# Devoir Surveillé 03

Le 22 Novembre
durée 3h30

*Documents écrits, électroniques, calculatrices et téléphones portables interdits*
*La plus grande attention sera apportée à la qualité de la rédaction, syntaxe et orthographe comprise.*
*Numérotez les copies et questions et soulignez ou encadrez les résultats.*
*Vous pouvez utiliser les résultats des questions en amont d'une question donnée en la référençant clairement*
*Les calculs, même longs et fastidieux doivent figurer sur la copie*

*Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.*

## Exercice I

Toutes les variables aléatoires du texte sont définies sur un certain espace probabilisé $(\Omega, \mathscr{T}, \mathbb{P})$. On notera $\mathbb{E}(X)$, resp. $\mathbb{V}(X)$, l'espérance, resp. la variance d'une variable aléatoire réelle $X$ lorsque celle-ci existe.

Soit $a$ un nombre réel strictement supérieur à 1 et $f : \mathbb{R} \to \mathbb{R}$ la fonction définie par

$$\forall x \in \mathbb{R},\ f(x) = \frac{a}{x^{a+1}} 1\!\!1_{\{x \geq 1\}} = \begin{cases} \frac{a}{x^{a+1}} & \text{si } x \geq 1 \\ 0 & \text{sinon} \end{cases}$$

**1.** Donner l'allure du graphe de $f$ et montrer que $f$ est une densité de probabilité.

On suppose maintenant que $X$ est une variable aléatoire réelle de densité $f$.

**2.a.** Donner un intervalle $I \subset \mathbb{R}$, suffisamment restreint tel que $\mathbb{P}(X \in I) = 1$. Montrer que $X$ admet une espérance et donner sa valeur.

**2.b.** A quelle condition sur $a$, $X$ admet-elle une variance ? Donner sa valeur lorsque cette condition est remplie.

**3.a.** Déterminer $F_X$, la fonction de répartition de $X$.

**3.b.** Résoudre l'équation $F_X(x) = \frac{1}{2}$ d'inconnue $x \in \mathbb{R}$.

**3.c.** La fonction Python `np.random.rand()` retourne un nombre réel aléatoire tiré suivant la loi uniforme sur $]0,1[$. En utilisant cette fonction, construire une fonction Python `X(a)` retournant un nombre réel aléatoire tiré suivant la loi de $X$. On spécifiera la valeur de $a$ en paramètre.

**4.** On fixe $a > 1$ et on considère, pour $n \in \mathbb{N}$, $n \geq 2$, une famille $X_1, \ldots, X_n$ de variables aléatoires indépendantes de même loi que $X$. On pose alors

$$T_n = \min(X_1, \ldots, X_n)$$

**4.a.** Montrer que $T_n$ est une variable aléatoire réelle à densité et déterminer la loi de $T_n$.

**4.b.** $T_n$ admet-elle une espérance ? Quelle est sa valeur ?

**4.c.** A quelle condition sur $a$ et $n$, $T_n$ admet-elle une variance ? Quelle est alors sa valeur ?

**4.d.** Proposer deux fonctions Python `T1(n,a)` et `T2(n,a)` permettant de tirer au sort un nombre réel en suivant la loi de $T_n$. `n` et `a` sont les paramètres $n$ et $a$ du texte. Les deux fonctions doivent être construites sur des principes différents.

**5.** On pose $Z = \ln(X)$.

**5.a.** Donner un intervalle $J \subset \mathbb{R}$, suffisamment restreint tel que $\mathbb{P}(Z \in J) = 1$. Déterminer la loi de $Z$.

**5.b.** Donner l'espérance et la variance de $Z$.

**6.** Soient $X_1$ et $X_2$ deux variables indépendantes suivant la loi de $X$, $Z_1 = \ln X_1$, $Z_2 = \ln X_2$.
**6.a.** Donner une densité de $W = Z_1 + Z_2$. On rappelle que si $U$ et $V$ sont deux variables aléatoires réelles indépendantes de densités respectives $f_U$ et $f_V$ alors leur somme $W$ est une variable aléatoire réelle ayant pour densité $f_W$ définie par

$$\forall w \in \mathbb{R},\ f_W(w) = \int_{-\infty}^{+\infty} f_U(u) f_V(w-u)\, du$$

**6.b.** En déduire qu'une densité de $P = X_1.X_2$ est donnée par

$$\forall p \in \mathbb{R},\ f_P(p) = a^2 \frac{\ln(p)}{p^{a+1}} 1\!\!1_{\{p \geq 1\}}$$

**6.c.** Quelle est l'espérance de $P$ ?
**6.d.** Quelle est sa variance dans le cas $a > 2$ ?
**6.e.** Ecrire une fonction Python `P(a)` permettant de tirer au sort un nombre réel en suivant la loi de $P$. `a` est le paramètre $a$.

# Exercice II

Le but de l'exercice est de proposer une modélisation probabiliste de l'évolution dans le temps de la hauteur d'un lac. La figure 1 montre les hauteurs annuelles moyennes du lac Huron aux Etats-Unis de 1875 à 1972.

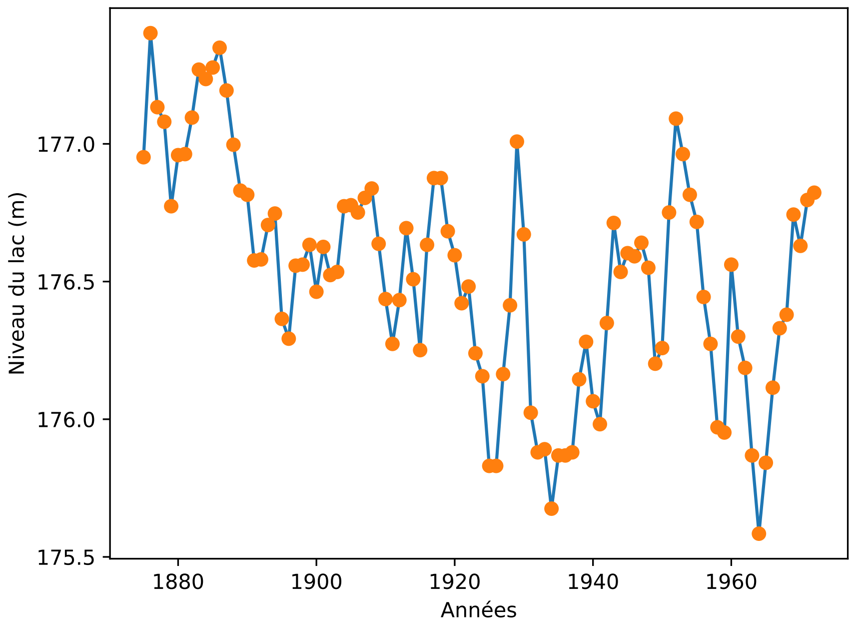


FIGURE 1 – Niveaux annuels du lac Huron exprimés en mètres.

Les variables aléatoires ci-après sont toutes définies sur un même espace probabilisé $(\Omega, \mathscr{T}, \mathbb{P})$. On note, sous réserve d'existence, $\mathbb{E}(X)$, resp. $\mathbb{V}(X)$ l'espérance, resp. la variance d'une variable aléatoire réelle $X$.

On modélise les observations d'un phénomène qui dépend du temps comme des réalisations de variables aléatoires $X_n$ où $n$ désigne un entier *relatif*.

Par exemple, si on modèlise la hauteur d'un lac tel que le lac Huron à l'année $n$ par la variable $X_n$, le graphique de la figure 1 montre donc les valeurs $(x_{1875}, \ldots, x_{1972})$ « données par le hasard » aux variables $(X_{1875}, \ldots, X_{1972})$ dans le cas du Lac Huron.

Le point d'indicer par les entiers relatifs est de ne pas fixer d'origine des temps. Les données du lac pour les années de 1900 à 1997 ou pour les années de $-340$ à $-243$ sont tenues pour similaires à celles représentées sur le graphe. Cela mène à la notion de *processus stationnaire*, définie ci-après.

