Math ECG 2. 2023-2024

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

Programme de colles n°10

Semaine du 04/03 au 08/03

Semaine du 04/03 au 08/03

Programme

- **Chapitre 15.** Intégralité.
- Reprise du **Concours blanc n°2**.
- **Chapitre 16.** Tout en douceur à partir de Jeudi.

Questions de cours

- 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 ?

- Construire, à l'aide de l'inégalité de Bienaymé-Tchebychev, un intervalle (symétrique) de confiance au risque $\alpha \in]0, 1[$ pour le paramètre p d'une loi de Bernoulli $\mathcal{B}(p)$ dont le milieu est la moyenne empirique d'un $n-$échantillon.

- Quel est l'estimateur du maximum de vraisemblance pour le paramètre λ d'une loi de Poisson ? On démontrera le résultat en étudiant la fonction de vraisemblance
$$L: \lambda \mapsto \prod_{i=1}^{n} P(X_i = x_i),$$
où $(X_1, ..., X_n)$ est un $n-$échantillon de la loi $\mathcal{P}(\lambda)$ et $(x_1, x_2, .., x_n) \in \mathbb{N}^n$ est une observation arbitraire fixée de l'échantillon.

- (À partir de Jeudi) Montrer que $\{M \in \mathcal{M}_3(\mathbb{R}) : {}^tM = -M\}$ est un sous-espace vectoriel de $\mathcal{M}_3(\mathbb{R})$ et en expliciter une base puis la dimension.

Suggestion d'exercices

Exercice 1. Soit X d'espérance m et de variance σ^2. On considère alors un $n-$échantillon $(X_1, ..., X_n)$ de X et on pose $T_n = X_1 + X_2 + ... + X_n$. On suppose que σ^2 est connue, mais pas m.

(1) Déterminer $E(T_n)$ et $V(T_n)$.
(2) Montrer que, pour tout $\varepsilon > 0$
$$P(|T_n - nm| > n\varepsilon) \leq \frac{\sigma^2}{n\varepsilon}.$$
(3) Quel est le risque de l'intervalle de confiance
$$\left[\frac{T_n}{n} - \varepsilon;\ \frac{T_n}{n} + \varepsilon\right]$$
pour m?

Exercice 2. On cherche à estimer le paramètre $\theta > 0$ d'une loi uniforme $\mathcal{U}([0,\theta])$. On dispose pour cela d'un $n-$échantillon de cette loi et on note $Y_n = \max(X_1, ..., X_n)$.

(1) Montrer que, pour tout réel $x > 1$, on a
$$P\left(Y_n \in \left[\theta; \frac{\theta}{x}\right]\right) = 1 - \left(\frac{1}{x}\right)^n.$$

(2) Soit $\alpha \in]0;1[$. On note $g_n(\alpha) = n\left(\left(\frac{1}{\alpha}\right)^{1/n} - 1\right)$.

Montrer que
$$I_{n,\alpha} = \left[Y_n, Y_n + \frac{Y_n}{n} g_n(\alpha)\right]$$
est un intervalle de confiance au niveau $1-\alpha$ pour θ.

Exercice 3. 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 4. 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 $T_n = \sqrt{\dfrac{n}{p^2 q}}(Y_n - p)$.

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

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

(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, $P_{[S_n=j]}(X_1 = 0) = \left(1 - \frac{1}{n}\right)^j$.

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

(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.

Exercice 6. Soit a un réel de $]0, 2[$ et f la fonction définie par :

$$f(x) = \begin{cases} \frac{x}{a} e^{-x^2/2a}, & \text{si } x \geq 0 \\ 0, & \text{si } x < 0 \end{cases}$$

(1) Rappeler l'expression intégrale ainsi que la valeur du moment d'ordre 2 d'une variable aléatoire Z suivant la loi normale de paramètres 0 et a.

(2) (a) Montrer que la fonction f est une densité de probabilité.
Dans la suite de l'exercice, on considère une variable aléatoire X de densité f.

(b) Déterminer la fonction de répartition F_X de X.
(c) Donner le tableau des variations des courbes de f et de F_X pour $a \in]0, 2[$ quelconque et tracez-les pour $a = 1$.
(d) Montrer que X admet une espérance et la calculer.

(3) On considère la variable aléatoire Y définie par : $Y = \frac{X^2}{2a}$. Déterminer la loi de Y.

(4) On suppose désormais que le paramètre a est inconnu et on souhaite l'estimer. Soit n un entier naturel supérieur ou égal à 1. On considère un échantillon $(X_1, \dots, X_n)$ composé de variables aléatoires indépendantes ayant toutes la même loi que X.

On note S_n la variable aléatoire définie par $S_n = \dfrac{1}{2n}\displaystyle\sum_{k=1}^{n} X_k^2$.

(a) Montrer que S_n est un estimateur sans biais et convergent de a.
(b) Proposer un intervalle de confiance de a au niveau de risque $\alpha \in]0,1[$ à l'aide de S_n.

(5) Soit la variable aléatoire Z définie par $Z(\omega) = \begin{cases} \dfrac{1}{X(\omega)-1}, & \text{si } X(\omega) \neq 1 \\ 0, & \text{sinon} \end{cases}$

Déterminer la fonction de répartition de Z.

Exercice 7. (À partir de Jeudi)
Écrire les matrices (dans les bases canoniques correspondantes) de

$$f_1 : \ M \in \mathcal{M}_2(\mathbb{R}) \longmapsto M - 2\,{}^t M \in \mathbb{R}; \qquad f_2 : \ P \in \mathbb{R}_3[x] \longmapsto (x-1)P' - 3P \in \mathbb{R}_3[x].$$

Exercice 8. (Pour les colles à partir de Jeudi)
On considère l'application

$$\begin{array}{rccc} f : & \mathbb{R}_3[x] & \to & \mathbb{R}_3[x] \\ & P & \mapsto & Q(x) = P(x+1) - P(x) \end{array}$$

(1) Montrer que f est linéaire.
(2) Soit $P \in \mathbb{R}_0[x]$. Calculer $f(P)$. f est-elle injective ?
(3) Déterminer une base de l'image de f. f est-elle surjective ?
(4) Déterminer un polynôme de $\mathbb{R}_3[x]$ n'ayant pas d'antécédent par f.
(5) Déterminer une base du noyau de f.
(6) Quelle est la matrice de f dans la base canonique ?