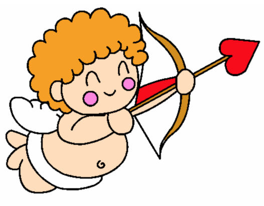
Math ECG 2. 2023-2024

Mathématiques Appliquées - F. Gaunard
http://frederic.gaunard.com
ENC Bessières, Paris 17e.



Devoir Maison n°4

Solution disponible le 21/02

Exercice 1

Une urne contient 2 boules : une blanche et une rouge. On effectue dans cette urne une succession de tirages en respectant la règle suivante : si la boule tirée est blanche, on l'enlève du jeu et on ajoute une boule rouge dans l'urne. Si la boule tirée est rouge, on l'enlève du jeu et on ajoute une boule blanche dans l'urne.

Pour tout $i \in \mathbb{N}$, on note N_i le nombre de boules blanches contenues dans l'urne après le tirage i.

(1) Que vaut, pour tout $i \in \mathbb{N}$, $N_i(\Omega)$?
(2) Justifier que N_i est une chaîne de Markov. Représenter alors, en justifiant rigoureusement, son graphe de transition et préciser la matrice de transition A associée.
(3) Écrire une fonction Python d'en-tête `def simul_N(i)` qui calcule et renvoie une simulation de N_i.
(4) Déterminer, s'il existe, l'état stable de la chaîne.
(5) Calculer A^2 puis A^3. En déduire l'expression, pour tout entier $i \in \mathbb{N}^*$, de A^i.
(6) Déterminer la loi, pour tout $i \in \mathbb{N}$, de N_i.
(7) La chaîne $(N_i)_{i\geq 0}$ est-elle convergente ?

Exercice 2

Soit X une variable aléatoire.
On appelle *médiane* de X tout réel m qui vérifie les deux conditions :

$$P(X \leq m) \geq \frac{1}{2} \text{ et } P(X \geq m) \geq \frac{1}{2}.$$

On suppose que X suit la loi exponentielle de paramètre $\lambda > 0$.

(1) (a) Montrer que X admet une unique médiane m que l'on calculera.
  (b) Soit M la fonction définie sur $\mathbb{R}$, à valeurs réelles, telle que : $\forall x \in \mathbb{R}$, $M(x) = E(|X - x|)$.
  Étudier les variations de la fonction M sur $\mathbb{R}$ et montrer que m est l'unique point en lequel M atteint son minimum.

(2) On suppose que le paramètre λ est inconnu. Soit α un réel vérifiant $0 < \alpha < 1$.
Pour n entier de $\mathbb{N}^*$, soit $(X_1, X_2, \ldots, X_n)$ un n-échantillon de variables aléatoires indépendantes et de même loi que X. On pose pour tout $n \in \mathbb{N}^*$: $Z_n = \min(X_1, X_2, \ldots, X_n)$.

  (a) Quelle est la loi de Z_n?
  (b) Établir l'existence de deux réels c et d tels que :

$$P\left(Z_n \leq \frac{c}{\lambda}\right) = \frac{\alpha}{2} \quad \text{et} \quad P\left(Z_n \geq \frac{d}{\lambda}\right) = \frac{\alpha}{2}.$$

  (c) En déduire un intervalle de confiance du paramètre m au niveau de confiance $1 - \alpha$.

Exercice 3

Soient $n \in \mathbb{N}^*$ et $p \in]0;1[$. On considère une variable aléatoire $N \hookrightarrow \mathcal{B}(n,p)$ et une suite $(U_k)_{k\geq 0}$ de variables aléatoires indépendantes de même loi $\mathcal{U}([0;1])$. On pose alors

$$T_n = \begin{cases} \min(U_1, U_2, ...U_k), & \text{si } [N=k] \text{ est réalisé avec } k \neq 0 \\ U_0, & \text{si } [N=0] \text{ est réalisé} \end{cases}$$

Étudier la convergence en loi de $(T_n)_{n\geq 0}$.
On observera que le n dont dépend T_n intervient dans la loi de N.

Problème*

Ce problème n'est pas facile. Il s'adresse aux étudiant.e.s qui préparent les épreuves du Top 3.
Soit $N \in \mathbb{N}^*$ un entier fixé. On note, dans tout ce problème, $E = [\![1, N]\!]$.

Partie 1 : Préliminaires

(1) Montrer que, pour tout $x > -1$, on a $\ln(1+x) \leq x$ avec égalité si et seulement si $x = 0$.
(2) Soient $(p_n)_{n\in E}$ et $(q_n)_{n\in E}$ deux distributions de probabilités. À l'aide de l'inégalité précédente, montrer que

$$\sum_{n=1}^{N} p_n \ln(q_n) \leq \sum_{n=1}^{N} p_n \ln(p_n)$$

et que l'égalité a lieu si et seulement si, pour tout $n \in E$, on a $p_n = q_n$.

