Recueil d'exercices 1. Corrigé.

Exercice 1 (Problème du collectionneur de coupons).

1. Soit $k \in \{1, \dots, n\}$. On observe que $X_{n,k}$ représente le temps qu'il faut attendre pour tirer un coupon différent des $k-1$ collectés précédemment. Ainsi, $X_{n,k}$ suit une loi géométrique de paramètre $p_{n,k} = 1 - (k-1)/n$, c'est-à-dire que
$$\forall r \in \mathbb{N}^* \qquad \mathbb{P}(X_{n,k} = r) = (1 - p_{n,k})^{r-1} p_{n,k}.$$
En particulier, on a alors $\mathbb{E}[X_{n,k}] = 1/p_{n,k}$ et $\mathrm{Var}(X_{n,k}) = (1-p_{n,k})/p_{n,k}^2$. De plus, à n fixé, les variables aléatoires $X_{n,k}$ sont indépendantes. Sachant que $T_n = X_{n,1} + \ldots + X_{n,n}$, on en déduit que
$$\mathbb{E}[T_n] = \sum_{k=1}^{n} \mathbb{E}[X_{n,k}] = \sum_{k=1}^{n} \frac{1}{p_{n,k}} = \sum_{k=1}^{n} \frac{n}{n-k+1} = n \sum_{\ell=1}^{n} \frac{1}{\ell} \sim n \ln n,$$
pour $n \to \infty$. De même, en utilisant en outre l'indépendance des $X_{n,k}$, on a
$$\mathrm{Var}(T_n) = \sum_{k=1}^{n} \mathrm{Var}(X_{n,k}) \leq \sum_{k=1}^{n} \frac{1}{p_{n,k}^2} = \sum_{k=1}^{n} \left(\frac{n}{n-k+1}\right)^2 = n^2 \sum_{\ell=1}^{n} \frac{1}{\ell^2} = \mathrm{O}(n^2).$$

2. Soit $\varepsilon > 0$. D'après la question précédente, il existe un entier $n_0 \geq 1$ tel que pour tout entier $n \geq n_0$,
$$\left| \frac{\mathbb{E}[T_n]}{n \ln n} - 1 \right| \leq \frac{\varepsilon}{2}.$$
On en déduit que
$$\left| \frac{T_n}{n \ln n} - 1 \right| > \varepsilon \qquad \Longrightarrow \qquad \left| \frac{T_n}{n \ln n} - \frac{\mathbb{E}[T_n]}{n \ln n} \right| \geq \left| \frac{T_n}{n \ln n} - 1 \right| - \left| \frac{\mathbb{E}[T_n]}{n \ln n} - 1 \right| > \frac{\varepsilon}{2}.$$
L'inégalité de Chebyshev conduit donc à
$$\mathbb{P}\left(\left| \frac{T_n}{n \ln n} - 1 \right| > \varepsilon \right) \leq \mathbb{P}\left(|T_n - \mathbb{E}[T_n]| > \frac{\varepsilon}{2} n \ln n \right) \leq \frac{4 \mathrm{Var}(T_n)}{(\varepsilon n \ln n)^2} = \mathrm{O}\left(\frac{1}{(\ln n)^2} \right).$$
On en déduit finalement que
$$\frac{T_n}{n \ln n} \xrightarrow[n \to \infty]{\mathbb{P}} 1.$$

Exercice 2 (Polynômes de Bernstein).

1. Soit $S_n(x) \sim \mathrm{Bin}(n, x)$. Alors $\mathbb{P}(S_n(x) = k) = \binom{n}{k} x^k (1-x)^{n-k}$ pour tout $k \in \mathbb{N}$. Ceci donne
$$\mathbb{E}[f(S_n(x)/n)] = \sum_{k \in \mathbb{N}} f(k/n) \mathbb{P}(S_n(x) = k) == \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} = B_n(x).$$