On rappelle que la covariance entre deux variables aléatoires réelles $X$ et $Y$ est définie, lorsque ces variables admettent une variance, par l'une des deux formules :

$$\mathbb{C}ov(X,Y) = \mathbb{E}[(X-\mathbb{E}(X)).(Y-\mathbb{E}(Y))] = \mathbb{E}(X.Y) - \mathbb{E}(X).\mathbb{E}(Y). \tag{+}$$

On dit que la suite de variables aléatoires $(X_n)_{n\in\mathbb{Z}}$ est un *processus stationnaire* de paramètres $\mu$ et $\gamma_X$ lorsque :

$$\forall n \in \mathbb{Z},\ \mathbb{E}(X_n) = \mu \text{ et } \forall n \in \mathbb{Z}, \forall p \in \mathbb{N}, \mathbb{C}ov(X_n, X_{n+p}) = \gamma_X(p),$$

où $\mu$ désigne une constante indépendante de $n$, $\mathbb{C}ov$ est définie dans (+) et $\gamma_X$ est une fonction $\mathbb{N} \to \mathbb{R}$ indépendante de $n$.

## Partie A
Préliminaires

**A.1.** Si $X$ et $Y$ sont deux variables aléatoires réelles admettant une variance, montrer que

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\mathbb{C}ov(X,Y) \text{ et } \mathbb{V}(X-Y) = \mathbb{V}(X) + \mathbb{V}(Y) - 2\,\mathbb{C}ov(X,Y)$$

**A.2.** Si $W_1, \dots, W_n$ est une famille de $n$ variables aléatoires réelles 2 à 2 non corrélées, *i.e.*

$$\forall i, j \in \{1, \dots, n\}, i \neq j \Rightarrow \mathbb{C}ov(W_i, W_j) = 0,$$

montrer que

$$\mathbb{V}(\sum_{i=1}^{n} W_i) = \sum_{i=1}^{n} \mathbb{V}(W_i)$$

**A.3.** Soit $(X_n)_{n\in\mathbb{Z}}$ un processus stationnaire de paramètres $\mu$ et $\gamma_X$.
**A.3.a.** Montrer que $\forall n \in \mathbb{Z}$, $\mathbb{V}(X_n) = \gamma_X(0)$.
**A.3.b.** Montrer que $\forall n \in \mathbb{Z}, \forall p \in \mathbb{N}$, $\mathbb{C}ov(X_{n-p}, X_n) = \gamma_X(p)$.
**A.3.c.** Montrer que $\forall m, n \in \mathbb{Z}$, $\mathbb{C}ov(X_n, X_m) = \gamma_X(|m-n|)$.

## Partie B

**B.1.** Soient $(Z_n)_{n\in\mathbb{Z}}$ des variables aléatoires réelles indépendantes et identiquement distribuées, *i.e.* ayant toutes même loi, d'espérance $m$ et de variance $\sigma^2$.
**B.1.a.** Calculer $\mathbb{C}ov(Z_n, Z_{n+p})$ lorsque $p = 0$ et lorsque $p$ est un entier naturel non nul.
**B.1.b.** $(Z_n)_{n\in\mathbb{Z}}$ est-il un processus stationnaire ?
**B.2.** On dit qu'un processus stationnaire $(W_n)_{n\in\mathbb{Z}}$ est un bruit blanc d'espérance nulle et de variance $\sigma^2$, ce que l'on note $(W_n) \sim \mathscr{BB}(0, \sigma^2)$ lorsque

$$\forall n \in \mathbb{N}, \mathbb{E}(W_n) = 0,\ \gamma_W(0) = \sigma^2 \text{ et } \forall p \in \mathbb{N}^*,\ \gamma_W(p) = 0.$$

On considère à présent une suite, indicée par $\mathbb{Z}$, $(X_n)_{n\in\mathbb{Z}}$ de variables aléatoires réelles vérifiant :

$$\forall n \in \mathbb{Z}, X_n = W_n + \theta W_{n-1}, \tag{R}$$

où $\theta$ est un nombre réel non nul et $(W_n) \sim \mathscr{BB}(0, \sigma^2)$.
**B.2.a.** Calculer $\mathbb{E}(X_n)$ pour $n \in \mathbb{Z}$.

**B.2.b.** Calculer, pour $n \in \mathbb{Z}$, $\mathbb{C}ov(X_n, X_{n+p})$ lorsque $p = 0$, $p = 1$ et lorsque $p$ est un entier naturel $\geq 2$.
**B.2.c.** $(X_n)_{n\in\mathbb{Z}}$ est-il un processus stationnaire ? Si oui, quels sont ses paramètres ?
**B.3.** Soit $(X_n)_{n\in Z}$ un processus stationnaire de paramètres $\mu$ et $\gamma_X$, solution de l'équation suivante :

$$\forall n \in \mathbb{Z}, X_n - \varphi X_{n-1} = W_n, \qquad (\star)$$

où $\varphi$ est un nombre réel non nul tel que : $-1 < \varphi < 1$ et $(W_n) \sim \mathscr{BB}(0, \sigma^2)$.

On suppose de plus que la suite $(\gamma_X(p))_{p\in\mathbb{N}}$ est majorée, en valeur absolue, par une constante $\Gamma$.
**B.3.a.** Pour $n \in \mathbb{Z}$, calculer $\mathbb{E}(X_n)$.
**B.3.b.** Montrer que pour tout entier relatif $n$, pour tout entier naturel $p$,

$$X_n - \varphi^{p+1}.X_{n-(p+1)} = \sum_{k=0}^{p} \varphi^k.W_{n-k} \qquad (\star_p)$$

**B.3.c.**
**3.c.i.** En prenant la variance de $(\star_p)$, montrer que pour tout $p \in \mathbb{N}$

$$(1 + \varphi^{2p+2}).\gamma_X(0) - 2.\varphi^{p+1}\gamma_X(p+1) = \sigma^2.\sum_{k=0}^{p} \varphi^{2k}.$$

**3.c.ii.** En déduire, par un passage à la limite dûment justifié, que

$$\gamma_X(0) = \frac{\sigma^2}{1-\varphi^2}.$$

**3.c.iii.** En déduire que

$$\forall p \in \mathbb{N},\ \gamma_X(p+1) = \frac{\varphi^{p+1}\sigma^2}{1-\varphi^2}.$$

**B.4.** Soit $(X_n)_{n\in Z}$ un processus stationnaire de paramètres $\mu$ et $\gamma_X$, solution de l'équation suivante :

$$\forall n \in \mathbb{Z}, X_n - \varphi X_{n-1} = m + W_n, \qquad (\star\star)$$

où $m \in \mathbb{R}$, $\varphi$ est un nombre réel non nul tel que : $-1 < \varphi < 1$ et $(W_n) \sim \mathscr{BB}(0, \sigma^2)$.

On suppose de plus que la suite $(\gamma_X(p))_{p\in\mathbb{N}}$ est bornée.
**B.4.a.** Pour $n \in \mathbb{Z}$, calculer $\mu = \mathbb{E}(X_n)$.
**B.4.b.** Pour $n \in \mathbb{Z}$, on pose $\tilde{X}_n = X_n - \mu$. Vérifier que $(\tilde{X}_n)_{n\in\mathbb{Z}}$ est un processus stationnaire de paramètres $\tilde{\mu} = 0$ et $\gamma_{\tilde{X}} = \gamma_X$.
**B.4.c.** Déduire de B.3 une formule pour $\gamma_X$.

⋆⋆⋆

On admettra pour la suite la majoration

$$\forall N \in \mathbb{N}^*, \left| \frac{1}{N} \sum_{p=0}^{N-1} (N-p)\gamma_X(p) \right| \leq \frac{3\sigma^2}{(1-\varphi)^2.(1-\varphi^2)} \qquad (\star\star\star)$$

## Partie C

On souhaite à présent utiliser les notions et résultats de la partie B pour modéliser les niveaux annuels du lac Huron représentés dans la figure 1.

