

STATISTIQUE MATHÉMATIQUE – EXAMEN

M1 Mathématiques Fondamentales, Université Paris-Saclay
2024-2025

Vendredi 18 avril 2025
9h – 12h

Une feuille A4 recto-verso manuscrite est autorisée en tant que support. La calculatrice n'est pas autorisée.

Avant de commencer :

- *Le sujet comporte deux problèmes totalement indépendants. Il n'est pas attendu de traiter l'intégralité du sujet pour avoir la note maximale.*
- *Des résultats pourront être admis d'une question sur l'autre, à la condition de l'écrire clairement.*
- *Les questions marquées d'une astérisque sont des questions bonus, dans laquelle toute trace de recherche même non aboutie pourra rapporter des points. Elles sont bien sûr facultatives.*
- *Une attention particulière sera portée aux questions d'interprétation, à la rigueur et à la précision de la rédaction.*

Bon courage, et faites vous plaisir !

Problème I – Modèle linéaire avec bruits corrélés

On se place dans un modèle linéaire *non supposé gaussien*, où Y , vecteur aléatoire de \mathbb{R}^n , s'écrit

$$Y = X\theta + \varepsilon, \quad (1)$$

avec $X \in \mathbb{R}^{n \times p}$ une matrice déterministe, $\theta \in \mathbb{R}^p$ un vecteur de paramètres inconnus, et ε un vecteur aléatoire de \mathbb{R}^n centré. *Contrairement au cas standard, on suppose ici que le bruit ε a une matrice de covariance $\Sigma \in \mathbb{R}^{n \times n}$ supposée symétrique définie positive, qui n'est pas forcément de la forme $\sigma^2 I_n$.* On suppose que Σ est connue.

1. Montrer que sous ces nouvelles hypothèses, on a toujours :

le modèle (1) est identifiable en $\theta \iff X^T X$ inversible.

Solution. On raisonne exactement comme dans le cours. Pour le sens indirect, supposons $X^T X$ inversible. Alors si $Y = X\theta + \varepsilon = X\theta' + \varepsilon'$, en passant à l'espérance, $X\theta = X\theta'$ et partant $X^T X(\theta - \theta') = 0$ et $\theta = \theta'$. Pour le sens direct, on raisonne par contraposée : si $X^T X$ n'est pas inversible, alors comme $\ker(X^T X) = \ker(X)$, X n'est pas injective. Pour $\eta \in \ker(X) \setminus \{0\}$, on a une symétrie $\theta \mapsto \theta + \eta$ qui rend le modèle non identifiable.

Dans toute la suite de cet exercice, nous nous plaçons sous les hypothèses d'identifiabilité de la question précédente. On note :

$$\hat{\theta} := \arg \min_{\theta \in \mathbb{R}^p} \|Y - X\theta\|^2,$$

avec $\|\cdot\|$ norme euclidienne usuelle sur \mathbb{R}^n . On rappelle qu'il est unique par hypothèse, et dans cet exercice il est appelé *estimateur des moindres carrés ordinaire*.

2. Rappeler la forme close de $\hat{\theta}$, et déterminer l'espérance puis la matrice de covariance de $\hat{\theta}$ en fonction de θ , Σ et X . *Solution.* On a d'après le cours que

$$\hat{\theta} = (X^T X)^{-1} X^T Y,$$

ce qui entraîne

$$\mathbb{E}[\hat{\theta}] = (X^T X)^{-1} X^T \mathbb{E}[Y] = (X^T X)^{-1} X^T X\theta = \theta$$

et

$$\text{Var}[\hat{\theta}] = (X^T X)^{-1} X^T \text{Var}[Y] X (X^T X)^{-1} = (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1}$$

Notons $\|\cdot\|_{\Sigma}$ la norme euclidienne sur \mathbb{R}^n définie pour tout $u \in \mathbb{R}^n$ par $\|u\|_{\Sigma}^2 = u^T \Sigma^{-1} u$. On considère le nouveau problème d'optimisation :

$$\arg \min_{\theta \in \mathbb{R}^p} \|Y - X\theta\|_{\Sigma}^2. \quad (2)$$

3. Montrer que $X^T \Sigma^{-1} X$ est symétrique définie positive. On pourra noter $S \in \mathbb{R}^{n \times n}$ la matrice symétrique définie positive telle que $S^2 = \Sigma^{-1}$ (dont on admettra l'existence et l'unicité). *Solution.* La symétrie est évidente. Soit $u \in \mathbb{R}^p$. On a $u^T X^T \Sigma^{-1} X u = u^T X^T S^2 X u = \|S X u\|^2 \geq 0$, et qui vaut 0 si et seulement si $S X u = 0$ si et seulement si $X u = 0$ (car S est inversible), si et seulement si $u = 0$ par identifiabilité (X est injective).

