ENSTA MA101
Lundi 29 juin 2015 (3h)

Exercice 1 (Modèle gaussien)**.** On souhaite estimer la précision d'un instrument de mesure. Pour cela on effectue pour chaque objet 2 mesures. On note X_n et Y_n les 2 mesures de l'objet n. On suppose que $(X_n, n \in \mathbb{N}^*)$ et $(Y_n, n \in \mathbb{N}^*)$ sont des suites indépendantes de variables aléatoires indépendantes, et pour tout $n \in \mathbb{N}^*$, les variables aléatoires X_n et Y_n ont même loi gaussienne $\mathcal{N}(\mu_n, \sigma^2)$ où $\mu = (\mu_n, n \in \mathbb{N}^*)$ est une suite de réels et $\sigma > 0$ représente la précision de l'instrument de mesure. On note $\theta = (\mu, \sigma^2)$ le paramètre de dimension infinie du modèle.

Soit $Z_n = (X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ le vecteur aléatoire de taille $2n$, $n \in \mathbb{N}^*$. On note $\hat{\sigma}_n^2$ l'estimateur du maximum de vraisemblance de σ^2, quand il existe, construit à partir de Z_n. On note $p_n(x, y; \theta)$, avec $x, y \in \mathbb{R}^n$, la vraisemblance associée à Z_n. La vraisemblance p_n est une fonction de θ est plus précisément de $(\mu_k, 1 \leq k \leq n)$ et σ^2.

1. Calculer la log-vraisemblance $\ell_n(x, y; \theta)$, avec $x, y \in \mathbb{R}^n$.
2. Calculer les estimateurs du maximum de vraisemblance de $(\mu_k, 1 \leq k \leq n)$ et de σ^2 construit à partir de Z_n.
3. Déterminer la loi de $X_k - Y_k$. Puis montrer que la suite $(\hat{\sigma}_n^2, n \geq 1)$ converge p.s. vers une limite que l'on précisera. L'estimateur du maximum de vraisemblance de σ^2 est-il convergent ?
4. Construire un estimateur convergent de σ^2.

Dans cet exemple [1], la dimension du paramètre croit avec le nombre de données. Ce phénomène est responsable du mauvais comportement des estimateurs du maximum de vraisemblance.

△

Exercice 2 (Loi de Benford et détection de fraude)**.** Une variable aléatoire, C, à valeurs dans $\{1, \ldots, 9\}$ suit la loi de Benford [2] [3] si pour tout $k \in \{1, \ldots, 9\}$:

$$\mathbb{P}(C = k) = \log_{10}\left(1 + \frac{1}{k}\right). \tag{1}$$

La loi de Benford modélise le premier chiffre significatif de nombres aléatoires dont "la loi (de la mantisse) est invariante par multiplication" (voir la question III.3) ; elle approche également la loi du premier chiffre significatif d'un produit de nombreuses variables aléatoires continues indépendantes et de même loi (voir la question III.7). Par exemple des séries de valeurs dépendant d'une unité de mesure (données économiques (prix d'actions, audit financier, comptabilité nationale, ...), constantes physiques, ...) peuvent être modélisées par une loi invariante par multiplication dans la mesure où le choix de l'unité de référence est arbitraire. Cette propriété est parfois utilisée pour la détection de fraudes sur les données économiques, voir la partie IV.

---

1. J. Neyman and E. Scott. Consistent estimates based on partially consistent observations. *Econometrica*, vol. 16(1), pp. 1-32 (1948).
2. S. Newcomb. Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, vol. 4(1), pp. 39-40 (1881).
3. F. Benford. The law of anomalous numbers. *Proc. Am. Philos. Soc.* vol. 78 (4), pp. 551-572 (1938).

