ECE 2 - Année 2018-2019
Lycée français de Vienne
Mathématiques - F. Gaunard
http://frederic.gaunard.com

Quinzaine de colle n°10

Période du 04/03 au 15/03

Semaine du 04/03 au 08/03

Programme

- **Chapitre 12.** Intégralité.
- Révisions générales

Questions de cours

- Soit (X_n) une suite v.a.i.i.d, où $X_n \hookrightarrow \mathcal{U}([0;1])$. Montrer que

$$\sqrt{\frac{12}{n}}\sum_{k=1}^{n} X_n - \sqrt{3n} \xrightarrow{\mathcal{L}} Z, \quad Z \hookrightarrow \mathcal{N}(0,1)$$

- Définition d'un estimateur non biaisé de θ à partir d'un $n-$échantillon $(X_1, ..., X_n)$ d'une variable X dépendant du paramètre θ. Définition d'estimateur asymptotiquement sans biais. Risque quadratique. Estimateur convergent.
- Définition d'un intervalle de confiance au niveau $1-\alpha$ pour θ.

Planche d'exercices

(1) Soit (X_n) une suite v.a.i.i.d de loi commune $\mathcal{U}(]0;1])$. Pour tout $n \in \mathbb{N}^*$, on note

$$Y_n = e^{\sqrt{n}}\left(\prod_{k=1}^{n} X_i\right)^{1/\sqrt{n}}, \qquad L_n = \ln(Y_n).$$

(a) Déterminer la loi de $-\ln(X_1)$. En déduire $E(\ln(X_1))$ et $V(\ln(X_1))$.
(b) À l'aide du TCL, montrer que (L_n) converge en loi vers une variable dont on donnera la loi.
(c) Que vaut $Y_n(\Omega)$?
(d) Soit $t < 0$. Que vaut $\lim_{n\to+\infty} P(Y_n \leq t)$?
(e) On définit la fonction F sur $\mathbb{R}$ par

$$F(x) = \begin{cases} 0, & \text{si } x \leq 0 \\ \Phi(\ln(x)), & \text{si } x > 0 \end{cases}$$

(i) Montrer que F est la fonction de répartition d'une variable aléatoire Z.
(ii) Montrer que (Y_n) converge en loi vers Z.

(2) Proposer un estimateur sans biais du paramètre λ d'une loi de Poisson à partir d'un $n-$échantillon de celle-ci. Quel est son risque quadratique?

(3) (German Tank Problem)

La méthode que nous allons présenter fut utilisée par les forces alliées durant la seconde guerre mondiale afin d'estimer l'étendue de la production allemande de tanks. Les forces de l'Axe étaient alors repérées par des numéros de séries consécutifs.

Dans notre modélisation, les allemands disposent de N tanks numérotés de 1 à N. Les force alliées observent aléatoirement, uniformément et "avec remise" n numéros de séries $(X_1, ..., X_n)$ et cherchent à estimer le paramètre N.

On pose $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ et $S_n = \max(X_1, ..., X_n)$.

(a) Déterminer $E(\overline{X}_n)$ et en déduire un estimateur T_n de N sans biais en fonction de $\overline{X}_n$.
(b) Calculer le risque quadratique de T_n et montrer que cet estimateur est convergent.
(c) Calculer, pour tout $k \in \mathbb{N}$, $P(S_n \leq k)$.
(d) Soit Y un v.a à valeurs dans $[\![1; N]\!]$. Montrer que

$$E(Y) = \sum_{k=1}^{N} P(Y \geq k).$$

(e) Montrer alors que

$$E(S_n) = N - \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n.$$

(f) Vérifier que, pour tout $k \in [\![0; N-1]\!]$,

$$0 \leq \left(\frac{k}{N}\right)^N \leq \int_{k/N}^{(k+1)/N} t^n \mathrm{d}t.$$

(g) En déduire que l'estimateur S_n est asymptotiquement sans biais.