On propose de modéliser ces observations par un processus stationnaire $(X_n)_{n\in\mathbb{Z}}$, de paramètres $\mu$ et $\gamma_X$ à estimer en fonction des données, vérifiant :

$$\forall n \in \mathbb{Z}, X_n - \varphi X_{n-1} = m + W_n, \qquad (\star\star)$$

où $(W_n) \sim \mathscr{B}\mathscr{B}(0, \sigma^2)$ et $m \in \mathbb{R}$, $\varphi \in ]-1, +1[$, $\sigma^2$ sont des paramètres eux aussi à estimer en fonction des données.

On suppose de plus que la suite $(\gamma_X(p))_{p\in\mathbb{N}}$ est bornée, *i.e.* est majorée, en valeur absolue, par une constante.

Les relations obtenues dans la partie B, c'est à dire

$$m = (1-\varphi).\mu,\ \gamma_X(0) = \frac{\sigma^2}{1-\varphi^2} \text{ et } \gamma_X(1) = \varphi.\gamma_X(0)$$

montrent que pour estimer $m$, $\varphi$, $\sigma^2$, il suffit de savoir estimer $\mu$, $\gamma_X(0)$ et $\gamma_X(1)$.

Une fois ces quantités estimées à partir des données, on pourra vérifier la cohérence du modèle en estimant les quantités $\gamma_X(p)$ pour $p \geq 2$ et en vérifiant que les relations supplémentaires

$$\forall p \geq 2, \frac{\gamma_X(p+1)}{\gamma_X(p)} = \varphi$$

sont correctes, au moins de manière approximative.

Les données dont nous disposons sont une liste $x = (x_1, \ldots, x_N)$ des niveaux (en mètres) du lac Huron sur $N$ années consécutives. Comme il a été dit en introduction, cette liste est censée être la liste des valeurs des variables aléatoires $(X_1, \ldots, X_N)$ extraite du processus stationnaire $(X_n)_{n\in\mathbb{Z}}$ pour la réalisation propre au lac Huron.

**C.1.** Estimation de $\mu$.

On estime $\mu$, l'espérance commune des variables $X_n$, à partir des données en posant

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

Le nombre $\hat{\mu}$ est donc la valeur de la variable aléatoire

$$M_N = \frac{1}{N}\sum_{n=1}^{N} X_n$$

pour la réalisation propre au lac Huron.

**C.1.a.** Quelle est l'espérance de la variable aléatoire $M_N$ ?

**C.1.b.** Montrer que la variance de la variable aléatoire $M_N$ est

$$\Sigma^2 = \frac{1}{N^2}\left(\sum_{n=1}^{N}\mathbb{V}(X_n) + 2\sum_{1\leq n<m\leq N}\mathbb{C}ov(X_n, X_m)\right) = \frac{1}{N}\gamma_X(0) + \frac{2}{N^2}\sum_{p=1}^{N-1}(N-p)\gamma_X(p)$$

**C.1.c.** En déduire, en appliquant l'inégalité de Bienaymé–Tchebycheff et l'inégalité $(\star\star\star)$, que, pour tout $\varepsilon > 0$

$$\mathbb{P}(|M_N - \mu| > \varepsilon) \leq \frac{6\sigma^2}{N.(1-\varphi)^2.(1-\varphi^2)\varepsilon^2}$$

puis que, pour tout $\varepsilon > 0$,

$$\lim_{N\to+\infty} \mathbb{P}(|M_N - \mu| > \varepsilon) = 0$$

**C.2.** Pour estimer $\sigma^2$, $\varphi$ et vérifier que $(X_n)$ a les propriétés d'un processus stationnaire, on se propose d'estimer $\gamma_X$ à partir de $X_1, X_2, \ldots, X_N$, en posant, pour $0 \leq p < N$ :

$$\Gamma_N(p) = \frac{1}{N-p} \sum_{n=1}^{N-p} (X_n - M_N).(X_{n+p} - M_N)$$

Sur notre jeu de données, les quantités $\gamma_X(p)$ seront évaluées par les quantités $\hat{\gamma}_X(p)$, valeurs des variables aléatoires $\Gamma_N(p)$ pour la réalisation propre au lac Huron *i.e.* , pour $0 \leq p < N$,

$$\hat{\gamma}_X(p) = \frac{1}{N-p} \sum_{n=1}^{N-p} (x_n - \hat{\mu}).(x_{n+p} - \hat{\mu})$$

On va seulement montrer que lorsque $p = 0$, $\Gamma_N(0)$ est un estimateur asymptotiquement sans biais de $\gamma_X(0)$ c'est-à-dire que $\mathbb{E}(\Gamma_N(0))$ tend vers $\gamma_X(0)$ lorsque $N$ tend vers l'infini. On pourrait faire de même lorsque $p > 0$.

**C.2.a.** Montrer que

$$\Gamma_N(0) = \frac{1}{N} \sum_{n=1}^{N} X_n^2 - M_N^2 = \frac{1}{N} \sum_{n=1}^{N} X_n^2 - \left( \frac{1}{N} \sum_{n=1}^{N} X_n \right)^2$$

**C.2.b.** Montrer que sous les hypothèses faites, *i.e.* $(X_n)$ est un processus stationnaire de paramètres $\mu$ et $\gamma_X$,

$$\mathbb{E}(\Gamma_N(0)) = \gamma_X(0) + \mu^2 - \frac{1}{N^2} \left( N.(\gamma_X(0) + \mu^2) + 2 \sum_{p=1}^{N-1} (N-p)(\gamma_X(p) + \mu^2) \right).$$

**C.2.c.** A partir de l'équation précédente, montrer que :

$$\mathbb{E}(\Gamma_N(0)) = \gamma_X(0) - \frac{1}{N^2} \left( N.\gamma_X(0) + 2 \sum_{p=1}^{N-1} (N-p).\gamma_X(p) \right).$$

**C.2.d.** En utilisant $(\star\star\star)$, montrer que :

$$\lim_{N \to +\infty} \mathbb{E}(\Gamma_N(0)) = \gamma_X(0).$$

**C.3.** Informatique.

**C.3.a.** Ecrire une fonction Python `mu_gamma01(x)` qui prend en argument les $N$ valeurs mesurées $x =$ `x` et retourne, en triplet dans cet ordre, les valeurs calculées des estimateurs $\hat{\mu}(p)$, $\hat{\gamma}_X(0)$ et $\hat{\gamma}_X(1)$.

**C.3.b.** Ecrire une fonction Python `m_phi_sigma2(x)` qui prend en argument les $N$ valeurs mesurées $x =$ `x` et retourne, en triplet, dans cet ordre, les valeurs calculées de $m$, $\varphi$ et $\sigma^2$.

Indication:On pourra utiliser la fonction `mu_gamma01(x)`.

**C.3.c.** Ecrire une fonction Python `W()` qui retourne une valeur aléatoire uniformément distribuée sur un intervalle, d'espérance nulle, de variance 1.

**C.3.d.** Ecrire une fonction Python `simule_lac(N,m,phi,sigma2)` qui retourne $N$ valeurs consécutives d'une suite des hauteurs d'un lac (avec les paramètres de la récurrence $m =$ `m`, $\varphi =$ `phi` et $\sigma^2 =$ `sigma2`) simulée aléatoirement suivant le modèle décrit en $(\star\star)$.

Indication:Pour initier la récurrence, on pourra simuler $X_0$ d'espérance $\mu$, de variance $\frac{\sigma^2}{1-\varphi^2}$ à l'aide la fonction `W()`. On pourra ensuite simuler les $W_n$, d'espérance 0, de variance $\sigma^2$ toujours à l'aide la fonction `W()`.

