Statistique 3: Examen **A rendre le:** 7 Mai 2020

Exercice 1 M-estimateurs convexes (8 points)

Soit $\mathcal{X}$ un espace mesurable, muni d'une probabilité P, et soit $X_1, X_2, \dots$ une suite de variables aléatoires i.i.d. sur $\mathcal{X}$ de loi P. Soit Θ un ouvert convexe de $\mathbb{R}^d$ ($d \geq 1$) et $\phi : \mathcal{X} \times \Theta \to \mathbb{R}$ une fonction telle que:

- Pour tout $\theta \in \Theta$, $\phi(\cdot, \theta)$ est mesurable et intégrable par-rapport à P;
- Pour tout $x \in \mathcal{X}$, $\phi(x, \cdot)$ est convexe.

Pour tout entier $n \geq 1$, on définit

$$\begin{aligned} \Phi_n \colon \Theta &\to \mathbb{R} \\ \theta &\mapsto \frac{1}{n}\sum_{i=1}^{n} \phi(X_i, \theta) \end{aligned}$$

et on définit aussi la fonction

$$\begin{aligned} \Phi \colon \Theta &\to \mathbb{R} \\ \theta &\mapsto \mathbb{E}\phi(X_1, \theta). \end{aligned}$$

On suppose que Φ est minorée et qu'elle atteint son minimum. On note $\Theta^* = \{\theta^* \in \Theta : \forall \theta \in \Theta, \Phi(\theta) \geq \Phi(\theta^*)\}$ et on suppose que Θ^* est un compact.

1. Montrer qu'avec probabilité 1, $\hat{\Phi}_n$ est aussi minorée et atteint son minimum, pour n assez grand.
2. (question facultative) Pouvez-vous trouver un exemple de fonction ϕ, de loi P et d'ensemble Θ tels que Φ est minorée, mais Θ^* n'est pas compact et Φ_n est presque sûrement non minorée ?
3. (question facultative) Montrer qu'il existe une suite de variables aléatoires $(\hat{\theta}_n)_{n\geq 1}$ telle qu'avec probabilité 1, $\hat{\theta}_n$ est un minimiseur de Φ_n pour n assez grand.
4. Montrer que $\mathsf{d}(\hat{\theta}_n, \Theta^*) \xrightarrow[n\to\infty]{\text{p.s.}} 0$.

Exercice 2 Une inégalité de concentration (12 points)

Soit $n \geq 1$ et $\mathcal{X}_1, \dots, \mathcal{X}_n$ des espaces mesurables. Pour tout $i \in \{1, \dots, n\}$, soit X_i une variable aléatoire à valeurs dans $\mathcal{X}_i$. On suppose que $X_1, \dots, X_n$ sont définies sur un même espace de probabilité et qu'elle sont indépendantes.

Soit $f : \mathcal{X}_1 \times \ldots \times \mathcal{X}_n \to \mathbb{R}$ une fonction mesurable. On suppose l'existence de réels strictement positifs $c_1, \ldots, c_2$ tels que pour tout $i \in \{1, \ldots, n\}$, pout tout $(x_1, \ldots, x_n) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, et pour tout $s, t \in \mathcal{X}_i$,

$$f(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, s, x_{i+1}, \ldots, x_n) \leq c_i.$$

On cherche alors à démontrer une inégalité de concentration pour la variable aléatoire $Y = f(X_1, \ldots, X_n)$.

1. Montrer que Y est intégrable.
2. Pour $i = 1, \ldots, n$, on pose $Y_i = \mathbb{E}[Y|X_1, \ldots, X_i]$.
   a) Vérifier que $Y - \mathbb{E}Y = \sum_{i=1}^n (Y_i - Y_{i-1})$, après avoir préalablement défini Y_0.
   b) Vérifier que pour tout $i = 1, \ldots, n$, $Y_{i-1} = \mathbb{E}[Y_i|X_1, \ldots, X_{i-1}]$ (où l'espérance conditionnelle sachant aucune variable sera comprise comme l'espérance au sens usuel).
3. Pour tout $i = 1, \ldots, n$, on pose

$$\begin{aligned} g_i \colon \mathcal{X}_1 \times \ldots \times \mathcal{X}_i &\to \mathbb{R} \\ (x_1, \ldots, x_i) &\mapsto \mathbb{E} f(x_1, \ldots, x_i, X_{i+1}, \ldots, X_n). \end{aligned}$$