4. Prouver que le problème (2) admet une unique solution appelée *estimateur des moindres carrés généralisé*, notée $\hat{\theta}_g$, qui vaut

$$\hat{\theta}_g = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y.$$

Solution. Notons, pour $\theta \in \mathbb{R}^p$, $f(\theta) = \|Y - X\theta\|_{\Sigma}^2 = (Y - X\theta)^T \Sigma^{-1} (Y - X\theta)$. On a que f est de classe C^∞ , faisons comme dans le cours et calculons son gradient

$$\begin{aligned} \nabla_{\theta} f(\theta) &= \nabla_{\theta} (Y - X\theta)^T \Sigma^{-1} (Y - X\theta) \\ &= \nabla_{\theta} (Y^T \Sigma^{-1} Y - 2(X^T \Sigma^{-1} Y)^T \theta + \theta^T X^T \Sigma^{-1} X \theta) \\ &= -2X^T \Sigma^{-1} Y + 2X^T \Sigma^{-1} X \theta, \end{aligned}$$

qui s'annule en $\hat{\theta}_g = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$, et l'inverse a bien un sens grâce à la question précédente. Justifions que ce point critique est un minimum global. Cela est établi en prenant la hessienne de f qui n'est autre que

$$Hf(\theta) = 2X^T \Sigma^{-1} X,$$

qui est symétrique définie positive d'après la question précédente. La fonction f est donc convexe : elle admet un minimum global en $\hat{\theta}_g = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$.

5. Déterminer l'espérance puis la matrice de covariance de $\hat{\theta}_g$ en fonction de θ, Σ et X .

Solution. On a

$$\mathbb{E}[\hat{\theta}_g] = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \mathbb{E}[Y] = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} X \theta = \theta$$

et

$$\text{Var}[\hat{\theta}_g] = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \text{Var}[Y] \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} = (X^T \Sigma^{-1} X)^{-1}.$$

6. On note \preceq l'ordre partiel sur les matrices symétriques défini dans le cours. En utilisant la matrice S définie à la question 3., montrer que

$$X(X^T \Sigma^{-1} X)^{-1} X^T \preceq \Sigma$$

puis en déduire

$$\text{Var}(\hat{\theta}_g) \preceq \text{Var}(\hat{\theta}).$$

Pour la première inégalité, on pourra prendre $v^T X(X^T \Sigma^{-1} X)^{-1} X^T v$ pour $v \in \mathbb{R}^n$, écrire $v = SS^{-1}v$ et reconnaître une certaine projection orthogonale.

Solution. Pour tout $v \in \mathbb{R}^n$,

$$\begin{aligned} v^T X(X^T \Sigma^{-1} X)^{-1} X^T v &= v^T X(X^T S^T S X)^{-1} X^T v \\ &= (S^{-1}v)^T \underbrace{SX((SX)^T SX)^{-1}(SX)^T}_{=\Pi} S^{-1}v, \end{aligned}$$

où Π est la projection orthogonale sur $\text{Im}(SX)$. En particulier, on a $\|\Pi w\| \leq \|w\|$ pour tout $w \in \mathbb{R}^n$, ce qui implique, par Cauchy-Schwarz,

$$(S^{-1}v)^T \Pi S^{-1}v = \langle S^{-1}v, \Pi S^{-1}v \rangle \leq \|S^{-1}v\| \|\Pi S^{-1}v\| \leq \|S^{-1}v\|^2 = v^T (S^{-1})^2 v = v^T \Sigma v.$$

D'où $X(X^T \Sigma^{-1} X)^{-1} X^T \preceq \Sigma$. On applique l'inégalité des formes quadratiques à

$X(X^T X)^{-1}w \in \mathbb{R}^n$ avec $w \in \mathbb{R}^p$. Cela donne

$$\begin{aligned} w^T(X^T X)^{-1}X^T X(X^T \Sigma^{-1} X)^{-1}X^T X(X^T X)^{-1}w &\leq w^T(X^T X)^{-1}X^T \Sigma X^T X(X^T X)^{-1}w \\ \iff w^T(X^T \Sigma^{-1} X)^{-1}w &\leq w^T(X^T X)^{-1}X^T \Sigma X^T X(X^T X)^{-1}w \\ \iff w^T \text{Var}(\hat{\theta}_g)w &\leq w^T \text{Var}(\hat{\theta})w, \end{aligned}$$

Ce qui conclut la question.

7. A quel théorème du cours le dernier résultat de la question 6 vous fait-il penser ? En s'inspirant de celui-ci, formulez une généralisation de ce théorème au cas où les bruits sont corrélés. On ne demande aucune démonstration. *Solution.* Cela évoque le théorème de Gauss-Markov qui donne un résultat d'optimalité de l'estimateur des moindres carrés dans l'ensemble des estimateurs sans biais et linéaires en Y . Ici, le théorème étendu au cas où les bruits sont corrélés serait de la forme suivante. Dans le modèle linéaire $Y = X\theta + \varepsilon$ identifiable avec $\text{Var}(\varepsilon) = \Sigma$ connue, $\hat{\theta}_g$ est un estimateur sans biais de θ et linéaire en Y . Pour tout autre estimateur $\tilde{\theta}$ de θ sans bruit et linéaire en Y , on a