2. On a pour tout $\varepsilon > 0$, $x \in [0,1]$ et $n \in \mathbb{N}^*$,
$$\begin{aligned}
|B_n(x) - f(x)| &= |\mathbb{E}[f(S_n(x)/n) - f(x)]| \\
&\leq \mathbb{E}[|f(S_n(x)/n) - f(x)|] \\
&\leq \mathbb{E}[|f(S_n(x)/n) - f(x)| \mathbb{1}_{|S_n(x)/n - x| \leq \varepsilon}] + \mathbb{E}[|f(S_n(x)/n) - f(x)| \mathbb{1}_{|S_n(x)/n - x| > \varepsilon}] \\
&=: T_1 + T_2.
\end{aligned}$$
Fixons $\delta > 0$. Puisque f est continue et $[0,1]$ est compact, f est uniformément continue. Il existe alors $\varepsilon > 0$ tel que $|f(x) - f(y)| < \delta$ pour tout $|x - y| < \varepsilon$. En particulier, $T_1 < \mathbb{E}[\delta] = \delta$. De plus, f est bornée. L'inégalité de Chebychev donne alors
$$\begin{aligned}
T_2 \leq \|f\|_\infty \mathbb{E}[\mathbb{1}_{|S_n(x)/n - x| > \varepsilon}] &= \|f\|_\infty \mathbb{P}(|S_n(x)/n - x| > \varepsilon) \\
&\leq \|f\|_\infty \varepsilon^{-1} \mathrm{Var}(S_n(x)/n) \\
&= \|f\|_\infty \varepsilon^{-1} x(1-x)/n \leq \|f\|_\infty \varepsilon^{-1}/(4n).
\end{aligned}$$
Par ce qui précède, on a pour $n \geq (4\delta\varepsilon\|f\|_\infty)^{-1}$ et tout $x \in [0,1]$,
$$|B_n(x) - f(x)| \leq 2\delta.$$
Ceci montre que $B_n(x)$ tend uniformément vers $f(x)$ quand $n \to \infty$.

Exercice 3 (Variables aléatoires symétriques et fonctions caractéristiques).

1. On a pour toute variable aléatoire réelle X, $\varphi_{-X}(t) = \varphi_X(-t) = \varphi_X(t)^*$, où z^* est le conjugué complexe de $z \in \mathbb{C}$. Par conséquent,

$$X \overset{\text{loi}}{=} -X \iff \varphi_{-X}(t) = \varphi_X(t)\,\forall t \in \mathbb{R} \iff \varphi_X(t) = \varphi_X(t)^*\,\forall t \in \mathbb{R} \iff \varphi_X(t) \in \mathbb{R}\,\forall t \in \mathbb{R}.$$

2. La variable Z est symétrique et indépendante de X. Donc le couple $(-Z, Y)$ a même loi que (Z, Y). Donc $-ZY$ et ZY ont même loi. De plus, on a pour toute fonction $f : \mathbb{R} \to \mathbb{R}$ mesurable,

$$\mathbb{E}[f(X)] = \mathbb{E}[f(ZY)] = \frac{1}{2}(\mathbb{E}[f(Y)] + \mathbb{E}[f(-Y)]).$$

Par conséquent, on a $\mu_X = \frac{1}{2}(\mu_Y + \mu_{-Y})$ et en particulier, $\varphi_X = \frac{1}{2}(\varphi_Y + \varphi_{-Y}) = \frac{1}{2}(\varphi_Y + \varphi_Y^*) = \Re\varphi_Y$.

3. On a $\frac{1}{2}e^{-|x|} = \frac{1}{2}(e^{-x}\mathbf{1}_{x\geq 0} + e^{x}\mathbf{1}_{x<0})$. La fonction caractéristique de la loi exponentielle de paramètre 1 est $t \mapsto (1 - it)^{-1}$. Par la dernière partie de l'exercice, on a

$$\varphi_X(t) = \Re\frac{1}{1 - it} = \Re\frac{1 + it}{|1 - it|^2} = \frac{1}{1 + t^2}.$$

4. Soit X comme dans la dernière partie. Comme φ_X est intégrable sur $\mathbb{R}$, on a par la formule d'inversion de Fourier,

$$\frac{1}{2}e^{-|x|} = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi_X(t)\,dt = \frac{1}{2}\int_{\mathbb{R}} e^{-itx}\frac{1}{\pi(1 + t^2)}\,dt = \frac{1}{2}\int_{\mathbb{R}} e^{itx}\frac{1}{\pi(1 + t^2)}\,dt.$$

Par conséquent, la fonction caractéristique de la loi de Cauchy de paramètre 1 est la fonction $t \mapsto e^{-|t|}$. Si X_1, X_2 sont de loi de Cauchy de paramètre 1, alors

$$\varphi_{\frac{X_1+X_2}{2}}(t) = \varphi_{X_1}(t/2)^2 = e^{-2|t/2|} = e^{-|t|} = \varphi_{X_1}(t).$$

Par conséquent, la loi de $\frac{X_1+X_2}{2}$ est celle de X_1.

