
Examen MAO Probabilités et Statistiques (15 juin 2021, 9h30-12h)

Lorsque l'énoncé vous demande de **coder**, vous avez le choix entre :

1. Écrire du code Python. Dans ce cas on pourra supposer qu'on a appelé les modules habituels :

```
import numpy as np
import scipy as sp
import matplotlib.pyplot as plt
import numpy.random as rnd
import scipy.stats as sts
```

Vous ne perdrez pas de point si vous vous trompez dans les noms des fonctions et leurs options.

2. Écrire du pseudo-code, par exemple :

```

FONCTION comptage(p)
  n = 0
  POUR k de 1 à 10
    u = Uniforme([0, p])
    SI u < p
      n = n + 1
  RENVOYER n
```

Exercice 1.— Test d'exponentialité

On dispose d'un n -échantillon X_1, \dots, X_n (indépendantes et de même loi) à valeurs dans $]0, \infty[$. On souhaite tester l'hypothèse H_0 : « il existe $\lambda > 0$ tel que la loi des X_i est une exponentielle de paramètre λ », contre l'hypothèse H_1 : « la loi des X_i est intégrable mais n'est pas une loi exponentielle ».

1. Sous l'hypothèse H_0 , donner l'estimateur du maximum de vraisemblance pour λ , que l'on note $\hat{\lambda}_n$. Justifier qu'il est fortement consistant.

Correction.— La vraisemblance s'écrit

$$V(\lambda) = \lambda^n \prod_{i=1}^n \exp(-\lambda X_i).$$

Son log vaut $n \log \lambda - n \bar{X}_n \lambda$, qui est maximal en $\hat{\lambda}_n = \frac{1}{\bar{X}_n}$.

Cet estimateur est fortement consistant car la loi des grands nombres implique que \bar{X}_n converge p.s. vers $\mathbb{E}[X_1] = \frac{1}{\lambda}$.

2. On considère la fonction de répartition empirique des X_i , notée $\hat{F}_X(t)$, ainsi que la fonction de répartition d'une loi exponentielle de paramètre $\hat{\lambda}_n$, notée $F_{\hat{\lambda}_n}(t)$. Écrire ces quantités en fonction des X_i .

Correction.— Pour $t \geq 0$, $\hat{F}_X(t) = \frac{1}{n} \sum 1_{X_i \geq t}$ et $F_{\hat{\lambda}_n}(t) = 1 - \exp(-t/\bar{X}_n)$

3. On définit

$$D_n = \sup_{t \geq 0} \left| \hat{F}_X(t) - F_{\hat{\lambda}_n}(t) \right|.$$

Montrer que sous l'hypothèse H_0 , la loi de D_n ne dépend pas de λ . On pourra utiliser le fait que λX_i suit une loi exponentielle de paramètre 1, et faire un changement de variable sur t .

Correction.—

$$\begin{aligned} D_n &= \sup_{t \geq 0} \left| \frac{1}{n} \sum 1_{X_i \geq t} - \left(1 - \exp\left(-\frac{t}{\bar{X}_n}\right) \right) \right| \\ &= \sup_{t \geq 0} \left| \frac{1}{n} \sum 1_{\lambda X_i \geq \lambda t} - \left(1 - \exp\left(-\frac{\lambda t}{\lambda \bar{X}_n}\right) \right) \right| \\ &= \sup_{u \geq 0} \left| \frac{1}{n} \sum 1_{Y_i \geq u} - \left(1 - \exp\left(-\frac{u}{\bar{Y}_n}\right) \right) \right| \end{aligned}$$

où l'on a posé $u = \lambda t$ et $Y_i = \lambda X_i$. On reconnaît alors la quantité D_n écrite pour les Y_i , qui sont des variables exponentielles de paramètre 1. La dernière quantité a une loi qui ne dépend pas de λ .

4. Supposons que l'on connaisse les quantiles $(q_{n,\beta})_{\beta \in]0,1[}$ de la loi de D_n . Soit $\alpha \in]0,1[$, proposer un test de niveau α de l'hypothèse H_0 contre H_1 .

Correction.— On peut prendre, par exemple, $\phi(X_1, \dots, X_n) = 1_{D_n \notin [0, q_{n,1-\alpha}]}$

5. Pour $\beta \in]0,1[$ fixé, proposer une manière d'estimer le quantile $q_{n,\beta}$. On ne cherchera pas à prouver la consistance de cet estimateur.