$$\text{Var}(\hat{\theta}_g) \preceq \text{Var}(\tilde{\theta}).$$

8. On rappelle que Σ est connue. On suppose dans cette question que le modèle est gaussien, i.e. que $\varepsilon \sim \mathcal{N}(0, \Sigma)$. Proposez un intervalle de confiance de probabilité de couverture $1 - \alpha$ pour θ_1 , le premier coefficient de θ . On cherchera un intervalle de diamètre le plus petit possible (bien qu'il ne soit pas demandé de prouver une quelconque optimalité). *Solution.* Dans ce modèle, $\hat{\theta}_g$ ayant une meilleure variance que $\hat{\theta}$, nous allons baser notre intervalle sur $\hat{\theta}_g$. Notons que par stabilité linéaire des gaussiennes, nous avons classiquement, d'après la question 5.,

$$(\hat{\theta}_g)_1 = e_1^T \hat{\theta}_g \sim \mathcal{N}(e_1^T \theta, e_1^T (X^T \Sigma^{-1} X)^{-1} e_1) = \mathcal{N}(\theta_1, [(X^T \Sigma^{-1} X)^{-1}]_{1,1}),$$

donc

$$\frac{(\hat{\theta}_g)_1 - \theta_1}{\sqrt{[(X^T \Sigma^{-1} X)^{-1}]_{1,1}}} \sim \mathcal{N}(0, 1)$$

et un intervalle de confiance de probabilité de couverture $1 - \alpha$ est donné par

$$\left[(\hat{\theta}_g)_1 \pm q_{1-\alpha/2} \sqrt{[(X^T \Sigma^{-1} X)^{-1}]_{1,1}} \right],$$

avec q_β le quantile d'ordre β de la loi $\mathcal{N}(0, 1)$. Notons que Σ est connue donc cet intervalle ne dépend bien que de paramètres connus.

- 9*. On suppose dans cette question que le modèle est gaussien, que la matrice Σ est inconnue mais qu'on a $\|\Sigma\|_{\text{op}} \leq R^2$ avec $R > 0$. Comment pourrait-on répondre à la question 8 ?
- 10*. On ne suppose plus que le modèle est gaussien. Si Σ est inconnue, comment pourrait-on l'estimer ?

Problème II – Plantons des étoiles

"Il faut porter encore en soi un chaos, pour pouvoir mettre au monde une étoile dansante"
(Nietzsche)

Dans ce problème, pour tout $n \geq 1$, $[n]$ désigne l'ensemble $\{1, \dots, n\}$. On note \mathbb{G}_n l'ensemble des graphes sur l'ensemble de sommets $V = [n]$, non orientés, simples et sans boucle¹. Quand le graphe $g \in \mathbb{G}_n$ dont on parle est clair dans le contexte, on note $u \sim v$ (resp. $u \not\sim v$) si $u \in [n]$ et $v \in [n]$ sont reliés (resp. non reliés) dans g par une arête.

Pour $g \in \mathbb{G}_n$, on dit qu'un couple $\mathcal{E} = (u, W)$ où $u \in [n]$ et $W \subseteq [n] \setminus \{u\}$ est une k -étoile de G centrée en u , et l'on note $\mathcal{E} \in G$, si $|W| = k$ et si

$$\forall w \in W, \quad u \sim w.$$

Nous allons considérer des graphes aléatoires². Pour $n \geq 1$, et deux paramètres $p \in [0, 1]$ et $k \geq 0$, où p et k dépendent éventuellement de n , on note :

- $\mathcal{G}(n, p)$ la loi d'un graphe aléatoire G dans \mathbb{G}_n où, indépendamment pour toute paire $\{u, v\}$ de sommets avec $u \neq v \in [n]$, $u \sim v$ (resp. $u \not\sim v$) avec probabilité p (resp. $1 - p$).
- $\mathcal{G}(n, p, k)$ le modèle où G est obtenu de la façon suivante. On tire d'abord $H \sim \mathcal{G}(n, p)$, puis, indépendamment de H , on tire un noeud $u \in [n]$ uniformément au hasard, et un sous ensemble $W \subset [n] \setminus \{u\}$ de cardinal k uniformément au hasard. On note $\mathcal{E}(G) = (u, W)$ et on forme le graphe G en ajoutant à H les arêtes nécessaires de sorte que $\mathcal{E}(G) = (u, W)$ soit une k -étoile de G . On dit qu'on a planté la k -étoile $\mathcal{E}(G)$ dans G .

Notons que par définition, $\mathcal{G}(n, p, 0) = \mathcal{G}(n, p)$. Le but de ce problème est de détecter la présence de k -étoiles plantées dans des graphes, autrement dit, d'étudier le test d'hypothèses suivant :

$$\mathcal{H}_0 : G \sim \mathcal{G}(n, p) \quad \text{contre} \quad \mathcal{H}_1 : G \sim \mathcal{G}(n, p, k). \quad (\star)$$

Lorsque les paramètres n, p, k sont clairs dans le contexte, on note \mathbb{P}_0 (resp. \mathbb{P}_1) la distribution de G sous \mathcal{H}_0 (resp. sous \mathcal{H}_1).