5. On a :

$$\begin{aligned}
\varphi_{X_1X_2}(t) &= \int\left(\int e^{-itxy}e^{-\frac{y^2}{2}}\frac{dy}{\sqrt{2\pi}}\right)e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}}\\
&= \int \varphi_{X_2}(tx)e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}}\\
&= \int e^{-(t^2+1)\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}}\\
&= \frac{1}{\sqrt{1 + t^2}}\int e^{-\frac{y^2}{2}}\frac{dy}{\sqrt{2\pi}} \qquad \text{(chgt. de var. } x \mapsto y/\sqrt{1 + t^2}\text{)}\\
&= \frac{1}{\sqrt{1 + t^2}}.
\end{aligned}$$

Puis $\varphi_{X_1X_2+X_3X_4}(t) = \frac{1}{1+t^2}$ et la question de la partie 3 montre que $X_1X_2 + X_3X_4$ suit la loi de densité $\frac{1}{2}e^{-|x|}$ sur $\mathbb{R}$.

Exercice 4 (Loi bêta). Puisque βG_1 et βG_2 suivent les lois $\Gamma(\alpha_1, 1)$ et $\Gamma(\alpha_2, 1)$, il suffit de considérer $\beta = 1$. On a alors pour une fonction f mesurable bornée,

$$
\begin{aligned}
&\mathbb{E}[f(G_1/(G_1+G_2))] \\
&= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty \int_0^\infty f(\tfrac{x}{x+y}) x^{\alpha_1-1} y^{\alpha_2-1} e^{-(x+y)}\, dxdy \\
&= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty \left(\int_0^\infty f(\tfrac{1}{1+z}) x^{\alpha_1+\alpha_2-1} z^{\alpha_2-1} e^{-x(1+z)}\, dz \right) dx && y \to xz \\
&= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty f(\tfrac{1}{1+z}) z^{\alpha_2-1} dz \int_0^\infty x^{\alpha_1+\alpha_2-1} e^{-x(1+z)}\, dx && \text{Fubini} \\
&= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty f(\tfrac{1}{1+z}) z^{\alpha_2-1} (1+z)^{-\alpha_1-\alpha_2}\, dz \int_0^\infty w^{\alpha_1+\alpha_2-1} e^{-w}\, dw && x \to w/(1+z) \\
&= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty f(\tfrac{1}{1+z}) z^{\alpha_2-1} (1+z)^{-\alpha_1-\alpha_2}\, dz \\
&= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^1 f(t) (\tfrac{1-t}{t})^{\alpha_2-1} t^{\alpha_1+\alpha_2} t^{-2}\, dt && \tfrac{1}{1+z} \to t \\
&= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^1 f(t) t^{\alpha_1-1} (1-t)^{\alpha_2-1}\, dt.
\end{aligned}
$$

La v.a. $G_1/(G_1+G_2)$ suit alors la loi de densité sur $(0,1)$,

$$\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} t^{\alpha_1-1}(1-t)^{\alpha_2-1}.$$

Dans le cas $\alpha_1 = \alpha_2 = 1$ (variables exponentielles), on obtient la loi uniforme sur $(0,1)$. Dans le cas $\alpha_1 = \alpha_2 = 1/2$ (loi de $N_1^2/(N_1^2+N_2^2)$ pour N_1, N_2 des gaussiennes centrées indépendantes), on obtient la loi de densité $1/(\pi\sqrt{t(1-t)})$. Celle-ci s'appelle la loi de l'arc sinus, car sa fonction de répartition est

$$F(x) = \int_0^x \frac{1}{\pi\sqrt{t(1-t)}}\, dt = \frac{1}{\pi} \arcsin(2t-1) + \frac{1}{2} = \frac{2}{\pi} \arcsin(\sqrt{x}).$$

Exercice 5.

1. On a pour tout $m > n$,

$$\mathbb{E}[|Y_m - Y_n|] = \mathbb{E}\left[\sum_{k=n+1}^m X_1 \cdots X_k\right] = \sum_{k=n+1}^m \mathbb{E}[X_1 \cdots X_k] = \sum_{k=n+1}^m 2^{-k} \le \sum_{k=n+1}^\infty 2^{-k} = 2^{-n}.$$

   Par conséquent, $(Y_n)_{n\ge 0}$ est une suite de Cauchy dans L^1 qui est un espace de Banach, et donc Y_n converge dans L^1 vers une limite Y.