   Vérifier que pour tout $i = 1, \ldots, n$, $Y_i = g_i(X_1, \ldots, X_i)$.
4. Soit $i \geq 1$. Fixons $(x_1, \ldots, x_{i-1}) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_{i-1}$ et posons $Z_i = g_i(x_1, \ldots, x_{i-1}, X_i)$.
   a) Montrer que Z_i est sous-Gaussienne, de paramètre d'échelle $c_i^2/4$ (si vous ne parvenez qu'à majorer le paramètre d'échelle par c_i^2, admettez le résultat).
   b) En déduire que pour tout $\lambda > 0$, avec probabilité 1,

$$\mathbb{E}\left[e^{\lambda(Y_i - Y_{i-1})}|X_1, \ldots, X_{i-1}\right] \leq e^{\lambda^2 c_n^2/8}$$

5. Conclure que pour tout $\lambda > 0$,

$$\mathbb{E}\left[e^{\lambda(Y - \mathbb{E}Y)}\right] \leq e^{\frac{\lambda^2 \sum_{i=1}^n c_i^2}{8}}.$$

   (On montrera d'abord que $\mathbb{E}\left[e^{\lambda(Y-\mathbb{E}Y)}\right] \leq e^{\lambda^2 c_n^2/8} \mathbb{E}\left[e^{\lambda \sum_{i=1}^{n-1}(Y_i - Y_{i-1})}\right]$ *puis on procèdera par récurrence)*
6. Montrer que pour tout $t \geq 0$,

$$\mathbb{P}[|f(X_1, \ldots, f(X_n) - \mathbb{E}f(X_1, \ldots, X_n)| \geq t] \leq 2 \exp\left(\frac{-2t^2}{\sum_{i=1}^n c_i^2}\right).$$

7. A l'aide du résultat précédent, retrouver:
   a) l'inégalité de Hoeffding;
   b) l'inégalité de concentration pour les fonctions Lipschitziennes appliquées aux variables aléatoires bornées.

8. Applications:
   a) Triangles dans des graphes aléatoires: soit G_n un graphe aléatoire d'Erdös-Rényi, de taille $n \geq 3$ et de paramètre $p \in (0,1)$, et soit T_n le nombre de triangles dans G_n. Vérifier que $\mathbb{E}[T_n] = \frac{n(n-1)(n-2)}{6}p^3$ et montrer que

$$T_n = \frac{n^3p^3}{6} + O_{\mathbb{P}}(n^2).$$

   b) Profondeur statistique de Tukey: soient $X_1, \ldots, X_n$ des vecteurs aléatoires i.i.d. dans $\mathbb{R}^d$ $(n, d \geq 1)$. Pour tout $x \in \mathbb{R}^d$, on définit la profondeur de x, dans le nuage de points $X_1, \ldots, X_n$, de la manière suivante:

$$D_n(x) = \inf \left\{ \frac{\#\{i = 1, \ldots, n : X_i \in H\}}{n} : H \in \mathcal{H}, x \in H \right\},$$

   où $\mathcal{H}$ est l'ensemble des demi-espaces affines fermés de $\mathbb{R}^d$. Autrement dit, $D_n(x)$ est le nombre minimal de poins de l'échantillon étant inclus dans un demi-espace fermé contenant x.

   i – D'après vous, pourquoi $D_n(x)$ est-elle appelée *la profondeur* du point x dans le nuage ?

   ii – Soit $x \in \mathbb{R}^d$ quelconque. Montrer que pour tout $\delta \in (0,1)$,

$$\mathbb{E}D_n(x) - \sqrt{\frac{\log(2/\delta)}{2n}} \leq D_n(x) \leq \mathbb{E}D_n(x) + \sqrt{\frac{\log(2/\delta)}{2n}}$$

   avec probabilité au moins $1 - \delta$.

Exercice 3 Une inégalité de concentration uniforme en n (7+∞ points)

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires i.i.d., centrées et sous-Gaussiennes, de paramètre d'échelle $\sigma^2 > 0$.

