

HEC-ESCP – Mathématiques II

Mathématiques approfondies

Corrigé du sujet adapté – 2019

Les questions et sous-questions conservent la numérotation de l'énoncé. Les notations $\mathbb{E}$ et $\mathbb{V}$ désignent l'espérance et la variance. La transposée d'une matrice A est notée ${}^tA$.

Partie I – Fonction logistique et lois logistiques

On pose $\Lambda(x) = 1/(1 + e^{-x})$ et $\lambda = \Lambda'$. La loi logistique standard est définie par la densité λ ; la loi $\mathcal{L}(r, s)$, pour $s > 0$, est celle de $r + sZ$ lorsque Z suit la loi logistique standard.

Corrigé – Partie I – Fonction logistique et lois logistiques.

1. La fonction logistique et son point fixe.

- a) La fonction Λ est continue sur $\mathbb{R}$. Elle est strictement croissante : e^{-x} décroît strictement, et la prise de l'inverse d'un nombre strictement positif renverse l'ordre. Ses limites valent 0 en $-\infty$ et 1 en $+\infty$. Le théorème des valeurs intermédiaires et la stricte croissance donnent donc une bijection de $\mathbb{R}$ sur $]0, 1[$. Pour $y \in]0, 1[$, la résolution de $y = \Lambda(x)$ donne

$$e^{-x} = \frac{1-y}{y}, \quad x = \ln\left(\frac{y}{1-y}\right).$$

Ainsi

$$\Lambda : \mathbb{R} \longrightarrow]0, 1[\text{ est bijective, } \quad \Lambda^{-1}(y) = L(y) = \ln\left(\frac{y}{1-y}\right).$$

- b) La dérivation donne

$$\Lambda'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} = \Lambda(x)(1 - \Lambda(x)).$$

Comme $0 < \Lambda(x) < 1$ et $u(1-u) = 1/4 - (u - 1/2)^2$, on a

$$0 < \Lambda'(x) \leq \frac{1}{4} \quad (x \in \mathbb{R}).$$

- c) Posons $h(x) = \Lambda(x) - x$. Cette fonction est continue et dérivable, avec $h'(x) = \Lambda'(x) - 1 \leq -3/4 < 0$. Elle est donc strictement décroissante sur $\mathbb{R}$. De plus $h(0) = 1/2 > 0$ et $h(1) = \Lambda(1) - 1 < 0$. Le théorème des valeurs intermédiaires donne un zéro dans $]0, 1[$, et la stricte décroissance assure qu'il est unique sur $\mathbb{R}$:

$$\exists! x_0 \in \mathbb{R}, \quad \Lambda(x_0) = x_0, \quad 0 < x_0 < 1.$$

- d) On a $-1 < h'(x) \leq -3/4$, donc $|h'(x)| \leq 1$. Si $x \neq x_0$, le théorème des accroissements finis sur le segment joignant x_0 et x fournit un réel ξ entre ces deux points pour lequel

$$h(x) - h(x_0) = h'(\xi)(x - x_0).$$

Comme $h(x_0) = 0$, la majoration de la dérivée donne $|h(x)| \leq |x - x_0|$. Pour $x = x_0$, les deux membres sont nuls. Ainsi

$$\forall x \in \mathbb{R}, \quad |\Lambda(x) - x| \leq |x - x_0|.$$

2. Approximation par dichotomie.

- a) Les valeurs initiales $a = 0$ et $b = 1$ encadrent x_0 , avec $h(a) > 0$ et $h(b) < 0$. À chaque étape, si $h(c) > 0$, la décroissance de h donne $c < x_0$: il faut remplacer a par c . Sinon, on a $x_0 \leq c$ et on remplace b par c . Le point fixe reste dans l'intervalle. Le script complété, avec le choix de précision justifié en b), est

```
import numpy as np
```

```
def Lambda(x):
    return 1/(1+np.exp(-x))
```

```
a = 0
```

```
b = 1
```

```
eps = 2*10**(-4)
```

```
while b-a > eps:
```

```
    c = (a+b)/2
```

```
    if Lambda(c) > c:
```

```
        a = c
```

```
    else:
```

```
        b = c
```

```
x0 = (a+b)/2
```

La longueur de l'intervalle est divisée par 2 à chaque passage ; la boucle finit donc pour tout $\text{eps} > 0$.

Instructions manquantes : $a = c$ puis $b = c$.

- b) Notons $\hat{x}_0$ la valeur du milieu renvoyé par le programme. Puisque le vrai point fixe appartient à $[a, b]$,

$$|\hat{x}_0 - x_0| \leq \frac{b - a}{2} \leq \frac{\text{eps}}{2}.$$

Il suffit donc d'imposer $0 < \text{eps} \leq 2 \times 10^{-4}$. Un choix suffisant est

$$\text{eps} = 2 \times 10^{-4} \implies |\hat{x}_0 - x_0| \leq 10^{-4}.$$

- c) L'inégalité de la question 1.d appliquée au nombre calculé donne

$$|\Lambda(\hat{x}_0) - \hat{x}_0| \leq |\hat{x}_0 - x_0| \leq 10^{-4}.$$

L'instruction `Lambda(x0)-x0` renvoie donc un nombre proche de 0, dont la valeur absolue est au plus 10^{-4} avec le choix précédent de eps .