A – Premières questions

1. En donnant des arguments succints et informels, à votre avis, comment la difficulté du problème du test d'hypothèses (\star) varie :
 - à p fixé, lorsque k augmente ?
 - à k fixé, lorsque p augmente ?

Solution. Plus une étoile est grande, plus elle est facile à détecter. On s'attend à ce que le problème soit plus facile lorsque k augmente, à p fixé. Plus il y a d'arêtes dans G qui sont dues au 'hasard', plus l'étoile est difficile à détecter. On s'attend à ce que le problème soit plus difficile lorsque p augmente, à k fixé.

Dans toute la suite, on considère le régime où p dépend de n et s'écrit $p = p(n) = \lambda/n$ avec λ réel strictement positif.

2. Pour une telle paramétrisation de p , le modèle $\mathcal{G}(n, p)$ est dit *de degré moyen d'ordre constant en n* . Justifier cette appellation. *Solution.* Calculons l'espérance du degré d'un noeud quelconque $u \in [n]$. Il y a $n - 1$ voisins possibles, chacun avec probabilité λ/n , soit en espérance $(n - 1)\lambda/n$ qui tend vers λ lorsque $n \rightarrow +\infty$ (une constante, donc).

¹c'est-à-dire que toutes les arêtes de tels graphes sont de la forme $\{u, v\}$ avec $u, v \in [n]$ et $u \neq v$.

²si cela vous semble loin du contexte du cours, observez que l'on peut parfaitement encoder un graphe de \mathbb{G}_n par un vecteur de $\mathbb{R}^{\binom{n}{2}}$ (et même de $\{0, 1\}^{\binom{n}{2}}$). Un graphe aléatoire peut ainsi être vu comme un vecteur aléatoire.

Pour tout $1 \leq k \leq n$, on note $\mathcal{E}(n, k)$ l'ensemble des k -étoiles du graphe complet³ de taille n .

3. Montrer que

$$|\mathcal{E}(n, k)| = n \binom{n-1}{k}.$$

Solution. Il faut choisir le centre de l'étoile (n choix), puis les k voisins du centre, parmi les $n-1$ voisins possibles. Cela donne bien $|\mathcal{E}(n, k)| = n \binom{n-1}{k}$.

B – Calcul de L

4. On rappelle que $p = p(n) = \lambda/n$. Montrer que pour tout graphe $g \in \mathbb{G}_n$,

$$\mathbb{P}_1(G = g) = \frac{1}{|\mathcal{E}(n, k)|} \left(\frac{\lambda}{n}\right)^{-k} \mathbb{P}_0(G = g) \sum_{\mathcal{E} \in \mathcal{E}(n, k)} \mathbf{1}_{\mathcal{E} \in g}.$$

On pourra commencer par calculer $\mathbb{P}_0(G = g)$, puis écrire

$$\mathbb{P}_1(G = g) = \sum_{\mathcal{E} \in \mathcal{E}(n, k)} \mathbb{P}_1(G = g | \mathcal{E}(G) = \mathcal{E}) \mathbb{P}_1(\mathcal{E}(G) = \mathcal{E}).$$

Solution. Notons $e(g)$ le nombre d'arêtes d'un graphe $g \in \mathbb{G}_n$. Par indépendance des arêtes, on a que

$$\mathbb{P}_0(G = g) = p^{e(g)} (1-p)^{\binom{n}{2} - e(g)} = \left(\frac{\lambda}{n}\right)^{e(g)} \left(1 - \frac{\lambda}{n}\right)^{\binom{n}{2} - e(g)}.$$

En suivant l'indication, on écrit

$$\begin{aligned} \mathbb{P}_1(G = g) &= \sum_{\mathcal{E} \in \mathcal{E}(n, k)} \mathbb{P}_1(G = g | \mathcal{E}(G) = \mathcal{E}) \mathbb{P}_1(\mathcal{E}(G) = \mathcal{E}) \\ &= \sum_{\mathcal{E} \in \mathcal{E}(n, k)} \mathbf{1}_{\mathcal{E} \in g} \underbrace{\left(\frac{\lambda}{n}\right)^{e(g)-k}}_{k \text{ arêtes de } \mathcal{E} \text{ fixées}} \left(1 - \frac{\lambda}{n}\right)^{\binom{n}{2} - e(g)} \underbrace{\frac{1}{|\mathcal{E}(n, k)|}}_{\text{proba uniforme sur } \mathcal{E}(n, k)} \\ &= \frac{1}{|\mathcal{E}(n, k)|} \left(\frac{\lambda}{n}\right)^{-k} \mathbb{P}_0(G = g) \sum_{\mathcal{E} \in \mathcal{E}(n, k)} \mathbf{1}_{\mathcal{E} \in g}. \end{aligned}$$

Pour tout graphe $g \in \mathbb{G}_n$, on note

$$L(g) := \frac{\mathbb{P}_1(G = g)}{\mathbb{P}_0(G = g)}$$

et $X_k(g)$ le nombre de k -étoiles de g .