Notations

Pour $x \geq 0$, on note $\lfloor x \rfloor$ la partie entière de x, c'est à dire l'unique entier $k \in \mathbb{N}$ tel que $k \leq x < k+1$. On définit la partie fractionnaire de x par

$$\{x\} = x - \lfloor x \rfloor.$$

On a $\{x\} \in [0, 1[$. Voir les exemples du tableau 1. Pour $q \geq 0$, on pose $x_{\{q\}} = \{x + q\}$. On rappelle que pour tout $q \geq 0$:

$$\{x + q\} = \{x + \{q\}\}. \tag{2}$$

Par convention ψ_X est la fonction caractéristique de la variable aléatoire réelle X.

I Loi invariante par translation

Soit X une variable aléatoire à valeurs dans $[0, 1[$. On note $X_{\{q\}}$ la partie fractionnaire de $X + q$. On dit que la loi de X est invariante par translation si pour tout $q \geq 0$, $X_{\{q\}}$ a même loi que X. D'après (2), la loi de X est invariante par translation si pour tout $q \in [0, 1[$, $X_{\{q\}}$ a même loi que X.

1. Montrer que pour tout $q \in [0, 1[$:

$$X_{\{q\}} = X + q - \mathbf{1}_{\{X \geq 1-q\}}. \tag{3}$$

2. En prenant, après l'avoir justifié, l'espérance dans (3), montrer que si la loi de X est invariante par translation alors $1 - X$ est de loi uniforme sur $[0, 1[$.
3. Soit F la fonction de répartition de X et $F_{\{q\}}$ celle de $X_{\{q\}}$ pour $q \in [0, 1[$. On suppose F continue. Montrer, en décomposant suivant les évènements $\{X < 1 - q\}$ et $\{X \geq 1 - q\}$ que :

$$F_{\{q\}}(x) = \begin{cases} F(x + 1 - q) - F(1 - q), & x \in [0, q], \\ F(x - q) + 1 - F(1 - q), & x \in [q, 1]. \end{cases}$$

4. Déduire des deux questions précédentes que la loi de X est invariante par translation si et seulement si X est de loi uniforme sur $[0, 1[$.

II Convergence en loi de la partie fractionnaire de sommes de variables aléatoires

Soit X une variable aléatoire réelle continue.

1. Soit $u \in \mathbb{R}^*$. Montrer que $\mathbb{P}(\cos(uX) = 1) = 0$. En déduire que $\psi_X(u) \neq 1$.
2. Déduire de la question précédente que $\psi_X(u) \neq \mathrm{e}^{i\theta}$ pour tout $\theta \in \mathbb{R}$ et $u \in \mathbb{R}^*$.
3. Montrer que $|\psi_X(u)| < 1$ pour tout $u \in \mathbb{R}^*$.

On suppose que X est à valeurs dans $[0, 1[$. Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires indépendantes de même loi que X. On note $S_n = \left\{\sum_{j=1}^n X_j\right\}$ la partie fractionnaire de la somme $\sum_{j=1}^n X_j$.

4. Montrer que pour $k \in \mathbb{Z}$, $\psi_X(2\pi k)^n = \psi_{S_n}(2\pi k)$.

Soit W une variable aléatoire à valeurs dans $[0, 1[$ telle que $\mathbb{P}(W = 0) = 0$. On admet qu'une suite de variables aléatoires $(W_n, n \in \mathbb{N}^*)$ à valeurs dans $[0, 1[$ converge en loi vers W si pour tout $k \in \mathbb{Z}$, $\lim_{n \to +\infty} \psi_{W_n}(2\pi k) = \psi_W(2\pi k)$.

