TD5. Variables aléatoires réelles et inégalités classiques

1. Fonction génératrice Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace de probabilités. Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire telle que $X(\Omega) \subset \mathbb{N}$.

a) Montrer que pour tout réel $s \in [0,1]$, la fonction s^X est une variable aléatoire intégrable. *On rappelle que par convention,* $0^0 = 1$.

On appelle *fonction génératrice* de X la fonction

$$\begin{aligned} G_X : [0,1] &\longrightarrow \mathbb{R} \\ s &\longmapsto \mathbb{E}[s^X]. \end{aligned}$$

b) Montrer que G_X est une fonction positive croissante. Calculer ses valeurs en 0 et en 1.

c) Calculer la fonction G_X lorsque X suit l'une des lois suivantes :
   i) Bernoulli de paramètre $p \in [0,1]$,
   ii) binomiale de paramètres $n \geq 0$ et $p \in [0,1]$,
   iii) géométrique de paramètre $p \in]0,1[$,
   iv) Poisson de paramètre $\lambda > 0$.

d) Que peut-on dire de deux variables aléatoires à valeurs entières qui ont la même fonction génératrice ?

Solution de l'exercice 1.

a) Soit $s \in (0,1]$. L'hypothèse $X(\Omega) \subset \mathbb{N}$ est inutile pour la mesurabilité, et pour l'intégrabilité on peut se contenter de $X(\Omega) \subset \mathbb{R}^+$. Comme la fonction $x \in \mathbb{R} \mapsto s^x \in \mathbb{R}$ est continue, le fait que X est une variables aléatoire entraine que s^X aussi (proposition 2.2.2b du poly). Comme elle est bornée (par 1), il s'agit d'une variable aléatoire intégrable. Pour $s = 0$, on a $s^X = \mathbb{1}_{\{X=0\}}$, qui est aussi une variable aléatoire, elle aussi bornée donc intégrable.

b) Soient s et t des réels tels que $0 \leq s \leq t \leq 1$. Pour tout $\omega \in \Omega$, on a $0 \leq s^{X(\omega)} \leq t^{X(\omega)} \leq 1$. Par positivité de l'espérance (théorème 2.3.1b), on en déduit que $0 \leq G_X(s) \leq G_X(t) \leq 1$. G_X est donc positive et croissante. Enfin, $G_X(0) = \mathbb{P}(X = 0)$ et $G_X(1) = 1$.

c) Soit X une variable aléatoire de loi :
   i) Bernoulli de paramètre $p \in [0,1]$. La fonction génératrice de X est

$$G_X(s) = \mathbb{E}[s^X] = s^0\mathbb{P}(X = 0) + s^1\mathbb{P}(X = 1) = 1 - p + ps.$$

ii) binomiale de paramètres $n \geq 0$ et $p \in [0,1]$. La fonction génératrice de X est

$$G_X(s) = \sum_{k=0}^{n} \binom{n}{k} s^k p^k (1-p)^{n-k} = (1-p+sp)^n.$$

iii) géométrique de paramètre $p \in]0,1[$. La fonction génératrice de X est

$$G_X(s) = \sum_{n \geq 0} s^n p(1-p)^n = \frac{p}{1-s(1-p)}.$$

Notons que son rayon de convergence est égal à $\frac{1}{p}$.

iv) Poisson de paramètre $\lambda > 0$. La fonction génératrice de X est

$$G_X(s) = e^{-\lambda} \sum_{n \geq 0} s^n \frac{\lambda^n}{n!} = e^{-\lambda + s\lambda} = e^{\lambda(s-1)}.$$

d) Deux variables aléatoires à valeurs entières qui ont la même fonction génératrice ont même loi (et la réciproque est triviale). Pour le voir, on remarque que ces fonctions sont C^∞ et que la dérivée d'ordre n de G_X est $G_X^{(n)}(s) = \mathbb{E}\left[\frac{X!}{(X-n)!} s^{X-n} \mathbb{1}_{\{X \geq n\}}\right]$. En particulier, $G_X^{(n)}(0) = n!\mathbb{P}(X=n)$, ce qui permet de retrouver la loi à partir de la fonction génératrice.

2. Soit X une variable aléatoire. Déterminer pour quelles valeurs de $\lambda \in \mathbb{R}$ la variable $e^{\lambda X}$ est intégrable et calculer $\mathbb{E}[e^{\lambda X}]$ dans chacun des cas suivants :

a) X suit la loi uniforme sur un intervalle $[a,b]$,

b) X suit la loi exponentielle de paramètre $\theta > 0$,

c) X suit la loi normale $\mathcal{N}(0,1)$.

