Math ECG 2. 2022-2023

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

Quinzaine de colle n°10

Période du 13/03 au 24/03

Semaine du 13/03 au 17/03

Programme

- **Chapitre 14.** Intégralité.
- Reprise du Concours Blanc.
- Révisions générales. *On posera nécessairement un exercice sur un chapitre traité en début d'année.*

Questions de cours

- Soit $(X_1, ..., X_n)$ un $n-$échantillon de $\mathcal{B}(p)$. On note $\overline{X}_n$ la moyenne empirique de l'échantillon. Montrer que, pour tout $\varepsilon > 0$,
$$P\left(\left|\overline{X}_n - p\right|\right) \leq \frac{1}{4n\varepsilon^2}.$$
En déduire, en fonction de n et α un intervalle de confiance centré en $\overline{X}_n$ au risque α pour p.

Suggestions d'exercices

Exercice 1. 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 2. 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 3. 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 $P(-a_\alpha \le T \le a_\alpha) = 1-\alpha$.
 (b) Pour n assez grand on considère alors que $P(-a_\alpha \le T_n \le a_\alpha) = 1-\alpha$. En déduire que :
$$P\left(Y_n - a_\alpha p\sqrt{\frac{q}{n}} \le p \le Y_n + a_\alpha p\sqrt{\frac{q}{n}}\right) = 1-\alpha.$$
 (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 :
$$P\left(Y_n - \frac{2a_\alpha}{3\sqrt{3n}} \le p \le Y_n + \frac{2a_\alpha}{3\sqrt{3n}}\right) \ge 1-\alpha.$$
 (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 4. Soit p un réel de $]0;1[$ et $q = 1-p$. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, de même loi de Bernoulli telle que :
$$\forall k \in \mathbb{N}^*, \qquad P(X_k = 1) = p, \quad \text{et} \quad P(X_k = 0) = q.$$
Pour n entier de $\mathbb{N}^*$, on définit pour tout $k \in [\![1;n]\!]$, la variable aléatoire $Y_k = X_k + X_{k+1}$.

(1) (a) Calculer pour tout $k \in [\![1;n]\!]$, $\mathrm{Cov}(Y_k, Y_{k+1})$.
 (b) Montrer que
$$0 < \mathrm{Cov}(Y_k, Y_{k+1}) < \frac{1}{4}.$$

(2) Calculer, pour tout couple (k,l) tel que $1 \le k < l \le n$, $\mathrm{Cov}(Y_k, Y_l)$.

(3) On note ε un réel strictement positif fixé. Montrer que
$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} Y_k - 2p\right| > \varepsilon\right) = 0.$$

Exercice 5. On joue à pile ou face jusqu'à obtenir "pile". S'il a fallu n jets de la pièce, on tire ensuite une boule au hasard dans une urne contenant des boules numérotées de 1 à n. Soit X le numéro de la boule obtenue.
Simuler la variable aléatoire X.
Après avoir déterminé l'ensemble des valeurs possibles de X, écrire $P(X = k)$ sous forme d'une série.
Déterminer $P(X = 1)$.

Exercice 6. On considère une suite d'épreuves indépendantes, la probabilité du succès étant p ($p \in]0, 1[$) à chaque épreuve.
Soit $n \in \mathbb{N}^*$, on note X_n la v.a.r. égale au nombre d'essais pour obtenir n succès.
Simuler la v.a.r. X_n.
Déterminer la loi de X_n.
Pour $n > 1$, on pose $Y = \dfrac{n-1}{X_n - 1}$. Déterminer $E(Y)$.

Exercice 7. Soit pour $n \in \mathbb{N}^*$,

$$P_n(x) = x^n + x - 1.$$

(1) Montrer que, pour tout $n \in \mathbb{N}^*$, il existe un unique $x_n \in]0, 1[$ tel que $P_n(x_n) = 0$.
(2) Étudier la monotonie de la suite (x_n)
(3) Montrer que la suite (x_n) converge vers un réel que l'on précisera.
(4) Montrer que, pour tout $n \in \mathbb{N}^*$: $\quad 1 - x_n \leq \dfrac{\ln n}{n}$.