5. Montrer que $(S_n, n \in \mathbb{N})$ converge en loi vers U de loi uniforme sur $[0, 1[$.

III Premier chiffre significatif d'une v.a. de loi invariante par multiplication

La notation scientifique de $y > 0$ correspond à la représentation $y = m\,10^\ell$ où le couple $(m, \ell) \in [1, 10[\times\mathbb{Z}$ est unique. Le nombre décimal m, que l'on notera $m(y)$ pour indiquer qu'il est une fonction de y, est appelé mantisse de y et sa partie entière, notée $c(y) = \lfloor m(y) \rfloor$, à valeurs dans $\{1, \dots, 9\}$ correspond au premier chiffre significatif de y, voir les exemples du tableau 1. Par construction on a $\log_{10}(m(y)) \in [0, 1[$, où pour $x > 0$ on note $\log_{10}(x) = \log(x)/\log(10)$ le logarithme de x en base 10.

y	0.00234	98.7	π
$\lfloor y \rfloor$	0	98	3
$\{y\}$	0.00234	0.7	0.14...
$m(y)$	2.34	9.87	π
$c(y)$	2	9	3

TABLE 1 – Partie entière $\lfloor y \rfloor$, partie fractionnaire $\{y\}$, mantisse $m(y)$ et premier chiffre significatif $c(y)$ de $y > 0$.

Soit U une variable aléatoire de loi uniforme sur $[0, 1[$. On pose $M = 10^U$ et $C = \lfloor M \rfloor$.

1. Montrer que C est de loi de Benford.
2. Pour $y > 0$ et $\theta > 0$, montrer que $\log_{10}(m(\theta y))$ est la partie fractionnaire de $\log_{10}(m(y)) + \log_{10}(m(\theta))$.
3. Soit Y une variable aléatoire strictement positive. Montrer, en utilisant la question I.4, que $m(\theta Y)$ a même loi que $m(Y)$ pour tout $\theta > 0$ si et seulement si $m(Y)$ a même loi que $M = 10^U$. En déduire que, dans ce cas, le premier chiffre significatif de Y, $c(Y)$, est de loi de Benford.

Soit V une variable aléatoire continue strictement positive. Soit $(V_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires indépendantes et de même loi que V. On note $Y_n = \prod_{k=1}^n V_n$, $M_n = m(Y_n)$ la mantisse de Y_n et $C_n = c(Y_n)$ son premier chiffre significatif.

4. (ADMIS) Montrer que $X = \log_{10}(m(V))$ est une variable aléatoire continue.
5. Vérifier que $\log_{10}(M_n)$ est la partie fractionnaire de $\sum_{k=1}^n \log_{10}(m(V_k))$.
6. Montrer à l'aide de la question II.5 que $(M_n, n \in \mathbb{N}^*)$ converge en loi vers $M = 10^U$.
7. Montrer que $(C_n, n \in \mathbb{N}^*)$ converge en loi vers C.

IV Données macro-économiques au sein de l'Union Européenne

On considère les données macro-économiques (en Euro) des pays de l'Union Européenne de l'Agence Européenne de Statistique [4] utilisées pour le calcul des critères de déficit des pays membres [5]

On dispose de la statistique du χ^2 d'adéquation à la loi de Benford du premier chiffre significatif des données macro-économiques pour les pays de la zone Euro pour chacune des années de 1999 à 2009.

1. Écrire la statistique du test du χ^2 d'adéquation à la loi de Benford lorsque l'on dispose de n données.

4. `http://ec.europa.eu/eurostat/home`
5. B. Rauch, M. Göttsche, G. Brähler and S. Engel. Fact and fiction in EU-governmental economic data. *German Economic Review*, vol. 12(3), pp. 243-255 (2011).

2. Donner le nombre de degré de liberté du test du χ^2 d'adéquation à la loi de Benford et la zone de rejet du test au niveau de α à l'aide du tableau 2 pour α égal à 5%.

d	8	9	10
$q = 95\%$	15.51	16.92	18.31
$q = 99\%$	20.09	21.67	23.21

Table 2 – Quantile d'ordre q de la loi du χ^2 à d degré de liberté.

3. On observe environ $n \simeq 140$ données par an pour chacun des pays de la zone Euro sur la période 1999-2009 (11 ans). Pour des raisons de forte variabilité annuelle, les auteurs de l'étude proposent de prendre la moyenne des statistiques de test du χ^2 qui sont calculées chaque année sur les 11 années pour chacun des pays, voir le tableau 3. Quelle est votre conclusion ?

pays	Grèce	Autriche	Allemagne	Italie	Espagne	France	Portugal	Pays Bas
χ^2	17.74	15.25	12.37	12.37	11.44	11.36	10.19	7.83

Table 3 – Statistique du χ^2 d'adéquation à la loi de Benford (moyenne sur 11 ans) pour quelques pays de la zone Euro.

