TD 4

Les exercices marqués d'une étoile ne sont pas prioritaires.

Exercice 1. Soit (X_n) une suite de variables aléatoires telle que pour tout $n \geq 1$, X_n suit la loi uniforme sur $\{k/n,\ 1 \leq k \leq n\}$. Montrer que la suite (X_n) converge en loi vers la loi uniforme sur $[0,1]$.

Avec la définition de la convergence en loi : soit f continue bornée. On a

$$\begin{aligned}\int f d\mathbb{P}_{X_n} &= \frac{1}{n}\sum_{k=1}^{n} f(k/n) \\ &= \sum_{k=1}^{n}\int_{k/n}^{(k+1)/n} f(k/n)dx \xrightarrow[n\to\infty]{} \int_0^1 f(x)dx\end{aligned}$$

par la convergence des sommes de Riemann (on peut mettre $f(k/n) - f(x) + f(x)$ et utiliser continue donc uniformément continue sur le compact $[0,1]$.)

Avec les fonctions caractéristiques. On a

$$\begin{aligned}\mathbb{E}[\exp(itX_n)] &= \frac{1}{n}\sum_{k=1}^{n}\exp(itk/n) \\ &= \frac{1}{n}\sum_{k=1}^{n}\big(\exp(it/n)\big)^k \\ &= \frac{1}{n}\frac{\exp(nit/n)-1}{\exp(it/n)-1}\end{aligned}$$

Le numérateur converge vers $\exp(it) - 1$. Le dénominateur est équivalent à it/n, donc multiplié par n il converge vers it. On a donc

$$\mathbb{E}[\exp(itX_n)] \xrightarrow[n\to\infty]{} \frac{\exp(it)-1}{it}$$

qui est bien la fonction caractéristique de la loi uniforme sur $[0,1]$.

Avec la fonction de répartition : si $0 \leq t \leq 1$, $F_{X_n}(t) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{(t\geq k/n)} = \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{(tn\geq k)} = [tn]/n$ qui converge vers t lorsque n tend vers l'infini.

Exercice 2. Soit $\lambda > 0$. On considère pour tout entier n une variable aléatoire T_n de loi géométrique de paramètre $p_n = \lambda/n$. Montrer que T_n/n converge en loi vers une variable aléatoire de loi exponentielle de paramètre λ.

On rappelle que la fonction caractéristique de la loi géométrique de paramètre p est $\Phi(t) = \frac{p\exp(it)}{1-(1-p)\exp(it)}$. La fonction caractéristique de T_n/n est donc

$$\Phi_n(t) = \frac{\frac{\lambda}{n}\exp(it/n)}{1-(1-\frac{\lambda}{n})\exp(it/n)}.$$

On fait un développement limité des exponentielles à l'ordre 1 pour obtenir

$$\Phi_n(t) = \frac{\frac{\lambda}{n}(1+it/n+o(1/n))}{1-(1-\frac{\lambda}{n})(1+it/n+o(1/n))} = \frac{\lambda(1+it/n+o(1/n))}{\lambda-it+o(1)}$$

qui tend vers $\lambda/(\lambda - it)$ qui est bien la fonction caractéristique de la loi exponentielle de paramètre λ.

Exercice 3.* Soit (U_n) une suite de variables aléatoires indépendantes et de même loi uniforme sur $[0,1]$. Soit A et B deux boréliens de $\mathbb{R}_+$ disjoints et de mesure de Lebesgue respective finie et strictement positive. On définit les deux suites de variables aléatoires discrètes par

$$X_n = \sum_{k=1}^{n} \mathbb{1}_A(nU_k), \qquad Y_n = \sum_{k=1}^{n} \mathbb{1}_B(nU_k).$$

1. Montrer que la suite (X_n) converge en loi vers une loi de Poisson dont on précisera le paramètre.

   Calculons la fonction caractéristique de X_n. On a

$$\begin{aligned}
\mathbb{E}[\exp(itX_n)] &= \mathbb{E}[\exp\big(it\sum_{k=1}^{n} \mathbb{1}_A(nU_k)\big)] \\
&= \mathbb{E}[\prod_{k=1}^{n} \exp\big(it\mathbb{1}_A(nU_k)\big)]
\end{aligned}$$