5. Dédurre de la question précédente que pour tout graphe $g \in \mathbb{G}_n$,

$$L(g) = \frac{X_k(g)}{\mathbb{E}_0[X_k(G)]}.$$

³pour rappel, le graphe complet de taille n est le graphe de \mathbb{G}_n dans lequel $u \sim v$ pour tous $u \neq v \in [n]$.

Solution. Notons que $X_k(g) = \sum_{\mathcal{E} \in \mathcal{E}(n,k)} \mathbb{1}_{\mathcal{E} \in g}$. Par linéarité de l'espérance, $\mathbb{E}_0[X_k(G)] = |\mathcal{E}(n,k)| \left(\frac{\lambda}{n}\right)^k$. Ainsi, d'après la question précédente,

$$L(g) = \frac{\mathbb{P}_1(G = g)}{\mathbb{P}_0(G = g)} = \frac{1}{|\mathcal{E}(n,k)|} \left(\frac{\lambda}{n}\right)^{-k} \sum_{\mathcal{E} \in \mathcal{E}(n,k)} \mathbb{1}_{\mathcal{E} \in g} = \frac{X_k(g)}{\mathbb{E}_0[X_k(G)]}$$

6. Si l'on fixe un niveau α , quelle est la forme d'un test ϕ uniformément plus puissant pour le problème (\star) ? On ne demande pas de calculer les constantes qui apparaissent, mais juste de donner sa forme globale. Au vu de la question 5, en quoi ce test ϕ a-t-il une expression intuitive pour répondre au problème (\star) ? *Solution.* D'après le théorème de Neyman-Pearson, un test (randomisé) UPP de niveau α est donné par

$$\phi(G) = \mathbb{1}_{L(G) > t} + c \mathbb{1}_{L(G) = t}$$

où t et c sont calibrées de sorte à satisfaire $\mathbb{E}_0[\phi(G)] = \alpha$. On ne détaille pas ces constantes, mais notons que d'après la forme de L établie à la question précédente, ϕ se réécrit sous la forme

$$\phi(G) = \mathbb{1}_{X_g(G) > t} + c \mathbb{1}_{X_g(G) = t}.$$

Ce test est donc simplement basé sur le comptage du nombre de k -étoiles dans G , ce qui est tout à fait naturel pour détecter leur présence !

Dans toute la suite, on dit que

- La détection forte d'une k -étoile est réalisable si pour tout $n \geq 1$ il existe un test $\phi_n : \mathbb{G}_n \rightarrow \{0, 1\}$ tel que

$$\mathbb{P}_0(\phi_n = 1) + \mathbb{P}_1(\phi_n = 0) \xrightarrow{n \rightarrow +\infty} 0.$$

- La détection faible d'une k -étoile est impossible si pour toute famille de tests $(\phi_n)_{n \geq 1}$ avec $\phi_n : \mathbb{G}_n \rightarrow \{0, 1\}$, on a

$$\mathbb{P}_0(\phi_n = 1) + \mathbb{P}_1(\phi_n = 0) \xrightarrow{n \rightarrow +\infty} 1.$$

C – Un résultat positif

7. Montrer que

$$\mathbb{E}_0[X_k(G)] \leq n \frac{\lambda^k}{k!}.$$

Solution. Nous avons déjà établi que $\mathbb{E}_0[X_k(G)] = |\mathcal{E}(n,k)| \left(\frac{\lambda}{n}\right)^k$, il suffit de conclure en utilisant la majoration classique $\binom{n}{k} \leq \frac{n^k}{k!}$, ce qui donne bien

$$\mathbb{E}_0[X_k(G)] = n \binom{n-1}{k} \left(\frac{\lambda}{n}\right)^k \leq n \frac{\lambda^k}{k!} \left(\frac{n-1}{n}\right)^k \leq n \frac{\lambda^k}{k!}.$$

8. Soit $\varepsilon > 0$. On suppose que $n \geq 3$ et que

$$k = k(n) \geq (1 + \varepsilon) \frac{\log n}{\log \log n}.$$

Montrer qu'alors la détection forte d'une k -étoile est réalisable pour un test ϕ_n très simple que vous explicitez.

On commencera par montrer que sous ces conditions, $\mathbb{P}_0(X_k(G) \geq 1) \xrightarrow[n \rightarrow \infty]{} 0$.

