

   Comparer avec le résultat de l'exercice précédent.

7. **Échanges entre deux urnes**. On considère deux urnes contenant chacune $N$ boules : l'urne $U_1$ contient initialement $N$ boules blanches et l'urne $U_2$ contient initialement $N$ boules rouges. À chaque temps $n \in \mathbb{N}$, on tire une boule au hasard dans $U_1$ et une boule au hasard dans $U_2$ et on les échange. Soit $X_n$ le nombre de boules blanches dans l'urne $U_1$ au temps $n$. Calculer par récurrence $\mathbb{E}[X_n]$.

8. **Somme aléatoire de lois de Bernoulli.** Soit $N$ une variable aléatoire de loi de Poisson $\mathcal{P}(\lambda)$, et $X_1, X_2, \ldots$ une suite de variables de Bernoulli $\mathcal{B}(p)$ indépendantes entre elles et indépendantes de la variable $N$. On pose

$$X = \sum_{i=1}^{N} X_i.$$

Calculer $\mathbb{P}[N = n \text{ et } X = k]$ pour $k \leq n$. En déduire que $X$ et $N - X$ sont des variables indépendantes, et donner leurs lois.

9. **Lois puissances.** Soit $X$ une variable aléatoire à valeurs dans $\mathbb{N}^*$, avec

$$\mathbb{P}[X = k] = \frac{C}{k^\alpha}$$

pour tout $k \geq 1$, $\alpha$ étant une constante strictement positive. Discuter suivant la valeur de $\alpha$ :
— de l'existence d'une telle variable aléatoire ;
— de l'existence de l'espérance $\mathbb{E}[X]$.

10. **Lois hypergéométriques.** On considère une population de $N$ individus, avec deux types d'individus : $M$ individus de type 1, et $N-M$ individus de type 2. On tire $n \leq \min(M, N-M)$ personnes au hasard parmi les $N$ individus, et on note $X$ le nombre d'individus de type 1 parmi cette sous-population de $n$ personnes. Ce nombre peut être n'importe quel entier entre 0 et $n$.

   (a) Montrer que la loi de $X$ est donnée par la formule

   $$\mathbb{P}[X = k] = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \quad \text{pour tout } k \in [\![0, n]\!] \,.$$

   On dit que c'est la loi hypergéométrique de paramètres $N$, $M$ et $n$.

   (b) En déduire la formule de Vandermonde :

   $$\sum_{k=0}^{N} \binom{M}{k}\binom{N-M}{n-k} = \binom{N}{n}.$$

   (c) Calculer l'espérance et la variance de $X$. On utilisera la formule

   $$k\binom{n}{k} = n\binom{n-1}{k-1}$$

   pour tout $n \geq k$.

   (d) Application : on tire deux mains de 2 cartes pour deux joueurs de poker, et 3 cartes que l'on pose faces visibles sur le tapis (*flop*). Le joueur $A$ a un coeur parmi ses 2 cartes, et il y a 2 coeurs parmi les cartes visibles. La suite du jeu (*turn* et *river*) révèle deux cartes en plus des 3 déjà visibles. Soit $X$ le nombre de coeurs parmi ces 2 nouvelles cartes. Donner la loi de $X$, et calculer la probabilité pour que $X = 2$ et ainsi que le joueur ait une couleur.

   (e) Étant donnée une variable $X$ de loi hypergéométrique $\mathcal{H}(N, M, n)$, on fixe $n$ et on suppose que $M = pN$ et que $N$ tend vers l'infini, $p$ étant un paramètre entre 0 et 1. Calculer la limite de $\mathbb{P}[X = k]$ pour tout $k \in [\![0, n]\!]$.

11. **La marche aléatoire simple.** On considère une particule dont la position $P_n$ à tout temps $n$ est un entier $k \in \mathbb{Z}$, et qui se déplace par sauts de taille 1. À chaque instant, la particule a une probabilité $\frac{1}{2}$ de sauter vers le haut, et une probabilité $\frac{1}{2}$ de sauter vers le bas, tous les sauts étant indépendants. Autrement dit,

   $$P_{n+1} = X_{n+1} + P_n,$$

   où $(X_n)_{n\in\mathbb{N}}$ est une suite de variables indépendantes, avec

   $$\mathbb{P}[X_i = +1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$$

   pour tout $i$.