   On utilise l'indépendance des U_k.

$$\begin{aligned}
\mathbb{E}[\exp(itX_n)] &= \prod_{k=1}^{n} \mathbb{E}[\exp\big(it\mathbb{1}_A(nU_k)\big)] \\
&= \Big(\int_0^1 \exp(it\mathbb{1}_A(nu))du\Big)^n \\
&= \Big(\frac{1}{n}\int_0^n \exp(it\mathbb{1}_A(v))dv\Big)^n \\
&= \Big(\frac{1}{n}\int_{[0,n]\cap A} \exp(it)dv + \frac{1}{n}\int_{[0,n]\cap A^c} dv\Big)^n \\
&= \big(\exp(it)\frac{\lambda([0,n]\cap A)}{n} + \frac{\lambda([0,n]\cap A^c)}{n}\big)^n \\
&= (1 + (\exp(it)-1)\frac{\lambda([0,n]\cap A)}{n})^n
\end{aligned}$$

   où λ est la mesure de Lebesgue sur $\mathbb{R}$. Cette quantité converge quand n tend vers l'infini vers $\exp\big((\exp(it)-1)\lambda(A)\big)$ qui est la fonction caractéristique de la loi de Poisson de paramètre $\lambda(A)$.

2. Montrer que le couple (X_n, Y_n) converge en loi vers un couple de variables aléatoires indépendantes.

   Comme (Y_n) est construit de la même façon que (X_n), on a également la convergence en loi de Y_n vers la loi de Poisson de paramètre $\lambda(B)$. Calculons la fonction caractéristique du couple (X_n, Y_n). On a

$$\begin{aligned}
\mathbb{E}[\exp(itX_n + isY_n)] &= \mathbb{E}[\exp\big(it\sum_{k=1}^{n} \mathbb{1}_A(nU_k) + is\sum_{k=1}^{n} \mathbb{1}_B(nU_k)\big)] \\
&= \mathbb{E}[\prod_{k=1}^{n} \exp\big(it\mathbb{1}_A(nU_k) + is1_B(nU_k)\big)]
\end{aligned}$$

   On utilise l'indépendance des U_k, et leur identité en loi.

$$\mathbb{E}[\exp(itX_n + isY_n)]$$

$$
\begin{aligned}
&= \prod_{k=1}^{n} \mathbb{E}[\exp\left(it\mathbb{1}_A(nU_k) + is\mathbb{1}_B(nU_k)\right)] \\
&= \Big(\int_0^1 \exp(it\mathbb{1}_A(nu))\exp(is\mathbb{1}_B(nu))du\Big)^n \\
&= \Big(\frac{1}{n}\int_0^n \exp(it\mathbb{1}_A(v))\exp(is\mathbb{1}_B(v))dv\Big)^n \\
&= \Big(\frac{1}{n}\int_{[0,n]\cap A} \exp(it)\exp(is\mathbb{1}_B(v))dv + \frac{1}{n}\int_{[0,n]\cap A^c} \exp(it\mathbb{1}_B(v))dv\Big)^n
\end{aligned}
$$

On utilise maintenant que A et B sont disjoints

$$
\begin{aligned}
&\mathbb{E}[\exp(itX_n + isY_n)] \\
&\quad = \Big(\frac{1}{n}\int_{[0,n]\cap A} \exp(it)dv + \frac{1}{n}\int_{[0,n]\cap B} \exp(is)dv + \frac{1}{n}\int_{[0,n]\cap A^c\cap B^c} dv\Big)^n \\
&\quad = \big(\exp(it)\frac{\lambda([0,n]\cap A)}{n} + \exp(is)\frac{\lambda([0,n]\cap B)}{n} + \frac{\lambda([0,n]\cap A^c\cap B^c)}{n}\big)^n \\
&\quad = \Big(1 + (\exp(it)-1)\frac{\lambda([0,n]\cap A)}{n} + (\exp(is)-1)\frac{\lambda([0,n]\cap B)}{n}\Big)^n
\end{aligned}
$$

