Examen partiel— mercredi 11 mars 2026

Exercice 1 — Soient $(X_1, \ldots, X_n)$ des variables aléatoires réelles de loi $\mathcal{N}(\mu, \sigma^2)$, les deux paramètres étant inconnus. Construire un intervalle de confiance pour μ, de niveau de confiance $1 - \alpha$.

Exercice 2 — On observe n variables aléatoires $U_i = (X_i, Y_i, Z_i)$ iid selon une loi P, et on cherche à tester l'hypothèse nulle H_0 :

$$H_0 : P = \text{Unif}(\mathbb{S}^2)$$

où $\mathbb{S}^2$ est la sphère unité de $\mathbb{R}^3$, c'est-à-dire l'ensemble des vecteurs de $\mathbb{R}^3$ de norme 1 :

$$\mathbb{S}^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}.$$

1. On note Σ la matrice de covariance d'une variable aléatoire uniforme sur la sphère, $U \sim \text{Unif}(\mathbb{S}^2)$. Montrer que sous H_0,

$$\sqrt{n}\frac{\sum_{i=1}^n U_i}{n} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, \Sigma).$$

2. Montrer que $\Sigma = \frac{1}{3} I_3$. Indice : traiter séparément les termes diagonaux et les termes non-diagonaux.
3. En déduire un test de niveau asymptotique $1 - \alpha$ de H_0.

Exercice 3 — On observe une suite d'indicateurs économiques $x_0, x_1, \ldots, x_T$, que l'on suppose liés par une relation temporelle linéaire : $\forall t \in \{1, \ldots, T\}$,

$$x_{t+1} = \alpha t + \beta x_t + \varepsilon_{t+1},$$

où les résidus ε_t sont iid de loi $\mathcal{N}(0, \sigma^2)$.

1. Écrire le modèle sous forme matricielle.
2. Donner l'estimateur des moindres carrés de α et β.
3. En posant $\gamma = 1 - \beta$ et $m = \alpha/\gamma$, le modèle peut s'écrire sous la forme d'un *retour à la moyenne* :

$$x_{t+1} = \gamma \times mt + (1 - \gamma) \times x_t + \varepsilon_{t+1}.$$

   (a) Expliquer le terme de "retour à la moyenne".
   (b) Proposer un estimateur de m et de γ. Sont-ils biaisés ? Sont-ils convergents lorsque $T \to \infty$?
   (c) Proposer un test de l'hypothèse H0 : $\gamma = 0$. Que signifie cette hypothèse ?

Exercice 4 — La *loi gaussienne tronquée* en $\tau \in \mathbb{R}$ est la loi d'une variable aléatoire $\mathscr{N}(0,1)$, conditionnée à être supérieure à τ. On notera $\mathscr{Q}_\tau$ cette loi. L'objectif de cet exercice est d'étudier le problème de *détecter la troncation* : autrement dit, étant donné un échantillon iid $X_1, \ldots, X_n$ et un τ fixé, de détecter si leur loi commune est $P_0 = \mathscr{N}(0,1)$ (hypothèse nulle H0) ou bien $P_1 = \mathscr{Q}(\tau)$ (hypothèse alternative H1).

1. Écrire la loi de l'échantillon sous l'hypothèse nulle et sous l'hypothèse alternative.
2. Calculer l'erreur totale[1] $\alpha_{\tau,n}$ du meilleur test possible et décrire ce test. Comment se comporte cette erreur totale lorsque $n \to \infty$ et τ est fixé ?
3. Dans cette question, on fixe un $\delta > 0$ aussi petit que l'on veut.
   (a) Donner l'équivalent de $\mathbb{P}(\mathscr{N}(0,1) < -t)$ lorsque $t \to \infty$.
   (b) Montrer que lorsque $n \to \infty$, il est essentiellement impossible de détecter une troncation plus petite que $-(1+\delta)\sqrt{2\ln(n)}$, au sens où le meilleur test possible a une erreur totale proche de 100%.
   (c) Montrer que lorsque $n \to \infty$, il est possible de détecter une troncation plus grande que $-(1-\delta)\sqrt{2\ln(n)}$, avec une erreur totale proche de 0%.
4. Lors d'une grande enquête, on demande à $n = 100$ personnes leur quotient intellectuel. Dans la population des personnes interrogées, cet indicateur numérique est conçu pour être distribué selon une loi $\mathscr{N}(100, 15^2)$. Sur les $n = 100$ personnes interrogées, le plus petit QI déclaré est 79. Quelqu'un aurait-il menti ? Quelle est la p-valeur du test ?