3. Densité logistique standard.

- a) La fonction $\lambda(x) = e^{-x}/(1 + e^{-x})^2$ est continue et positive sur $\mathbb{R}$. Pour des bornes finies $a < b$, l'identité $\lambda = \Lambda'$ donne

$$\int_a^b \lambda(t) dt = \Lambda(b) - \Lambda(a).$$

Les limites de Λ donnent alors la convergence de l'intégrale sur toute la droite et $\int_{-\infty}^{+\infty} \lambda(t) dt = 1$. Donc

$$\lambda \text{ est une densité, } F_Z(x) = \Lambda(x) \text{ pour la loi logistique standard.}$$

- b) En multipliant le numérateur et le dénominateur par e^{-2x} , on obtient $\lambda(-x) = \lambda(x)$: la densité est paire. Ses deux premières dérivées sont

$$\lambda'(x) = \frac{e^{-x}(e^{-x} - 1)}{(1 + e^{-x})^3}, \quad \lambda''(x) = \frac{e^{-x}(e^{-2x} - 4e^{-x} + 1)}{(1 + e^{-x})^4}.$$

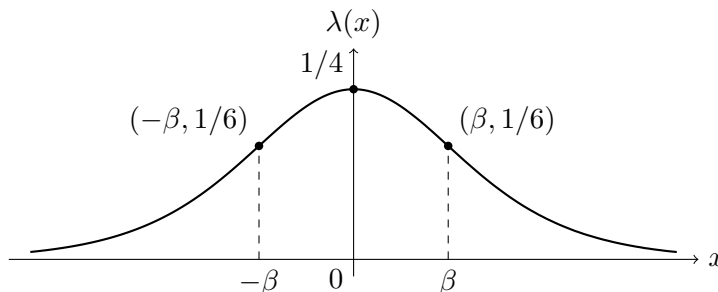
Ainsi λ est croissante sur $] -\infty, 0]$ et décroissante sur $[0, +\infty[$, avec maximum $\lambda(0) = 1/4$. Sa limite en $+\infty$ est 0 ; la parité donne la même limite en $-\infty$. Pour le signe de λ'' , posons $z = e^{-x} > 0$. Le polynôme $z^2 - 4z + 1$ s'annule en $2 - \sqrt{3}$ et $2 + \sqrt{3}$; il est négatif entre ses racines et positif à l'extérieur. Comme $(2 - \sqrt{3})(2 + \sqrt{3}) = 1$, en posant $\beta = \ln(2 + \sqrt{3}) > 0$, on obtient

$$\lambda''(x) > 0 \text{ pour } |x| > \beta, \quad \lambda''(x) < 0 \text{ pour } |x| < \beta.$$

Les deux changements de signe donnent deux points d'inflexion. Pour une racine z de $z^2 - 4z + 1$, on a $(1 + z)^2 = 6z$, donc $\lambda(x) = 1/6$ à ces points. Ainsi

$$\lambda \text{ est paire, points d'inflexion } \left(-\ln(2 + \sqrt{3}), \frac{1}{6}\right) \text{ et } \left(\ln(2 + \sqrt{3}), \frac{1}{6}\right).$$

La courbe est convexe sur les deux intervalles extérieurs, concave entre ces points, et admet l'axe des abscisses comme asymptote aux deux extrémités. Son allure est la suivante.



4. Moments et simulation de $\mathcal{L}(r, s)$.

- a) Pour $x \geq 0$, on a $0 < \lambda(x) \leq e^{-x}$. Pour tout entier $m \geq 0$, l'intégrale $\int_0^{+\infty} x^m e^{-x} dx$ converge : sur un segment, la fonction est continue, et pour x assez grand, $x^m e^{-x} \leq e^{-x/2}$ par les croissances comparées. La parité de λ et la comparaison donnent donc

$$\int_{-\infty}^{+\infty} |x|^m \lambda(x) dx < +\infty.$$

Ainsi Z possède des moments de tout ordre entier. En particulier son espérance existe et vaut 0, car $x\lambda(x)$ est impaire ; les deux intégrales convergent absolument. Si $Y = r + sZ$, alors pour $m \geq 1$,

$$|Y|^m \leq (|r| + s|Z|)^m \leq 2^m (|r|^m + s^m |Z|^m).$$

La borne assure l'existence de ses moments ; par linéarité, $\mathbb{E}(Y) = r + s\mathbb{E}(Z) = r$. Donc

$$Y \sim \mathcal{L}(r, s) \text{ possède des moments de tout ordre, } \mathbb{E}(Y) = r.$$

- b) Si U suit la loi uniforme sur $]0, 1[$, la croissance de $L = \Lambda^{-1}$ donne, pour $x \in \mathbb{R}$,

$$\mathbb{P}(L(U) \leq x) = \mathbb{P}(U \leq \Lambda(x)) = \Lambda(x).$$

Ainsi $L(U)$ suit la loi logistique standard, et $r + sL(U)$ suit $\mathcal{L}(r, s)$. Avec les imports usuels `numpy as np` et `numpy.random as rd`, on écrit

```
def grandlogis(n, p, r, s):
    X = np.zeros((n, p))
    for i in range(n):
        for j in range(p):
            u = rd.random()
            while u == 0 or u == 1:
                u = rd.random()
            X[i, j] = r+s*np.log(u/(1-u))
    return X
```

La boucle de contrôle garde l'argument du logarithme strictement positif. Chaque coefficient utilise un nouveau tirage indépendant. Ainsi

`grandlogis(n,p,r,s)` renvoie un tableau $n \times p$ de réalisations indépendantes de $\mathcal{L}(r, s)$.