Cette quantité converge quand n tend vers l'infini vers $\exp\left((\exp(it)-1)\lambda(A)) + (\exp(is)-1)\lambda(B))\right) = \exp\left((\exp(it)-1)\lambda(A)\right)\exp\left((\exp(is)-1)\lambda(B)\right)$ qui est le produit des fonctions caractéristique de la loi de Poisson de paramètre $\lambda(A)$ et de celle de paramètre $\lambda(B)$. D'où la convergence du couple vers un couple de lois de Poisson indépendantes, puisque la fonction caractéristique est de forme produit.

Exercice 4. Soit (X_n) une suite de variables aléatoires indépendantes de même loi uniforme sur l'intervalle $[0,1]$. Soit $Y_n = \min\{X_1, \cdots, X_n\}$.

1. Calculer la fonction de survie de Y_n. En déduire sa fonction de répartition.

   $\mathbb{P}(Y_n > t) = \mathbb{P}(\min\{X_1,\cdots,X_n\} > t) = \mathbb{P}(X_1 > t, \cdots, X_n > t) = \mathbb{P}(X_1 > t)^n = (1-t)^n$ pour $0 \leq t \leq 1$ en utilisant l'indépendance et l'identité en loi. La fonction de répartition est donc $\mathbb{P}(Y_n \leq t) = 1 - \mathbb{P}(Y_n > t) = 1 - (1-t)^n$ pour $0 \leq t \leq 1$.

2. Montrer que nY_n converge en loi vers la loi exponentielle de paramètre 1.

   On utilise la caractérisation de la convergence en loi par les fonctions de répartition. On a $\mathbb{P}(nY_n \leq t) = \mathbb{P}(Y_n \leq t/n) = 1 - (1-t/n)^n$ qui converge vers $1 - \exp(-t)$, la fonction de répartition de la loi exponentielle de paramètre 1.

Soit (N_n) une suite de variables aléatoires indépendantes de loi binomiale de paramètre respectifs (n,p) avec $0 < p < 1$, indépendante de la suite (X_n). Soit $Z_n = nY_{N_n}$.

3. Calculer la fonction de répartition de Z_n.

   On a
   $\mathbb{P}(Z_n \leq t) = \mathbb{P}(nY_{N_n} \leq t) = \mathbb{P}(Y_{N_n} \leq t/n) = \sum_{k=0}^n \mathbb{P}(N_n = k, Y_k \leq t/n)$. On utilise l'indépendance de (N_n) et (X_n) pour avoir
   $\mathbb{P}(Z_n \leq t) = \sum_{k=0}^n \mathbb{P}(N_n = k)\mathbb{P}(Y_k \leq t/n) = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k}(1-(1-t/n)^k)$
   $= 1 - \sum_{k=0}^n C_n^k \big(p(1-t/n)\big)^k (1-p)^{n-k}$
   $= 1 - \big(p(1-t/n) + 1 - p\big)^n = 1 - (1-tp/n)^n$.

4. Montrer que Z_n converge en loi vers la loi exponentielle de paramètre p.

   La fonction de répartition de Z_n converge vers $1 - \exp(-tp)$ qui est la fonction de répartition de la loi exponentielle de paramètre p.

Exercice 5.* Soit (X_n) et (Y_n) deux suites de variables aléatoires réelles définies sur le même espace et convergeant en loi respectivement vers les variables aléatoires X et Y.

1. On suppose que pour tout n, X_n est indépendant de Y_n et que X est indépendant de Y. Montrer que $X_n + Y_n$ converge en loi vers $X + Y$.

   Par indépendance, on a $\Phi_{X_n+Y_n} = \Phi_{X_n}\Phi_{Y_n}$ qui tend vers $\Phi_X\Phi_Y = \Phi_{X+Y}$.