Correction.— On considère $(Y_i^{(j)})_{1 \leq i \leq n, 1 \leq j \leq N}$ iid exponentielles de paramètre 1. Pour chaque j , on calcule $D_n^{(j)}$ obtenue à partir des $(Y_i^{(j)})_{1 \leq i \leq n}$. On dispose ainsi de N réalisations indépendantes de la loi demandée. On les trie par ordre croissant : $D_n^{[1]} \leq D_n^{[2]} \leq \dots \leq D_n^{[N]}$, et on renvoie $\hat{q}_{n,\beta} = D_n^{[\beta N]}$.

Il est possible de montrer que cet estimateur est fortement consistant (voir Rivoirard et Stoltz, section 8.5). Établir un intervalle de confiance est toutefois plus difficile...

6. Pour une certaine suite a_n qui tend vers l'infini, on définit $Z_n = a_n D_n$. Comment choisir a_n pour espérer avoir une convergence en loi de Z_n sous H_0 ? On ne demande pas une preuve concrète mais des justifications convaincantes.

Quel est alors le comportement de Z_n sous H_1 ?

Correction.— On propose $a_n = \sqrt{n}$. En effet, sous H_0 , par le théorème de Kolmogorov-Smirnov, $\sqrt{n} \|\hat{F}_X - F_\lambda\|_\infty$ converge en loi. Cela indique que $\|\hat{F}_X - F_\lambda\|_\infty$ est d'ordre $\frac{1}{\sqrt{n}}$; par inégalité triangulaire, si on montre que $\|F_\lambda - F_{\hat{\lambda}_n}\|_\infty$ est du même ordre, on aura une justification convaincante.

Avec un peu d'analyse, on montre que $\|F_\lambda - F_{\hat{\lambda}_n}\|_\infty \leq \frac{|\lambda - \hat{\lambda}_n|}{\lambda}$. Par le théorème central limite, $\sqrt{n} \left(\frac{1}{\hat{\lambda}_n} - \frac{1}{\lambda} \right) \rightarrow_{\mathcal{L}} \mathcal{N}(0, 1/\lambda^2)$, d'où $\sqrt{n} \frac{\lambda - \hat{\lambda}_n}{\lambda_n} \rightarrow \mathcal{N}(0, 1)$. Comme $\hat{\lambda}_n$ converge vers λ (p.s.), par Slutsky $\sqrt{n} \frac{\lambda - \hat{\lambda}_n}{\lambda} \rightarrow \mathcal{N}(0, 1)$. Ainsi, $\|F_\lambda - F_{\hat{\lambda}_n}\|_\infty$ semble être d'ordre $\frac{1}{\sqrt{n}}$.

Dans ce cas, sous H_1 , par hypothèse d'intégrabilité $\hat{\lambda}_n$ converge p.s. vers $\frac{1}{\mathbb{E}[X_1]}$, donc pour tout $t > 0$, $F_{\hat{\lambda}_n}(t)$ converge p.s. vers $1 - \exp(-t/\mathbb{E}[X_1])$. D'autre part, par Glivenko-Cantelli (ou juste par la loi des grands nombres), $\hat{F}_X(t)$ converge p.s. vers $F(t)$ (la fonction de répartition de la loi de X_i en t). Comme cette loi n'est pas exponentielle, il existe t tel que $F(t) \neq 1 - \exp(-t/\mathbb{E}[X_1])$, et donc p.s. pour tout n assez grand, $D_n \geq \frac{1}{2} |F(t) - (1 - t/\mathbb{E}[X_1])| > 0$. Cela implique que $Z_n \rightarrow \infty$ p.s. sous l'hypothèse H_1 .

Exercice 2.— Loi de Poisson conditionnée

Soit un réel $\lambda > 0$ et deux entiers naturels $a < b$. On souhaite simuler la loi de Poisson de paramètre λ conditionnée à être dans $[a, b]$. Autrement dit, on veut simuler une variable Y telle que, pour tout $k \in \mathbb{N} \cap [a, b]$,

$$\mathbb{P}(Y = k) = \mathbb{P}(X = k \mid a \leq X \leq b)$$

où X suit une loi de Poisson de paramètre λ .

