TD - Probabilités et Chaînes de Markov

Dans toute la suite, $(\Omega, \mathcal{F}, \mathbb{P})$ désigne un espace probabilisé.

Rappel 1 (Loi forte des grands nombres)**.** Soit (X_n) une suite de variables aléatoires indépendantes identiquement distribuées (i.i.d.) intégrables (i.e. $\mathbb{E}[|X_1|] < +\infty$). La moyenne empirique converge presque sûrement vers $\mathbb{E}[X_1]$:

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{p.s.} \mathbb{E}[X_1].$$

Exercice 1 (Illustrations LGN)**.** Expliciter le résultat de la LGN ci-dessus pour les lois classiques suivantes :

(i) la loi de Bernoulli $b(p)$ de paramètre $p \in]0,1[$

(ii) la loi binomiale $B(n,p)$ de paramètres $n \in \mathbb{N}^*$ et $p \in]0,1[$

(iii) la loi géométrique $\mathcal{G}(p)$ de paramètre $p \in]0,1[$

(iv) la loi uniforme $\mathcal{U}([0,1])$

(v) la loi exponentielle $\mathcal{E}(\alpha)$ de paramètre $\alpha > 0$

(vi) la loi normale $\mathcal{N}(m, \sigma^2)$ de paramètres $m \in \mathbb{R}$ et $\sigma > 0$

(vii) la loi de Cauchy, de densité $x \mapsto \frac{1}{\pi}\frac{1}{1+x^2}$ sur $\mathbb{R}$.

Illustrer numériquement (avec le module `stats` de Python par exemple) les résultats obtenus ici.

Exercice 2 (Méthode de Monte-Carlo)**.** Soient f une fonction et (X_n) une suite de variables aléatoires réelles i.i.d. de loi μ. Justifier le principe suivant, appelé méthode de Monte-Carlo :

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow[n\to\infty]{p.s.} \mathbb{E}[f(X_1)] = \int_{\mathbb{R}} f \,\mathrm{d}\mu.$$

Appliquer cette méthode pour déterminer numériquement une valeur approchée des intégrales suivantes :

$$\int_3^5 x^2 \ln x \,\mathrm{d}x, \quad \int_0^1 4\sqrt{1-x^2}\,\mathrm{d}x, \quad \int_{\mathbb{R}} e^{-x^2}\sin x \,\mathrm{d}x, \quad \int_{[-1,1]^3} \mathbb{1}_{x^2+2y^2+3z^2\leq 1}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

Rappel 2 (Théorème Central Limite)**.** Soit (X_n) une suite de variables aléatoires indépendantes identiquement distribuées (i.i.d.) de carré intégrable (i.e. $\mathbb{E}[X_1^2] < +\infty$). Pour $m = \mathbb{E}[X_1]$ et $\sigma = \sqrt{\mathrm{Var}(X_1)}$:

$$\sqrt{n}\,\frac{\overline{X_n} - \mathbb{E}[X_1]}{\sigma} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Autrement dit, pour les fonctions de répartition :

$$\forall x \in \mathbb{R},\ \mathbb{P}\left(\sqrt{n}\,\frac{\overline{X_n} - \mathbb{E}[X_1]}{\sigma} \leq x\right) \xrightarrow[n\to\infty]{} F_{\mathcal{N}(0,1)}(x).$$

Exercice 3 (TCL)**.** Illustrer numériquement le résultat ci-dessus pour les lois proposées dans le premier exercice. On représentera pour cela la distribution empirique sous forme d'un histogramme auquel on superposera la densité de la loi.

Rappel 3 (Intervalle de confiance)**.** Soit $\alpha \in]0,1[$ donné. On appelle intervalle de confiance de niveau $1-\alpha$ pour un paramètre θ tout intervalle I tel que :

$$\forall \theta \in \Theta,\ \mathbb{P}_\theta\left(\theta \in I\right) \geq 1 - \alpha.$$

Exercice 4 (Intervalles de confiance : Cas gaussien)**.** Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de loi $\mathcal{N}(\theta, 1)$. L'objectif de l'exercice est de donner un intervalle de confiance de niveau 0.95 pour le paramètre θ (inconnu).