△

Correction

Exercice 1 1. Pour $x = (x_1, \ldots, x_n)$ et $y = (y_1, \ldots, y_n)$, la vraisemblance est, en utilisant l'indépendance des variables aléatoires :

$$p_n(x, y; \theta) = (2\pi\sigma^2)^{-n} \exp\left(-\sum_{k=1}^n \frac{(x_k - \mu_k)^2}{2\sigma^2} - \sum_{k=1}^n \frac{(y_k - \mu_k)^2}{2\sigma^2} \right).$$

La log-vraisemblance est donc :

$$\ell_n(x, y; \theta) = -n \log(2\pi\sigma^2) - \sum_{k=1}^n \frac{(x_k - \mu_k)^2}{2\sigma^2} - \sum_{k=1}^n \frac{(y_k - \mu_k)^2}{2\sigma^2}.$$

2. On annule les dérivées. Pour $1 \leq k \leq n$, l'égalité $\partial\ell_n/\partial\mu_k = 0$ donne :

$$\frac{x_k - \mu_k}{\sigma^2} + \frac{y_k - \mu_k}{\sigma^2} = 0$$

soit $\mu_k = (x_k + y_k)/2$. Par ailleurs ℓ_n est un polynôme de degré 2 en μ_k dont le coefficient principal est strictement négatif. La log-vraisemblance est donc maximale en $\mu_k = (x_k + y_k)/2$. Elle est alors égale à :

$$\tilde{\ell}_n(x, y, \sigma^2) = -n \log(2\pi\sigma^2) - \frac{1}{4\sigma^2} \sum_{k=1}^n (x_k - y_k)^2.$$

On calcule la dérivée $\partial\tilde{\ell}_n/\partial\sigma^2$:

$$\frac{\partial\tilde{\ell}_n}{\partial\sigma^2}(x, y, \sigma^2) = -\frac{n}{\sigma^2} + \frac{1}{4\sigma^4} \sum_{k=1}^n (x_k - y_k)^2 = -\frac{n}{\sigma^4} \left(\sigma^2 - \frac{1}{4n} \sum_{k=1}^n (x_k - y_k)^2 \right).$$

Elle s'annule donc pour σ^2 égal à $v_n = \sum_{k=1}^n (x_k - y_k)^2/4n$. De plus $\partial\tilde{\ell}_n/\partial\sigma^2$ est négative (resp. positive) strictement pour $\sigma^2 < v_n$ (resp. $\sigma^2 > v_n$). On en déduit donc que $\tilde{\ell}_n$, et donc ℓ_n est maximal en $\sigma^2 = v_n$. On en déduit que l'estimateur du maximum de vraisemblance de μ_k, pour $1 \leq k \leq n$ est $\hat{\mu}_k = (X_k + Y_k)/2$ et que l'estimateur du maximum de vraisemblance de σ^2 est :

$$\hat{\sigma}_n^2 = \frac{1}{4n} \sum_{k=1}^n (X_k - Y_k)^2.$$

3. On a $\mathbb{E}[X_k - Y_k] = 0$ et, par indépendance, $\mathrm{Var}(X_k - Y_k) = \mathrm{Var}(X_k) + \mathrm{Var}(Y_k) = 2\sigma^2$. Comme X_k est indépendante de Y_k, on en déduit que (X_k, Y_k) est un vecteur gaussien et donc $X_k - Y_k$ est de loi gaussienne $\mathcal{N}(0, 2\sigma^2)$.