Exercice 8. Soit f la fonction définie sur $\mathbb{R}^{+*} \times \mathbb{R}^{+*}$, à valeurs réelles, par :

$$f(x, y) = \frac{x^2 + xy + \sqrt{y}}{x\sqrt{y}}$$

(1) Montrer que f est de classe $\mathcal{C}^2$ sur $\mathbb{R}^{+*} \times \mathbb{R}^{+*}$.
(2) Déterminer les points critiques de f.
(3) Quelle est la nature de ces points critiques ?

Exercice 9. Pour tout entier naturel n, on pose

$$u_n = \prod_{k=0}^{n} \left(1 + \frac{1}{2^k}\right).$$

(1) Donner, sous forme d'entiers ou de fractions simplifiées, les valeurs de u_0, u_1 et u_2.

(2) (a) Montrer que, pour tout entier naturel n, on a : $u_n \geq 2$.
(b) Exprimer u_{n+1} en fonction de u_n puis en déduire les variations de la suite (u_n).
(c) Établir que, pour tout réel x strictement supérieur à -1, on a : $\ln(1 + x) \leq x$.
(d) En déduire, pour tout entier naturel n, un majorant de $\ln(u_n)$.

(3) En utilisant les questions précédentes, montrer que la suite (u_n) converge vers un réel $\ell \in [2, \mathrm{e}^2]$.

(4) On se propose dans cette question de déterminer la nature de la série de terme général $(\ell - u_n)$.
(a) Justifier que la suite $(\ln(u_n))_{n\in\mathbf{N}}$ converge et que l'on a

$$\ln(\ell) = \sum_{k=0}^{+\infty} \ln\left(1 + \frac{1}{2^k}\right).$$

(b) Montrer que, pour tout $n \in \mathbb{N}$, on a

$$\ln\left(\frac{\ell}{u_n}\right) = \sum_{k=n+1}^{+\infty} \ln\left(1 + \frac{1}{2^k}\right).$$

(c) Vérifier, en utilisant le résultat de la question 3a), que
$$\forall n \in \mathbb{N}, \qquad 0 \leq \ln\left(\frac{\ell}{u_n}\right) \leq \frac{1}{2^n}.$$
(d) Déduire de la question précédente que, pour tout $n \in \mathbb{N}$,
$$0 \leq \ell - u_n \leq \ell\left(1 - \mathrm{e}^{-\frac{1}{2^n}}\right).$$
(e) Justifier que, pour tout réel x, on a $1 - \mathrm{e}^{-x} \leq x$. En déduire que, pour tout $n \in \mathbb{N}$,
$$0 \leq \ell - u_n \leq \frac{\ell}{2^n}.$$
Conclure quant à la nature de la série de terme général $(\ell - u_n)$.

Exercice 10. (*Inégalité de Cantelli)

On considère une variable aléatoire X admettant une espérance, notée m, et une variance, notée σ^2.

(1) Soient $\alpha \geq 0$ et $\lambda \geq 0$ fixés.

(a) Montrer que $P(X - m \geq \alpha) = P(X - m + \lambda \geq \alpha + \lambda)$.
(b) Déterminer $E((X - m + \lambda)^2)$.
(c) En déduire que
$$P(X - m \geq \alpha) \leq \frac{\sigma^2 + \lambda^2}{(\alpha + \lambda)^2}.$$
(d) Étudier les variations (et extremums) sur $[0; +\infty[$ de
$$\varphi : \ \lambda \mapsto \frac{\sigma^2 + \lambda^2}{(\alpha + \lambda)^2}.$$
(e) En déduire alors que pour un bon choix de λ, on a l'inégalité de *Cantelli*,
$$P(X - m \geq \alpha) \leq \frac{\sigma^2}{\sigma^2 + \alpha^2}.$$
(f) Avec la même méthode, montrer que
$$P(X - m \leq -\alpha) \leq \frac{\sigma^2}{\sigma^2 + \alpha^2}.$$

(2) Montrer qu'on a
$$P(|X - m| \geq \alpha) \leq \frac{2\sigma^2}{\sigma^2 + \alpha^2}.$$
Cette inégalité est-elle meilleure que celle de Bienaymé-Tchebychev?