On rappelle une propriété vue en cours : si $(U_n)_{n \geq 1}$ est une suite iid de variables uniformes sur $[0, 1]$, alors la variable aléatoire

$$T = \inf\{n \geq 0 \mid U_1 \times \dots \times U_{n+1} < e^{-\lambda}\} \quad (1)$$

est finie p.s. et suit une loi de Poisson de paramètre λ .

1. **Coder** une fonction qui simule une variable de Poisson de paramètre λ en utilisant la formule (1).

Correction.—

```
def poisson(lam):
    n = 0
    prod = rnd.random()
    while prod >= np.exp(-lam):
        prod *= rnd.random()
        n += 1
    return n
```

2. À l'aide de la méthode du rejet, **coder** une fonction qui simule la variable aléatoire Y .

Correction.—

```
def varY(lam,a,b):
    while True:
        x=poisson(lam)
        if a <= x and x <= b:
            return x
```

On suppose désormais et jusqu'à la fin du problème que $\lambda \leq a$ et $2\lambda \leq b$.

3. On admet le résultat suivant : si X suit une loi de Poisson de paramètre λ , et si $n \geq \lambda$,

$$\mathbb{P}(X \geq n) \leq e^{-\lambda} \left(\frac{\lambda e}{n} \right)^n. \quad (2)$$

On appelle *complexité* de l'algorithme précédent l'espérance E_1 du nombre de d'appels à la fonction de la question 1. Montrer que celle-ci vérifie

$$E_1 \geq e^\lambda \left(\frac{a}{\lambda e} \right)^a.$$

Correction.— Le nombre d'appels est une variable géométrique de paramètre $\mathbb{P}(a \leq X \leq b)$, donc son espérance est $1/\mathbb{P}(a \leq X \leq b) \geq 1/\mathbb{P}(a \leq X) \geq e^\lambda \left(\frac{a}{\lambda e} \right)^a$.

On voit que cette complexité est plus grande qu'une exponentielle en a ; l'algorithme devient donc inutilisable si a devient trop grand. On va donc tenter d'utiliser un algorithme de Metropolis-Hastings.

On considère l'espace $E = [a, b] \cap \mathbb{N}$, et on note π la loi de Y . On cherche à construire une chaîne de Markov qui admette π comme unique loi invariante.

On commence avec la matrice stochastique

$$Q = \begin{pmatrix} 0 & 1 & & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & & \\ & \frac{1}{2} & 0 & \frac{1}{2} & & \\ & & \ddots & \ddots & \ddots & \\ & & & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & & & 1 & 0 \end{pmatrix}$$

où la première ligne/colonne représente l'état a , la deuxième l'état $a+1$, etc., la dernière l'état b . Les coefficients non représentés valent 0.

4. Montrer que cette matrice vérifie les hypothèses de l'algorithme de Metropolis-Hastings, et que celui-ci donne (dans les notations du cours)

$$\begin{aligned} \forall x \in [a, b-1] \cap \mathbb{N}, \quad \alpha(x, x+1) &= \min \left(1, \frac{\lambda Q(x+1, x)}{(x+1)Q(x, x+1)} \right) \\ \alpha(x+1, x) &= \min \left(1, \frac{(x+1)Q(x, x+1)}{\lambda Q(x+1, x)} \right). \end{aligned}$$

Correction.— On constate que $Q(x, y) \neq 0 \Leftrightarrow Q(y, x) \neq 0 (\Leftrightarrow y = x \pm 1)$. De plus, $\pi(x) = \frac{1}{\mathbb{P}(a \leq X \leq b)} e^{-\lambda} \frac{x^\lambda}{x!}$. On en déduit $\frac{\pi(x)}{\pi(x+1)} = \frac{x+1}{\lambda}$. Ensuite il s'agit juste de la formule du cours pour $\alpha \dots$

5. En déduire que la matrice de transition P de la chaîne de Markov de Metropolis-Hastings s'écrit

$$P = \begin{pmatrix} * & \frac{\lambda}{2(a+1)} & & & & & \\ \frac{1}{2} & * & \frac{\lambda}{2(a+2)} & & & & \\ & \frac{1}{2} & * & \frac{\lambda}{2(a+3)} & & & \\ & & & \ddots & & & \\ & & & \frac{1}{2} & * & \frac{\lambda}{2(b-1)} & \\ & & & & \frac{1}{2} & * & \frac{\lambda}{b} \\ & & & & & 1 & 0 \end{pmatrix}.$$

On ne cherchera pas à exprimer les coefficients diagonaux indiqués par une $*$.