   Les variables aléatoires $(X_k - Y_k, k \in \mathbb{N}^*)$ sont indépendantes de même loi gaussienne $\mathcal{N}(0, 2\sigma^2)$ et intégrables. On déduit de la loi forte des grands nombres que la suite $(\hat{\sigma}_n^2, n \in \mathbb{N}^*)$ converge p.s. vers $\mathbb{E}[(X_1 - Y_1)^2]/4 = \sigma^2/2$. L'estimateur du maximum de vraisemblance de σ^2 n'est donc pas convergent.

4. On déduit de la réponse à la question précédente que l'estimateur $2\hat{\sigma}_n^2$ est un estimateur convergent de σ^2.

Exercice 2 **I Loi invariante par translation**

1. Soit $x \in [0,1[$ et $q \in [0,1[$. Si $x+q < 1$, alors $\{x+q\} = x+q$. Si $x+q \geq 1$, alors $1 \leq x+q < 2$ et donc $\{x+q\} = x+q-1$. On a donc :

$$x_{\{q\}} = (x+q)\mathbf{1}_{\{x+q<1\}} + (x+q-1)\mathbf{1}_{\{x+q\geq 1\}} = x+q-\mathbf{1}_{\{x+q\geq 1\}} = x+q-\mathbf{1}_{\{x\geq 1-q\}}.$$

   On en déduit le résultat.

2. Les variables aléatoires étant bornées elles sont intégrables. On en déduit :

$$\mathbb{E}[X_{\{q\}}] = \mathbb{E}[X] + q - \mathbb{P}(X \geq 1-q).$$

   Si $X_{\{q\}}$ a même loi que X alors elles ont même espérance. On déduit de la précédente égalité que $\mathbb{P}(1-X \leq q) = \mathbb{P}(X \leq 1-q) = q$, pour tout $q \in [0,1[$. Ceci assure que $1-X$ est de loi uniforme sur $[0,1[$.

3. Soit $x \in [0,1]$. On a :

$$\begin{aligned}
F_{\{q\}}(x) &= \mathbb{P}(X_{\{q\}} \leq x) \\
&= \mathbb{P}(X+q \leq x, X < 1-q) + \mathbb{P}(X+q-1 \leq x, X \geq 1-q) \\
&= \mathbb{P}(X \leq x-q) + \mathbb{P}(X \leq x+1-q) - \mathbb{P}(X < 1-q) \\
&= F(x-q)\mathbf{1}_{\{x\geq q\}} + F(x+1-q) - F(1-q) \\
&= (1+F(x-q))\mathbf{1}_{\{x\geq q\}} + F(x+1-q)\mathbf{1}_{\{x<q\}} - F(1-q),
\end{aligned}$$

   où l'on a utilisé la continuité de F pour la quatrième égalité et pour la cinquième égalité le fait que $F(y) = 1$ pour $y \geq 1$ car X est à valeurs dans $0,1[$.

4. Si X est de loi uniforme alors d'après la question précédente, on a $F_{\{q\}}(x) = x$ pour $x \in [0,1]$. Ceci assure que $X_{\{q\}}$ est de loi uniforme sur $[0,1[$ pour tout $q \in [0,1[$.

   Si la loi de X est invariante par translation, alors d'après la question I.2, $1-X$ est de loi uniforme sur $[0,1[$ et donc X est de loi uniforme sur $[0,1[$.

   Donc la loi de X est invariante par translation si et seulement si X est de loi uniforme sur $[0,1[$.

II Convergence en loi de la partie fractionnaire de sommes de variables aléatoires