Solution. Suivons l'indication : on travaille à n et sous l'hypothèse sur k , $k \rightarrow \infty$. On a par Stirling,

$$\begin{aligned} \mathbb{P}_0(X_k(G) \geq 1) &\leq n \frac{\lambda^k}{k!} = \exp(\log n + k \log \lambda - \log(k!)) \\ &= \exp(\log n + k \log \lambda - k \log(k) + k + o(k)) \\ &= \exp(\log n - k \log(k) + O(k)). \end{aligned}$$

Il reste à voir que comme $k \geq (1+\varepsilon) \frac{\log n}{\log \log n}$, on a $k \log k \geq (1+\varepsilon) \frac{\log n}{\log \log n} \log \left(\frac{\log n}{\log \log n} \right) \geq (1+\varepsilon/2) \log n$ pour n assez grand. Pour n assez grand, on a donc $\mathbb{P}_0(X_k(G) \geq 1) \leq \exp(-(\varepsilon/2) \log n + O(k))$. On a envie de dire que ce dernier terme tend bien vers 0. Pour être tout à fait rigoureux, il faut comparer k à $\log n$. C'est vrai par exemple si $k > (\varepsilon/4) \log n$. Si $k \geq (\varepsilon/4) \log n$ alors l'hypothèse est encore plus forte et le terme dominant est $k \log k$ domine k et dépasse $C \log n \log \log n$ pour une certaine constante C , qui domine bien les termes en $\log n$. Dans tous les cas, le terme de droite tend vers 0.

D – Une inégalité informationnelle

Soient P_0 and P_1 deux mesures de probabilités définies sur un espace au plus dénombrable Γ . On notera, $P_0(\gamma) = P_0(\{\gamma\})$ pour tout $\gamma \in \Gamma$, et de même pour P_1 . La distance en variation totale entre P_0 et P_1 est définie par

$$d_{\text{TV}}(P_0, P_1) := \sup_{B \subseteq \Gamma} (P_1(B) - P_0(B)).$$

9. Montrer que le sup dans la définition ci-dessus est atteint pour $B = B_{\text{opt}}$ où

$$B_{\text{opt}} := \{\gamma \in \Gamma, P_1(\gamma) > P_0(\gamma)\}$$

et en déduire que

$$d_{\text{TV}}(P_0, P_1) = \frac{1}{2} \sum_{\gamma \in \Gamma} |P_1(\gamma) - P_0(\gamma)|.$$

Solution. Considérons un ensemble $B \subseteq \Gamma$. On va contrôler

$$\Delta := P_1(B_{\text{opt}}) - P_0(B_{\text{opt}}) - (P_1(B) - P_0(B))$$

et montrer que cette quantité est positive (pas besoin de strictement). On a

$$\begin{aligned} \Delta &= \sum_{\gamma \in B_{\text{opt}}} (P_1(\gamma) - P_0(\gamma)) - \sum_{\gamma \in B} (P_1(\gamma) - P_0(\gamma)) \\ &= \sum_{\gamma \in B_{\text{opt}} \setminus B} (P_1(\gamma) - P_0(\gamma)) - \sum_{\gamma \in B \setminus B_{\text{opt}}} (P_1(\gamma) - P_0(\gamma)). \end{aligned}$$

Mais comme $P_1(\gamma) - P_0(\gamma) > 0$ si et seulement si $\gamma \in B_{\text{opt}}$, on a que

$$\sum_{\gamma \in B_{\text{opt}} \setminus B} (P_1(\gamma) - P_0(\gamma)) \geq 0 \quad \text{et} \quad \sum_{\gamma \in B \setminus B_{\text{opt}}} (P_1(\gamma) - P_0(\gamma)) \leq 0.$$

Cela montre que $\Delta \geq 0$, et prouve le résultat voulu.

10. Montrer que, pour X variable aléatoire à valeurs dans Γ ,

$$\inf_{\phi} [P_0(\phi(X) = 1) + P_1(\phi(X) = 0)] = 1 - d_{\text{TV}}(P_0, P_1),$$

où l'inf est pris sur toutes les fonctions mesurables $\phi : \Gamma \rightarrow \{0, 1\}$. Donner une interprétation de ce résultat. *Solution.* Soit $\phi : \mathcal{X} \rightarrow \{0, 1\}$ mesurable. Les ensembles $B := \phi^{-1}(\{0\})$ et $B^c = \phi^{-1}(\{1\})$ sont mesurables et par définition de d_{TV} ,

$$d_{\text{TV}}(P_0, P_1) \geq P_1(B^c) - P_0(B^c) = 1 - P_1(B) - P_0(B^c) = 1 - P_1(\phi(X) = 0) - P_0(\phi(X) = 1)$$

ce qui entraîne

$$1 - d_{\text{TV}}(P_0, P_1) \leq P_1(B^c) - P_0(B^c) = 1 - P_1(B) - P_0(B^c) = P_1(\phi(X) = 0) + P_0(\phi(X) = 1).$$

Cette borne est atteinte lorsque $\phi = \mathbb{1}_{B_{\text{opt}}}$ d'après la question précédente. Ce résultat montre que $1 - d_{\text{TV}}(P_0, P_1)$ est la meilleure somme des erreurs de type I et II possible pour tester P_0 contre P_1 . En d'autres termes, cette quantité $1 - d_{\text{TV}}(P_0, P_1)$ quantifie la difficulté de tester P_0 contre P_1 .