- c) On produit un grand nombre N de simulations standard, par exemple avec `grandlogis(N,1,0,1)`, puis on calcule leur variance empirique :

```
Z = grandlogis(100000, 1, 0, 1)
```

```
variance = np.var(Z)
```

Pour des réalisations $Z_1, \dots, Z_N$, le calcul correspond à

$$V_N = \frac{1}{N} \sum_{i=1}^N Z_i^2 - \left(\frac{1}{N} \sum_{i=1}^N Z_i \right)^2.$$

La question a) assure notamment un moment d'ordre 4. La loi faible des grands nombres s'applique donc à Z_i et à Z_i^2 : la première moyenne converge en probabilité vers 0, et la seconde vers $\mathbb{E}(Z^2)$. Par composition par le carré et différence de suites convergeant en probabilité, V_N converge en probabilité vers $\mathbb{E}(Z^2) = \mathbb{V}(Z)$. Ainsi

`np.var(grandlogis(N,1,0,1))`
fournit une approximation de $\mathbb{V}(Z)$ pour N grand.

5. Logarithme d'un quotient d'exponentielles.

- a) Posons $X_1 = \ln U_1$ et $X_2 = -\ln U_2$. Les deux variables sont indépendantes, chacune étant une fonction d'une des deux exponentielles indépendantes. Pour $t \in \mathbb{R}$,

$$\mathbb{P}(\ln U_1 \leq t) = \mathbb{P}(U_1 \leq e^t) = 1 - e^{-e^t}.$$

Cette fonction de répartition est de classe $\mathcal{C}^1$; sa dérivée donne une densité de X_1 , $h(t) = e^t e^{-e^t}$. La transformation affine $X_2 = -\ln U_2$ donne la densité $k(t) = h(-t) = e^{-t} e^{-e^{-t}}$. La somme $X_1 + X_2 = \ln(U_1/U_2)$ a, par le théorème de convolution, une densité donnée par

$$\begin{aligned} q(x) &= \int_{-\infty}^{+\infty} h(x-t)k(t) dt \\ &= e^x \int_{-\infty}^{+\infty} e^{-2t} \exp(-(1+e^x)e^{-t}) dt. \end{aligned}$$

Pour x fixé, le changement de variable $v = e^{-t}$ est de classe $\mathcal{C}^1$ et bijectif de $\mathbb{R}$ sur $]0, +\infty[$. Avec $dt = -dv/v$ et l'inversion des bornes, il transforme l'intégrale en une intégrale impropre de même nature :

$$q(x) = e^x \int_0^{+\infty} v e^{-(1+e^x)v} dv.$$

Pour en établir la convergence et la valeur, fixons $b = 1 + e^x > 0$. Sur $[0, B]$, les fonctions $U(v) = v$ et $V(v) = -e^{-bv}/b$ sont de classe $\mathcal{C}^1$, avec $U'(v) = 1$ et $V'(v) = e^{-bv}$. L'intégration par parties donne

$$\begin{aligned} \int_0^B v e^{-bv} dv &= \left[-\frac{v}{b} e^{-bv} \right]_0^B + \frac{1}{b} \int_0^B e^{-bv} dv \\ &= -\frac{B}{b} e^{-bB} + \frac{1 - e^{-bB}}{b^2} \rightarrow \frac{1}{b^2}, \end{aligned}$$

par les croissances comparées. La convolution converge donc pour tout x , et

$$q(x) = \frac{e^x}{(1 + e^x)^2} = \frac{e^{-x}}{(1 + e^{-x})^2} = \lambda(x).$$

Par conséquent,

$\ln \left(\frac{U_1}{U_2} \right) \text{ suit la loi logistique standard.}$

b) Si V est uniforme sur $]0, 1[$, alors pour $t \geq 0$,

$$\mathbb{P}(-\ln V > t) = \mathbb{P}(V < e^{-t}) = e^{-t}.$$

Ainsi $-\ln V$ suit la loi exponentielle de paramètre 1. Deux tirages uniformes indépendants fournissent donc deux exponentielles indépendantes, dont le logarithme du quotient se simule ainsi :

```
def logis_standard():
    v1 = rd.random()
    while v1 == 0 or v1 == 1:
        v1 = rd.random()
    v2 = rd.random()
    while v2 == 0 or v2 == 1:
        v2 = rd.random()
    u1 = -np.log(v1)
    u2 = -np.log(v2)
    return np.log(u1/u2)
```

$\text{logis_standard() simule la loi logistique standard.}$

Partie II – Variance de la loi logistique standard

On utilise l'identité admise dans l'énoncé pour calculer la somme des inverses des carrés, puis la variance de la loi logistique standard. Cette loi a pour densité $\lambda(x) = e^{-x}/(1 + e^{-x})^2$ et pour espérance 0, d'après la partie I.

Corrigé – Partie II – Variance de la loi logistique standard.**8. Somme de la série des inverses des carrés.**

- a) Sur $[0, \pi/2]$, la fonction $x \mapsto x - \sin x$ a pour dérivée $1 - \cos x \geq 0$ et vaut 0 en 0. La fonction $x \mapsto \tan x - x$ a pour dérivée $1/\cos^2 x - 1 \geq 0$ et vaut aussi 0 en 0. Ainsi $\sin x \leq x \leq \tan x$. Pour $0 < x < \pi/2$, les trois nombres sont strictement positifs ; on peut prendre leurs inverses puis leurs carrés. Comme $1/\tan^2 x = 1/\sin^2 x - 1$, on obtient

$$\sin x \leq x \leq \tan x, \quad \frac{1}{\sin^2 x} - 1 \leq \frac{1}{x^2} \leq \frac{1}{\sin^2 x}.$$