2. Soit X et Y deux variables aléatoires indépendantes de même loi de Bernoulli de paramètre $1/2$. On pose, pour tout $n \geq 1$, $X_n = X + 1/n$ et $Y_n = (1 - X) - 1/n$. Montrer que les suites (X_n) et (Y_n) convergent en loi respectivement vers X et Y mais que le couple (X_n, Y_n) ne converge pas en loi vers (X, Y).

   Les suites (X_n) et (Y_n) convergent presque sûrement vers X et $1 - X$ respectivement, d'où la convergence en loi en remarquant que Y et $1 - X$ ont la même loi.
   La suite $X_n + Y_n$ est constante égale à 1, donc elle converge en loi vers 1 qui n'est pas la loi de $X + Y$. donc le couple ne converge pas en loi.

Exercice 6.* Soit (X_n) une suite de variables aléatoires réelles définies sur le même espace probabilisé, indépendantes et de même loi de fonction de répartition F. On définit pout tout $n \geq 1$ les variables $I_n = \min_{1\leq i\leq n}\{X_i\}$ et $M_n = \max_{1\leq i\leq n}\{X_i\}$.

1. Exprimer la fonction de répartition de I_n et M_n en fonction de F.

   Par indépendance et identité en loi, on a $\mathbb{P}(I_n > t) = \mathbb{P}(X_1 > t, \ldots, X_n > t) = (1 - F(t))^n$. Donc $F_{I_n}(t) = 1 - (1 - F(t))^n$.
   $\mathbb{P}(M_n \leq t) = \mathbb{P}(X_1 \leq t, \ldots, X_n \leq t) = F(t)^n$. Donc $F_{M_n}(t) = F(t)^n$.

2. Etudier la convergence simple des fonctions de répartition de I_n et M_n.

   $F_{I_n}(t) = 1 - (1 - F(t))^n$ tend vers 0 si $F(t) = 0$ et vers 1 si $0 < F(t) \leq 1$.
   $F_{M_n}(t) = F(t)^n$ tend vers 0 si $0 \leq F(t) < 1$ et vers 1 si $F(t) = 1$.

3. Soit $t_- = \inf\{t; F(t) > 0\}$. Que dire de la convergence en loi de I_n si $t_- > -\infty$, si $t_- = -\infty$?

   Si $t_- > -\infty$, alors F_{I_n} converge vers $\mathbb{1}_{[t_-,+\infty[}$ en tout point de continuité de cette fonction. C'est la fonction de répartition de la masse de Dirac en t_-, donc (I_n) converge en loi vers δ_{t_-}.
   Si $t_- = -\infty$, alors F_{I_n} converge vers 1 en tout point. Ce n'est pas une fonction de répartition, donc il n'y a pas de convergence en loi de la suite (I_n).

4. Soit $t_+ = \inf\{t; F(t) = 1\}$. Que dire de la convergence en loi de M_n si $t_+ < +\infty$, si $t_+ = +\infty$?

   Si $t_+ < +\infty$, alors F_{M_n} converge vers $\mathbb{1}_{[t_+,+\infty[}$ en tout point de continuité de cette fonction. C'est la fonction de répartition de la masse de Dirac en t_+, donc (M_n) converge en loi vers δ_{t_+}.
   Si $t_+ = +\infty$, alors F_{M_n} converge vers 0 en tout point. Ce n'est pas une fonction de répartition, donc il n'y a pas de convergence en loi de la suite (M_n).

Exercice 7. Soit (X_n) une suite de variables aléatoires réelles définies sur le même espace probabilisé, indépendantes et de même loi de fonction de répartition F. On appelle mesure empirique de $(X_1, \ldots, X_n)$ la loi de probabilité (aléatoire)

$$\mathbb{P}_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k}.$$

1. Calculer $\mathbb{P}_n(]-\infty,t])$ pour tout $t\in\mathbb{R}$.

   $\mathbb{P}_n(]-\infty,t]) = n^{-1}\sum_{k=1}^n \delta_{X_k}(]-\infty,t]) = n^{-1}\sum_{k=1}^n \mathbb{1}_{]-\infty,t]}(X_k)$.

