ENS de Lyon 2nd semestre 2018-2019
L3 - Intégration et Probabilités

DM 1

Exercice 1 *Réordonnement*

1. (a) p.s., $Y_k^{(n)} \leq x \iff \#\{i \in [\![1;n]\!] : X_i \leq x\} \geq k \iff \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}} \geq k$.

   (b) Remarquons que pour tout $x \in \mathbb{R}$, les variables aléatoires $(\mathbf{1}_{\{X_i \leq x\}})_i$ sont des variables aléatoires i.i.d. $\mathcal{L}^1$ de loi de Bernoulli de paramètre $\mathbb{P}(X_i \leq x) = F_X(x)$. D'où d'après la loi forte des grands nombres,

$$\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}} \underset{n\to\infty}{\longrightarrow} F_X(x) \quad \text{p.s.}$$

   Soit $t \in]0;1[$, soit $\varepsilon > 0$,

$$\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i \leq G_X(t)+\varepsilon\}} \underset{n\to\infty}{\longrightarrow} F_X(G_X(t)+\varepsilon) = F_X(\inf\{x : F_X(x) \geq t\} + \varepsilon) > t \quad \text{p.s.}$$

   l'inégalité stricte venant de la stricte croissance de F_X sur $\mathbb{R}$, et de la continuité à droite de F_X. Ainsi presque sûrement pour n assez grand, on a $\sum_{i=1}^n \mathbf{1}_{\{X_i \leq G_X(t)+\varepsilon\}} > tn$, et la somme étant entière ceci implique p.s. pour n assez grand, $\sum_{i=1}^n \mathbf{1}_{\{X_i \leq G_X(t)-\varepsilon\}} \geq \lfloor tn \rfloor$. D'après la question (a), p.s. pour n assez grand on a $Y_{\lfloor tn \rfloor}^{(n)} \leq G_X(t)+\varepsilon$, et donc p.s. $\limsup_n Y_{\lfloor tn \rfloor}^{(n)} \leq G_X(t)+\varepsilon$. Le réel ε étant arbitraire, et en considérant une suite dénombrable de ε tendant vers 0, on obtient

$$\limsup_{n\to\infty} Y_{\lfloor tn \rfloor}^{(n)} \leq G_X(t) \quad \text{p.s.}$$

   En considérant $\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i \leq G_X(t)-\varepsilon\}}$, on montre par le même raisonnement que $\liminf_n Y_{\lfloor tn \rfloor}^{(n)} \geq G_X(t)$ p.s. d'où la conclusion de la question.
   Si $t = 0$ ou $t = 1$, il suffit de remarque que de toute façon p.s., pour tout $k \in [\![1;n]\!]$ on a $Y_k^{(n)} \geq G_X(0)$ et $Y_k^{(n)} \leq G_X(1)$ et d'utiliser l'une ou l'autre des conclusions précédentes sur la limsup ou la liminf de la suite.

2. (a) On a en utilisant la borne d'union puis le théorème de transfert (la loi jointe de (X_i, X_j) étant $\mathrm{Leb}_{[0,1]} \otimes \mathrm{Leb}_{[0,1]}$)

$$\begin{aligned}\mathbb{P}(\exists i,j \in [\![1;n]\!] : Y_i^{(n)} = Y_j^{(n)}) &= \mathbb{P}\Big(\bigcup_{1\leq i<j\leq n} \{X_i = X_j\}\Big) \\ &\leq \sum_{1\leq i<j\leq n} \mathbb{P}(X_i = X_j) = \frac{n(n-1)}{2}\int_{\mathbb{R}^2} \mathbf{1}_{\{x=y\}}\,\mathrm{d}x\,\mathrm{d}y = 0,\end{aligned}$$

   le borélien $\{(x,y) \in \mathbb{R}^2 : x = y\}$ étant de mesure nulle.

   (b) On a

$$\begin{aligned}&\mathbb{P}((Y_1^{(n)},\dots,Y_n^{(n)}) = (X_{\sigma(1)},\dots,X_{\sigma(n)})) \\ &= \mathbb{P}(X_{\sigma(1)} < \dots < X_{\sigma(n)}) = \int_0^1 \mathbf{1}_{\{x_{\sigma(1)}<\dots<x_{\sigma(n)}\}}\,\mathrm{d}x_1 \dots \mathrm{d}x_n = \int_0^1 \mathbf{1}_{\{x_1<\dots<x_n\}}\,\mathrm{d}x_1 \dots \mathrm{d}x_n.\end{aligned}$$