- b) Pour $1 \leq k \leq n$, le nombre $k\pi/(2n+1)$ appartient à $]0, \pi/2[$. En sommant les inégalités précédentes et en utilisant l'identité admise,

$$\frac{2n(n+1)}{3} - n \leq \frac{(2n+1)^2}{\pi^2} \sum_{k=1}^n \frac{1}{k^2} \leq \frac{2n(n+1)}{3}.$$

Ainsi

$$\frac{n(2n-1)}{3} \leq \frac{(2n+1)^2}{\pi^2} \sum_{k=1}^n \frac{1}{k^2} \leq \frac{2n(n+1)}{3}.$$

- c) Après multiplication par $\pi^2/(2n+1)^2$, les deux bornes tendent vers $\pi^2/6$. Le théorème d'encadrement montre que la suite des sommes partielles converge, ce qui établit à la fois la convergence et la valeur de la série :

$$\sum_{k=1}^{+\infty} \frac{1}{k^2} = \frac{\pi^2}{6}.$$

9. Calcul de la variance logistique.

- a) Le moment d'ordre 2 existe par la question 4. La densité est paire et l'espérance est nulle, donc

$$\mathbb{V}(Z) = \mathbb{E}(Z^2) = 2 \int_0^{+\infty} x^2 \lambda(x) dx.$$

Pour $b > 0$, les fonctions $u(x) = x^2$ et $v(x) = -1/(1 + e^x)$ sont de classe $\mathcal{C}^1$ sur $[0, b]$, avec $u'(x) = 2x$ et $v'(x) = \lambda(x)$. Une intégration par parties sur ce segment donne

$$\int_0^b x^2 \lambda(x) dx = -\frac{b^2}{1 + e^b} + 2 \int_0^b \frac{x}{1 + e^x} dx.$$

Le terme au bord tend vers 0, puisque $0 \leq b^2/(1 + e^b) \leq b^2 e^{-b} \rightarrow 0$. L'intégrale de droite converge : son intégrande est continu en 0 et, pour $x \geq 0$, il est compris entre 0 et xe^{-x} , dont l'intégrale converge. En passant à la limite,

$$\mathbb{V}(Z) = 4 \int_0^{+\infty} \frac{xe^{-x}}{1 + e^{-x}} dx.$$

b) Pour $q \geq 0$, l'identité géométrique finie est

$$\frac{1}{1+q} = \sum_{k=0}^n (-1)^k q^k + \frac{(-1)^{n+1} q^{n+1}}{1+q}.$$

On l'applique à $q = e^{-x}$, puis on multiplie par xe^{-x} . L'intégrale de chaque terme converge : pour tout $a > 0$, celle de xe^{-ax} converge, et la valeur absolue du reste est majorée par $xe^{-(n+2)x}$. On peut donc intégrer cette somme *finie* et obtenir

$$\begin{aligned} \int_0^{+\infty} \frac{xe^{-x}}{1+e^{-x}} dx &= \sum_{k=0}^n (-1)^k \int_0^{+\infty} xe^{-(k+1)x} dx + I_n, \\ I_n &= (-1)^{n+1} \int_0^{+\infty} \frac{xe^{-(n+2)x}}{1+e^{-x}} dx. \end{aligned}$$

c) Calculons d'abord l'intégrale pour $a > 0$. Sur $[0, b]$, les fonctions $u(x) = x$ et $v(x) = -e^{-ax}/a$ sont de classe $\mathcal{C}^1$, de dérivées 1 et e^{-ax} . L'intégration par parties donne

$$\int_0^b xe^{-ax} dx = -\frac{be^{-ab}}{a} + \frac{1-e^{-ab}}{a^2}.$$

En faisant tendre b vers $+\infty$, on obtient $\int_0^{+\infty} xe^{-ax} dx = 1/a^2$. Par conséquent,

$$|I_n| \leq \int_0^{+\infty} xe^{-(n+2)x} dx = \frac{1}{(n+2)^2} \rightarrow 0.$$

L'égalité de b), avec ce calcul de chacune des intégrales, devient

$$\int_0^{+\infty} \frac{xe^{-x}}{1+e^{-x}} dx = \sum_{k=0}^n \frac{(-1)^k}{(k+1)^2} + I_n.$$

On passe à la limite dans cette égalité entre réels, ce qui donne

$$I_n \rightarrow 0, \quad \int_0^{+\infty} \frac{xe^{-x}}{1+e^{-x}} dx = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)^2}.$$

d) La série des valeurs absolues converge par 8.c). On sépare donc les indices pairs et impairs :

$$\begin{aligned} \sum_{j=1}^{+\infty} \frac{(-1)^{j-1}}{j^2} &= \sum_{j=1}^{+\infty} \frac{1}{j^2} - 2 \sum_{\ell=1}^{+\infty} \frac{1}{(2\ell)^2} \\ &= \frac{\pi^2}{6} - \frac{1}{2} \frac{\pi^2}{6} = \frac{\pi^2}{12}. \end{aligned}$$

La formule de a) fournit alors

$$\mathbb{V}(Z) = 4 \frac{\pi^2}{12} = \frac{\pi^2}{3}.$$

10. Deux intégrales contenant le logarithme.

a) Au voisinage de 0, on a $\sqrt{x} |\ln x| \rightarrow 0$ et $\sqrt{x} (\ln x)^2 \rightarrow 0$. Il existe donc $\eta \in]0, 1[$ tel que, pour $0 < x \leq \eta$,

$$|\ln x|e^{-x} \leq \frac{1}{\sqrt{x}}, \quad (\ln x)^2 e^{-x} \leq \frac{1}{\sqrt{x}}.$$