1. Rappeler pourquoi pour tout $\delta \in (0,1)$, et pour tout $n \geq 1$,

$$\mathbb{P}\left[|\bar{X}_n| \leq \sigma\sqrt{\frac{2\log(2/\delta)}{n}}\right] \geq 1 - \delta.$$

2. En utilisant une borne d'union, montrer que pour tout $\delta \in (0,1)$,

$$\mathbb{P}\left[\forall n \geq 1, |\bar{X}_n| \leq \sigma\sqrt{\frac{2\log(\pi^2n^2/(3\delta))}{n}}\right] \geq 1 - \delta.$$

3. Une telle borne, uniforme en n, peut être intéressante en apprentissage en ligne. Par exemple, dans le problème de bandit, à partir de chaque nouvelle observation, on met peut mettre à jour des intervalles de confiance, et on souhaite que tous ces intervalles soient valides, i.e., que chacun contienne le paramètre d'intérêt avec la bonne probabilité. En revanche, pour un niveau δ fixé, la taille des intervalles de confiance donnés par la borne précédente est de l'ordre de $\sqrt{(\log n)/n}$, et on peut se demander si ces intervalles ne sont pas conservateurs. Nous allons montrer, dans cette question, qu'on peut en fait obtenir des intervalles de taille de l'ordre de $\sqrt{(\log \log n)/n}$, en appliquant un outil plus fin que la borne d'union. Fixons une suite $(t_n)_{n\geq 1}$ décroissante de réels strictement positifs.

   a) Pour tout entier $k \geq 0$, soit $E_k = \{2^k, 2^k+1, \ldots, 2^{k+1}-1\}$. Pour $n \geq 1$, on note $S_n = X_1 + \ldots + X_n$. Montrer que

   $$\mathbb{P}[\exists n \geq 1, |\bar{X}_n| > t_n] \leq \sum_{k=0}^{\infty} \mathbb{P}[\max_{n \in E_k} |S_n| > 2^k t_{2^{k+1}}].$$

   b) On admettra l'inégalité de Doob pour les martingales suivante:

   $$\forall m \geq 1, \forall t \geq 0, \mathbb{P}[\max_{1 \leq i \leq m} |S_i| > t] \leq 2e^{-\frac{t^2}{2\sigma^2 m}}.$$

   En déduire que, pour tout $k \in \mathbb{N}$,

   $$\mathbb{P}[\max_{n \in E_k} |S_n - S_{2^k}| > 2^{k-1} t_{2^{k+1}}] \leq 2e^{-\frac{2^k t_{2^{k+1}}^2}{8\sigma^2}}.$$

   c) Vérifier que, pour tout $k \in \mathbb{N}$,

   $$\mathbb{P}[|S_{2^k}| > 2^{k-1} t_{2^{k+1}}] \leq 2e^{-\frac{2^k t_{2^{k+1}}^2}{8\sigma^2}}.$$

   d) En utilisant le fait que, pour tout $k \geq 0$,

   $$\max_{n \in E_k} |S_n| \leq |S_{2^k}| + \max_{n \in E_k} |S_n - S_{2^k}|,$$

   déduire que

   $$\mathbb{P}[\max_{n \in E_k} |S_n| > 2^k t_{2^{k+1}}] \leq 4e^{-\frac{2^k t_{2^{k+1}}^2}{8\sigma^2}}.$$

   e) On choisit $t_n = 4\sigma \sqrt{\frac{\log\left(\frac{2\pi^2 (\log_2 n)^2}{3\delta}\right)}{n}}$, pour tout $n \geq 1$, où $\delta \in (0,1)$ et $\log_2$ est le logarithme en base 2. Vérifier que la suite $(t_n)_{n\geq 2}$ est bien définie et qu'elle est décroissante.

f) En déduire qu'avec probabilité au moins $1-\delta$,

$$\forall n \geq 2, \quad |\bar{X}_n| \leq 4\sigma \sqrt{\frac{\log\left(\frac{2\pi^2(\log_2 n)^2}{3\delta}\right)}{n}}.$$

Ce résultat peut s'interpréter comme une loi du logarithme itéré uniforme, non asymptotique. La taille des intervalles de confiances obtenus est, cette fois-ci, de l'ordre de $\sqrt{(\log\log n)/n}$, qui est la vitesse qu'on retrouve dans la loi du logarithme itéré standard, ce qui prouve que cette taille, asymptotiquement, ne peut être améliorée.

4. (∞ points) Supposons que $X_1, X_2, \dots$ sont i.i.d., Gaussiennes, de moyenne $\mu \in \mathbb{R}$ et de variance $\sigma^2 > 0$. Proposer une suite d'estimateurs $(\hat{\mu}_n)_{n\geq 1}$ telle que, quelle que soit la valeur de μ, on a, pour tout $\delta \in (0,1)$,

$$\mathbb{P}\left[\forall n \geq 1, |\hat{\mu}_n - \mu| \geq t(n,\delta)\right] \geq 1-\delta,$$

où, pour tout $\delta \in (0,1)$, $t(n,\delta)$ est une suite de réels positifs tendant vers zéro strictement plus vite que $\sqrt{(\log\log n)/n}$.