1. Soit $u \neq 0$. On a $\{\cos(uX) = 1\} = \{X \in (2\pi/u)\mathbb{Z}\}$. Comme $(2\pi/u)\mathbb{Z}$ est dénombrable et que X est continue, on en déduit que $\mathbb{P}(\cos(uX) = 1) = 0$. Comme p.s. $\cos(uX) < 1$, on a $\mathbb{E}[\cos(uX)] < 1$. Comme $\psi_X(u) = \mathbb{E}[\cos(uX)] + i\mathbb{E}[\sin(uX)]$, on en déduit que $\psi_X(u) \neq 1$.
2. Soit $u \neq 0$. La variable aléatoire $X-(\theta/u)$ est continue donc d'après la question précédente, $\psi_X(u)\,\mathrm{e}^{-i\theta} = \psi_{X-(\theta/u)}(u) \neq 1$. On en déduit que $\psi_X(u) \neq \mathrm{e}^{i\theta}$.
3. Soit $u \neq 0$. Comme $|\psi_X(u)| \leq 1$ et $|\psi_X(u)| \neq 1$ d'après la question précédente, on en déduit que $|\psi_X(u)| < 1$.
4. Soit $k \in \mathbb{Z}$. On a, en utilisant l'indépendance et le fait que $\mathrm{e}^{i2\pi km} = 1$ pour $m \in \mathbb{Z}$:

$$\psi_X(2\pi k)^n = \mathbb{E}\left[\mathrm{e}^{i2\pi k\sum_{j=1}^n X_j}\right] = \mathbb{E}\left[\mathrm{e}^{i2\pi kS_n + i2\pi k\lfloor\sum_{j=1}^n X_j\rfloor}\right] = \mathbb{E}\left[\mathrm{e}^{i2\pi kS_n}\right] = \psi_{S_n}(2\pi k).$$

Soit W une variable aléatoire à valeurs dans $[0,1[$ telle que $\mathbb{P}(W = 0) = 0$. Montrons qu'une suite de variables aléatoires $(W_n, n \in \mathbb{N}^*)$ à valeurs dans $[0,1[$ converge en loi vers W si pour tout $k \in \mathbb{Z}$, $\lim_{n\to+\infty} \psi_{W_n}(2\pi k) = \psi_W(2\pi k)$.

Soit g une fonction définie sur $\mathbb{R}$ continue périodique de période 1. Pour tout $\varepsilon > 0$, il existe m et un polynôme trigonométrique d'ordre m, $P_\varepsilon(x) = \sum_{k=-m}^{m} a_k \, \mathrm{e}^{i2\pi kx}$ tel que $\sup_{x\in[0,1]} |g(x) - P_\varepsilon(x)| \leq \varepsilon$. On en déduit que :

$$|\mathbb{E}[g(W_n)] - \mathbb{E}[g(W)]| \leq 2\varepsilon + |\mathbb{E}[P_\varepsilon(W_n)] - \mathbb{E}[P_\varepsilon(W)]| \leq 2\varepsilon + \sum_{k=-m}^{m} |a_k||\psi_{W_n}(i2\pi k) - \psi_W(i2\pi k)|.$$

On en déduit que $\limsup_{n\to+\infty} |\mathbb{E}[g(W_n)] - \mathbb{E}[g(W)]| \leq 2\varepsilon$, et comme $\varepsilon > 0$ et arbitraire il vient $\lim_{n\to+\infty} |\mathbb{E}[g(W_n)] - \mathbb{E}[g(W)]| = 0$.

Pour terminer la démonstration, il faut étendre la convergence précédente à des fonctions g continues bornées et non forcément périodique. Soit g une fonction définie sur $\mathbb{R}$ continue bornée. On pose $c = g(1) - g(0)$. Pour $\varepsilon \in]0, 1/2[$, on définit la fonction continue h_ε sur $\mathbb{R}$ par :

$$h_\varepsilon(x) = \frac{c}{\varepsilon}|x - (1-\varepsilon)|\mathbf{1}_{\{|x-1|\leq\varepsilon\}}.$$

Ainsi la fonction $g + h_\varepsilon$ est une fonction continue sur $[0, 1]$ avec $(g + h_\varepsilon)(0) = (g + h_\varepsilon)(1)$. En particulier, elle peut être étendue sur $\mathbb{R}$ en une fonction continue périodique de période 1. On déduit de ce qui précède que :

$$\lim_{n\to+\infty} \mathbb{E}[(g + h_\varepsilon)(W_n)] = \mathbb{E}[(g + h_\varepsilon)(W)] = \mathbb{E}[g(W)] + \mathbb{E}[h_\varepsilon(W)].$$