   (a) On suppose $P_0 = 0$. En fonction de $n \geq 0$, quelles sont les valeurs possibles pour $P_n$ ? Calculer pour ces valeurs $\mathbb{P}[P_n = k]$.

(b) On fixe un temps impair $t = 2n+1$, et on cherche à calculer la probabilité pour que $X_{2n+1} = 1$ et pour que tous les $X_{i\leq 2n}$ soient négatifs ou nuls. Autrement dit, on veut calculer la probabilité de l'événement

"la marche aléatoire atteint $k = 1$ pour la première fois au temps $t = 2n + 1$".

On note $C_n$ le nombre d'excursions négatives de taille $2n$, c'est-à-dire le nombre de chemins $x_0, x_1, \dots, x_{2n}$ avec $x_0 = x_{2n} = 0$ et tous les $x_{i\leq 2n}$ négatifs ou nuls. Montrer que $C_n$ vérifie l'équation de récurrence :

$$C_{n+1} = \sum_{k=0}^{n} C_k C_{n-k}.$$

En déduire que si $C(z) = \sum_{k=0}^{\infty} C_k\, z^k$, alors $C(z) = z\,(C(z))^2 + 1$.

(c) Résoudre l'équation vérifiée par $C(z)$, et montrer que $C_n = \frac{1}{n+1}\binom{2n}{n}$ pour tout $n$.

(d) Soit $T_1$ le temps aléatoire nécessaire pour que $(X_n)_{n\in\mathbb{N}}$ atteigne le niveau 1. Calculer $\mathbb{P}[T_1 = 2n + 1]$, puis $\mathbb{P}[T_1 < +\infty]$.

12. **L'inégalité de Fortuin–Kasteleyn–Ginibre.** Soit $f, g : \{0,1\}^n \to \mathbb{R}$ deux fonctions croissantes, c'est-à-dire que pour tous vecteurs $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$ de $\{0,1\}^n$ tels que $x_1 \leq y_1, \dots, x_n \leq y_n$, on ait

$$f(x) \leq f(y) \quad ; \quad g(x) \leq g(y).$$

On considère $n$ variables $X_i$ indépendantes, chacune suivant une loi de Bernoulli $\mathcal{B}(p)$. Montrer que les variables aléatoires $f(X_1, \dots, X_n)$ et $g(X_1, \dots, X_n)$ ont une covariance positive :

$$\mathrm{cov}(f(X_1, \dots, X_n), g(X_1, \dots, X_n)) \geq 0.$$

On pourra raisonner par récurrence sur $n$.

13. **Théorème de Turan.** Soit $G$ un graphe, c'est-à-dire une paire $(V, E)$ avec $V$ ensemble de sommets, et $E$ ensemble d'arêtes $\{v, w\}$ avec $v \neq w$ et $v, w \in V$. Si $v \in V$, on note $\deg(v)$ le nombre de sommets connectés à $v$, c'est-à-dire le nombre de paires $\{v, w\} \in E$ avec $w \in V$. Un ensemble de *sommets indépendants* de $G$ est un ensemble $I \subset V$ qui ne contient pas de sommets connectés.

Dans ce qui suit, on ordonne les sommets $v_1, \dots, v_n$ de $G$ aléatoirement, et on construit récursivement des ensembles $\emptyset = I_0, I_1, \dots, I_n$ en ajoutant $v_k$ à $I_{k-1}$ si $v_k$ n'est connecté à aucun sommet de $I_{k-1}$.

(a) Montrer que tous les ensembles $I_k$ sont des ensembles de sommets indépendants.

(b) Calculer la probabilité pour que $I_n$ contienne le sommet $v_k$, puis l'espérance du nombre de sommets de $I_n$.

(c) En déduire que $G$ admet au moins un ensemble de sommets indépendants avec plus de

$$\sum_{v\in V} \frac{1}{\deg(v) + 1}$$

sommets.