Solution de l'exercice 2.

a) Pour tout $\lambda \in \mathbb{R}$, la variable aléatoire $e^{\lambda X}$ est bornée (lorsque X suit une loi uniforme sur $[a,b]$), et donc intégrable. De plus,

$$\mathbb{E}[e^{\lambda X}] = \int_a^b (b-a)^{-1} e^{\lambda x} dx = (\lambda(b-a))^{-1} [e^{\lambda x}]_a^b = (\lambda(b-a))^{-1} [e^{\lambda b} - e^{\lambda a}].$$

b)

$$E[e^{\lambda X}] = \int_0^{+\infty} \theta e^{(\lambda - \theta)x} dx = \begin{cases} +\infty \text{ si } \lambda \geq \theta, \\ \frac{\theta}{\theta - \lambda} \text{ si } \lambda < \theta. \end{cases}$$

c) Pour la loi normale, en faisant le changement de variable $y = x - \lambda$, il vient $\lambda x - x^2/2 = -y^2/2 + \lambda^2/2$ et on trouve ainsi

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\lambda x - x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2 + \lambda^2/2} dy = e^{\lambda^2/2}.$$

3. Montrer qu'une variable aléatoire positive dont l'espérance est nulle est nulle presque sûrement.

Solution de l'exercice 3. Soit $X : (\Omega, \mathscr{F}, \mathbb{P}) \to \mathbb{R}_+$ une variable aléatoire réelle positive. Il s'agit de montrer que si $\mathbb{E}[X] = 0$, alors $\mathbb{P}(X > 0) = 0$. Pour tout $n \geq 1$, définissons un événement $A_n \in \mathscr{F}$ en posant $A_n = \{X \geq \frac{1}{n}\}$. La suite d'événements $(A_n)_{n\geq 1}$ est croissante et vérifie $\bigcup_{n\geq 1} A_n = \{X > 0\}$. On en déduit que $\mathbb{P}(X > 0)$ est la limite des $\mathbb{P}(X \geq \frac{1}{n})$ lorsque n tend vers l'infini.

En applicant l'inégalité de Markov, on obtient alors $\mathbb{P}(X \geq \frac{1}{n}) \leq n\mathbb{E}(X) = 0$ d'où $\mathbb{P}(X \geq \frac{1}{n}) = 0$ pour tout n et en passant à la limite, $\mathbb{P}(X > 0) = 0$.

4. Calculer la fonction de répartition de la variable aléatoire X définie sur l'espace $([0,1], \mathscr{B}_{[0,1]}, \mathsf{Leb})$ par

$$\forall t \in [0,1],\ X(t) = \begin{cases} -\log(2t) & \text{si } t \leq \frac{1}{2} \\ 0 & \text{sinon.} \end{cases}$$

Existe-t-il une partie dénombrable $D \subset \mathbb{R}$ telle que $X \in D$ presque sûrement ? La loi de X admet-elle une densité ?

Solution de l'exercice 4. On calcule la fonction de répartition de X. Pour tout $x < 0$, on a $\mathbb{P}(X \leq x) = 0$. Pour $x = 0$, on trouve $\mathbb{P}(X \leq 0) = \mathsf{Leb}(]\frac{1}{2}, 1]) = \frac{1}{2}$. Pour $x > 0$, on a $\mathbb{P}(X \leq x) = \mathsf{Leb}([\frac{1}{2}e^{-x}, 1]) = 1 - \frac{1}{2}e^{-x}$. Finalement,

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{2} + \frac{1}{2}(1 - e^{-x}) & \text{si } x \geq 0. \end{cases}$$

Ainsi, $F_X = \frac{1}{2}(F_{\delta_0} + F_\eta)$, où $\eta(dx) = e^{-x}\mathbb{1}_{\mathbb{R}_+}(x)\ dx$ est la loi exponentielle de paramètre 1.

Pour tout $x \in \mathbb{R}$, on a $\mathbb{P}(X = x) = 0$ si $x \neq 0$ et $\mathbb{P}(X = 0) = \frac{1}{2}$. Soit D une partie dénombrable de $\mathbb{R}$. Si 0 appartient à D, on a $\mathbb{P}(X \in D) = \frac{1}{2}$, sinon $\mathbb{P}(X \in D) = 0$. Dans aucun cas on n'a $\mathbb{P}(X \in D) = 1$. La variable aléatoire X n'est donc pas une variable aléatoire discrète.

Par ailleurs, le singleton $\{0\}$, qui est de mesure de Lebesgue nulle, a une masse $\frac{1}{2}$ sous la loi de X. Ainsi, la loi de X n'est pas absolument continue par rapport à la mesure de Lebesgue, et n'admet en particulier pas de densité.