1. On rappelle que l'erreur totale est la somme de l'erreur de première espèce et de l'erreur de seconde espèce.

♡ Correction ♡

Exercice 2.

C'est juste le test de Student classique comme dans le cours.

Exercice 2.

La première question est simplement le théorème central-limite pour la loi uniforme sur la sphère. Cette loi est bornée, donc elle est L^2. Elle est évidemment centrée : pour le voir, on peut raisonner par symétrie en se disant que la loi uniforme sur la sphère est symétrique par rapport à l'origine, donc $\mathbb{E}[U] = -\mathbb{E}[U] = 0$. On a donc bien

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \xrightarrow[n\to\infty]{\text{loi}} \mathscr{N}(0, \mathrm{Var}(U)).$$

Pour calculer la matrice de covariance d'une loi uniforme sur la sphère, il y a beaucoup de façons de faire. Je propose d'utiliser la représentation suivante : $\xi \sim \mathscr{N}(0, I_P)$ alors $\xi/|\xi|$ est uniforme sur la sphère, puisque que la loi normale standard est invariante par les rotations. Ainsi,

$$\Sigma_{1,1} = \mathbb{E}\left[\frac{\xi_1^2}{|\xi|^2}\right].$$

On voit qu'on peut échanger le rôle de ξ_1 et ξ_2 sans changer la loi de ξ, donc on a aussi

$$\Sigma_{2,2} = \Sigma_{1,1} = \Sigma_{3,3}.$$

En sommant les trois on voit que

$$\Sigma_{1,1} + \Sigma_{2,2} + \Sigma_{3,3} = \mathbb{E}\left[\frac{\xi_1^2 + \xi_2^2 + \xi_3^2}{|\xi|^2}\right] = \mathbb{E}\left[\frac{|\xi|^2}{|\xi|^2}\right] = 1.$$

On en déduit immédiatement qu'on a $\Sigma_{i,i} = 1/3$. Quant aux termes non diagonaux, prenons $\Sigma_{1,2}$. On a

$$\Sigma_{1,2} = \mathbb{E}\left[\frac{\xi_1\xi_2}{|\xi|^2}\right].$$

Comme la loi de ξ_1 est symétrique, on peut remplacer ξ_1 par $-\xi_1$ et donc

$$\Sigma_{1,2} = -\mathbb{E}\left[\frac{\xi_1\xi_2}{|\xi|^2}\right] = -\Sigma_{1,2}.$$

Donc c'est zéro, et c'est la même chose pour les autres termes non diagonaux.

Pour faire un test de H_0, on peut par exemple prendre la norme au carré. D'après la question 1, on a

$$\left|\frac{1}{\sqrt{n} \times \sqrt{1/3}} \sum_{i=1}^{n} U_i\right|^2 \xrightarrow[n\to\infty]{\text{loi}} \chi^2(3).$$

Ainsi,

$$\frac{3}{n}\left\{\left(\sum_{i=1}^n X_i\right)^2+\left(\sum_{i=1}^n Y_i\right)^2+\left(\sum_{i=1}^n Z_i\right)^2\right\}\xrightarrow[n\to\infty]{\text{loi}}\chi^2(3).$$

En notant C cette statistique, on rejettera H_0 si $C>\kappa_{3,1-\alpha}$ où $\kappa_{3,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2(3)$.

Exercice 3.