# Correction DS 03

**Correction Ex.–1** Soit $a$ un nombre réel strictement supérieur à 1 et $f : \mathbb{R} \to \mathbb{R}$ la fonction définie par

$$\forall x \in \mathbb{R},\ f(x) = \frac{a}{x^{a+1}} 1\!\!1_{\{x \geq 1\}} = \begin{cases} \frac{a}{x^{a+1}} & \text{si } x \geq 1 \\ 0 & \text{sinon} \end{cases}$$

**1.**

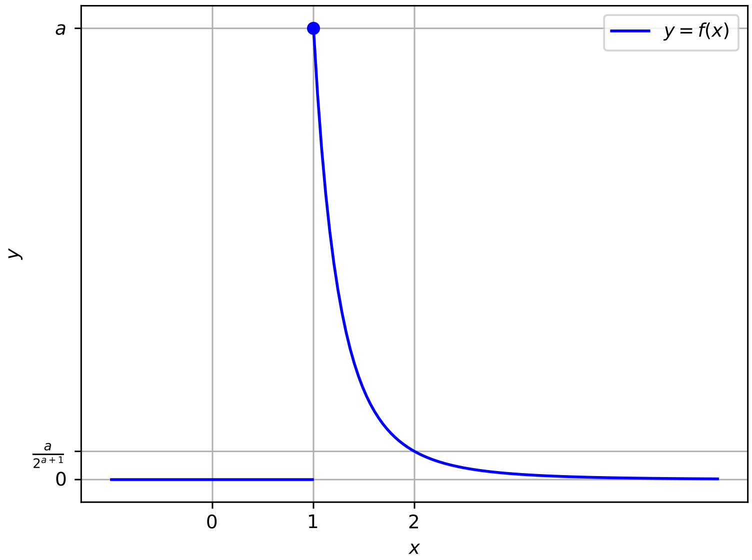


FIGURE 2 – Le graphe de $f$.

— $f$ est clairement positive sur $\mathbb{R}$
— $f$ est continue sur $]-\infty, 1[$ et sur $]1, +\infty[$, considérer son intégrale généralisée sur $\mathbb{R}$ à un sens.
— On a, le crochet étant à prendre au sens des limites,

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_1^{+\infty} \frac{a}{x^{a+1}}\, dx = \left[-x^{-a}\right]_1^{+\infty} = 1$$

— $f$ est donc une densité de probabilité

On suppose maintenant que $X$ est une variable aléatoire réelle de densité $f$.

**2.a.** $X$ est presque-sûrement à valeurs dans $I = [1, +\infty[$ car $f = 0$ sur $\mathbb{R} \setminus I$ et $f > 0$ sur $I$.

$X$ admet une espérance si

$$\int_{-\infty}^{+\infty} |x| f(x)\, dx = \int_1^{+\infty} x.\frac{a}{x^{a+1}}\, dx$$

est convergente. Or, on a, le crochet étant à prendre au sens des limites, comme $a > 1$,

$$\int_1^{+\infty} x.\frac{a}{x^{a+1}}\, dx = \left[-\frac{a}{a-1} x^{1-a}\right]_1^{+\infty} = \frac{a}{a-1}$$

Donc $X$ admet une espérance, et on a

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x.f(x)\, dx = \int_1^{+\infty} x.\frac{a}{x^{a+1}}\, dx = \frac{a}{a-1}$$

**2.b.** $X$ admet une variance si

$$\int_{-\infty}^{+\infty} |x|^2 f(x)\, dx = \int_1^{+\infty} x^2 . \frac{a}{x^{a+1}}\, dx$$

est convergente. Or, on a, le crochet étant à prendre au sens des limites, comme $a > 1$, si $a \neq 2$,

$$\int_1^{+\infty} x^2 . \frac{a}{x^{a+1}}\, dx = \left[ -\frac{a}{a-2} x^{2-a} \right]_1^{+\infty} = \begin{cases} +\infty & \text{si } 1 < a < 2 \\ \frac{a}{a-2} & \text{si } a > 2 \end{cases}$$

Dans le cas $a = 2$, l'intégration fait apparaitre un ln et l'intégrale diverge vers $+\infty$.

Donc $X$ admet une variance si et seulement si $a > 2$, et on a

$$\mathbb{E}(X^2) = \frac{a}{a-2}$$

et

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{a}{a-2} - \left( \frac{a}{a-1} \right)^2 = \frac{a}{(a-2)(a-1)^2} ((a-1)^2 - a(a-2)) = \frac{a}{(a-2)(a-1)^2}$$

**3.a.** $X$ est (p.s.) à valeurs dans $[1, +\infty[$ du fait de la nullité de $f$ sur $]-\infty, 1[$, donc, pour $x \leq 1$,

$$F_X(x) = \mathbb{P}(X \leq x) = 0$$

Pour $x > 1$,

$$F_X(x) = \mathbb{P}(X \leq x) = \int_1^x \frac{a}{t^{a+1}}\, dt = \left[ -t^{-a} \right]_{t=1}^{t=x} = 1 - \frac{1}{x^a}$$

En résumé, on a

$$\forall x \in \mathbb{R},\ F_X(x) = \begin{cases} 0 & \text{si } x \leq 1 \\ 1 - \frac{1}{x^a} & \text{si } x \geq 1 \end{cases}$$

**3.b.** L'équation $F_X(x) = \frac{1}{2}$ d'inconnue $x \in \mathbb{R}$ n'a pas de solution $x \leq 1$. Résolvons la pour $x \geq 1$, on a alors

$$F_X(x) = \frac{1}{2} \Leftrightarrow \frac{1}{x^a} = \frac{1}{2} \Leftrightarrow x = 2^{\frac{1}{a}}$$

**3.c.** On construit cette fonction en construisant la réciproque $Q_X$, la *fonction des quantiles* de $X$, de $F_X$ restreinte à $]1, +\infty[$ :

$$Q_X = F_X^{-1} : ]0, 1[ \to ]1, +\infty[$$

définie par (pour $u \in ]0, 1[$, la résolution de l'équation $F_X(x) = u$ en $x = (1-u)^{-\frac{1}{a}}$ se fait comme précédemment)

$$\forall u \in ]0, 1[,\ Q_X(u) = F_X^{-1}(u) = (1-u)^{-\frac{1}{a}}$$

On a alors (cours), si $U \sim \mathscr{U}_{]0,1[}$, $X = F_X^{-1}(U)$ a pour fonction de répartition $F_X$.

```
import numpy as np
def Fmoins1(u,a):
    return (1-u)**(-1/a)
def X(a):
    """
    Une simulation de X avec paramètre a
    """
    return Fmoins1(np.random.rand(),a)
```

**4.** On fixe $a > 1$ et on considère, pour $n \in \mathbb{N}$, $n \geq 2$, une famille $X_1, \ldots, X_n$ de variables aléatoires indépendantes de même loi que $X$. On pose alors

$$T_n = \min(X_1, \ldots, X_n)$$

**4.a.** On détermine $F_{T_n}$, la fonction de répartition de $T_n$. Comme $X_1, \ldots, X_n$ sont à valeurs $\geq 1$, $T_n$ est aussi à valeurs $\geq 1$ et donc

$$\forall t \in ]-\infty, 1[, F_{T_n}(t) = 0$$

Soit $t \geq 1$, on a

$$\begin{aligned} 1 - F_{T_n}(t) &= \mathbb{P}(T_n > t) = \mathbb{P}(X_1 > t \text{ et } \ldots \text{ et } X_n > t) \\ &\overset{\text{indep } X_k}{=} \mathbb{P}(X_1 > t) \ldots \mathbb{P}(X_n > t) \\ &= (1 - F_X(t))^n = \frac{1}{t^{n.a}} \end{aligned}$$

On a donc

$$\forall t \in \mathbb{R}, F_{T_n}(t) = \begin{cases} 0 & \text{si } t \leq 1 \\ 1 - \frac{1}{t^{n.a}} & \text{si } t \geq 1 \end{cases}$$

$T_n$ est donc une variable aléatoire réelle à densité : elle a pour densité $f$ avec $a' = n.a$.
**4.b.** D'après 2.a, $T_n$ admet une espérance car $n.a > 1$ et

$$\mathbb{E}(T_n) = \frac{na}{na - 1}$$

**4.c.** D'après 2.b, $T_n$ admet une variance si et seulement si $n.a > 2$. Sachant que $n \geq 2$ et $a > 1$, cette condition est toujours remplie. On a alors

$$\mathbb{V}(T_n) = \frac{na}{(na-1)^2(na-2)}$$

**4.d.** Pour construire les deux fonctions Python `T1(n,a)` et `T2(n,a)` permettant de tirer au sort un nombre réel en suivant la loi de $T_n$. On peut