Partie 2 : Estimateur du maximum de vraisemblance pour une matrice de transition

On considère une chaine de Markov $(X_k)_{k\geq 0}$ à N états, dont la loi initiale X_0 est connue et dont on cherche à *estimer* les coefficients de la matrice de transition $A = (p_{i,j})_{(i,j)\in E^2}$.

Notamment, on a, pour tout $k \in \mathbb{N}$, $X_k(\Omega) = E$ et, pour tout $(i,j) \in E^2$,

$$p_{i,j} = P_{[X_k=i]}(X_{k+1} = j) = p_{i,j}.$$

On suppose qu'on dispose alors d'une trajectoire de longueur $n+1$ de la chaîne (avec $n \in \mathbb{N}^*$), c'est dire des variables $(X_0, X_1, ..., X_n)$.
Attention, contrairement au contexte classique des problèmes d'estimation, il ne s'agit pas d'un $n-$échantillon d'une même variable aléatoire.
Considérant alors une observation arbitraire $(x_1, ..., x_n) \in E^n$ des n derniers termes de cette trajectoire et une observation initiale $x_0 \in E$, on introduit la **fonction de vraisemblance** (conditionnelle à $X_0 = x_0$) comme la fonction L définie sur l'ensemble des matrices stochastiques de taille N par

$$L(A) = \mathrm{P}_{[X_0=x_0]}\left([X_1 = x_1] \cap [X_2 = x_2] \cap ... \cap [X_n = x_n]\right).$$

On observera que la matrice A contient les probabilités de transition de la chaîne et que la fonction L dépend bien des coefficients de A.
On note également, pour $(i,j) \in E^2$,

$$N_n(i,j) = \mathrm{Card}\left(\{k \in [\![0, n-1]\!] : \ X_k = i \text{ et } X_{k+1} = j\}\right)$$

et

$$M_n(i) = \mathrm{Card}\left(\{k \in [\![0, n-1]\!] : \ X_k = i\}\right).$$

(3) Expliquer ce que représentent les quantités $N_n(i,j)$ et $M_n(i)$.
(4) Justifier que, pour tout $i \in E$, on a :

$$\sum_{j=1}^{N-1} N_n(i,j) = M_n(i).$$

(5) Montrer, à l'aide de la convention $0^0 = 1$, que,

$$L(A) = \prod_{i=1}^{N} \prod_{j=1}^{N} p_{i,j}^{N_n(i,j)}$$

puis que

$$\ln(L(A)) = \sum_{i=1}^{N} \sum_{j=1}^{N} N_n(i,j) \ln(p_{i,j}).$$

(6) Montrer, à l'aide de la Question (2), que, pour tout $i \in E$ tel que $M_n(i) \neq 0$, on a :

$$\sum_{j=1}^{N} \frac{N_n(i,j)}{M_n(i)} \ln(p_{i,j}) \leq \sum_{j=1}^{N} \frac{N_n(i,j)}{M_n(i)} \ln\left(\frac{N_n(i,j)}{M_n(i)}\right).$$

On note alors, pour tout couple $(i,j) \in E^2$, $\hat{p}_{i,j} = \begin{cases} \dfrac{N_n(i,j)}{M_n(i)}, & \text{si } M_n(i) \neq 0 \\ 0, & \text{sinon} \end{cases}$ et $\hat{A} = (\hat{p}_{i,j})_{(i,j)\in E^2}$.

(7) Montrer alors que, pour toute matrice A stochastique de $\mathcal{M}_N(\mathbb{R})$, on a

$$L(A) \leq L(\hat{A}).$$

L'estimateur $\hat{A}$ est alors l'estimateur du maximum de vraisemblance pour la matrice de transition de la chaîne de Markov.

(8) Python.

(a) Recopier et compléter la fonction ci-dessous qui prend en argument le nombre d'états N de la chaîne et la trajectoire observée `X=`$[x_0, x_1, ..., x_n]$ et renvoie une estimation de A.

```
def EMV_A(N, X) :
    n=len(X)
    A=np.zeros((N,N))
    M=np.zeros(N)
    for k in range(n-1):
        A[....][....]+=1
    for i in range(N) :
        M[i]=....
        if M[i] !=0 :
            A[i]=[A[i][j]/M[i] for j in range(N)]
    return A
```

(b) Que devrait raisonnablement renvoyer la commande `EMV_A(3, rd.randint(1, 4, 10000))`?