2. On décompose

$$Y_n = \sum_{k=1}^n X_1 \cdots X_k = X_1 \left(1 + \sum_{k=2}^n X_2 \cdots X_k\right).$$

   Par le même argument que dans la dernière partie, la suite $Y_n' = \sum_{k=2}^n X_2 \cdots X_k$ converge dans L^1 vers une limite Y', si bien que

$$Y = X_1(1+Y').$$

   Puisque Y_n' est égale en loi à Y_{n-1}, les limites Y' et Y sont également égales en loi. De plus, puisque Y_n' est indépendante de X_1 pour tout n, la limite Y' est également indépendante de X_1 (*exo : pourquoi ?*). Il vient que $Y \overset{\text{loi}}{=} X(1+Y)$ avec $X \sim \mathrm{Unif}(0,1)$ indépendante de Y.

3. Par la dernière partie, on a pour tout $t \in \mathbb{R}$,

$$\begin{aligned}
\varphi(t) &= \mathbb{E}[e^{itX(1+Y)}] \\
&= \int_{[0,1]\times\mathbb{R}} e^{itx(1+y)}\,dx \otimes \mu_Y(dy) \\
&= \int_0^1 e^{itx}\left(\int_{\mathbb{R}} e^{itxy}\mu_Y(dy)\right)dx && \text{par Fubini} \\
&= \int_0^1 e^{itx}\varphi(tx)\,dx.
\end{aligned}$$

4. On a pour $t > 0$, par un changement de variables $x \mapsto x/t$,

$$\varphi(t) = \int_0^1 e^{itx}\varphi(tx)\,dx = \frac{1}{t}\int_0^t e^{ix}\varphi(x)\,dx,$$

si bien que pour tout $t \geq 0$,

$$t\varphi(t) = \int_0^t e^{ix}\varphi(x)\,dx$$

(c'est vrai pour $t = 0$ car $0\varphi(0) = 0$). Il vient que la fonction $f(t) = t\varphi(t)$ satisfait à l'équation différentielle

$$f'(t) = \frac{e^{it}}{t}f(t),$$

qui admet comme solution générale

$$f(t) = C\exp\left(\int_1^t e^{is}/s\,ds\right).$$

Puisque $1/t = \exp(-\ln t) = \exp(-\int_1^t 1/s\,ds)$, il vient que

$$\varphi(t) = C\exp\left(\int_1^t (e^{is}-1)/s\,ds\right).$$

Avec la condition $\varphi(0) = 1$, cela donne

$$\varphi(t) = \exp\left(\int_0^t (e^{is}-1)/s\,ds\right).$$

Puisque $\varphi(t) = \varphi(-t)^*$, cela donne l'expression suivante pour tout $t \in \mathbb{R}$ (avec $\int_0^t := -\int_t^0$ pour $t < 0$) :

$$\varphi(t) = \exp\left(\int_0^t (e^{is}-1)/|s|\,ds\right).$$

Exercice 6 (Théorème de Glivenko-Cantelli).

1. On rappelle que

$$F_n(x) = \frac{1}{n}\mathrm{Card}\{j \in \{1,\dots,n\} \mid X_j \leq x\}. \tag{1}$$

En particulier, $F_n(x)$ prend ses valeurs dans $\{0, 1/n, 2/n, \dots, (n-1)/n, 1\}$. Donc, pour tout $k \in \{0,\dots,n\}$,

$$\left\{F_n(x) = \frac{k}{n}\right\} = \bigcup_{\substack{J\subseteq\{1,\dots,n\}\\ \mathrm{Card}\,J=k}}\left(\bigcap_{j\in J}\{X_j \leq x\} \cap \bigcap_{j\in\{1,\dots,n\}\setminus J}\{X_j > x\}\right).$$

Comme les $X_1,\dots,X_n$ sont des variables aléatoires, cela assure que $F_n(x)$ est aussi une variable aléatoire.