La fonction h_ε définie sur $[1/2, 3/2]$ est continue et vérifie $h_\varepsilon(1/2) = h_\varepsilon(3/2) = 0$. Elle peut donc être étendue uniquement en une fonction $\tilde{h}_\varepsilon$ continue périodique de période 1 sur $\mathbb{R}$, telle que $h_\varepsilon = \tilde{h}_\varepsilon$ sur $[1/2, 3/2]$. On a également $h_\varepsilon \leq \tilde{h}_\varepsilon \leq |c|$. On en déduit :

$$\limsup_{n\to+\infty} \mathbb{E}[h_\varepsilon(W_n)] \leq \lim_{n\to+\infty} \mathbb{E}[\tilde{h}_\varepsilon(W_n)] = \mathbb{E}[\tilde{h}_\varepsilon(W)] \leq |c| \left(\mathbb{P}(W < \varepsilon) + \mathbb{P}(W > 1-\varepsilon)\right).$$

On a également $\mathbb{E}[h_\varepsilon(W)] \leq |c|\mathbb{P}(W > 1-\varepsilon)$. Comme $\mathbb{P}(W = 0) = 0$, on en déduit que pour tout $\eta > 0$, il existe $\varepsilon \in]0, 1/2[$ tel que $|c|(\mathbb{P}(W < \varepsilon) + \mathbb{P}(W > 1-\varepsilon)) < \eta$. Donc pour n assez grand :

$$|\mathbb{E}[g(W_n)] - \mathbb{E}[g(W)]| \leq \eta + \mathbb{E}[h_\varepsilon(W_n)] + \mathbb{E}[h_\varepsilon(W)] \leq 4\eta.$$

Ceci assure que $\lim_{n\to+\infty} \mathbb{E}[g(W_n)] = \mathbb{E}[g(W)]$ et donc $(W_n, n \in \mathbb{N}^*)$ converge en loi vers W.

5. On déduit des questions précédentes que pour $k \in \mathbb{Z}^*0$, $\lim_{n\to+\infty} \psi_{S_n}(2\pi k) = 0$ et $\psi_{S_n}(0) = 1$. On a également $\psi_U(0) = 1$ et pour $k \in \mathbb{Z}^*$:

$$\psi_U(2\pi k) = \int_0^1 \mathrm{e}^{iu2\pi k} \, du = \frac{\mathrm{e}^{i2\pi k} - 1}{2\pi k} = 0.$$

Comme $\mathbb{P}(U = 0) = 0$, on en déduit que $(S_n, n \in \mathbb{N}^*)$ converge en loi vers U.

III Premier chiffre significatif d'une v.a. de loi invariante par multiplication

1. On a pour $k \in \{1, \ldots, 9\}$:

$$\mathbb{P}(C = k) = \mathbb{P}(10^U \in [k, k+1[) = \mathbb{P}(U \in [\log_{10}(k), \log_{10}(k+1)[) = \log_{10}(k+1) - \log_{10}(k).$$

Donc C est de loi de Benford.

2. Soit $y > 0$ et $\theta > 0$. On remarque que $m(y)m(\theta) \in [1, 100[$ et donc que :

$$m(\theta y) = \begin{cases} m(\theta)m(y) & \text{si } m(\theta)m(y) < 10, \\ m(\theta)m(y)/10 & \text{si } m(\theta)m(y) \geq 10. \end{cases}$$

On pose $q = \log_{10}(m(\theta))$ et $x = \log_{10}(m(y))$. Il vient :

$$\log_{10}(m(\theta y)) = \begin{cases} x + q & \text{si } x + q < 1, \\ x + q - 1 & \text{si } x + q \geq 1. \end{cases}$$

Autrement dit $\log_{10}(m(\theta y)) = \{x + q\}$.