— Soit utiliser `X` avec comme paramètre $n.a$

— Soit tirer au sort $n$ variables $X_k$ en utilisant $n$ fois `X(a)` et en prenant le minimum de ces valeurs. (C'est la définition de $T_n$).

```python
def T1(n,a):
    """
    Une simulation de T_n:  première méthode
    """
    return X(n*a)
def T2(n,a):
    """
    Une simulation de T_n:  deuxième méthode, def originelle de T_n
    """
    Min=0 # de toute facon le min final sera >=1
    for k in range(n): #on fait n tirages X(a), on garde la plus petite valeur
        Xk=X(a)
        if Xk < Min:
            Min=Xk
    return Min
```

**5.** On pose $Z = \ln(X)$.

**5.a.** D'après 2.a, comme $X$ est à valeurs $\geq 1$, $Z$ est à valeurs $\geq 0$. Donc, pour $z < 0$,

$$F_Z(z) = \mathbb{P}(Z \leq z) = 0$$

Pour $z \geq 0$, on a

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(\ln X \leq z) = \mathbb{P}(X \leq e^z) \overset{(e^z \geq 1)}{=} 1 - \frac{1}{(e^z)^a} = 1 - e^{-a.z}$$

On reconnait la fonction de répartition d'une v.a exponentielle de param. $a$ et donc $Z \sim \mathscr{E}(a)$.
**5.b.** On a donc (cours)

$$\mathbb{E}(Z) = \frac{1}{a} \text{ et } \mathbb{V}(Z) = \frac{1}{a^2}$$

**6.** Soient $X_1$ et $X_2$ deux variables indépendantes suivant la loi de $X$, $Z_1 = \ln X_1$, $Z_2 = \ln X_2$.
**6.a.** $Z_1$ et $Z_2$ sont deux variables indépendantes (lemme des coalitions, fonctions de « blocs » de variables indépendantes), de loi $\mathscr{E}(a)$.

En utilisant la formule de convolution, on calcule, pour $w \in \mathbb{R}$,

$$f_W(w) = \int_{-\infty}^{+\infty} a.e^{-az} 1\!\!1_{\{z \geq 0\}} a.e^{-a(w-z)} 1\!\!1_{\{w-z \geq 0\}} \, dz = a^2 e^{-a.w} \int_{-\infty}^{+\infty} 1\!\!1_{\{0 \leq z \leq w\}} \, dz$$

On a donc

$$f_W(w) = \begin{cases} 0 & \text{si } w \leq 0 \\ a^2 e^{-a.w}.w & \text{si } w \geq 0 \end{cases}$$

**6.b.** Appliquons la méthode de la formule de transfert générique pour déterminer une densité de $P = X_1.X_2 = e^W$. Soit $h : \mathbb{R} \to \mathbb{R}$ une fonction telle que $h(P)$ admet une espérance. On a

$$\mathbb{E}(h(P)) \quad = \quad \mathbb{E}(h(e^W)) = \int_0^{+\infty} h(e^w) a^2 e^{-a.w}.w \, dw$$

Effectuons le changement de variable $\mathscr{C}^1$ strictement monotone de $]0, +\infty[$ sur $]1, +\infty[$, $p = e^w$, $w = \ln(p)$, $dw = \frac{1}{p}\, dp$, pour obtenir

$$\mathbb{E}(h(P)) \quad = \quad \int_1^{+\infty} h(p) a^2 p^{-a}.\ln(p).\frac{1}{p} \, dp$$

ce qui montre qu'une densité de $P$ est donnée par

$$\forall p \in \mathbb{R},\ f_P(p) = a^2 \frac{\ln(p)}{p^{a+1}} 1\!\!1_{\{p \geq 1\}}$$

**6.c.** On a, par indépendance de $X_1$ et $X_2$ que

$$\mathbb{E}(P) = \mathbb{E}(X_1.X_2) = \mathbb{E}(X_1).\mathbb{E}(X_2) = \left(\frac{a}{a-1}\right)^2$$

**6.d.** Dans le cas $a > 2$, on a, par indépendance de $X_1$ et $X_2$ que

$$\mathbb{E}(P^2) = \mathbb{E}(X_1^2.X_2^2) = \mathbb{E}(X_1^2).\mathbb{E}(X_2^2) = \left(\frac{a}{a-2}\right)^2$$

et donc, par KOENIG–HUYGHENS,

$$\mathbb{V}(P) = \left(\frac{a}{a-2}\right)^2 - \left(\frac{a}{a-1}\right)^4$$

On peut travailler cette expression mais, on n'obtient pas forcément d'expression très intéressante..

**6.e.** On utilise simplement la définition de $P$ et surtout pas un calcul d'inverse de fonction de répartition.

```
def P(a):
    """
    Une simulation de X_1.X_2 : on utilise la def
    """
    return X(a)*X(a)
```

**Correction Ex.–2**

**Partie A**
Préliminaires

**A.1.** *c.f.* Cours
**A.2.** *c.f.* Cours ou une récurrence à partir de la question précédente.
**A.3.** Soit $(X_n)_{n\in\mathbb{Z}}$ un *processus stationnaire* de paramètres $\mu$ et $\gamma_X$.
**A.3.a.** Soit $n \in \mathbb{Z}$. On a $\mathbb{V}(X_n) = \mathbb{C}ov(X_n, X_{n+0}) = \gamma_X(0)$.
**A.3.b.** Soit $n \in \mathbb{Z}$, $p \in \mathbb{N}$, en appliquant la définition de processus stationnaire avec $n-p$ à la place de $n$, on obtient

$$\mathbb{C}ov(X_{n-p}, X_n) = \mathbb{C}ov(X_{(n-p)}, X_{(n-p)+p}) = \gamma_X(p)$$

**A.3.c.** Soit $m,n \in \mathbb{Z}$. En posant $p = |m-n| \in \mathbb{N}$, alors, suivant que $m < n$ ou $m \geq n$, on a $m = n-p$ ou $m = n+p$ et donc, dans tous les cas, $\mathbb{C}ov(X_n, X_{n\pm p}) = \gamma_X(p) = \gamma_X(|m-n|)$.

**Partie B**

**B.1.** Soient $(Z_n)_{n\in\mathbb{Z}}$ des variables aléatoires réelles indépendantes et identiquement distribuées, *i.e.* ayant toutes même loi, d'espérance $m$ et de variance $\sigma^2$.
**B.1.a.** Soit $n \in \mathbb{Z}$.

— Lorsque $p = 0$, $\mathbb{C}ov(Z_n, Z_{n+p}) = \mathbb{C}ov(Z_n, Z_n) = \mathbb{V}(Z_n) = \sigma^2$ ;
— lorsque $p > 0$, $\mathbb{C}ov(Z_n, Z_{n+p}) = 0$ car $Z_n$ et $Z_{n+p}$ sont supposées indépendantes.

**B.1.b.** $(Z_n)_{n\in\mathbb{Z}}$ est un processus stationnaire, de paramètres $\mu = m$ et

$$\gamma_Z(p) = \begin{cases} 0 & \text{si p=0} \\ \sigma^2 & \text{si p>0} \end{cases}$$

**B.2.** On considère à présent :

$$\forall n \geq 1, X_n = W_n + \theta W_{n-1}, \tag{R}$$

où $\theta$ est un nombre réel non nul et $(W_n) \sim \mathscr{BB}(0, \sigma^2)$.
**B.2.a.** Soit $n \in \mathbb{Z}$. En appliquant l'espérance à la relation (R), on obtient par linéarité de l'espérance,

$$\mathbb{E}(X_n) = \mathbb{E}(W_n) + \theta.\mathbb{E}(W_{n-1}) = 0$$

**B.2.b.**

— Soit $n \in \mathbb{Z}$ et $p = 0$, on a alors

$$\mathbb{C}ov(X_n, X_n) = \mathbb{E}(X_n^2) = \mathbb{E}(W_n^2 + \theta^2 W_{n-1}^2 + 2\theta W_{n-1}.W_n)$$

et donc, par linéarité de l'espérance

$$\mathbb{C}ov(X_n, X_n) = \mathbb{E}(W_n^2) + \theta^2\mathbb{E}(W_{n-1}^2) + 2\theta.\mathbb{E}(W_{n-1}.W_n)$$