Le modèle s'écrit $Y=X\theta+\varepsilon$, où

$$Y=\begin{pmatrix}x_1\\x_2\\\vdots\\x_T\end{pmatrix}\in\mathbb{R}^{T,1},\quad X=\begin{pmatrix}0&x_0\\1&x_1\\\vdots&\vdots\\T-1&x_{T-1}\end{pmatrix}\in\mathbb{R}^{T,2},\quad \theta=\begin{pmatrix}\alpha\\\beta\end{pmatrix}\in\mathbb{R}^2,\quad \varepsilon=\begin{pmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_T\end{pmatrix}\in\mathbb{R}^T.$$

et $\varepsilon=(\varepsilon_1,\dots,\varepsilon_T)^\top$ sont des vecteurs colonnes, et $\varepsilon\sim\mathcal{N}(0,\sigma^2 I_T)$. L'estimateur des moindres carrés de θ est donné par

$$\hat\theta=(X^\top X)^{-1}X^\top Y.$$

Calculons exactement ces termes. On sait que

$$\sum_{t=0}^{T-1}t^2=\frac{(T-1)T(2T-1)}{6}.$$

Si on note $S=(T-1)(2T-1)/6$ et $m_0=\frac{1}{T}\sum_{t=0}^{T-1}tx_t$ et $v_0=\frac{1}{T}\sum_{t=0}^{T-1}x_t^2$, on a donc

$$X^\top X=T\times\begin{pmatrix}S&m_0\\m_0&v_0\end{pmatrix}$$

et donc

$$(X^\top X)^{-1}=\frac{1}{TD}\begin{pmatrix}v_0&-m_0\\-m_0&S\end{pmatrix}$$

où $D=Sv_0-m_0^2=\det(X^\top X)$. D'autre part, en notant $m_1=\frac{1}{T}\sum_{t=1}^{T}(t-1)x_t$ et $c_1=\frac{1}{T}\sum_{t=0}^{T-1}x_tx_{t+1}$, on a

$$X^\top Y=T\times\begin{pmatrix}m_1\\c_1\end{pmatrix}.$$

Donc en faisant la multiplication matricielle, on obtient la formule suivante pour $\hat\theta$:

$$\begin{aligned}\hat\theta&=(X^\top X)^{-1}X^\top Y\\&=\frac{1}{D}\begin{pmatrix}v_0&-m_0\\-m_0&S\end{pmatrix}\begin{pmatrix}m_1\\c_1\end{pmatrix}\\&=\frac{1}{D}\begin{pmatrix}v_0m_1-m_0c_1\\Sc_1-m_0m_1\end{pmatrix}.\end{aligned}$$

En revenant aux formules exactes pour m_0, v_0, m_1, c_1, on obtient finalement

$$\hat{\alpha} = \frac{\left(\sum_{t=1}^{T}(t-1)x_t\right)\left(\sum_{t=0}^{T-1} x_t^2\right) - \left(\sum_{t=0}^{T-1} tx_t\right)\left(\sum_{t=0}^{T-1} x_t x_{t+1}\right)}{S\left(\sum_{t=1}^{T} x_t^2\right) - \left(\sum_{t=0}^{T-1} tx_t\right)^2}$$

et

$$\hat{\beta} = \frac{S\left(\sum_{t=0}^{T-1} x_t x_{t+1}\right) - \left(\sum_{t=0}^{T-1} tx_t\right)\left(\sum_{t=1}^{T}(t-1)x_t\right)}{S\left(\sum_{t=1}^{T} x_t^2\right) - \left(\sum_{t=0}^{T-1} tx_t\right)^2}.$$

Le terme $\gamma(mt)+(1-\gamma)x_t$ est une moyenne pondérée entre une tendance qui ne dépend que de t, à savoir mt (un comportement moyen) et la valeur x_t. Lorsque x_t s'éloigne beaucoup de la tendance, la moyenne pondérée a tendance à forcer x_{t+1} à revenir vers la tendance.