2. Montrer que $\mathbb{P}_n(]-\infty,t])$ converge lorsque n tend vers l'infini et donner sa limite.

   Les variables aléatoires $\mathbb{1}_{]-\infty,t]}(X_k)$ sont indépendantes, de même loi et intégrables d'espérance $\mathbb{E}[\mathbb{1}_{]-\infty,t]}(X_k)] = \mathbb{P}(X_k\le t) = F(t)$, à t fixé. La loi des grands nombres dit que $\mathbb{P}_n(]-\infty,t]) = n^{-1}\sum_{k=1}^n \mathbb{1}_{]-\infty,t]}(X_k)$ converge $\mathbb{P}$-presque sûrement quand n tend vers l'infini vers $F(t)$.

3. En déduire que $\mathbb{P}$-presque sûrement la suite $(\mathbb{P}_n)$ converge en loi vers la loi de fonction de répartition F.

   On a pour tout t, $\mathbb{P}(\mathbb{P}_n(]-\infty,t])\to F(t))=1$. On veut
   $\mathbb{P}(\forall t\in\mathbb{R}$ point de continuité de $F, \mathbb{P}_n(]-\infty,t])\to F(t))=1$. Pour conclure, il faut montrer qu'on peut choisir l'ensemble de mesure pleine sur lequel on a la convergence indépendamment de t. On a $\mathbb{P}(\forall t\in\mathbb{Q}$ point de continuité de $F, \mathbb{P}_n(]-\infty,t])\to F(t))=1$ car l'union d'ensembles de mesure nulle reste de mesure nulle, et c'est suffisant pour avoir le résultat par continuité à droite de F et $\mathbb{P}_n(]-\infty,t])$.

Exercice 8. Quel est le nombre minimum de lancers nécessaires avec une pièce équilibrée pour que la fréquence de "Pile" observées soit comprise entre 0.4 et 0.6 avec probabilité 0.95 ?

Soit X_n le résultat du n-ème lancer : $X_n=1$ si on a obtenu pile, 0 sinon. La fréquence observée de pile est $S_n/n = 1/n\sum_{k=1}^n X_k$. Les (X_n) sont des variables aléatoires indépendantes et de même loi de Bernoulli de paramètre $p=1/2$ qui admet un moment d'ordre 2 et a pour variance 1/4. On peut donc utiliser le théorème central limite et utiliser l'approbation normale. Soit $Y\sim\mathcal{N}(0,1)$, on a

$$\begin{aligned}\mathbb{P}(0.4\le S_n/n\le 0.6) &= \mathbb{P}(|S_n/n-1/2|\le 0.1)\\ &= \mathbb{P}(\sqrt{n/\mathsf{Var}(X)}|S_n/n-1/2|\le 0.1\sqrt{n/\mathsf{Var}(X)})\\ &\simeq \mathbb{P}(|Y|\le 0.2\sqrt{n})\\ &= 2\Phi(0.2\sqrt{n})-1.\end{aligned}$$

Donc $0.2\sqrt{n}=1.96$ et $n=(1.96/0.2)^2=96.04$. Il faut donc 97 lancers.

Exercice 9. Le jour d'une élection, on effectue un sondage à la sortie des urnes sur un échantillon de 1000 personnes, 52% répondent avoir voté pour A et 48% avoir voté pour B.

1. Donner un intervalle de confiance à 95% de la proportion de votes en faveur de A

   Soit p la probabilité de votes pour un candidat et $\overline{X}_n$ la fréquence observée des votes sur une population de taille n. On a vu en cours que l'intervalle de confiance de niveau α est

$$\left[\overline{X}_n - \Pi^{-1}\Big(\frac{1+\alpha}{2}\Big)\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}};\overline{X}_n + \Pi^{-1}\Big(\frac{1+\alpha}{2}\Big)\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right]$$

   Pour A, on obtient $[0.52-1.96\sqrt{0.52\times 0.48/1000}; 0.52+1.96\sqrt{0.52\times 0.48/1000}] = [0.489, 0.551]$

2. Donner un intervalle de confiance à 95% de la proportion de votes en faveur de B

   Pour B, on obtient
   $[0.48-1.96\sqrt{0.52\times 0.48/1000}; 0.48+1.96\sqrt{0.52\times 0.48/1000}] = [0.449, 0.511]$

3. Le sondage permet-il d'annoncer la victoire de A avec 95% de certitude ?

   Non car les intervalles de confiance se recoupent.