11. On suppose que $P_0(\gamma) > 0$ pour tout $\gamma \in \Gamma$. Pour tout $\gamma \in \Gamma$ on note

$$\ell(\gamma) := \frac{P_1(\gamma)}{P_0(\gamma)}.$$

Soit X une variable aléatoire à valeurs dans Γ . On note E_0 l'espérance relative à P_0 . En utilisant la question 9, montrer que

$$2d_{\text{TV}}(P_0, P_1) \leq \sqrt{E_0[\ell^2(X)] - 1}.$$

Solution. La question 9 montre que $2d_{\text{TV}}(P_0, P_1)$ s'écrit aussi

$$2d_{\text{TV}}(P_0, P_1) = \sum_{\gamma \in \Gamma} |P_1(\gamma) - P_0(\gamma)| = \sum_{\gamma \in \Gamma} P_0(\gamma) |\ell(\gamma) - 1| = E_0[|\ell(X) - 1|] \leq \sqrt{E_0[(\ell(X) - 1)^2]},$$

par Cauchy-Schwarz. Notons de plus que $E_0[(\ell(X) - 1)^2] = E_0[\ell(X)^2] - 2E_0[\ell(X)] + 1$ et on conclut en montrant que $E_0[\ell(X)] = 1$. En effet,

$$E_0[\ell(X)] = \sum_{\gamma \in \Gamma} P_0(\gamma) \frac{P_1(\gamma)}{P_0(\gamma)} = \sum_{\gamma \in \Gamma} P_1(\gamma) = 1.$$

E – Un résultat négatif

Soit $\varepsilon > 0$. On suppose dans toute cette partie que $n \geq 3$ et que

$$k = k(n) \leq (1 - \varepsilon) \frac{\log n}{\log \log n}.$$

12. Rappelons que X_k est définie dans la partie B. En écrivant

$$\mathbb{E}_0[X_k(G)^2] = M_1 + M_2 + M_3$$

avec

$$M_1 := \sum_{\mathcal{E}=(u,W) \in \mathcal{E}(n,k)} \sum_{\mathcal{E}'=(u,W') \in \mathcal{E}(n,k)} \mathbb{P}_0(\mathcal{E} \in G, \mathcal{E}' \in G),$$

$$M_2 := \sum_{\mathcal{E}=(u,W) \in \mathcal{E}(n,k)} \sum_{\substack{\mathcal{E}'=(u',W') \in \mathcal{E}(n,k) \\ u' \in W}} \mathbb{P}_0(\mathcal{E} \in G, \mathcal{E}' \in G),$$

et

$$M_3 := \sum_{\mathcal{E}=(u,W) \in \mathcal{E}(n,k)} \sum_{\substack{\mathcal{E}'=(u',W') \in \mathcal{E}(n,k) \\ u' \notin W}} \mathbb{P}_0(\mathcal{E} \in G, \mathcal{E}' \in G),$$

montrer que

$$\mathbb{E}_0[X_k(G)^2] = \mathbb{E}_0[X_k(G)]^2(1 + o(1)),$$

où le $o(1)$ est pris lorsque $n \rightarrow +\infty$.

On montrera notamment que $M_1 = o(\mathbb{E}_0[X_k(G)]^2)$ et $M_2 = o(\mathbb{E}_0[X_k(G)]^2)$.

Solution. Comme on a toujours $\mathbb{E}_0[X_k(G)^2] \geq \mathbb{E}_0[X_k(G)]^2$, il nous faut une majoration de $\mathbb{E}_0[X_k(G)^2]$ pour prouver le résultat voulu. On a

$$\begin{aligned} \mathbb{E}_0[X_k(G)^2] &= \mathbb{E}_0 \left[\left(\sum_{\mathcal{E} \in \mathcal{E}(n,k)} \mathbb{1}_{\mathcal{E} \in G} \right)^2 \right] \\ &= \mathbb{E}_0 \left[\sum_{\mathcal{E} \in \mathcal{E}(n,k)} \sum_{\mathcal{E}' \in \mathcal{E}(n,k)} \mathbb{1}_{\mathcal{E} \in G, \mathcal{E}' \in G} \right] \\ &= \sum_{\mathcal{E} \in \mathcal{E}(n,k)} \sum_{\mathcal{E}' \in \mathcal{E}(n,k)} \mathbb{P}_0(\mathcal{E} \in G, \mathcal{E}' \in G) \\ &= M_1 + M_2 + M_3 \end{aligned}$$

comme annoncé dans l'indication, car les indices des sommations forment une partition de $\mathcal{E}(n,k) \times \mathcal{E}(n,k)$. Nous allons gérer les termes séparément. Notons que nous raisonnons à n grand et qu'en faisant les mêmes calculs qu'en question 8, on voit que dans ce régime de k ,

$$\mathbb{E}_0[X_k(G)] \sim n \frac{\lambda^k}{k!} = \exp(o(\log n) - k \log k + \log n) \geq \exp((\varepsilon/2) \log n + o(\log n)) \rightarrow +\infty.$$