Comme $\gamma = 1-\beta$, un estimateur sans biais de γ est donné par $\hat{\gamma} = 1-\hat{\beta}$. On connaît même sa loi, $\mathscr{N}(\gamma, \sigma^2\ell_\beta^2)$ où ℓ_β^2 est le second terme diagonal de $(X^\top X)^{-1}$ (on l'a calculé plus haut). Pour le terme de tendance m, on propose simplement

$$\hat{m} = \frac{\hat{\alpha}}{\hat{\gamma}} = \frac{\hat{\alpha}}{1-\hat{\beta}}.$$

Il n'y a plus de raison que ce terme soit sans biais. En fait, on pourrait calculer explicitement la loi de $\hat{m}$ car c'est le rapport de deux lois gaussiennes corrélées, mais la formule est un peu compliquée.

L'hypothèse $\gamma = 0$ signifie que le modèle ne dépend pas vraiment du temps et se comporte exclusivement comme une marche aléatoire, $x_{t+1} = x_t + \varepsilon_{t+1}$. Le test est simplement un test de $\beta = 1$ dans le modèle linéaire, donc en notant $\hat{\sigma}^2$ l'estimateur de la variance (somme des carrés des résidus divisée par $T-2$) alors on voit que

$$\frac{\hat{\beta}-1}{\sqrt{\hat{\sigma}^2\ell_\beta^2}} \sim \mathscr{T}(T-2).$$

En notant $t_{1-\alpha,T-2}$ le quantile symétrique d'ordre $1-\alpha$ de la loi $\mathscr{T}(T-2)$, on rejettera H_0 si

$$|\hat{\beta}-1| > t_{1-\alpha,T-2}\sqrt{\hat{\sigma}^2\ell_\beta^2}.$$

Exercice 4.

Notons

$$\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-x^2/2}dx$$

et $\Psi(t) = 1-\Phi(t)$ (c'est la fonction de survie). La densité de la gaussienne tronquée en τ est simplement

$$f_\tau(x) = \frac{1}{\sqrt{2\pi}\Psi(\tau)}e^{-x^2/2}\mathbf{1}_{x>\tau}.$$

Sous H1, la densité de l'échantillon est donc $p_1(x) = f_\tau(x_1) \times \cdots \times f_\tau(x_n)$, c'est-à-dire

$$p_1(x) = \frac{1}{(2\pi)^{n/2}\Psi(\tau)^n} e^{-\sum_{i=1}^n x_i^2/2} \mathbf{1}_{\min x_i \geqslant \tau}.$$

Le test optimal est donné par la région de rejet $p_1(x)/p_0(x) > 1$, c'est-à-dire après simplification, $\mathbf{1}_{\min x_i \geqslant \tau} > \Psi(\tau)^n$. Comme $\Psi(\tau) < 1$, cette inégalité est vérifiée si et seulement si $\min x_i \geqslant \tau$. Le test optimal consiste donc à rejeter l'hypothèse nulle si le minimum des X_i est plus grand que τ. Si un des x_i est plus petit que τ, on ne rejette pas l'hypothèse nulle.

L'erreur de première espèce est donc la probabilité que l'événement de rejet se produise sous H0, et c'est

$$\mathbb{P}(\min X_i \geqslant \tau) = \mathbb{P}(\forall i \in \{1, \dots, n\}, X_i \geqslant \tau) = \prod_{i=1}^n \mathbb{P}(X_i \geqslant \tau) = \Psi(\tau)^n.$$

L'erreur de seconde espèce est la probabilité que l'événement de rejet n'arrive pas sous H1. C'est impossible : sous H1, les X_i sont tronquées, donc le minimum ne peut pas être plus petit que τ. L'erreur totale est donc égale à

$$\alpha_{\tau,n} = \Psi(\tau)^n.$$

Lorsque $n \to \infty$, $\Psi(\tau)^n$ converge vers 0, donc l'erreur totale converge vers 0.

On veut maintenant vérifier que si $\tau = a_n := -(1+\delta)\sqrt{2\ln(n)}$, alors $\Phi(\tau)^n \to 1$, tandis que si $\tau = b_n := -(1-\delta)\sqrt{2\ln(n)}$, alors $\Phi(\tau)^n \to 0$.