4. Quel est le nombre minimal de personnes à sonder pour pouvoir annoncer la victoire de A avec 95% de certitude, en supposant toujours 52% de réponses favorables ?

   On doit avoir des intervalles disjoints, donc
   $0.48 + 1.96\sqrt{0.52 \times 0.48/n} < 0.52 - 1.96\sqrt{0.52 \times 0.48/n}$
   $\Leftrightarrow 2 \times 1.96\sqrt{0.52 \times 0.48/n} < 0.52 - 0.48$
   $\Leftrightarrow \sqrt{0.52 \times 0.48/n} < (0.52 - 0.48)/(2 \times 1.96)$
   $\Leftrightarrow n > 0.52 \times 0.48((2 \times 1.96)/(0.52 - 0.48))^2$
   $\Leftrightarrow n > 2397.1$ Le nombre minimum est 2398.

Exercice 10.* Quand une machine mettant des médicaments en sachets est bien réglée, le poids d'un sachet suit une loi de moyenne $m = 2.5g$ et d'écart-type $\sigma = 0.1$g. Lors d'un contrôle de fabrication, le poids d'un lot de 100 sachets s'est élevé à 242g. Peut-on conclure, avec 95% de certitude, à un dérèglement de la machine ?

On fait une approximation normale pour la loi d'un lot de 100 sachets en utilisant le TCL. Soit X_k le poids du k-ème sachet. Le poids total est $S_{100} = \sum_{k=1}^{100} X_k$. Les X_k sont indépendants et de même loi admettant un moment d'ordre 2, donc on peut utiliser le TCL. On obtient que $(\overline{X}_{100} - \mathbb{E}[X_1])\sqrt{100/\sigma^2}$ suit approximativement une loi normale centrée réduite. Un intervalle de confiance à 95% pour la moyenne est

$$\begin{aligned}
&\left[\overline{X}_{100} - 1.96\sqrt{\frac{\sigma^2}{100}}; \overline{X}_{100} + 1.96\sqrt{\frac{\sigma^2}{100}}\right] \\
&= [2.42 - 0.0196; 2.42 + 0.0196] \\
&= [2.40; 2.44]
\end{aligned}$$

m n'est pas dans cet intervalle, donc on peut affirmer avec 95% de certitude que la machine est déréglée.

Exercice 11. En France métropolitaine en 2014 on a recensé 399284 naissances de garçons et 381883 naissances de filles. Ces chiffres sont-ils compatibles avec l'hypothèse d'équi-probabilité des naissances avec 95% de certitude ?

Soit $X_n = 1$ si la n-ème naissance est une fille, 0 si c'est un garçon. Les X_n sont indépendantes et de même loi de Bernoulli de paramètre p. On peut utiliser l'approximation normale pour obtenir un intervalle de confiance à 95% pour p. Soit $n = 399284 + 381883 = 781167$ et $\overline{X}_n = 381883/n$. L'intervalle de confiance est

$$\begin{aligned}
&\left[\overline{X}_n - \Pi^{-1}\Big(\frac{1+\alpha}{2}\Big)\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}; \overline{X}_n + \Pi^{-1}\Big(\frac{1+\alpha}{2}\Big)\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right] \\
&= [0.4889 - 1.96 * 0.00057; 0.4889 + 1.96 * 0.00057] \\
&= [0.4883; 0.4894].
\end{aligned}$$

Cet intervalle ne contient pas 1/2, donc à 95% de certitude on peut rejeter l'hypothèse d'équi-probabilité des naissances.