Finalement, X n'est ni une variable discrète ni une variable à densité.

5. Soit X une variable aléatoire qui suit la loi normale $\mathcal{N}(0,1)$. Montrer que pour tout $n \in \mathbb{N}$, la variable aléatoire X^n est intégrable et calculer $\mathbb{E}[X^n]$. Vérifier que pour tout $n \geq 0$, $\mathbb{E}[X^n]$ est le nombre de manières d'apparier n points, c'est-à-dire le nombre de partitions de l'ensemble $\{1, \ldots, n\}$ par des paires.

Solution de l'exercice 5. La densité de la loi normale centrée réduite est la fonction $f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. On sait que pour tout $n \geq 1$, la fonction $x \mapsto |x|^n e^{-\frac{x^2}{2}}$ tend vers 0 lorsque x tend vers $+\infty$ ou $-\infty$. Soit $n \geq 1$ un entier. Puisque $x \mapsto |x|^{n+2}e^{-\frac{x^2}{2}}$ tend vers 0 en l'infini, on a $|x|^n e^{-\frac{x^2}{2}} = O(\frac{1}{x^2})$ en $+\infty$ et $-\infty$, si bien que l'intégrale $\int_{-\infty}^{+\infty} |x|^n e^{-\frac{x^2}{2}}\, dx$ converge. La loi normale centrée réduite admet donc des moments de tous les ordres.

Pour tout $n \geq 0$, posons

$$m_n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n e^{-\frac{x^2}{2}}\, dx.$$

Si n est impair, m_n est l'intégrale d'une fonction intégrable impaire, donc $m_n = 0$. Ceci peut se vérifier en faisant le changement de variable $y = -x$ qui donne la relation $m_n = -m_n$.

Pour $n = 0$, m_0 est l'intégrale de la densité d'une loi de probabilités, donc $m_0 = 1$. Soit $n \geq 2$ un entier pair. On écrit $n = 2p$. Une intégration par parties donne, pour tout $R > 0$,

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}} \int_{-R}^{+R} x^{2p} e^{-\frac{x^2}{2}}\, dx &= \frac{1}{\sqrt{2\pi}} \int_{-R}^{+R} \underbrace{x^{2p-1}}_{u} \underbrace{xe^{-\frac{x^2}{2}}}_{v'}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \left[-x^{2p-1}e^{-\frac{x^2}{2}}\right]_{-R}^{R} + (2p-1)\frac{1}{\sqrt{2\pi}} \int_{-R}^{+R} x^{2p-2} e^{-\frac{x^2}{2}}\, dx.
\end{aligned}$$

En faisant tendre R vers $+\infty$, on trouve la relation $m_{2p} = (2p-1)m_{2p-2}$, qu'on résout en $m_{2p} = (2p-1)(2p-3)\ldots 3.1$. Ce nombre est souvent noté $(2p)!!$ et vaut $\frac{(2p)!}{2^p p!}$.

Finalement, les moments de la loi normale centrée réduite sont donnés par

$$\forall n \geq 1,\ m_n = \begin{cases} 0 & \text{si } n \text{ est impair,} \\ (2p)!! = \frac{(2p)!}{2^p p!} & \text{si } n = 2p. \end{cases}$$

Pour apparier n points, il faut choisir avec lequel des $n-1$ autres éléments apparier le premier, puis il en reste $n-2$ à apparier pour lesquels on procède de même. On obtient la même équation de récurrence que précédemment, avec une unique possibilité si $n = 2$ et aucune si n est impair. Le nombre de manière d'apparier n points est donc égal au moment d'ordre n de la loi normale.

6. Soit $\theta > 0$ un réel. Soit X une variable aléatoire de loi exponentielle de paramètre θ. Montrer que pour tout entier $n \geq 1$, la variable aléatoire X^n est intégrable et calculer $\mathbb{E}[X^n]$. Donner une interprétation combinatoire de ce nombre lorsque $\theta = 1$.

Solution de l'exercice 6. Pour $n = 0$, on a évidemment $\mathbb{E}[X^0] = 1$. Soit $n \geq 1$. On intègre par parties (en dérivant le monôme et en primitivant l'exponentielle) :

$$\mathbb{E}[X^n] = \theta \int_0^{+\infty} x^n e^{-\theta x} dx = \theta \left[\frac{x^n e^{-\theta x}}{-\theta}\right]_0^{+\infty} + n \int_0^{+\infty} x^{n-1} e^{-\theta x} dx = \frac{n}{\theta}\mathbb{E}[X^{n-1}].$$

En raisonnant par récurrence, on obtient immédiatement $\mathbb{E}[X^n] = n!\theta^{-n}$.