Comme $\mathbb{E}(W_n) = \mathbb{E}(W_{n-1}) = 0$, on a, par C.3.b,

$$\mathbb{E}(W_n^2) = \mathbb{E}(W_{n-1}^2) = \gamma_W(0) = \sigma^2 \text{ et } \mathbb{E}(W_{n-1}.W_n) = \mathbb{C}ov(W_{n-1}, W_n) = \gamma_W(1) = 0$$

et donc

$$\mathbb{C}ov(X_n, X_n) = (1+\theta^2).\sigma^2$$

— Soit $n \in \mathbb{Z}$ et $p = 1$, on a alors, en développant $X_n.X_{n+1} = (W_n + \theta W_{n-1}).(W_{n+1} + \theta W_n)$,

$$\mathbb{C}ov(X_n, X_{n+1}) = \mathbb{E}(X_n.X_{n+1}) = \mathbb{E}(W_n.W_{n+1} + \theta^2 W_{n-1}.W_n + \theta W_{n-1}.W_{n+1} + \theta W_n^2)$$

et donc, par linéarité de l'espérance

$$\mathbb{C}ov(X_n, X_{n+1}) = \mathbb{E}(W_n.W_{n+1}) + \theta^2 \mathbb{E}(W_{n-1}.W_n) + \theta \mathbb{E}(W_{n-1}.W_{n+1}) + \theta \mathbb{E}(W_n^2)$$

Comme $\mathbb{E}(W_{n+1}) = \mathbb{E}(W_n) = \mathbb{E}(W_{n-1}) = 0$ et par C.3.b, on a

$$\mathbb{E}(W_n^2) = \gamma_W(0) = \sigma^2, \mathbb{E}(W_n.W_{n+1}) = \mathbb{E}(W_{n-1}.W_n) = \gamma_W(1) = 0 \text{ et } \mathbb{E}(W_{n-1}.W_{n+1}) = \gamma_W(2) = 0$$

et donc

$$\mathbb{C}ov(X_n, X_{n+1}) = \theta.\sigma^2$$

— Soit $n \in \mathbb{Z}$ et $p > 1$, on a alors, en développant $X_n.X_{n+p} = (W_n + \theta W_{n-1}).(W_{n+p} + \theta W_{n+p-1})$,

$$\mathbb{C}ov(X_n, X_{n+p}) = \mathbb{E}(X_n.X_{n+p}) = \mathbb{E}(W_n.W_{n+p} + \theta^2 W_{n-1}.W_{n+p-1} + \theta W_{n-1}.W_{n+p} + \theta W_n.W_{n+p-1})$$

et donc, par linéarité de l'espérance

$$\mathbb{C}ov(X_n, X_{n+p}) = \mathbb{E}(W_n.W_{n+p}) + \theta^2 \mathbb{E}(W_{n-1}.W_{n+p-1}) + \theta \mathbb{E}(W_{n-1}.W_{n+p}) + \theta \mathbb{E}(W_n.W_{n+p-1})$$

Le même argument que précédemment, par C.3.b, puis $p+1 > p > p-1 \geq 1$, donne donc que

$$\mathbb{C}ov(X_n, X_{n+p}) = \gamma_W(p) + \theta^2.\gamma_W(p) + \theta.\gamma_W(p+1) + +\theta.\gamma_W(p-1) = 0$$

**B.2.c.** En conclusion, la suite $(X_n)_{n\in\mathbb{Z}}$ est un processus stationnaire de paramètres $\mu = 0$ et

$$\gamma_X(p) = \begin{cases} (1+\theta^2).\sigma^2 & \text{si } p = 0 \\ \theta.\sigma^2 & \text{si } p = 1 \\ 0 & \text{si } p > 1 \end{cases}$$

**B.3.** Soit $(X_n)$ un processus stationnaire de paramètres $\mu$ et $\gamma_X$, solution de l'équation suivante :

$$\forall n \geq 1, X_n - \varphi X_{n-1} = W_n, \tag{$\star$}$$

où $\varphi$ est un nombre réel non nul tel que : $-1 < \varphi < 1$ et $(W_n) \sim \mathscr{BB}(0, \sigma^2)$.
**B.3.a.** Soit $n \in \mathbb{Z}$. En appliquant l'espérance à ($\star$), on obtient, par linéarité,

$$\mathbb{E}(X_n) - \varphi.\mathbb{E}(X_{n-1}) = \mathbb{E}(W_n) = 0.$$

Comme $\mathbb{E}(X_n) = \mu$ est indépendant de $n$, on obtient $(1-\varphi).\mu = 0$ et donc $\mu = 0$.
**B.3.b.** Ecrivons ($\star$) pour $n$, $n-1$,...,$n-p$ en multipliant à chaque fois par $\varphi$. On obtient alors

$$\begin{array}{rcrcl} X_n & - & \varphi.X_{n-1} & = & W_n \\ \varphi.X_{n-1} & - & \varphi^2.X_{n-2} & = & \varphi.W_{n-1} \\ \varphi^2.X_{n-1} & - & \varphi^3.X_{n-3} & = & \varphi^2.W_{n-2} \\ \varphi^3.X_{n-2} & - & \varphi^4.X_{n-4} & = & \varphi^2.W_{n-3} \\ \vdots & & \vdots & = & \vdots \\ \varphi^p.X_{n-p} & - & \varphi^{p+1}.X_{n-(p+1)} & = & \varphi^p.W_{n-p} \end{array}$$

En additionnant ces égalités, elles se télescopent pour donner

$$X_n - \varphi^{p+1}.X_{n-(p+1)} = \sum_{k=0}^{p} \varphi^k.W_{n-k}. \tag{$\star_p$}$$

Une démonstration "plus stricte" se fait par récurrence sur $p$ en posant, au rang $p \in \mathbb{N}$ :

$$(H_p) : \forall n \in \mathbb{Z}, X_n - \varphi^{p+1}.X_{n-(p+1)} = \sum_{k=0}^{p} \varphi^k.W_{n-k}$$

— (Init.) Au rang $p = 0$, il s'agit de la propriété $(\star)$ ;
— (Hérédité.) Supposons $(H_p)$ vraie au rang $p$. Soit $n \in \mathbb{Z}$. On a alors, en spécialisant $(H_p)$ à $n-1 \in \mathbb{Z}$,

$$X_{n-1} - \varphi^{p+1}.X_{n-(p+2)} = \sum_{k=0}^{p} \varphi^k.W_{n-1-k}.$$

En multipliant cette identité par $\varphi$ et en lui ajoutant $(\star)$, on obtient

$$X_n - \varphi.X_{n-1} + \varphi.X_{n-1} - \varphi^{p+2}.X_{n-(p+2)} = W_n + \sum_{k=0}^{p} \varphi^{k+1}.W_{n-1-k}$$

*i.e.* après changement d'indice $\ell = k+1$ et regroupement

$$X_n - \varphi^{p+2}.X_{n-(p+2)} = \sum_{\ell=0}^{p+1} \varphi^\ell.W_{n-\ell}.$$

Du fait que $n$ est quelconque $\in \mathbb{Z}$, on a donc obtenu $(H_{p+1})$.
— Par récurrence, $\forall p \in \mathbb{N}$, $H_p$, ce qui est la proposition demandée.