2. Notons G_n la fonction de répartition de Γ_n. Pour tout $x \in \mathbb{R}$,

$$G_n(x) = \Gamma_n((-\infty,x]) = \frac{1}{n}\sum_{j=1}^n \mathbf{1}_{X_j\leq x} = \frac{1}{n}\mathrm{Card}\{j \in \{1,\dots,n\} \mid X_j \leq x\} = F_n(x).$$

3. Dans le cas où $F(x) = x$ pour tout $x \in [0,1]$, les variables X_n suivent la loi uniforme sur l'intervalle $[0,1]$. De plus, à $x \in [0,1]$ fixé, en utilisant (1) et la loi forte des grands nombres, on obtient

$$\text{p.s.} \qquad F_n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{X_j \le x\}} \xrightarrow[n\to\infty]{} \mathbb{E}[\mathbb{1}_{\{X_1 \le x\}}] = F(x).$$

En d'autres termes,

$$\forall x \in [0,1] \quad \text{p.s.} \qquad \lim_{n\to\infty} F_n(x) = F(x). \tag{2}$$

4. Il s'agit d'échanger le "pour tout $x \in [0,1]$" et le "presque sûrement" dans (2), ce qui ne peut se faire directement car x parcourt un ensemble infini non dénombrable. Cependant, on peut déduire de (2) que l'événement

$$A = \left\{ \forall x \in [0,1] \cap \mathbb{Q} \quad \lim_{n\to\infty} F_n(x) = F(x) \right\}$$

est de probabilité un, car on se restreint alors à prendre x dans un ensemble dénombrable. Plaçons-nous sur l'événement A et prenons un réel $x \in [0,1]$. Par densité de $\mathbb{Q}$ dans $\mathbb{R}$, il existe une suite croissante $(a_p)_{p\in\mathbb{N}}$ de rationnels et une suite décroissante $(b_p)_{p\in\mathbb{N}}$ de rationnels qui convergent toutes deux vers x. Comme la fonction F_n est croissante, on a $F_n(a_p) \le F_n(x) \le F_n(b_p)$ pour tous entiers n et p. On obtient, en faisant tendre n vers l'infini,

$$F(a_p) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(b_p).$$

La fonction F étant continue sur $\mathbb{R}$, le membre de gauche et le membre de droite tend tous deux vers $F(x)$ quand $p \to \infty$. On en déduit que sur l'événement presque sûr A,

$$\forall x \in [0,1] \qquad \lim_{n\to\infty} F_n(x) = F(x).$$

5. Sur un événement de probabilité un (l'événement A par exemple), la suite de fonctions $(F_n)_{n\in\mathbb{N}^*}$ converge simplement vers F sur l'intervalle $[0,1]$. Comme F est continue et comme toutes les fonctions F_n sont croissantes, on peut appliquer le théorème de Dini pour conclure que la convergence est uniforme sur $[0,1]$. Elle est en fait uniforme sur $\mathbb{R}$, puisque $F_n(x) = F(x)$ pour tout réel $x \notin [0,1]$ et tout entier $n \ge 1$.

6. On a clairement pour tous réels $x \in \mathbb{R}$ et $u \in [0,1]$,

$$F^{-1}(u) \le x \qquad \Longleftrightarrow \qquad u \le F(x).$$

Donc, si U suit la loi uniforme sur $[0,1]$, on a pour tout réel $x \in \mathbb{R}$,

$$\{F^{-1}(U) \le x\} = \{U \le F(x)\} \qquad \text{et} \qquad \mathbb{P}(F^{-1}(U) \le x) = \mathbb{P}(U \le F(x)) = F(x),$$

ce qui prouve que $F^{-1}(U)$ est une variable aléatoire de fonction de répartition égale à F.

7. La question précédente assure que si $U_1, \dots, U_n$ désignent des variables aléatoires indépendantes et de loi uniforme sur $[0,1]$, alors $(F^{-1}(U_1), \dots, F^{-1}(U_n))$ a même loi que $(X_1, \dots, X_n)$. Donc, en désignant par $\stackrel{\mathcal{L}}{=}$ l'égalité en loi, on a

$$\begin{aligned} V_n &= \sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{X_j \le x\}} - F(x) \right| \\ &\stackrel{\mathcal{L}}{=} \sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{F^{-1}(U_j) \le x\}} - F(x) \right| = \sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{U_j \le F(x)\}} - F(x) \right| = \sup_{x\in\mathbb{R}} |F_n^{\mathrm{unif}}(F(x)) - F(x)|. \end{aligned}$$

Pour conclure, il suffit d'observer que cette dernière quantité est égale à

$$\sup_{t\in F(\mathbb{R})} |F_n^{\mathrm{unif}}(t) - t| \le \sup_{t\in[0,1]} |F_n^{\mathrm{unif}}(t) - t|,$$

ce qui tend presque sûrement vers zéro quand $n \to \infty$, d'après le résultat de la question 6.