L'intégrale de $x^{-1/2}$ converge en 0. Au voisinage de $+\infty$, pour $x \geq 1$, $0 \leq \ln x \leq x$, donc

$$|\ln x|e^{-x} \leq xe^{-x}, \quad (\ln x)^2 e^{-x} \leq x^2 e^{-x}.$$

Ces deux dernières intégrales convergent par comparaison aux moments de la loi exponentielle de paramètre 1. Sur tout segment inclus dans $]0, +\infty[$, les fonctions sont continues. Ainsi

$$\boxed{\int_0^{+\infty} |\ln x| e^{-x} dx < +\infty, \quad \int_0^{+\infty} (\ln x)^2 e^{-x} dx < +\infty.}$$

- b)** Pour U de loi exponentielle de paramètre 1, le théorème de transfert, appliqué aux moments dont l'existence vient d'être établie, donne

$$\mathbb{E}(\ln U) = I, \quad \mathbb{E}((\ln U)^2) = J, \quad \mathbb{V}(\ln U) = J - I^2.$$

Les variables $\ln U_1$ et $\ln U_2$ sont indépendantes, car U_1, U_2 le sont. La question 5.a) donne à $Z = \ln U_1 - \ln U_2$ la loi logistique standard. Les variances de variables indépendantes s'ajoutent, donc

$$\frac{\pi^2}{3} = \mathbb{V}(Z) = \mathbb{V}(\ln U_1) + \mathbb{V}(\ln U_2) = 2(J - I^2).$$

Ainsi

$$\boxed{J - I^2 = \frac{\pi^2}{6}.}$$

Partie III – Estimation à partir de données binaires

La proportion empirique estime $F(\theta)$. On inverse F lorsque cette proportion appartient à $]0, 1[$, en contrôlant séparément les événements où elle vaut 0 ou 1. Le théorème central limite et un encadrement direct de fonctions de répartition donnent ensuite la loi limite de l'estimateur centré.

Corrigé – Partie III – Estimation à partir de données binaires.

Dans cette partie, on pose $p = F(\theta)$ et $q = 1 - p$. Les espérances et probabilités sont calculées sous la loi $\mathbb{P}_\theta$ du modèle.

11. Bijection définie par F .

Puisque f est une densité continue, la fonction de répartition $F(x) = \int_{-\infty}^x f(t) dt$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et $F' = f$. Comme f est strictement positive, F est strictement croissante. Les limites d'une fonction de répartition sont

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

La continuité, la croissance stricte et ces limites permettent d'appliquer le théorème de la bijection :

$$F : \mathbb{R} \longrightarrow]0, 1[\text{ est une bijection.}$$

En particulier, $p = F(\theta) \in]0, 1[$ et sa réciproque F^{-1} est strictement croissante.

12. Théorème central limite pour la proportion empirique.

Les variables Y_j sont indépendantes, de même loi $\mathcal{B}(1, p)$, d'espérance p et de variance $pq > 0$. Le théorème central limite s'applique donc à leur moyenne :

$$\sqrt{n}(\bar{Y}_n - p) \xrightarrow{\mathcal{L}} \mathcal{N}(0, pq).$$

La variance limite est $pq = F(\theta)(1 - F(\theta))$. On utilisera également la linéarité de l'espérance et l'indépendance pour écrire

$$\mathbb{E}_\theta(\bar{Y}_n) = p, \quad \mathbb{V}_\theta(\bar{Y}_n) = \frac{pq}{n}.$$

Ainsi, pour tout $\delta > 0$, l'inégalité de Bienaymé-Tchebychev donne

$$\mathbb{P}_\theta(|\bar{Y}_n - p| \geq \delta) \leq \frac{pq}{n\delta^2}.$$

13. Construction et convergence de T_n .

- a) L'événement complémentaire de $E_n = [0 < \bar{Y}_n < 1]$ est la réunion disjointe de $[\bar{Y}_n = 0]$ et $[\bar{Y}_n = 1]$. Comme les observations sont des Bernoulli, ces événements correspondent respectivement à $Y_1 = \dots = Y_n = 0$ et $Y_1 = \dots = Y_n = 1$. L'indépendance donne

$$\mathbb{P}_\theta(E_n) = 1 - q^n - p^n \longrightarrow 1.$$

En effet, $0 < p, q < 1$, donc les deux suites géométriques tendent vers 0.

- b) i)** Sur E_n , $T_n = F^{-1}(\bar{Y}_n)$. La croissance stricte de F et de sa réciproque donne, pour tout $x \in \mathbb{R}$,

$$\{\omega \in E_n : T_n(\omega) \leq x\} = [\bar{Y}_n \leq F(x)] \cap E_n.$$

Sur E_n^c , on a $T_n = 0$. Par conséquent

$$[T_n \leq x] = \begin{cases} [\bar{Y}_n \leq F(x)] \cap E_n & \text{si } x < 0, \\ ([\bar{Y}_n \leq F(x)] \cap E_n) \cup E_n^c & \text{si } x \geq 0. \end{cases}$$

Les événements du membre de droite appartiennent à $\mathcal{A}$, car $\bar{Y}_n$ est une variable aléatoire. La tribu est stable par intersection, complémentation et union ; donc

$$[T_n \leq x] \in \mathcal{A} \quad (x \in \mathbb{R}).$$

- ii)** L'événement $[\bar{Y}_n \leq F(x)] \cap E_n$ est inclus dans $[T_n \leq x]$, et la différence éventuelle est contenue dans E_n^c . Par croissance et sous-additivité de la probabilité,

$$\begin{aligned} \mathbb{P}_\theta([\bar{Y}_n \leq F(x)] \cap E_n) &\leq \mathbb{P}_\theta(T_n \leq x) \\ &\leq \mathbb{P}_\theta([\bar{Y}_n \leq F(x)] \cap E_n) + 1 - \mathbb{P}_\theta(E_n). \end{aligned}$$