**B.3.c.**

**3.c.i.** Soit $p \in \mathbb{N}$, en prenant la variance de $(\star_p)$, on obtient que

$$\mathbb{V}(X_n - \varphi^{p+1}.X_{n-(p+1)}) = \mathbb{V}(\sum_{k=0}^{p} \varphi^k.W_{n-k})$$

et donc en utilisant A.1, A.2 (les $\varphi^k.W_{n-k}$ sont deux à deux non corrélées) et le caractère quadratique de la variance, il vient que

$$\mathbb{V}(X_n) + \varphi^{2p+2}.\mathbb{V}(X_{n-(p+1)}) - 2.\,\mathbb{C}ov(X_n, \varphi^{p+1}.X_{n-(p+1)}) = \sum_{\ell=0}^{p} \varphi^{2k}.\mathbb{V}(W_{n-k}).$$

Comme

$$\mathbb{V}(X_n) = \mathbb{V}(X_{n-(p+1)}) = \gamma_X(0),\ \mathbb{C}ov(X_n, \varphi^{p+1}.X_{n-(p+1)}) = \varphi^{p+1}.\gamma_X(p+1) \text{ et } \mathbb{V}(W_{n-k}) = \sigma^2,$$

on obtient finalement

$$(1+\varphi^{2p+2}).\gamma_X(0) - 2.\varphi^{p+1}\gamma_X(p+1) = \sigma^2.\sum_{k=0}^{p} \varphi^{2k}.$$

**3.c.ii.** Soulignons que $|\varphi| < 1$ par hypothèse. On reconnait à droite de cette identité une somme de termes géométrique valant $\sigma^2 \frac{1-\varphi^{2(p+1)}}{1-\varphi^2}$ dont la limite lorsque $p \to +\infty$ vaut $\frac{\sigma^2}{1-\varphi^2}$.

A gauche de l'identité, lorsque $p \to +\infty$,

$$(1+\varphi^{2p+2}).\gamma_X(0) \to \gamma_X(0)$$

et, comme on a supposé la suite $\gamma_X(p)$ bornée, $-2.\varphi^{p+1}\gamma_X(p+1) \to 0$.

Par opérations sur les limites et unicité de la limite, on en déduit que

$$\gamma_X(0) = \frac{\sigma^2}{1-\varphi^2}.$$

**3.c.iii.** Soit $p \in \mathbb{N}$, en reprenant le résultat de B.3.c.i et la valeur tout juste trouvée pour $\gamma_X(0)$, on a

$$(1+\varphi^{2p+2}).\frac{\sigma^2}{1-\varphi^2} - 2.\varphi^{p+1}\gamma_X(p+1) = \sigma^2.\frac{1-\varphi^{2(p+1)}}{1-\varphi^2}.$$

En réarrangeant cette égalité, on a

$$2.\varphi^{p+1}\gamma_X(p+1) = 2.\varphi^{2p+2}.\frac{\sigma^2}{1-\varphi^2}$$

et finalement, en simplifiant par $\varphi^{p+1}$, on obtient le résultat cherché, *i.e.*

$$\forall p \in \mathbb{N},\ \gamma_X(p+1) = \frac{\varphi^{p+1}\sigma^2}{1-\varphi^2}.$$

NB : le cas $\varphi = 0$ est le cas où $(X_n)$ est un bruit blanc, déjà traité en B.1 et B.2.

**B.4.** Soit $(X_n)_{n \in Z}$ un processus stationnaire de paramètres $\mu$ et $\gamma_X$, solution de l'équation suivante :

$$\forall n \in \mathbb{Z},\ X_n - \varphi X_{n-1} = m + W_n, \qquad (\star\star)$$

où $m \in \mathbb{R}$, $\varphi$ est un nombre réel non nul tel que : $-1 < \varphi < 1$ et $(W_n) \sim \mathscr{B}\mathscr{B}(0,\sigma^2)$.

On suppose de plus que la suite $(\gamma_X(p))_{p \in \mathbb{N}}$ est bornée.

**B.4.a.** Soit $n \in \mathbb{Z}$, en appliquant l'espérance à l'identité $(\star\star)$, on obtient $\mu - \varphi.\mu = m$ et donc

$$\mu = \mathbb{E}(X_n) = \frac{m}{1-\varphi}.$$

**B.4.b.** Pour $n \in \mathbb{Z}$, on pose $\tilde{X}_n = X_n - \mu$.

On a, pour $n \in \mathbb{Z}$,

— $\mathbb{E}(\tilde{X}_n) = \mathbb{E}(X_n) - \mu = 0$ ;

— et, pour $p \in \mathbb{N}$,

$$\mathbb{C}ov(\tilde{X}_n, \tilde{X}_{n+p}) = \mathbb{E}((X_n - \mu).(X_{n+p} - \mu)) = \mathbb{C}ov(X_n, X_{n+p}) = \gamma_X(p)$$

Ces quantités sont bien indépendantes de $n$ et donc $(\tilde{X}_n)_{n \in \mathbb{Z}}$ est un processus stationnaire de paramètres $\tilde{\mu} = 0$ et $\gamma_{\tilde{X}} = \gamma_X$.

**B.4.c.** De B.3 appliqué à $(\tilde{X}_n)$ ( l'hypothèse de "suite bornée" est remplie), on en déduit que

$$\forall p \in \mathbb{N},\ \gamma_X(p) = \varphi^p.\frac{\sigma^2}{1-\varphi^2}$$

$\star\star\star$

On admettra pour la suite la majoration

$$\forall N \in \mathbb{N}^*,\ \left|\frac{1}{N}\sum_{p=0}^{N-1}(N-p)\gamma_X(p)\right| \leq \frac{3\sigma^2}{(1-\varphi)^2.(1-\varphi^2)} \qquad (\star\star\star)$$

## Partie C

**C.1.a.** En prenant l'espérance dans la relation $M_N = \frac{1}{N}\sum_{n=1}^{N} X_n$, par linéarité de l'espérance et du fait que $\mathbb{E}(X_n) = \mu$, on obtient

$$\mathbb{E}(M_N) = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}(X_n) = \mu.$$

**C.1.b.** La variance de la variable aléatoire $M_N$ est

$$\Sigma^2 = \frac{1}{N^2}\left(\sum_{n=1}^{N}\mathbb{V}(X_n) + 2\sum_{1\leq n<m\leq N}\mathbb{C}ov(X_n, X_m)\right)$$

Ce qui s'obtient comme dans la question C.2, ou par récurrence sur la question C.1.

On observe que (C.3), $\mathbb{C}ov(X_n, X_m) = \gamma_X(m-n)$ lorsque $n < m$ et donc on peut réordonner la somme double par rapport aux valeurs de $p = m-n$ (ce qui revient à sommer sur les sous-diagonales d'un carré). Cette différence varie de 1 à $N-1$ et le nombre de termes sur la diagonale $p$ est $N-p$.

De même $\mathbb{V}(X_n) = \gamma_X(0)$ (cela correspond à la diagonale $p=0$) et donc, en réécrivant

$$\Sigma^2 = \frac{1}{N}\gamma_X(0) + \frac{2}{N^2}\sum_{p=1}^{N-1}(N-p)\gamma_X(p)$$

**C.1.c.** Soit $\varepsilon > 0$. En appliquant l'inégalité de BIENAYMÉ–TCHEBYCHEFF à $M_N$, on obtient

$$\mathbb{P}(|M_N-\mu| > \varepsilon) \leq \frac{\Sigma^2}{\varepsilon^2}$$

En reprenant la formule démontrée à la question précédente, on voit que

$$\Sigma^2 \leq \frac{1}{N}\gamma_X(0) + \Sigma^2 = \frac{2}{N}\left(\frac{1}{N}\sum_{p=0}^{N-1}(N-p)\gamma_X(p)\right)$$

et donc, en utilisant la majoration ($\star\star\star$), on obtient

$$\mathbb{P}(|M_N-\mu| > \varepsilon) \leq \frac{6\sigma^2}{N.(1-\varphi)^2.(1-\varphi^2)\varepsilon^2}.$$

Il est clair que le membre de gauche de cette inégalité tend vers 0 lorsque $N \to +\infty$ et donc, par le théorème de gendarmes (une proba. est $\geq 0$ !), on a

$$\lim_{N\to+\infty}\mathbb{P}(|M_N-\mu| > \varepsilon) = 0$$

**C.2.a.** La formule de KOENIG–HUYGHENS donne que pour toute famille de nombres $(x_n)_{1\leq n\leq N}$

$$\frac{1}{N}\sum_{n=1}^{N}(x_n-\bar{x})^2 = \frac{1}{N}\sum_{n=1}^{N}x_n^2 - \bar{x}^2$$

où $\bar{x} = \frac{1}{N}\sum_{n=1}^{N}x_n$. On a donc, en appliquant cette formule pour chaque $\omega \in \Omega$, en posant $x_n = X_n(\omega)$, que

$$\Gamma_N(0) = \frac{1}{N}\sum_{n=1}^{N}X_n^2 - (M_N)^2 = \frac{1}{n}\sum_{n=1}^{N}X_n^2 - \left(\frac{1}{N}\sum_{n=1}^{N}X_n\right)^2$$