(4) Soit (X_n) une suite v.a.i.i.d de loi $\mathcal{B}(p^2)$. On cherche à estimer p.
(a) Montrer que $E(\overline{X}_n) - E\left(\sqrt{\overline{X}_n}\right)^2 > 0$.
(b) En déduire que l'estimateur $T_n = \sqrt{\overline{X}_n}$ n'est pas sans biais.
(c) Soit $\varepsilon > 0$.
(i) Écrire l'inégalité de Bienaymé-Tchebychev appliquée à $\overline{X}_n$.
(ii) Montrer, en distinguant deux cas $x \leq y$ et $x > y$, que pour tous $x, y \in [0; 1]$, on a $|\sqrt{x} - \sqrt{x}| \leq \sqrt{|y - x|}$.
(iii) En déduite que l'estimateur T_n est convergent.

(5) Un sondage consiste à proposer l'affirmation $\mathcal{A}$ à certaines personnes d'une population donnée. Le sujet abordé étant délicat, on adopte la méthode suivante afin de mettre en confiance les personnes sondées pour qu'elles ne mentent pas. L'enquêteur dispose de 20 cartes, numérotées de 1 à 20. La personne sondée tire une carte au hasard et ne la montre pas à l'enquêteur. La règle est alors la suivante:
- si elle tire la carte 1, elle répond "vrai" si elle est d'accord avec l'affirmation $\mathcal{A}$ et 'faux" sinon.
- si elle tire un autre numéro, elle répond "vrai" si elle n'est pas d'accord avec l'affirmation $\mathcal{A}$ et 'faux" sinon.

Le but de l'enquête est d'évaluer la proportion p de gens de cette population qui sont réellement d'accord avec l'affirmation $\mathcal{A}$.

(a) On interroge une personne selon ce procédé et on considère l'événement V suivant :
V : la personne répond "vrai" On note $\theta = P(V)$.
En utilisant la formule des probabilités totales, exprimer θ en fonction de p. En déduire p en fonction de θ.

On rappelle que l'on ne connaît ni p, ni θ.

(a) On considère un échantillon aléatoire, de taille n, extrait de la population considérée et on note S_n, le nombre de réponses "vrai" obtenues. On suppose n assez grand pour pouvoir considérer que cet échantillonnage est assimilable à un tirage avec remise.
  (i) Donner la loi de S_n ainsi que son espérance et sa variance.
  (ii) Montrer que $\dfrac{S_n}{n}$, est un estimateur sans biais de θ.

(b) On suppose que sur 100 personnes interrogées, 23 personnes ont répondu "vrai".
  (i) Donner une estimation ponctuelle de θ et de p.
  (ii) Donner un intervalle de confiance à 95% de θ puis de p.
  On donne $\Phi(1,96) = 0,975$.

(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 $\overline{X}_n = \dfrac{1}{Y_n} = \frac{1}{n} S_n$.

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

(b) Pour tout $n \in \mathbb{N}^*$, on pose $T_n = \dfrac{Y_n - p}{\sqrt{\frac{p^2 q}{n}}}$, et on admet que T_n converge en loi vers une variable aléatoire T qui suit la loi normale centrée réduite.
  (i) Soit $\alpha \in \,]0\,;\,1[$ et a_α l'unique réel vérifiant $P([T > a_\alpha]) = \frac{\alpha}{2}$.
  Montrer que $P\left(-a_\alpha \leq T \leq a_\alpha\right) = 1 - \alpha$.
  (ii) Pour n assez grand on considère alors que $P\left(-a_\alpha \leq T_n \leq a_\alpha\right) = 1 - \alpha$. En déduire que :
$$P\left(Y_n - a_\alpha p\sqrt{\frac{q}{n}} \leq p \leq Y_n + a_\alpha p\sqrt{\frac{q}{n}}\right) = 1 - \alpha.$$
  (iii) Étudier la fonction $f : x \mapsto x\sqrt{1-x}$ sur l'intervalle $[0; 1]$.
  (iv) Déduire des deux questions précédentes que pour n assez grand :
$$P\left(Y_n - \frac{2a_\alpha}{3\sqrt{3n}} \leq p \leq Y_n + \frac{2a_\alpha}{3\sqrt{3n}}\right) \geq 1 - \alpha.$$
  (v) 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* $\frac{2}{45\sqrt{3}} \simeq 0,026$.