On rappelle l'équivalent suivant, valable lorsque $t \to +\infty$:

$$\Phi(-t) = \mathbb{P}(\mathscr{N}(0,1) < -t) \sim g(t) = \frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t}.$$

On voit très facilement que $g(a_n) \to 0$ lorsque $n \to \infty$, donc

$$\ln(\Psi(a_n)^n) = n \ln(1 - \Phi(a_n)) \sim -ng(a_n)$$

et en utilisant la définition de a_n, on voit que

$$ng(a_n) = n^{1-(1+\delta)^2} \frac{1}{a_n\sqrt{2\pi}}.$$

Comme $1-(1+\delta)^2 < 0$, cette quantité tend vers 0. On en déduit directement que $\alpha_{a_n,n} = \Psi(a_n)^n \to e^0 = 1$. L'erreur totale converge vers 100% : le test optimal va *systématiquement* rejeter l'hypothèse nulle, tout simplement parce que sous l'hypothèse nulle, le minimum des X_i est forcément plus grand que a_n.

Un raisonnement similaire montre que

$$\ln(\Psi(b_n)^n) \sim n^{1-(1-\delta)^2} \frac{1}{b_n\sqrt{2\pi}}.$$

Or, comme b_n est d'ordre $-\sqrt{\ln n}$ et que $1-(1-\delta)^2 > 0$, cette quantité tend vers $-\infty$. On en déduit que $\Psi(b_n)^n$ tend vers $e^{-\infty} = 0$.

Dans cet exercice, on vient en réalité de démontrer un résultat très puissant : si les X_i sont iid de loi $\mathscr{N}(0,1)$, alors leur maximum converge vers $\sqrt{2\ln(n)}$, au sens où

$$\mathbb{P}\left(\left|\frac{\max X_i}{\sqrt{2\ln(n)}}-1\right|>\delta\right)\to 0$$

et pareil pour le minimum qui converge vers $-\sqrt{2\ln(n)}$. En fait, le lemme de Borel-Cantelli peut être utilisé pour montrer une convergence presque sûre et les fluctuations sont d'ordre $\sqrt{\ln\ln n}$.

Par conséquent, sous l'hypothèse nulle, tronquer les X_i à une valeur encore plus petite que $-(1+\delta)\sqrt{2\ln(n)}$ est essentiellement impossible de détecter, puisqu'il n'y aura rien à tronquer. Réciprochement, comme le minimum va être très proche de $-\sqrt{2\ln(n)}$, tronquer les X_i à une valeur comme $-(1-\delta)\sqrt{2\ln(n)}$ se verra immédiatement.

Dans le cas d'une loi $\mathscr{N}(\mu,\sigma^2)$, le raisonnement ci-dessus montre que le minimum devrait être d'ordre $\mu-\sigma\sqrt{2\ln(n)}$. Avec $\mu=100$ et $\sigma=15$ et $n=100$, on a $\sqrt{2\ln(100)}=\sqrt{4(\ln 2+\ln 5)}\approx 2\sqrt{0.7+1.6}=2\sqrt{2.3}\approx 2\times 1.5=3$. Donc le minimum devrait être d'ordre

$$100-15\times 3=100-45=55.$$

79 est beaucoup plus grand, donc il est probable que parmi les gens interrogés, ceux qui ont un QI trop faible ont menti. La p-valeur est ici la probabilité pour que le minimum de n gaussiennes $\mathscr{N}(10,15^2)$ indépendantes soit plus grand que 79, c'est-à-dire

$$\mathbb{P}(\min X_i>79)=\mathbb{P}(X_1\geqslant 79)^{100}=(1-\Psi((79-100)/15))^{100}=(1-\Phi(-1.5))^{100}.$$

On sait que $\Phi(-1.5)$ est quelque part entre 5% et 10% (en fait c'est 8.6) donc sans surprise la p-valeur ci-dessus sera vraiment extrêmement petite, inférieure à 0.95^{100} ce qui est vraiment très proche de zéro.