Pour $\theta = 1$, $\mathbb{E}[X^n] = n!$ est le nombre de bijections d'un ensemble ayant n éléments dans lui même.

7. Soit $\lambda > 0$ un réel. Soit X une variable aléatoire de loi de Poisson de paramètre λ. Montrer que pour tout entier $k \geq 1$, la variable aléatoire $X(X-1)\dots(X-k+1)$ est intégrable et calculer son espérance. Calculer $\mathbb{E}[X^m]$ pour $m \in \{1,2,3,4\}$ et vérifier que pour chacune de ces valeurs de m, $\mathbb{E}[X^m]$ est le nombre de partitions d'un ensemble à m éléments lorsque $\lambda = 1$. *On peut démontrer que cette assertion est vraie pour tout $m \geq 1$.*

Solution de l'exercice 7. $Y_k := X(X-1)\dots(X-k+1)$ est une variable aléatoire positive, on peut donc calculer son espérance (éventuellement infinie, auquel cas elle n'est pas intégrable). En utilisant le fait que $Y_k = 0$ lorsque $X = 0, \dots, k-1$, on obtient

$$\mathbb{E}[Y_k] = e^{-\lambda} \sum_{i \geq k} \frac{i(i-1)\dots(i-k+1)\lambda^i}{i!} = e^{-\lambda}\lambda^k \sum_{i \geq k} \frac{\lambda^{i-k}}{(i-k)!} = \lambda^k < +\infty.$$

On sait que les Y_k permettent de retrouver les X^k par combinaison linéaire (famille échelonnée de polynômes, même si ici X désigne une variable aléatoire et pas une indéterminée). On trouve, en identifiant les coefficients

$$X = Y_1, \quad X^2 = Y_2 + Y_1, \quad X^3 = Y_3 + 3X^2 - 2X = Y_3 + 3Y_2 + Y_1,$$

$$X^4 = Y_4 + 6X^3 - 11X^2 + 6X = Y_4 + 6Y_3 + 7X^2 - 6X = Y_4 + 6Y_3 + 7Y_2 + Y.$$

On prend maintenant les espérances et on utilise la relation $\mathbb{E}[Y_k] = \lambda^k$ calculée plus haut pour obtenir les premiers moment de X :

$$\mathbb{E}[X] = \lambda, \quad \mathbb{E}[X^2] = \lambda^2 + \lambda, \quad \mathbb{E}[X^3] = \lambda^3 + 3\lambda^2 + \lambda, \quad \mathbb{E}[X^4] = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda.$$

Pour $\lambda = 1$, on obtient

$$\mathbb{E}[X] = 1, \quad \mathbb{E}[X^2] = 3, \quad \mathbb{E}[X^3] = 5, \quad \mathbb{E}[X^4] = 15.$$

On constate que pour ces 4 valeurs, $\mathbb{E}[X^m]$ est le nombre de partitions d'un ensemble à m éléments, et même que le coefficient devant λ^k est celui des partitions de cet ensemble en k sous-ensembles. Par exemple pour $m = 4$, on a : pour $k = 4$, une seule partition (composée de 4 singletons), pour $k = 3$, 6 partitions (composées d'une paire et de deux singletons), pour $k = 2$, 7 partitions (3 composées de deux paires et 4 composées d'un brelan et d'un singleton) et enfin pour $k = 1$ une seule (réduite à l'ensemble total).

8. Inégalité de Markov Soit $X : (\Omega, \mathscr{F}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire. Soit $a > 0$ un nombre réel.

a) Montrer que $\mathbb{P}(|X| \geq a) \leq \dfrac{\mathbb{E}(|X|)}{a}$.

b) Que peut-on dire de la proportion de la population qui gagne plus de dix fois le salaire moyen ?

Solution de l'exercice 8.

a) On vérifie facilement que pour tout $\omega \in \Omega$, $a\mathbb{1}_{\{|X(\omega)| \geq a\}} \leq |X(\omega)|$. Par positivité de l'espérance, on obtient en intégrant :

$$a\mathbb{P}(|X| \geq a) \leq \mathbb{E}(|X|).$$

Il suffit de diviser par a pour obtenir l'inégalité demandée.

b) En considérant que X est le salaire, et en choisissant $a = 10\mathbb{E}(|X|)$, on déduit de l'inégalité de Markov que la proportion de la population qui gagne plus de dix fois le salaire moyen est inférieure à $1/10$.