Math ECG 2. 2023-2024

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

Quinzaine de colle n°8

Semaine du 05/02 au 09/02

Semaine du 05/02 au 09/02

Programme

- **Chapitre 14.** Intégralité. *On essaiera de proposer un exercice complet sur les chaînes de Markov (réduction de la matrice de transition à faire), avec si possible un programme Python à écrire/compléter et une convergence de la chaîne à visualiser.*
- **Chapitre 15.** Estimation ponctuelle uniquement. On pourra faire chercher l'expression du maximum de vraisemblance ou faire calculer le biais de certains estimateurs.
  On propose néanmoins pour travailler, dans un second temps, des exercices avec aussi des questions sur les intervalles de confiance.

Questions de cours

- On considère un réel $p \in]0;1[$ et la matrice $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 2/3 & 1/3 \\ p & 1-p & 0 \end{pmatrix}$.
  i. Dessiner un graphe probabiliste qui aurait A pour matrice de transition.
  ii. On considère la chaîne de Markov $(X_k)_{k\geq 0}$ associée à ce graphe et on suppose que X_0 suit la loi uniforme sur $[\![1,3]\!]$. Écrire un programme qui, pour un entier naturel k et un réel p pris en arguments, simule et renvoie la variable X_k.
  iii. Déterminer l'état stable de la chaîne.
  iv. Déterminer la loi de X_1.

- On considère une chaîne de Markov $(X_n)_{n\geq 0}$ à deux états de matrice de transition $A = \begin{pmatrix} 1/3 & 2/3 \\ 1/4 & 3/4 \end{pmatrix}$.
  i. Écrire une fonction Python qui renvoie une simulation de X_k.
  ii. Déterminer l'état stable de la chaîne.
  iii. Montrer, à l'aide de considérations simples sur des suites usuelles, qu'il y a convergence en loi de (X_n) vers une v.a dont on précisera la loi, et ce, peu importe la loi initiale X_0.

- On considère un $n-$échantillon $(X_1, X_2, ..., X_n)$ d'une v.a $X \hookrightarrow \mathcal{B}(p)$ dont on cherche à estimer le paramètre p.
  i. Montrer que la moyenne empirique de l'échantillon fournit un estimateur non biaisé de p.
  ii. Est-il convergent ?

Suggestion d'exercices

Exercice 1. Deux urnes A et B contiennent initialement chacune deux boules numérotées 0 et 1. Leur contenu évolue lors d'une expérience aléatoire et est simulé par une liste Python.
Ainsi la ligne de code `A = [0, 1]` permet de voir qu'initialement, l'urne A contient une boule numérotée 0 et une autre numérotée 1.

On considère la fonction Python suivante qui permet de simuler une variable aléatoire X.

```
def simulX():
    A = [0, 1]
    B = [0;1]
    Y = rd.randint(0, 2, [4, 2])
    k = 0
    while np.sum(A) > 0 and k < 4:
        i = Y[k, 0]
        j = Y[k, 1]
        c = A[i]
        A[i] = B[j]
        B[j] = c
        k = k + 1
    return k
```

Expliquer le protocole ainsi simulé et donner la loi de X.

Exercice 2. Étudier la convergence en loi de (T_n) où T_n est la v.a qui peut être simulée par la fonction ci-dessous.

```
def simul_T(n) :
    U=[rd.random() for k in range(n)]
    N=np.sum[rd.binomial(1, p) for k in range(n)]
    if N==0 :
        return U[0]
    return(max([U[j] for j in range(N)]))
```

Exercice 3. 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.

Étudier la chaîne de Markov $(N_i)_{i \geq 0}$.

Exercice 4. Dans le cadre d'une étude sur la santé au travail, on a interrogé au hasard 500 salariés de différents secteurs et de différentes régions de France. 145 d'entre eux déclarent avoir déjà subi un harcèlement moral au travail.

(1) Donner une estimation ponctuelle de la proportion de salariés ayant déjà subi un harcèlement moral au travail.
(2) Donner une estimation de cette proportion par un intervalle de confiance à 90%.
(3) Si avec les mêmes données on calculait un intervalle de confiance à 95%, serait-il plus grand ou plus petit que celui trouvé à la question précédente ? (justifier sans calcul.)

Exercice 5. Soit X une variable aléatoire d'espérance m et de variance σ^2. On considère un $n-$échantillon $(X_1, X_2, ..., X_n)$ de X et on note

$$S_n = \sum_{k=1}^{n} X_k, \qquad \overline{X}_n = \frac{S_n}{n}.$$

(1) Expliquer pourquoi l'intervalle

$$\left[\overline{X}_n - \frac{\sigma}{\sqrt{n\alpha}}; \ \overline{X}_n + \frac{\sigma}{\sqrt{n\alpha}}\right]$$

fournit un intervalle de confiance (exact) pour m au risque α.

(2) Expliquer pourquoi, en notant $u_\alpha = \Phi^{-1}(1-\alpha/2)$, l'intervalle

$$\left[\overline{X}_n - \frac{\sigma}{\sqrt{n}}u_\alpha; \ \overline{X}_n + \frac{\sigma}{\sqrt{n}}u_\alpha\right]$$

fournit un intervalle de confiance asymptotique pour m au risque α.