Correction.— On distingue les cas : pour $x = a$, on a $\alpha(a, a+1) = \min(1, \frac{\lambda \times 1/2}{(a+1) \times 1}) = \min(1, \frac{\lambda}{2(a+1)}) = \frac{\lambda}{2(a+1)}$, d'où $P(a, a+1) = \alpha(a, a+1) \times Q(a, a+1) = \frac{\lambda}{2(a+1)}$.

Pour $x \in [a+1, b-2] \cap \mathbb{N}$, $\alpha(x, x+1) = \min(1, \frac{\lambda}{x+1}) = \frac{\lambda}{x+1}$ (car $\lambda \leq a \leq x+1$), d'où $P(x, x+1) = \alpha(x, x+1)Q(x, x+1) = \frac{\lambda}{2(x+1)}$.

Pour $x = b-1$, $\alpha(b-1, b) = \min(1, \frac{2\lambda}{b-1+1}) = \frac{2\lambda}{b}$ (car $2\lambda \leq b$), d'où $P(b-1, b) = \alpha(b-1, b)Q(b-1, b) = \frac{\lambda}{b}$.

On procède de même pour les entrées $P(x+1, x)$.

6. **Coder** une fonction qui prend en entrée a, b, λ, Y_0, n et renvoie une simulation de la chaîne de Markov (Y_0, Y_1, \dots, Y_n) de matrice de transition P .

Correction.—

```
def transition_P(x,a,b,lam):
    # effectue une transition de la chaîne de Markov depuis l'état x
    if x==b:
        return b-1
    u = rnd.random()
    if x==b-1:
        if u < lam/b:
            return b
        if u < lam/b + 1/2:
            return b-2
        return b-1
    if u < lam / (2*(x+1)):
        return x+1
    if x==a:
        return a
    if u < 1/2 + lam / (2*(x+1)):
        return x-1
    return x

def markov(a,b,lam,Y0,n):
    chaine = [Y0]
    Y=Y0
    for _ in range(n):
        Y=transition_P(Y,a,b,lam)
        chaine.append(Y)
    return chaine
```

7. Justifier que la chaîne de Markov de matrice de transition P est irréductible et apériodique.

Correction.— Il est clair que la chaîne est irréductible, par exemple, pour tous $x, y \in E$, partant de x , avec probabilité positive on peut aller en a en $(x-a)$ étapes, puis aller en y en $(y-a)$ étapes. Pour l'apériodicité, il suffit de voir que pour un certain x , $P(x, x) > 0$. C'est le cas pour $x = a$ par exemple (car $P(a, a) = \frac{2(a+1)-\lambda}{2(a+1)} > 0$ puisque $\lambda \leq a$).

8. On note $\text{Sp}(P)$ l'ensemble des valeurs propres de P , ainsi que

$$\rho := \sup \{|z|, z \in \text{Sp}(P) \setminus \{1\}\}.$$

Rappeler pourquoi $\rho < 1$.

Correction.— C'est le cas pour toute chaîne de Markov à espace d'états fini irréductible et apériodique (et c'est une conséquence du théorème de Perron-Frobenius).

9. On note ν_n la loi de Y_n , que l'on identifie à un vecteur ligne de longueur $b - a + 1$. On prend $\|\cdot\|$ une norme sur \mathbb{R}^{b-a+1} . Montrer que pour toute loi initiale ν_0 , il existe $C > 0$ tel que

$$\|\nu_n - \pi\| \leq C\rho^n.$$

Correction.— Attention, ici on parle de mesures et non de fonctions sur E . Pour voir comment elles évoluent dans la chaîne de Markov, il faut appliquer tP (et non P comme on l'a fait dans le cours pour des fonctions sur l'espace d'état). Une autre façon de le dire est que les mesures sont des vecteurs **ligne** et que, par exemple, $\nu_1 = \nu_0 P$. On va donc adapter les arguments du cours.