On en déduit aussi un encadrement utile :

$$\begin{aligned} \left| \mathbb{P}_\theta(T_n \leq x) - \mathbb{P}_\theta(\bar{Y}_n \leq F(x)) \right| &\leq \mathbb{P}_\theta(E_n^c) \\ &= p^n + q^n. \end{aligned}$$

En effet, les deux événements coïncident sur E_n et leurs parties situées dans E_n^c ont chacune une probabilité comprise entre 0 et $\mathbb{P}_\theta(E_n^c)$.

- c)** Si $x < \theta$, la croissance stricte de F donne $\delta = p - F(x) > 0$. Alors

$$\mathbb{P}_\theta(\bar{Y}_n \leq F(x)) \leq \mathbb{P}_\theta(|\bar{Y}_n - p| \geq \delta) \leq \frac{pq}{n\delta^2} \longrightarrow 0.$$

L'encadrement de b) ii) et $p^n + q^n \rightarrow 0$ donnent donc $\mathbb{P}_\theta(T_n \leq x) \rightarrow 0$. Si $x > \theta$, posons maintenant $\delta = F(x) - p > 0$. On a

$$1 - \mathbb{P}_\theta(\bar{Y}_n \leq F(x)) = \mathbb{P}_\theta(\bar{Y}_n > F(x)) \leq \frac{pq}{n\delta^2} \longrightarrow 0.$$

Le même encadrement donne alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\theta(T_n \leq x) = \begin{cases} 0 & \text{si } x < \theta, \\ 1 & \text{si } x > \theta. \end{cases}$$

- d)** Fixons $\varepsilon > 0$. L'événement $[|T_n - \theta| > \varepsilon]$ est inclus dans la réunion de $[T_n \leq \theta - \varepsilon]$ et de $[T_n > \theta + \varepsilon]$. Ainsi

$$\mathbb{P}_\theta(|T_n - \theta| > \varepsilon) \leq \mathbb{P}_\theta(T_n \leq \theta - \varepsilon) + 1 - \mathbb{P}_\theta(T_n \leq \theta + \varepsilon) \longrightarrow 0$$

par c). Ceci prouve directement la convergence en probabilité. De plus, T_n dépend des observations $(Y_1, \dots, Y_n)$ et de la fonction connue F , sans faire intervenir la valeur inconnue de θ . Donc

$$\begin{aligned} T_n &\xrightarrow{\mathbb{P}} \theta, \\ (T_n) &\text{ est une suite convergente d'estimateurs de } \theta. \end{aligned}$$

14. Loi limite de l'estimateur centré.

- a) i) La fonction f est continue et strictement positive sur $\mathbb{R}$, donc $x \mapsto 1/f(x)$ est continue en θ . Pour $\varepsilon > 0$, il existe $\delta > 0$ tel que $|x - \theta| < \delta$ implique $|1/f(x) - 1/f(\theta)| < \varepsilon$. En posant $\alpha = \delta/2 > 0$, on obtient

$$\boxed{\forall x \in [\theta - \alpha, \theta + \alpha], \quad \left| \frac{1}{f(x)} - \frac{1}{f(\theta)} \right| \leq \varepsilon.}$$

- ii) Considérons une réalisation de $[|T_n - \theta| \leq \alpha] \cap E_n$. Sur E_n , $F(T_n) = \bar{Y}_n$. Si $\bar{Y}_n = p = F(\theta)$, la bijectivité de F donne $T_n = \theta$ et la définition impose $U_n = 1/f(\theta)$; cette réalisation appartient donc à $B_n(\varepsilon)$. Sinon, $T_n \neq \theta$. La fonction F est continue sur le segment de bornes θ, T_n et dérivable sur son intérieur. Le théorème des accroissements finis donne un réel c situé entre ces deux bornes tel que

$$\bar{Y}_n - p = F(T_n) - F(\theta) = f(c)(T_n - \theta).$$

Il vient $U_n = 1/f(c)$. Comme $|T_n - \theta| \leq \alpha$, on a $c \in [\theta - \alpha, \theta + \alpha]$ et i) donne $|U_n - 1/f(\theta)| \leq \varepsilon$. Dans les deux cas,

$$\boxed{[|T_n - \theta| \leq \alpha] \cap E_n \subset B_n(\varepsilon).}$$

- b) En prenant les événements complémentaires de l'inclusion précédente,

$$\left[\left| U_n - \frac{1}{f(\theta)} \right| > \varepsilon \right] \subset [|T_n - \theta| > \alpha] \cup E_n^c.$$

La probabilité du premier événement à droite tend vers 0 par la question 13 d), et celle du second vaut $p^n + q^n \rightarrow 0$. La sous-additivité donne donc

$$\boxed{U_n \xrightarrow{\mathbb{P}} \frac{1}{f(\theta)}}.$$

Pour toute réalisation, on a également l'identité

$$\sqrt{n}(T_n - \theta) = U_n \sqrt{n}(\bar{Y}_n - p).$$

Si $\bar{Y}_n \neq p$, elle résulte de la définition du quotient. Si $\bar{Y}_n = p$, alors $p \in]0, 1[$ implique E_n , puis $T_n = F^{-1}(p) = \theta$: les deux membres sont donc nuls.