(3) **Application.** On suppose que $X \hookrightarrow \mathcal{B}(p)$ et on cherche à estimer p (qu'on ne connait pas). On ne connait donc *a fortiori* pas non plus la variance. Expliquer comment obtenir des intervalles de confiance (exact et asymptotique) pour m (au risque α) qui ne dépendent pas de σ.

Exercice 6. Soit $p \in]0\,;1[$, on considère une variable aléatoire X qui suit la loi géométrique de paramètre p. On pose $q = 1-p$. Pour tout $n \in \mathbb{N}^*$, on considère un $n-$échantillon $(X_1, \ldots, X_n)$ de la loi de X.

On pose $S_n = \sum\limits_{k=1}^{n} X_k$, $Y_n = \dfrac{n}{S_n}$ et $\overline{X}_n = \dfrac{1}{Y_n} = \frac{1}{n}S_n$.

(1) Montrer que $\overline{X}_n$ est un estimateur sans biais de $\frac{1}{p}$. Quel est son risque quadratique ?

(2) Pour tout $n \in \mathbb{N}^*$, on pose $T_n = \sqrt{\dfrac{n}{p^2q}}(Y_n - p)$.

On **admet** que T_n converge en loi vers une variable aléatoire T qui suit la loi normale centrée réduite.

(a) Soit $\alpha \in]0\,;1[$ et a_α l'unique réel vérifiant $P([T > a_\alpha]) = \frac{\alpha}{2}$.
Montrer que $P\left(-a_\alpha \leq T \leq a_\alpha\right) = 1-\alpha$.

(b) Pour n assez grand on considère alors que $P\left(-a_\alpha \leq T_n \leq a_\alpha\right) = 1-\alpha$. En déduire que :

$$P\left(Y_n - a_\alpha p\sqrt{\frac{q}{n}} \leq p \leq Y_n + a_\alpha p\sqrt{\frac{q}{n}}\right) = 1-\alpha.$$

(c) Étudier la fonction $f : x \mapsto x\sqrt{1-x}$ sur l'intervalle $[0;1]$.

(d) Déduire des deux questions précédentes que pour n assez grand :

$$P\left(Y_n - \frac{2a_\alpha}{3\sqrt{3n}} \leq p \leq Y_n + \frac{2a_\alpha}{3\sqrt{3n}}\right) \geq 1-\alpha.$$

(e) On suppose que $n = 900$, une réalisation de l'échantillon $(X_1, \ldots, X_{900})$ a donné la valeur 4 à $\overline{X}_{900}$.
Donner alors la réalisation y_{900} de la variable Y_{900}.
On se donne un niveau de risque $\alpha = 0,05$, le nombre $a_{0,05}$ vaut à peu près 2.
Trouver une fourchette pour p avec un niveau de confiance d'au moins $0,95$. *On donne* $\frac{2}{45\sqrt{3}} \simeq 0,026$.

Exercice 7. Dans cet exercice, on considère un $n-$échantillon $(X_1, ..., X_n)$ d'une variable aléatoire suivant une loi de Poisson de paramètre θ inconnu ($\theta \in]0, +\infty[$). On souhaite estimer le paramètre $\exp(-\theta)$.

On note, pour $k \in [\![1; n]\!]$, $S_k = \displaystyle\sum_{i=1}^{k} X_i$, puis on définit, pour tout $i \in [\![1; n]\!]$, la variable aléatoire Y_i par

$$Y_i = \begin{cases} 1, & \text{si } X_i = 0 \\ 0, & \text{sinon} \end{cases}, \quad \text{et on note} \quad \overline{Y_n} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

(1) Préciser, pour tout $k \in [\![1; n]\!]$, la loi de S_k et, pour tout $i \in [\![1; n]\!]$, la loi de Y_i.
(2) Donner la loi de $n\overline{Y}_n$, puis en déduire que $\overline{Y}_n$ est un estimateur sans biais de $\exp(-\theta)$.
(3) Déterminer le risque quadratique $r\left(\overline{Y}_n\right)$. En déduire que $\overline{Y}_n$ est un estimateur convergent.

(4) Montrer que pour tout j entier naturel, $P_{[S_n=j]}(X_1 = 0) = \left(1 - \dfrac{1}{n}\right)^j$.

On introduit alors, pour $n \in \mathbb{N}^*$, l'estimateur $T_n = \left(1 - \dfrac{1}{n}\right)^{S_n}$.

(5) Montrer que T_n admet une espérance et que c'est encore un estimateur sans biais de $\exp(-\theta)$.

(6) Montrer que T_n admet une variance vérifiant

$$V(T_n) = \exp(-2\theta)\left(\exp\left(\frac{\theta}{n}\right) - 1\right).$$

L'estimateur T_n est-il convergent?

(7) On souhaite comparer les performances de $\overline{Y_n}$ et T_n en tant qu'estimateurs de $\exp(-\theta)$.

(a) En utilisant le théorème des accroissements finis, démontrer que

$$1 \leq \frac{\exp(\theta) - 1}{\theta} \leq \exp(\theta).$$

(b) Soit la fonction $h : [0, 1] \to \mathbb{R}$ définie par

$$h(t) = t\exp(\theta) + (1 - t) - \exp(t\theta)$$

pour tout $t \in [0, 1]$. Étudier les variations de h.
(c) En déduire que

$$\exp\left(\frac{\theta}{n}\right) \leq \frac{\exp(\theta)}{n} + \frac{n-1}{n}.$$

(d) Quel estimateur parait alors le plus performant? On comparera les risques quadratiques.