(i) Justifier que, pour tout θ réel et $X_1, \ldots, X_n$ i.i.d. de loi $\mathcal{N}(\theta, 1)$:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \theta\right) \sim \mathcal{N}(0,1).$$

(ii) Pour $Y \sim \mathcal{N}(0,1)$, on sait que $\mathbb{P}(|Y| \leq 1.96) \geq 0.95$. Expliquer pourquoi l'intervalle suivant est bien un intervalle de confiance de niveau 0.95 :

$$\left[\overline{X_n} - \frac{1.96}{\sqrt{n}}, \overline{X_n} + \frac{1.96}{\sqrt{n}}\right].$$

(iii) Dans le cas plus général où la loi est $\mathcal{N}(\theta, \sigma)$ avec σ connu, justifier de la même manière que l'intervalle suivant est bien un intervalle de confiance de niveau 0.95 :

$$\left[\overline{X_n} - \frac{\sigma x_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \overline{X_n} + \frac{\sigma x_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right]$$

où l'on a choisi $x_{1-\frac{\alpha}{2}}$, tel que si $Y \sim \mathcal{N}(0,1)$, alors $\mathbb{P}(|Y| \leq x_{1-\frac{\alpha}{2}}) = 1 - \alpha$.

Rappel 4 (Intervalle de confiance asymptotique)**.** Soit $\alpha \in]0,1[$ donné. On appelle intervalle de confiance asymptotique de niveau $1-\alpha$ pour un paramètre θ tout intervalle I_n tel que :

$$\forall \theta \in \Theta, \; \mathbb{P}_\theta\left(\theta \in I_n\right) \xrightarrow[n\to\infty]{} 1 - \alpha.$$

Exercice 5 (Intervalles de confiance asymptotiques)**.** Tracer les intervalles de confiance asymptotiques, de la forme (1) obtenus pour les exemples de l'exercice 1 :

$$\left[\overline{X_n} - \frac{1.96\sigma}{\sqrt{n}}, \overline{X_n} + \frac{1.96\sigma}{\sqrt{n}}\right]. \tag{1}$$

Exercice 6 (Cas d'une proportion)**.** Soient $X_1, \ldots, X_n$ i.i.d. de loi de Bernoulli de paramètre θ inconnu. Déterminons un intervalle de confiance de niveau $1-\alpha$ pour θ. Le théorème central limite assure que :

$$\sqrt{n}\,\frac{\overline{X_n} - \theta}{\sqrt{\theta(1-\theta)}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

On peut montrer que cet argument mène à un intervalle de confiance asymptotique de la forme suivante :

$$\left[\frac{\overline{X_n} + \frac{r^2}{2n} - \frac{r}{\sqrt{n}}\sqrt{\frac{r^2}{4n} + \overline{X_n}(1-\overline{X_n})}}{1 + \frac{r^2}{n}}, \frac{\overline{X_n} + \frac{r^2}{2n} + \frac{r}{\sqrt{n}}\sqrt{\frac{r^2}{4n} + \overline{X_n}(1-\overline{X_n})}}{1 + \frac{r^2}{n}}\right]$$

où $r = x_{1-\frac{\alpha}{2}}$ comme précédemment. Une autre méthode, s'appuyant notamment sur le lemme de Slutsky, consiste à s'intéresser à la convergence en loi suivante :

$$\sqrt{n}\,\frac{\overline{X_n} - \theta}{\sqrt{\overline{X_n}(1-\overline{X_n})}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Justifier que l'on obtient alors un intervalle de confiance asymptotique de la forme (mêmes notations) :

$$\left[\overline{X_n} - r\sqrt{\frac{\overline{X_n}(1-\overline{X_n})}{n}}, \overline{X_n} + r\sqrt{\frac{\overline{X_n}(1-\overline{X_n})}{n}}\right]$$

Exercice 7. Posons $\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$ et considérons $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de loi μ.