- c) Posons $Z_n = \sqrt{n}(T_n - \theta)$ et $H_n(t) = \mathbb{P}_\theta(\sqrt{n}(\bar{Y}_n - p) \leq t)$. La question 12 donne, pour tout réel t ,

$$H_n(t) \longrightarrow \Phi\left(\frac{t}{\sqrt{pq}}\right),$$

où Φ est la fonction de répartition de la loi normale centrée réduite. Fixons $x \in \mathbb{R}$ et posons

$$t_n = \theta + \frac{x}{\sqrt{n}}, \quad a_n = \sqrt{n}(F(t_n) - F(\theta)).$$

Sur E_n , la croissance stricte de F donne exactement

$$Z_n \leq x \iff T_n \leq t_n \iff \bar{Y}_n \leq F(t_n) \iff \sqrt{n}(\bar{Y}_n - p) \leq a_n.$$

Les deux événements extrêmes coïncident sur E_n . Leur différence de probabilités est donc majorée en valeur absolue par $\mathbb{P}_\theta(E_n^c)$:

$$|\mathbb{P}_\theta(Z_n \leq x) - H_n(a_n)| \leq p^n + q^n.$$

La dérivabilité de F en θ donne $a_n \rightarrow f(\theta)x$. Pour $x = 0$, c'est immédiat car $a_n = 0$; pour $x \neq 0$,

$$a_n = x \frac{F(\theta + x/\sqrt{n}) - F(\theta)}{x/\sqrt{n}} \longrightarrow x f(\theta).$$

Notons $a = x f(\theta)$. Pour tout $\delta > 0$ et pour n assez grand, $a - \delta \leq a_n \leq a + \delta$. Par croissance de H_n ,

$$H_n(a - \delta) - (p^n + q^n) \leq \mathbb{P}_\theta(Z_n \leq x) \leq H_n(a + \delta) + (p^n + q^n).$$

Les fonctions de répartition aux deux seuils fixes convergent vers $\Phi((a - \delta)/\sqrt{pq})$ et $\Phi((a + \delta)/\sqrt{pq})$, et le terme d'erreur tend vers 0. Pour tout $\eta > 0$, la continuité de Φ permet de choisir δ de sorte que ces deux limites diffèrent de $\Phi(a/\sqrt{pq})$ de moins de $\eta/2$. Pour n assez grand, les erreurs de convergence des deux fonctions de répartition et le terme $p^n + q^n$ sont chacun inférieurs à $\eta/4$. L'encadrement donne alors $|\mathbb{P}_\theta(Z_n \leq x) - \Phi(a/\sqrt{pq})| < \eta$. Ainsi, pour tout réel x ,

$$\mathbb{P}_\theta(Z_n \leq x) \longrightarrow \Phi\left(\frac{x f(\theta)}{\sqrt{pq}}\right).$$

Comme $f(\theta) > 0$, la limite est la fonction de répartition d'une loi normale centrée de variance $pq/f(\theta)^2$. Par définition de la convergence en loi,

$$\boxed{\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{F(\theta)(1 - F(\theta))}{f(\theta)^2}\right)}.$$

Partie IV – Régression logistique

On observe des réponses binaires pour k niveaux des facteurs exogènes. Lorsque les colonnes de la matrice $M \in \mathcal{M}_{p,k}(\mathbb{R})$ engendrent $\mathbb{R}^p$, on établit l'unicité d'une approximation par moindres carrés, puis on construit des estimateurs convergents des composantes du paramètre α .

Corrigé – Partie IV – Régression logistique.**15. Moindres carrés et identification du paramètre.**

- a) Le rang de M étant égal à p , ses colonnes $x^{(1)}, \dots, x^{(k)}$ engendrent $\mathbb{R}^p$. Si U représente un vecteur $u \in \mathbb{R}^p$, les coefficients de ${}^tMU$ sont les produits scalaires $\langle x^{(i)}, u \rangle$. Ainsi

$${}^tMU = 0 \implies u \perp \text{Vect}(x^{(1)}, \dots, x^{(k)}) = \mathbb{R}^p \implies U = 0.$$

L'application représentée par tM est donc injective. Si $M{}^tMU = 0$, alors

$$0 = {}^tUM{}^tMU = ({}^tMU){}^tMU = \|{}^tMU\|^2.$$

On obtient ${}^tMU = 0$, puis $U = 0$. La matrice carrée $M{}^tM$, de taille p , a donc un noyau réduit à $\{0\}$, d'où

$$\boxed{M{}^tM \text{ est inversible.}}$$

- b) Fixons $H \in \mathcal{M}_{k,1}(\mathbb{R})$ et posons

$$U_0 = (M{}^tM)^{-1}MH, \quad R_0 = {}^tMU_0 - H.$$

La définition de U_0 donne

$$MR_0 = M{}^tMU_0 - MH = 0.$$

Pour toute matrice colonne $U \in \mathcal{M}_{p,1}(\mathbb{R})$, écrivons $V = U - U_0$. On a

$${}^tMU - H = R_0 + {}^tMV.$$

Le produit scalaire des deux termes du membre de droite est nul, puisque

$$({}^tMV)R_0 = {}^tVMR_0 = 0.$$

En développant le carré de la norme, on obtient donc

$$\|{}^tMU - H\|^2 = \|R_0\|^2 + \|{}^tM(U - U_0)\|^2 \geq \|R_0\|^2.$$

Le second terme s'annule exactement lorsque ${}^tM(U - U_0) = 0$, soit, par l'injectivité établie en a), lorsque $U = U_0$. Ainsi le minimum est atteint en un seul point :

$$\boxed{U_0 = (M{}^tM)^{-1}MH \text{ est l'unique minimiseur,} \\ \min_{U \in \mathcal{M}_{p,1}(\mathbb{R})} \|{}^tMU - H\|^2 = \|{}^tMU_0 - H\|^2.}$$

Cette norme au carré est précisément l'unique coefficient de la matrice scalaire figurant dans l'énoncé.

