# Statistique Mathématique

Examen — 5 mai 2021

*Durée : 3 heures. Documents et calculatrices interdits. Prenez soin de bien justifier vos réponses. Les trois parties sont indépendantes, mais il est conseillé de les traiter dans l'ordre. Les questions marquées d'un astérisque ($\star$) sont plus difficiles, il est normal qu'elles demandent plus d'efforts. Bonne chance !*

## Modélisation de paysage politique

Nous allons considérer dans cet examen la modélisation d'un paysage politique. Les citoyens d'un pays imaginaire se regroupent progressivement en partis politiques, formant à chaque instant $n$ une *partition aléatoire* de $\{1, \ldots, n\}$, l'ensemble des citoyens. On note cette partition aléatoire $B_n$ : un parti politique est donc un ensemble de citoyens et $B_n$ un ensemble de partis. Le nombre total de partis politiques au temps $n$ est donc donné par $|B_n|$, le cardinal de $B_n$. Par exemple, pour $n = 5$, nous pourrions observer

$$B_n = \{\{1,3\},\{4\},\{2,5\}\}, \qquad |B_n| = 3 \,.$$

On convient que $B_1 = \{\{1\}\}$ et on note $b \in B_n$ un élément de la partition $B_n$, c'est-à-dire un parti politique. Nous modélisons le processus d'évolution du paysage politique de la manière suivante : le citoyen $n+1$ rejoint chaque parti existant avec probabilité $|b|/(n+\theta)$ ou crée son propre parti avec probabilité $\theta/(n+\theta)$, où $\theta \in \mathbb{R}_+$ est un paramètre inconnu. Il est impossible pour un citoyen de quitter le parti auquel il appartient.

## Partie I : Préliminaires (4 points)

Dans cette première partie, nous établissons les propriétés fondamentales de notre modèle.

I.1. Vérifiez que la loi de $B_{n+1}$ sachant $B_n$ est bien définie.

I.2. Comment va évoluer $(B_n)_{n\geq 1}$ si $\theta = 0$ ? Si $\theta = +\infty$ ? On appelle $\theta$ le *coefficient d'ambition.* Cette nomenclature vous paraît-elle appropriée ?

I.3. Tracez sur un même graphique plusieurs trajectoires de $|B_n|$ pour $\theta = \theta_1$ avec $\theta_1$ proche de 0. Répétez ce dessin pour un $\theta = \theta_2 \gg \theta_1$.

I.4. Le modèle proposé vous paraît-il réaliste ?

## Partie II : Étude d'un estimateur de $\theta$ (9 points)

On suppose dans cette partie que $\theta > 0$. Nous allons tenter d'estimer $\theta$ à partir **d'une réalisation** de la suite $B_1, \ldots, B_n$. Nous étudions dans un premier temps l'estimateur défini par $\hat{\theta}_n^{(1)} := |B_n| / \log n$.

II.1. Montrez que $|B_n|$ s'écrit $\xi_1 + \cdots + \xi_n$, où $\xi_1, \ldots, \xi_n$ sont des variables aléatoires de Bernoulli indépendantes. On précisera le paramètre de chacune de ces Bernoulli.

II.2. Déduire de la question II.1. que

$$\mathbb{E}\left[|B_n|\right] = \sum_{k=0}^{n-1} \frac{\theta}{\theta+k} \quad \text{et} \quad \mathrm{Var}\left(|B_n|\right) = \sum_{k=1}^{n-1} \frac{\theta k}{(\theta+k)^2} \,.$$

II.3. Montrez que $\mathbb{E}\left[|B_n|\right] - \theta \log n = \mathcal{O}(1)$ lorsque $n \to +\infty$, où les constantes dans le $\mathcal{O}$ peuvent dépendre de $\theta$. Que pouvez-vous dire du biais de $\hat{\theta}_n$ ?

II.4. Montrez que

$$\frac{\mathrm{Var}\left(|B_n|\right)}{\theta \log n} = 1 + o(1) \,.$$

($\star$) II.5. En utilisant les questions II.3. et II.4., montrez que l'estimateur $\hat{\theta}_n^{(1)}$ est consistant.

II.6. En utilisant la même technique de preuve, peut-on montrer que $\hat{\theta}_n^{(1)}$ est fortement consistant ?

II.7. On admet le résultat suivant :

$$\frac{|B_n| - \mathbb{E}[|B_n|]}{\sqrt{\operatorname{Var}(|B_n|)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

lorsque $n \to +\infty$. Montrez que $\mathbb{E}[|B_n|]$ et $\operatorname{Var}(|B_n|)$ peuvent être remplacés par $\theta \log n$ dans cette limite.

II.8. En utilisant la question précédente, proposez un intervalle de confiance asymptotique de niveau $1-\alpha$ pour $\theta$.

## Partie III : Micro-partis (7 points)

Dans cette dernière partie, on étudie la distribution des micro-partis (les partis politiques comptant un seul membre) et on en déduit un autre estimateur de $\theta$. On suppose toujours $\theta > 0$ et on note $C_n$ le nombre de micro-partis au temps $n$.

III.1. Décrire la loi de $C_{n+1}$ conditionnellement à $C_n$.

III.2. Soit $a \in \{1, \ldots, n\}$. Déduire de la question précédente que

$$\mathbb{E}[C_{n+1} \mid C_n = a] = \frac{a(n+\theta-1)+\theta}{\theta+n} .$$

III.3. En utilisant la question précédente, montrez que

$$\mathbb{E}[C_n] = \frac{n\theta}{n+\theta-1} .$$

($\star$) III.4. Montrez que

$$\operatorname{Var}(C_n) = \frac{n(n-1)(n-2+2\theta)\theta}{(n+\theta-2)(n+\theta-1)^2} .$$

III.5. Proposez un nouvel estimateur $\hat{\theta}_n^{(2)}$ pour le paramètre $\theta$.

III.6. Quel estimateur conseillez-vous d'utiliser ? Justifiez votre réponse.