3. On pose $X = \log_{10}(m(Y))$ et $q = \log_{10}(m(\theta)) \in [0, 1[$. D'après la question précédente, on a $\log_{10}(m(\theta Y)) = X_{\{q\}}$. On en déduit que $m(Y)$ a même loi que $m(\theta Y)$ pour tout $\theta > 0$ si et seulement si X a même loi que $X_{\{q\}}$ pour tout $q \in [0, 1[$. On déduit de la question I.4 que X a même loi que U et donc $m(Y)$ a même loi que $M = 10^U$.
4. Soit f la densité de V. Pour $0 \leq a < b < 1$, on a :

$$\begin{aligned} \mathbb{P}(X \in [a, b]) = \mathbb{P}(m(V) \in [10^a, 10^b]) &= \sum_{\ell \in \mathbb{N}} \mathbb{P}(V \in [10^{a+\ell}, 10^{b+\ell}]) \\ &= \int \sum_{\ell \in \mathbb{N}} \mathbf{1}_{[10^{a+\ell}, 10^{b+\ell}]}(v) f(v)\, dv \\ &= \int \sum_{\ell \in \mathbb{N}} \mathbf{1}_{[a+\ell, b+\ell]}(x) \log(10) 10^x\, f(10^x)\, dx \\ &= \int \log(10) \mathbf{1}_{[a,b]}(x) \sum_{\ell \in \mathbb{N}} 10^{x+\ell}\, f(10^{x+\ell})\, dx. \end{aligned}$$

Donc X est une variable aléatoire continue de densité $\log(10)\mathbf{1}_{[0,1[}(x) \sum_{\ell \in \mathbb{N}} 10^{x+\ell}\, f(10^{x+\ell})$.
5. Cela découle directement de la question III.2. On déduit de la question II.5 que la suite $(\log_{10}(M_n), n \in \mathbb{N}^*)$ converge en loi vers $\log_{10}(M)$. Comme la fonction $x \mapsto 10^x$ est continue, on en déduit que $(M_n, n \in \mathbb{N}^*)$ converge en loi vers M.
6. L'ensemble des points de discontinuité de la fonction $x \mapsto \lfloor x \rfloor$ est $\mathbb{Z}$. Comme M est une variable aléatoire continue (de densité $\mathbf{1}_{[1,10[}(v)/(v \log(10))$) on a $\mathbb{P}(M \in \mathbb{Z}) = 0$. Donc on en déduit que $(C_n = \lfloor M_n \rfloor, n \in \mathbb{N}^*)$ converge en loi vers $C = \lfloor M \rfloor$.

IV Données macro-économiques de la Grèce

1. Les statistiques du χ^2 sont données par :

$$\zeta_n^{(1)} = n \sum_{k=1}^{9} \frac{(\hat{p}_k - p_k)^2}{\hat{p}_k} \quad \text{et} \quad \zeta_n^{(2)} = n \sum_{i=1}^{9} \frac{(\hat{p}_k - p_k)^2}{p_k},$$

où $p_k = \mathbb{P}(C = k)$ et $\hat{p}_k$ est le nombre d'occurrences de k parmi les n observations. Sous l'hypothèse nulle, H_0, les premiers chiffres significatifs sont indépendant et suivent la loi de Benford, et sous l'hypothèse alternative, H_1, les premiers chiffres significatifs sont indépendant et ne suivent pas la loi de Benford.

2. Le nombre de degré est 8 et la zone de rejet (pour les deux statistiques de test $\zeta_n^{(1)}$ et $\zeta_n^{(2)}$) est $[z, +\infty[$ avec z le quantile d'ordre $1 - \alpha$ de la loi $\chi^2(8)$ soit $z \simeq 15.51$.

3. La loi du premier chiffre significatif de données macro-économique est invariante par changement de numéraire, c'est donc la loi de Benford. Si on suppose que les données sont des réalisations de variables aléatoires indépendantes, alors on rejette H_0 pour la Grèce. Ceci peut suggérer que les données macro-économique de la Grèce ne sont pas fiables.