Pour justifier cette dernière égalité, on peut soit faire un changement de variable $y_i = x_{\sigma(i)}$, soit utiliser Fubini dans l'ordre qui nous arrange, obtenir

$$\int_0^1 \mathbf{1}_{\{x_{\sigma(1)}<\dots<x_{\sigma(n)}\}}\,\mathrm{d}x_1\dots\,\mathrm{d}x_n = \int_0^1 \mathbf{1}_{\{x_{\sigma(1)}<\dots<x_{\sigma(n)}\}}\,\mathrm{d}x_{\sigma(1)}\dots\,\mathrm{d}x_{\sigma(n)}$$

puis réétiqueter les variables muettes. Bref, la proba ne dépend pas de σ.
Maintenant, les événements $((Y_1^{(n)},\dots,Y_n^{(n)}) = (X_{\sigma(1)},\dots,X_{\sigma(n)}))_{\sigma\in\mathfrak{S}_n}$ étant tous disjoints à un événement de probabilité nulle près (d'après la question 2.(a)), et leur réunion étant un événement presque sûr, on a

$$\begin{aligned}1 &= \mathbb{P}\Big(\bigcup_{\sigma\in\mathfrak{S}_n}(Y_1^{(n)},\dots,Y_n^{(n)}) = (X_{\sigma(1)},\dots,X_{\sigma(n)})\Big)\\ &= \sum_{\sigma\in\mathfrak{S}_n}\mathbb{P}((Y_1^{(n)},\dots,Y_n^{(n)}) = (X_{\sigma(1)},\dots,X_{\sigma(n)})).\end{aligned}$$

Comme il y a $n!$ termes dans la somme, tous égaux, on a que chacun vaut $\frac{1}{n!}$.

(c)

$$\begin{aligned}\mathbb{E}[h(Y_1^{(n)},\dots,Y_n^{(n)})] &= \mathbb{E}[h(Y_1^{(n)},\dots,Y_n^{(n)})\sum_{\sigma\in\mathfrak{S}_n}\mathbf{1}_{\{X_{\sigma(1)}\le\dots\le X_{\sigma(n)}\}}]\\ &= \sum_{\sigma\in\mathfrak{S}_n}\mathbb{E}[h(X_{\sigma(1)},\dots,X_{\sigma(n)})\mathbf{1}_{\{X_{\sigma(1)}\le\dots\le X_{\sigma(n)}\}}]\\ &= \sum_{\sigma\in\mathfrak{S}_n}\int_0^1 h(x_{\sigma(1)},\dots,x_{\sigma(n)})\mathbf{1}_{\{0\le x_{\sigma(1)}<\dots<x_{\sigma(n)}\le 1\}}\,\mathrm{d}x_1\dots\,\mathrm{d}x_n,\end{aligned}$$

par théorème de transfert. On obtient le résultat encore une fois par changement de variable ou Fubini.

(d) Soit $k\in[\![1;n]\!]$, f une fonction borélienne positive, posons $h:(y_1,\dots,y_n)\mapsto f(y_k)$. Alors on a d'après Fubini,

$$\begin{aligned}&\mathbb{E}[f(Y_k^{(n)})] = \mathbb{E}[h(Y_1^{(n)},\dots,Y_n^{(n)})] = n!\int_0^1 f(y_k)\mathbf{1}_{\{0\le y_1<\dots<y_n\le 1\}}\,\mathrm{d}y_1\dots\,\mathrm{d}y_n\\ &= n!\int_{\mathbb{R}} f(y)\mathbf{1}_{\{0\le y\le 1\}}\Big(\int_{\mathbb{R}^{k-1}}\mathbf{1}_{\{0\le y_1\le\dots\le y_{k-1}\le y\}}\,\mathrm{d}y_1\dots\,\mathrm{d}y_{k-1}\Big)\Big(\int_{\mathbb{R}^{n-k}}\mathbf{1}_{\{y\le y_{k+1}\le\dots\le y_n\le 1\}}\,\mathrm{d}y_{k+1}\dots\,\mathrm{d}y_n\Big)\,\mathrm{d}y.\end{aligned}$$

Maintenant, par le changement de variable $y_j' = y_j/y$ et la question (b), on obtient

$$\begin{aligned}\int_{\mathbb{R}^{k-1}}\mathbf{1}_{\{0\le y_1\le\dots\le y_{k-1}\le y\}}\,\mathrm{d}y_1\dots\,\mathrm{d}y_{k-1} &= y^{k-1}\int_{\mathbb{R}^{k-1}}\mathbf{1}_{\{0\le y_1'\le\dots\le y_{k-1}'\le 1\}}\,\mathrm{d}y_1'\dots\,\mathrm{d}y_{k-1}'\\ &= y^{k-1}\mathbb{P}(X_1\le\dots\le X_{k-1}) = \frac{y^{k-1}}{(k-1)!}.\end{aligned}$$

