
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.
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 .
3. On définit

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

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 .

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 .
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.
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 ?

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

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

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

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

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

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)

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

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

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

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

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 .
7. Justifier que la chaîne de Markov de matrice de transition P est irréductible et apériodique.
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 $\rho < 1$.

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

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

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.

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 $U_1 \times \dots \times U_n \dots$