(i) Calculer $L_X(t) = \mathbb{E}[e^{tx}]$ et montrer que $\ln L_X(t) \leq \frac{t^2}{2}$.

(ii) Montrer que pour tout $\lambda > 0$,

$$\mathbb{P}\left(|\overline{X_n} - \mathbb{E}[X_1]| \geq r\right) = 2\mathbb{P}\left(\overline{X_n} \geq r\right) \leq 2\exp\left(-n(\lambda r + \ln L_X(\lambda))\right).$$

(iii) En déduire un intervalle de confiance (non asymptotique) pour le paramètre θ dans le modèle :

$$\mathbb{P}_\theta = \frac{1}{2}\delta_{\theta-1} + \frac{1}{2}\delta_{\theta+1}$$

Exercice 8 (Transformée de Laplace et Inégalité de Chernov)**.** On s'intéresse dans cet exercice à la transformée de Laplace de variables aléatoires. Si X est une variable réelle, on pose $L_X(t) = \mathbb{E}[e^{tX}]$.

(i) Montrer que le domaine de définition de la transformée ci-dessus est un intervalle contenant 0. Peut-il être réduit à $\{0\}$?

(ii) Calculer les transformées de Laplace (lorsque c'est possible) pour les lois $b(p)$, $\mathcal{E}(\lambda)$, $\mathcal{N}(m, \sigma^2)$.

(iii) Soit μ une loi fixée. Considérons $X_1, \ldots, X_n$ variables aléatoires i.i.d. de loi μ. Montrer que :

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1] \geq r\right) \leq \exp\left(-n \sup_{\lambda>0}\{\lambda(r + \mathbb{E}[X_1] - \ln L_X(\lambda)\right).$$

En particulier, on a montré le *principe de grandes déviations* suivant :

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1] \geq r\right) \leq \sup_{\lambda>0}\{\lambda(r + \mathbb{E}[X_1] - \ln L_X(\lambda)).$$

Rappel 5 (Chaînes de Markov)**.** Soit E un espace d'états. Une suite (X_n) à valeurs dans E est une chaîne de Markov si elle satisfait la propriété suivante :

$$\forall n \in \mathbb{N},\ \forall i_1, \ldots, i_{n+1},\ \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n).$$

Autrement dit, le futur ne dépend que du présent. La chaîne est de plus dite homogène si :

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_n = j \mid X_{n-1} = i).$$

La loi de X_0 est appelée loi initiale de la chaîne et, pour une chaîne de Markov homogène, $(p(i,j) = \mathbb{P}(X_n = j \mid X_{n-1} = i))$ est appelée matrice de transition.

On rappelle que, par récurrence, si μ désigne la loi initiale de la chaîne et p la matrice de transition :

$$\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) = \mu(i_0)p(i_0, i_1)\ldots p(i_{n-1}, i_n).$$

De même, on montre par récurrence que pour un espace d'états E fini, en notant μ_n la mesure de probabilité de X_n en ligne :

$$\mu_n = \mu_0 P^n.$$

Exercice 9 (Critère)**.** Soient (Y_n) une suite de variables aléatoires i.i.d. à valeurs dans un espace F, X_0 une variable aléatoire indépendante de Y et $f : E \times F \to E$ une application mesurable. Montrer que la relation de récurrence suivante :

$$X_0 \text{ donné et } X_{n+1} = f(X_n, Y_{n+1})$$

définit une chaîne de Markov (X_n) à valeurs dans E.

Exercice 10 (Lien Matrices-Graphe)**.** Donner les graphes des chaînes de Markov homogènes données par les matrices de transition ci-dessous :

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 3/10 & 1/10 & 6/10 \\ 0 & 1 & 0 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Exercice 11 (Mesure invariante)**.** Déterminer l'unique mesure de probabilité invariante de la matrice de transition suivante :

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}.$$

Exercice 12 (Mesure réversible)**.** Montrer que toute mesure m vérifiant $m_i p(i,j) = m_j p(j,i)$ pour tout (i,j), est invariante.