De la même manière, on obtient par le changement de variable $y_j' = (1-y_j)/(1-y)$

$$\int_{\mathbb{R}^{n-k}}\mathbf{1}_{\{y\le y_{k+1}\le\dots\le y_n\le 1\}}\,\mathrm{d}y_{k+1}\dots\,\mathrm{d}y_n = \frac{(1-y)^{n-k}}{(n-k)!}.$$

En concaténant ces trois derniers résultats, on conclut que

$$\mathbb{E}[f(Y_k^{(n)}] = n!\int_{\mathbb{R}} f(y)\mathbf{1}_{\{0\le y\le 1\}}n!\frac{y^{k-1}}{(k-1)!}\frac{(1-y)^{n-k}}{(n-k)!}\,\mathrm{d}y,$$

et donc $Y_k^{(n)}$ admet la densité suivante (qui est celle de la loi Bêta de paramètres k et $n-k+1$)

$$n!\mathbf{1}_{\{0\le y\le 1\}}\frac{y^{k-1}}{(k-1)!}\frac{(1-y)^{n-k}}{(n-k)!} = \mathbf{1}_{\{0\le y\le 1\}}k\binom{n}{k}y^{k-1}(1-y)^{n-k}.$$

Exercice 2 *Estimation*

1. $\mathbb{P}(|\mu_n - m| > x) \leq \frac{\mathrm{Var}(\mu_n)}{x}^2 = \frac{\sigma^2}{nx^2}$. On voit que $\delta(\alpha, n) = \sigma\sqrt{\frac{1}{\alpha n}}$ convient.
2. (a) Considérons d'abord le cas où X est centrée. Par convexité d'exponentielle et comme $X \in [a, b]$ presque sûrement, on a p.s. l'inégalité suivante :

$$e^{\lambda X} \leq \frac{X-a}{b-a}e^{\lambda b} + \frac{b-X}{b-a}e^{\lambda a}.$$

En prenant l'espérance puis en posant $\theta = \frac{-a}{b-a}$ et $u = \lambda(b-a)$ on obtient

$$\mathbb{E}[e^{\lambda X}] \leq \frac{-a}{b-a}e^{\lambda b} + \frac{b}{b-a}e^{\lambda a} = \theta e^{-(1-\theta)u} + (1-\theta)e^{\theta u} \leq e^{u^2/8} = e^{\lambda^2(b-a)^2/8}.$$

Quand X n'est plus centrée on applique à $X - m$ (qui appartient à $[a-m, b-m]$ p.s.) et on a toujours $\mathbb{E}[e^{\lambda(X-m)}] \leq e^{\lambda^2(b-a)^2/8}$.
On utilise ensuite l'inégalité de Chernoff. Soit $\lambda > 0, x > 0$.

$$\mathbb{P}((\mu_n - m) \geq x) \leq \mathbb{E}[e^{\lambda(\mu_n - m)}]e^{-\lambda x} = (\mathbb{E}[e^{\lambda(X_1 - m)/n}])^n e^{-\lambda x} \leq \exp\left(\frac{\lambda^2(b-a)^2}{8n} - \lambda x\right).$$

On optimise en $\lambda = 4xn/(b-a)^2$ qui est bien positif car $x > 0$, ce qui donne

$$\mathbb{P}((\mu_n - m) \geq x) \leq \exp(-2x^2/(b-a)^2).$$

C'est l'inégalité de Hoeffding. Le cas $x \leq 0$ est trivial car un événement est toujours de probabilité plus petite que 1.

(b) En appliquant l'inégalité précédente à $-X$, on obtient

$$\mathbb{P}((\mu_n - m) \leq -x) \leq \exp(-2x^2/(b-a)^2).$$

Par union on a l'inégalité bilatère

$$\mathbb{P}(|\mu_n - m| \geq x) \leq \mathbb{P}((\mu_n - m) \leq -x) + \mathbb{P}((\mu_n - m) \geq x) \leq 2\exp(-2nx^2/(b-a)^2).$$

On voit alors que $\delta(\alpha, n) = (b-a)\sqrt{\frac{\log(2/\alpha)}{2n}}$ donne un intervalle de confiance au niveau α.