Pour l'instant, supposons $k \rightarrow \infty$. On a tout d'abord, en notant ℓ le nombre de noeuds commun à W et W' ,

$$\begin{aligned} M_1 &\leq \sum_{\mathcal{E} \in \mathcal{E}(n,k)} (\lambda/n)^k \sum_{\ell=0}^k \binom{k}{\ell} \binom{n-k-1}{k-\ell} (\lambda/n)^{k-\ell} \\ &\leq \mathbb{E}_0[X_k(G)] \sum_{\ell=0}^k \binom{k}{\ell} n^{k-\ell} (\lambda/n)^{k-\ell} \\ &= \mathbb{E}_0[X_k(G)] (1 + \lambda)^k, \end{aligned}$$

et il reste à montrer que $(1 + \lambda)^k = o(\mathbb{E}_0[X_k(G)]) = o(n\lambda^k/k!)$. En passant au log et en utilisant Stirling, il suffit de montrer que

$$k \log(1 + \lambda) - k \log \lambda - \log n + k \log k - k \rightarrow -\infty,$$

ce qui est vrai car avec les hypothèses sur k a que ce terme est plus petit que $-(\varepsilon/2) \log n$ pour n assez grand. Cela est encore vrai si k ne tend pas vers $+\infty$. Passons à M_2 , qui est plus facile : notons que les couples $\mathcal{E}, \mathcal{E}'$ présents dans la somme n'ont pas d'arête en commun si $u \notin W'$, et exactement une si $u \in W'$. En comptant ces possibilités, on

obtient

$$\begin{aligned}
M_2 &\leq |\mathcal{E}(n, k)|(\lambda/n)^k k \left[\binom{n-2}{k-1} (\lambda/n)^{k-1} + \binom{n-2}{k} (\lambda/n)^k \right] \\
&\leq \mathbb{E}_0[X_k(G)](k^2 \lambda^{k-1}/k! + k \lambda^k/k!) \\
&\leq \mathbb{E}_0[X_k(G)] O\left(\frac{k^2}{n} \mathbb{E}_0[X_k(G)]\right) = o(\mathbb{E}_0[X_k(G)]^2).
\end{aligned}$$

Enfin, comme les paires d'étoiles dans M_3 n'ont aucune arête en commun,

$$\begin{aligned}
M_3 &= \sum_{\mathcal{E}=(u,W) \in \mathcal{E}(n,k)} \sum_{\substack{\mathcal{E}'=(u',W') \in \mathcal{E}(n,k) \\ u' \notin W}} \mathbb{P}_0(\mathcal{E} \in G) \mathbb{P}_0(\mathcal{E}' \in G) \\
&\leq \mathbb{E}_0[X_k(G)]^2.
\end{aligned}$$

On a donc montré que $\mathbb{E}_0[X_k(G)]^2 \leq \mathbb{E}_0[X_k(G)^2] \leq \mathbb{E}_0[X_k(G)]^2 + o(\mathbb{E}_0[X_k(G)]^2)$ ce qui conclut la question.

13. En déduire que la détection faible d'une k -étoile est impossible.

On pourra utiliser les questions 10 et 11.

Solution. Nous avons montré à la question précédente que $\mathbb{E}_0[X_k(G)^2] = \mathbb{E}_0[X_k(G)]^2 + o(\mathbb{E}_0[X_k(G)]^2)$ et d'après l'expression de L question 5, cela implique

$$\mathbb{E}_0[L(G)^2] = \frac{\mathbb{E}_0[X_k(G)^2]}{\mathbb{E}_0[X_k(G)]^2} \rightarrow 1.$$

En appliquant les résultats de la partie D à $\Gamma = \mathbb{G}_n$, $P_0 = \mathbb{P}_0$ et $P_1 = \mathbb{P}_1$, l'inégalité prouvée à la question 11 implique que, en explicitant les dépendances en n ,

$$d_{\text{TV}}(\mathbb{P}_{0,n}, \mathbb{P}_{1,n}) \leq \frac{1}{2} \sqrt{\mathbb{E}_{0,n}[L^2(G)] - 1} \rightarrow 0,$$

et par la question 10, que pour toute famille $(\phi_n)_{n \geq 1}$ de tests,

$$\mathbb{P}_0(\phi_n(G) = 1) + \mathbb{P}_1(\phi_n(G) = 0) \geq 1 - d_{\text{TV}}(\mathbb{P}_{0,n}, \mathbb{P}_{1,n}) \rightarrow 1.$$

Cela prouve que la détection faible d'une k -étoile est impossible.

- 14*. Avez-vous une idée d'application de ce problème ?

Grâce aux questions 8 et 13, nous avons établi que la difficulté du problème (\star) change brutalement selon que $k \geq (1 + \varepsilon)k^*(n)$ ou que $k \leq (1 - \varepsilon)k^*(n)$, avec $k^*(n) = \frac{\log n}{\log \log n}$. Ce phénomène est communément appelé *transition de phase* et $k^*(n)$ est appelé *seuil critique*.

- 15*. A votre avis, que devient ce seuil critique $k^*(n)$ si on étudie le problème (\star) cette fois-ci avec $p = p_0$ fixé constant dans $]0, 1[$?

Fin du sujet.