P étant π -réversible, elle est autoadjointe pour le produit scalaire $\langle \cdot, \cdot \rangle_\pi$ (voir le cours), et donc diagonalisable. C'est donc aussi le cas de tP . Notons ${}^t\pi, {}^t\pi_1, \dots, {}^t\pi_{d-1}$ une base de vecteurs propres de tP (autrement dit les π_i sont des vecteurs ligne); on rappelle que π est associé à la valeur propre 1, et que celle-ci est de multiplicité 1. On note aussi $1, \lambda_1, \dots, \lambda_{d-1}$ les valeurs propres associées. On décompose alors ν_0 sur cette base :

$$\nu_0 = a\pi + \sum_{i=1}^{d-1} a_i \pi_i.$$

Alors

$$\nu_n = \nu_0 P^n = a\pi + \sum_{i=1}^{d-1} a_i \lambda_i^n \pi_i.$$

Comme $\lambda_i \in]-1, 1[$, cette quantité tend vers $a\pi$. Comme de plus la somme des composantes vaut 1, nécessairement $a = 1$. On en déduit

$$\|\nu_n - \pi\| = \left\| \sum_{i=1}^{d-1} a_i \lambda_i^n \pi_i \right\| \leq \sum_{i=1}^{d-1} |a_i| |\lambda_i|^n \|\pi_i\| \leq \rho^n \sum_{i=1}^{d-1} |a_i| \|\pi_i\|.$$

10. On admet qu'il existe une constante $c > 0$ telle que $\rho \leq 1 - \frac{c}{(b-a)\mathbb{P}(Y=a)}$. Comparer les avantages et les inconvénients des deux méthodes de simulation de Y proposées.

Correction.— Pour l'algorithme markovien, on voit que $\|\nu_n - \pi\| \leq C\rho^n \leq C \left(1 - \frac{c}{(b-a)\mathbb{P}(Y=a)}\right)^n$. Si $(b-a)$ est grand¹, cette quantité est proche de $C \exp(-\frac{cn}{(b-a)\mathbb{P}(Y=a)}) \leq C \exp(-\frac{cn}{b-a})$, et donc la chaîne converge en un temps d'ordre $b-a$. Dans la plupart des cas, c'est bien mieux que l'algorithme précédent (à moins que $b-a$ soit lui-même d'ordre $e^\lambda \left(\frac{a}{\lambda e}\right)^a \dots$ mais si on ne tombe pas dans cet extrême, tout va bien). Cependant, l'algorithme de rejet fournit une simulation exacte de π , là où l'algorithme markovien ne donne qu'une approximation de la mesure. On pourrait utiliser un algorithme de couplage par le passé monotone pour palier ce défaut...

Exercice 3.— BONUS : Finir en queue de Poisson (à faire uniquement si vous avez essayé toutes les questions précédentes)

Démontrer l'inégalité (2). On pourra utiliser la propriété (1), puis appliquer une inégalité de Markov après avoir appliqué une fonction bien choisie à $U_1 \times \dots \times U_n \dots$

1. Si $b-a$ est petit, on n'a en fait pas besoin de l'algorithme de Metropolis-Hastings et on peut juste lister les $b-a+1$ valeurs possible et leur probabilité.

Correction.— Par la formule (1), et pour tout $t > 0$,

$$\begin{aligned}
 \mathbb{P}(X \geq n) &= \mathbb{P}\left(U_1 \times \cdots \times U_{n+1} \geq e^{-\lambda}\right) \\
 &\leq \mathbb{P}\left(U_1 \times \cdots \times U_n \geq e^{-\lambda}\right) \\
 &\leq \mathbb{P}\left((U_1 \times \cdots \times U_n)^t \geq e^{-t\lambda}\right) \\
 &\leq e^{t\lambda} \mathbb{E}\left[(U_1 \times \cdots \times U_n)^t\right] \\
 &\leq e^{t\lambda} \mathbb{E}[U_1^t]^n \\
 &\leq e^{t\lambda} \left(\frac{1}{t+1}\right)^n
 \end{aligned}$$

où on a calculé directement l'espérance de U_1^t ... On optimise sur t , on voit en dérivant que c'est minimal pour $t = \frac{n}{\lambda} - 1$ (qui est > 0 pour $n > \lambda$; si $n = \lambda$ l'inégalité (2) est immédiate). En appliquant en ce t , on obtient (2).