3. Pour des Bernoulli de paramètre inconnu on peut prendre $\sigma = 1/2$ dans 1. et $(b-a) = 1$ dans 2. On compare alors $\delta(\alpha, n) = \frac{1}{2\sqrt{\alpha n}}$ à $\delta(\alpha, n) = \sqrt{\frac{\log(2/\alpha)}{2n}}$. Pour $\alpha = 0.05$ on obtient $\approx 2, 2/\sqrt{n}$ contre $\approx 1.4/\sqrt{n}$. L'inégalité de Hoeffding est déjà meilleure. Quand $\alpha \to 0$, l'écart se creuse et l'inégalité de Hoeffding est infiniment meilleure.
4. (a)
   (b) D'après la question précédente,

$$\#\{i, 1 \leq i \leq k \text{ t.q. } \mu^{(i)} \geq \nu_n\} = \#\{i, 1 \leq i \leq k \text{ t.q. } \mu^{(i)} \leq \nu_n\} = k/2.$$

Si $|\nu_n - m| > x$, alors soit $\nu_n > m + x$ et on a au moins $k/2$ indices i où $\mu^{(i)} > m + x$, soit $\nu_n < m - x$ et on a au moins $k/2$ indices i où $\mu^{(i)} < m - x$. Dans les deux cas on a au moins $k/2$ indices i tels que $|\mu^{(i)} - m| > x$.

(c) De la question suivante on déduit

$$\begin{aligned}\mathbb{P}(|\nu_n - m| > x) &\leq \mathbb{P}(\#\{i, |\mu^{(i)} - m| > x\} \geq k/2) \\ &= \mathbb{P}\left(\sum_{i=1}^{k} \mathbf{1}_{|\mu^{(i)}-m|>x} \geq k/2\right).\end{aligned}$$

Posons $M = \mathbb{E}[1_{|\mu^{(i)}-m|>x}]$. Introduisant M des deux côtés et divisant par k, on obtient

$$\begin{aligned}\mathbb{P}(|\nu_n - m| > x) &\leq \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} \mathbf{1}_{|\mu^{(i)}-m|>x} - M \geq 1/2 - M\right) \\ &\leq \begin{cases}\exp(-2k(1/2-M)^2) & \text{si } (1/2-M) \geq 0 \text{ (Hoeffding)} \\ 1 & \text{sinon (c'est une proba !).}\end{cases} \\ &= \exp(-2k(1/2-M)_+^2)\end{aligned}$$

On peut maintenant utiliser Bienaymé-Tchébychev pour borner $M = \mathbb{P}(|\mu^{(i)} - m| > x) \leq \frac{\mathrm{Var}(\mu^{(i)})}{x^2} = \frac{\sigma^2}{\ell x^2}$.
On en déduit

$$\mathbb{P}(|\nu_n - m| > x) \leq \exp\left(-2\left(1/2 - \frac{\sigma^2}{\ell x^2}\right)_+^2\right)$$

(d) On a

$$\mathbb{P}\left(|\nu_n - m| > \sigma\sqrt{\frac{C\log(1/\alpha)}{n}}\right) \leq \exp\left(-2k\left(1/2 - \frac{n}{\ell C\log(1/\alpha)}\right)_+^2\right) \leq \alpha$$

Il faut donc choisir k et C tels que

$$k\left(1/2 - \frac{n}{\ell C\log(1/\alpha)}\right)_+^2 \geq \frac{\log(1/\alpha)}{2}.$$

Comme $n/\ell \approx k$ on voit que si $k \approx \log(1/\alpha)$ le contenu de la parenthèse est d'ordre constant, ce qui nous donne une chance. Posons $k = \lceil B\log(1/\alpha)\rceil$. On a

$$k\left(1/2 - \frac{n}{\ell C\log(1/\alpha)}\right)_+^2 \geq B\log(1/\alpha)\left(\frac{1}{2} - \frac{Bn}{Clk}\right)_+^2 \geq B\log(1/\alpha)\left(\frac{1}{2} - \frac{2B}{C}\right)^2.$$

Ceci est bien $\geq \frac{\log(1/\alpha)}{2}$ pour par exemple $B = 9$ et $C = 108$.

5. Si on veut estimer un million de paramètres à la fois (possible quand on fait du *machine learning*), qu'on n'a pas d'hypothèses sur les variables en dehors de la variance finie (possible quand on fait de la *finance*) et qu'on veut que tous soient bien estimés simultanément, alors par union, il faut estimer chaque paramètre au niveau $\alpha' = 5.10^{-8}$. On a alors des intervalles de largeur $\leq \sigma\sqrt{\frac{2000}{n}}$ et 200000 observations suffisent à faire ramener la largeur des intervalles à un dixième d'écart-type (possible quand on a du *big data*). Pour l'estimateur usuel, trop sensible aux valeurs extrêmes, il faudrait un milliard d'observations.