**C.2.b.** En prenant l'espérance de la formule précédente, on obtient, par linéarité de $\mathbb{E}$, que

$$\mathbb{E}(\Gamma_N(0)) = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}(X_n^2) - \mathbb{E}(M_N^2)$$

En utilisant le fait que (Notation et résultats de C.1.a et C.1.b)

$$\mathbb{E}(X_n^2) = \mathbb{V}(X_n) + \mu^2 = \gamma_X(0) + \mu^2 \text{ et } \mathbb{E}(M_N^2) = \mathbb{V}(M_N) + \mu^2 = \Sigma^2 + \mu^2$$

il vient que (on remplace $\Sigma^2$ par la formule trouvée en C.1.b.)

$$\mathbb{E}(\Gamma_N(0)) = \gamma_X(0) + \mu^2 - \frac{1}{N^2}\left(N.(\gamma_X(0)+\mu^2) + 2\sum_{p=1}^{N-1}(N-p)(\gamma_X(p)+\mu^2)\right).$$

**C.2.c.** Dans l'équation précédente, en développant, la quantité en facteur de $\mu^2$ est

$$1-\frac{1}{N^2}\left(N+2\sum_{p=1}^{N-1}(N-p)\right)=1-\frac{1}{N^2}\left(N+2\frac{N.(N-1)}{2}\right)=1-\frac{1}{N^2}\left(N^2\right)=0$$

On a bien sûr reconnu et utilisé la formule donnant la somme des entiers de 1 à $N-1$. De l'équation précédente, il reste donc :

$$\mathbb{E}(\Gamma_N(0))=\gamma_X(0)-\frac{1}{N^2}\left(N.\gamma_X(0)+2\sum_{p=1}^{N-1}(N-p).\gamma_X(p)\right).$$

**C.2.d.** Comme en C.1.c, par $(\star\star\star)$, lorsque $N\to+\infty$,

$$\frac{1}{N^2}\left(N.\gamma_X(0)+2\sum_{p=1}^{N-1}(N-p).\gamma_X(p)\right)=\frac{1}{N}\left(\gamma_X(0)+2\frac{1}{N}.\sum_{p=1}^{N-1}(N-p).\gamma_X(p)\right)\to 0$$

et donc :

$$\lim_{N\to+\infty}\mathbb{E}(\Gamma_N(0))=\gamma_X(0).$$

**C.3.** Informatique.
**C.3.a.** Il s'agit juste d'implémenter les formules de moyennes proposées, à la façon d'un calcul de moyenne, de variance ou de covariance usuel.

```
def mu_gamma01(x):
    """
    Calcul de la moyenne et des gamma(p), p=0,1
    """
    N = len(x)
    mu = 0
    for xi in x:
        mu += xi
    mu =mu/N
    gamma = [0]*2 #Prepare la liste des gamma
    for p in range(2):
        for n in range(N-p):
            gamma[p] += (x[n]-mu)*(x[n+p]-mu)
        gamma[p] = gamma[p]/(N-p)
    return mu,gamma[0],gamma[1]
```

**C.3.b.** Ici, c'est encore plus simple, on utilise juste les formules qui transforme les paramètres d'entrée en paramètres de sortie, le tout est d'avoir compris les relations exhibées dans l'introduction à cette méthode.

```
def m_phi_sigma2(x):
    """
    retourne les paramètres estimés du modèle
    """
    mu,gamma0,gamma1 = mu_gamma01(x)
    phi = gamma1/gamma0
    m =mu*(1-phi)
    sigma2 = gamma0*(1-phi**2)
    return m,phi,sigma2
```

**C.3.c.** Comme `numpy.random.rand()` retourne une variable $U$ uniforme sur $[0,1]$ et donc d'espérance $\frac{1}{2}$, de variance $\frac{1}{12}$, il suffit de prendre une fonction affine ad-hoc de $U$, *i.e.* $W=\sqrt{12}.(U-\frac{1}{2})$.

```
from numpy.random import rand
from numpy import sqrt
def W():
    """
    v.a uniforme sur intervalle esperance nulle, variance 1
    """
    return (rand()-0.5)*2*sqrt(3)
```

**C.3.d.** A près avoir mis les bons paramètres permettant de simuler des v.a. d'espérance et de variance adéquates à l'aide de `W()`, on met un place un calcul de suite récurrente avec, à l'initialisation et à chaque étape, un élément aléatoire.

```
def simule_lac(N,m,phi,sigma2):
    X = [0]*N
    sigma = sqrt(sigma2)
    s = sigma/sqrt(1-phi**2)
    mu =m/(1-phi)
    X[0] = mu + s*W()
    for n in range(1,N):
        X[n] = m + phi*X[n-1] + sigma*W()
    return X
```

⋆⋆⋆

En récupérant les données du lac Huron (cf Scripts corrigés complets), on peut grâce à ces fonctions, simuler un lac qui ressemble au lac Huron, pourvu que l'on croie que ce lac suive le modèle proposé. Peut-on s'en servir pour faire des prédictions ? Bonne question, ça me semble douteux. Ce serait comme observer un joueur de Monopoly, deviner sa façon de jouer et en déduire une prédiction pour reste de la partie.

Le modèle proposé dans le texte est bien connu en hydrologie (et il est utilisé dans beaucoup d'autres domaines), il s'appelle le modèle $AR-1$, comme auto-régressif d'ordre 1.

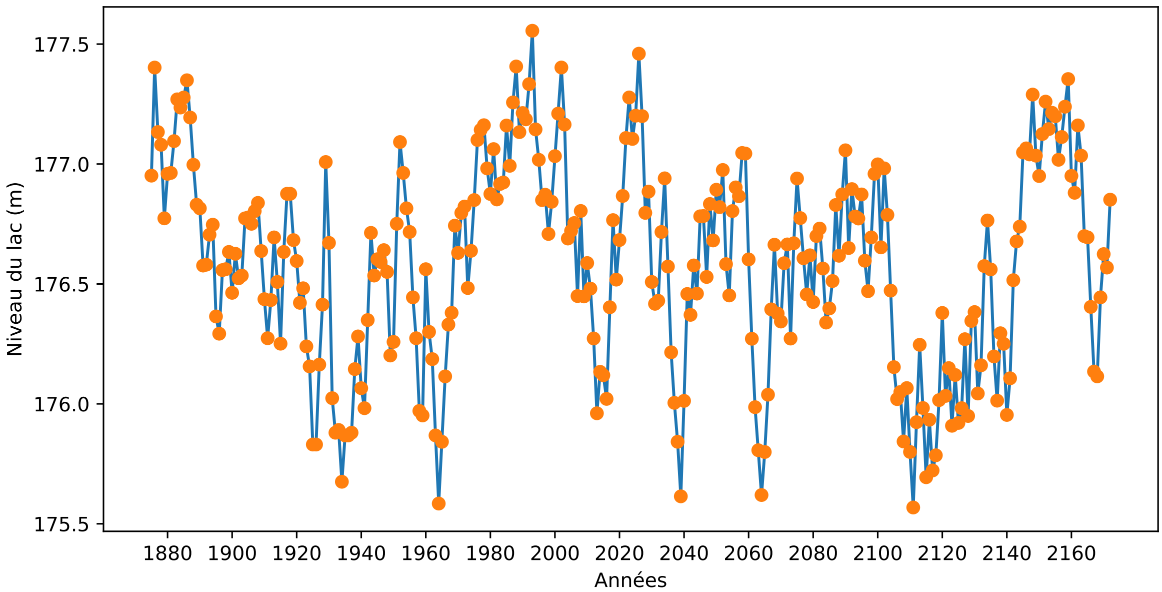


FIGURE 3 – « Prolongation » des données du lac Huron.