- c) Supposons maintenant $\text{rg}(M) < p$. L'espace $F = \text{Vect}(x^{(1)}, \dots, x^{(k)})$ est alors un sous-espace strict de $\mathbb{R}^p$. Son orthogonal a pour dimension $p - \dim F > 0$: il existe donc un vecteur non nul $v \in F^\perp$. Pour chaque i , on a

$$\langle \alpha + v, x^{(i)} \rangle = \langle \alpha, x^{(i)} \rangle + \langle v, x^{(i)} \rangle = \langle \alpha, x^{(i)} \rangle.$$

Les deux paramètres distincts α et $\alpha + v$ donnent les mêmes probabilités $\Lambda(\langle \alpha, x^{(i)} \rangle)$, donc les mêmes lois de toutes les variables $Y_{i,n}$. Plus généralement, si A' représente un autre paramètre α' , la bijectivité de Λ donne

$$\begin{aligned} \forall i, \quad \Lambda(\langle \alpha', x^{(i)} \rangle) &= \Lambda(\langle \alpha, x^{(i)} \rangle) \\ \iff {}^t M(A' - A) &= 0. \end{aligned}$$

Ainsi les lois observées déterminent seulement le paramètre à l'ajout près d'un élément du noyau de ${}^t M$. En particulier,

$$\boxed{\text{rg}(M) < p \implies \alpha \text{ n'est pas identifiable à partir de ces lois.}}$$

16. Construction d'estimateurs convergents.

- a) Posons, pour $1 \leq i \leq k$,

$$\theta_i = \langle \alpha, x^{(i)} \rangle.$$

Pour chaque indice i fixé, la suite $(Y_{i,n})_{n \geq 1}$ est constituée de variables indépendantes de loi de Bernoulli de paramètre $\Lambda(\theta_i)$. La fonction Λ est la fonction de répartition de la loi logistique standard, dont la densité est continue et strictement positive sur $\mathbb{R}$. On peut donc appliquer le résultat de la question 13 avec $F = \Lambda$ et $F^{-1} = L$:

$$\forall i \in \llbracket 1, k \rrbracket, \quad T_{i,n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta_i.$$

Fixons des réels $c_1, \dots, c_k$ et posons $C = \sum_{i=1}^k |c_i|$. Si $C = 0$, tous les coefficients sont nuls, et la somme des estimateurs ainsi que le paramètre à estimer sont identiquement nuls. Supposons donc $C > 0$. Pour tout $\varepsilon > 0$, si les k inégalités $|T_{i,n} - \theta_i| \leq \varepsilon/C$ sont satisfaites, alors

$$\left| \sum_{i=1}^k c_i T_{i,n} - \sum_{i=1}^k c_i \theta_i \right| \leq \sum_{i=1}^k |c_i| |T_{i,n} - \theta_i| \leq \varepsilon.$$

Il en résulte l'inclusion d'événements

$$\left[\left| \sum_{i=1}^k c_i (T_{i,n} - \theta_i) \right| > \varepsilon \right] \subset \bigcup_{i=1}^k \left[|T_{i,n} - \theta_i| > \frac{\varepsilon}{C} \right].$$

La majoration de la probabilité d'une réunion par la somme des probabilités donne

$$\mathbb{P} \left(\left| \sum_{i=1}^k c_i (T_{i,n} - \theta_i) \right| > \varepsilon \right) \leq \sum_{i=1}^k \mathbb{P} \left(|T_{i,n} - \theta_i| > \frac{\varepsilon}{C} \right) \xrightarrow[n \rightarrow +\infty]{} 0.$$

La dernière limite résulte de la convergence de chacun des k termes, le nombre k étant fixé. Les variables $T_{i,n}$, puis leurs combinaisons linéaires, sont des statistiques de l'échantillon qui ne dépendent pas des paramètres inconnus. On obtient donc

$$\boxed{\sum_{i=1}^k c_i T_{i,n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \sum_{i=1}^k c_i \langle \alpha, x^{(i)} \rangle,}$$

d'où une suite convergente d'estimateurs de ce paramètre.

b) Revenons à l'hypothèse $\text{rg}(M) = p$ et notons

$$B = (M^t M)^{-1} M = (b_{j,i})_{\substack{1 \leq j \leq p \\ 1 \leq i \leq k}}.$$

La matrice B est déterminée par les niveaux connus des facteurs. Par définition,

$$A_{j,n} = \sum_{i=1}^k b_{j,i} T_{i,n}.$$

La question a), appliquée aux coefficients de la j -ième ligne de B , montre que cette suite converge en probabilité vers $\sum_i b_{j,i} \theta_i$. Or la colonne $\Theta = {}^t(\theta_1, \dots, \theta_k)$ vaut ${}^t M A$. Par conséquent,

$$B\Theta = (M^t M)^{-1} M^t M A = A.$$

La j -ième composante du membre de gauche est donc α_j . Finalement,

$\forall j \in \llbracket 1, p \rrbracket, \quad A_{j,n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \alpha_j,$ $(A_{j,n})_{n \geq 1} \text{ est une suite convergente d'estimateurs de } \alpha_j.$
--