Examen du 4 mai 2018

Durée : 2 heures. Tous documents interdits.

Exercice 1. Soit $f:[0,1]\to[0,\infty)$ et $0<T_\infty\le\infty$ tels que $\{f\neq 0\}$ est dénombrable et

$$\#\{x : f(x)\in(0,t]\}<\infty \Longleftrightarrow t<T_\infty.$$

Soit (U_i) une suite de v.a. i.i.d. uniformes dans $[0,1]$. Pour tout $t>0$, on définit $F(t)$ la partition de $\mathbb{N}$ telle que

$$i \overset{F(t)}{\sim} j \Longleftrightarrow \forall x\in[U_i\wedge U_j, U_i\vee U_j], f(x)\notin(0,t].$$

Pour $k\ge 1$, on définit $F_k(t)$ la restriction de $F(t)$ à $[k]:=\{1,\dots,k\}$.

a) Rappeler la topologie usuelle sur l'espace des partitions de $\mathbb{N}$.

b) Que vaut $F(0+)$? Pour $t<T_\infty$ fixé, que peut-on dire de la loi de la partition $F(t)$? Dire pourquoi la limite $F(T_\infty-)$ existe. Cette partition peut-elle avoir des singletons ?

Soient $\beta>-1$ et $\alpha\in\mathbb{R}$. On définit $\mathcal{R}(t):=\{(x,f(x)) : f(x)\in(0,t]\}$. On suppose que $(\mathcal{R}(t);t\ge 0)$ est un processus de Markov et que conditionnellement à $\mathcal{R}(t)=\{(x_1,f(x_1)),\dots,(x_n,f(x_n))\}$, où $x_1<\dots<x_n$, chaque intervalle (x_{i-1},x_i) pour $i\in\{1,\dots,n+1\}$ (notant $x_0=0$ et $x_{n+1}=1$) se fragmente indépendamment à taux Δ_i^α avec $\Delta_i:=x_i-x_{i-1}$. Plus précisément, indépendamment pour chaque i, on ajoute un point à $\mathcal{R}(t)$ égal à $(x_{i-1}+Z_i\Delta_i, t+Y_i)$, où les variables aléatoires

$$Y_i\sim \mathrm{Exp}\,(\Delta_i^\alpha) \qquad \text{et} \qquad Z_i\sim \mathrm{Beta}\,(\beta+1,\beta+1)$$

sont toutes indépendantes.

c) Quelle est la loi du processus $(\#\mathcal{R}(t);t>0)$ lorsque $\alpha=0$?

d) Lorsque $\alpha<0$, montrer que

$$\mathbb{E}(T_\infty)\le\frac{1}{1-\mathbb{E}(Z_1^{-\alpha})}$$

En déduire que $\mathbb{P}(T_\infty<\infty)=0$ ou 1 selon que $\alpha\ge 0$ ou $\alpha<0$.

e) Soient $((V_i,X_i);i\ge 0)$ les atomes d'un processus de Poisson ponctuel sur $[0,1]\times[0,\infty)$ d'intensité la mesure de Lebesgue sur $[0,1]\times[0,\infty)$. Soit $f:[0,1]\to[0,\infty)$ définie par $f(t)=\sum_{i\ge1}X_i\mathbb{1}_{\{V_i\}}$. Montrer que f peut être générée par la procédure décrite précédemment pour certaines valeurs de α et β que l'on précisera.

Solution de l'exercice 1.

a) La topologie usuelle est la topologie générée par la distance ultramétrique

$$d(\pi,\pi')=(\min\{k:\pi_{|[k]}\neq\pi'_{|[k]}\})^{-1}$$

b) On voit facilement que $F(0+)$ est la partition grossière réduite au seul élément $\mathbb{N}$. Pour tout $t < T_\infty$, $F(t)$ est bien sûr échangeable. La limite $F(T_\infty-)$ existe car le processus $(F(t); t \in (0, T_\infty))$ est monotone (partitions de plus en plus fines). En effet, cela implique que $F_k(t)$ est stationnaire (constante pour $t \in (t_k, T_\infty)$ pour un certain t_k), ce qui donne la convergence dans la topologie usuelle. La partition $F(T_\infty-)$ a des singletons dès qu'il existe un intervalle ouvert I tel que $\{x \in I : f(x) \in (0, T_\infty)\}$ est dense dans I. Il est facile de voir qu'alors pour tout $i \in \mathbb{N}$ tel que $U_i \in I$, $\{i\} \in F(T_\infty-)$.

c) Lorsque $\alpha = 0$, le processus $(\#\mathcal{R}(t); t > 0)$ est un processus de Yule de paramètre 1.

d) On suppose que $\alpha < 0$. Pour $t \in (0, T_\infty)$, on définit $I_t := \min\{x : f(x) \in (0, t]\}$. On voit alors que conditionnellement à $I_t = x$, à taux x^α, I_t saute de x à Zx, où Z est comme dans l'énoncé. Ainsi, si T_n désigne le n-ième temps de saut de (I_t) (avec $T_0 = 0$),

$$I_{T_n} = \prod_{i=1}^{n} Z_i,$$

où les Z_i sont iid, et conditionnellement à $I_{T_n} = x$, $T_{n+1} - T_n$ suit la loi exponentielle de paramètre x^α, dont l'espérance est $x^{-\alpha}$. Or $T_\infty \leq \lim_n T_n = \sum_{n\geq 0}(T_{n+1} - T_n)$, donc

$$\mathbb{E}(T_\infty) \leq \sum_{n\geq 0} \mathbb{E}(T_{n+1} - T_n) = \sum_{n\geq 0} \mathbb{E}(I_{T_n}^{-\alpha}) = \sum_{n\geq 0} \mathbb{E}(\prod_{i=1}^{n} Z_i^{-\alpha}) = \sum_{n\geq 0} \mathbb{E}(Z_1^{-\alpha})^n = \frac{1}{1 - \mathbb{E}(Z_1^{-\alpha})}.$$

On peut donc en déduire que $\mathbb{P}(T_\infty < \infty) = 1$ dès que $\alpha < 0$. Pour $\alpha = 0$, nous savons que $\mathbb{P}(T_\infty = \infty) = 1$ puisqu'il s'agit du processus de Yule. Lorsque $\alpha > 0$, les temps d'attente sont plus longs que pour le processus de Yule, donc T_∞ croît stochastiquement avec α, ce qui donne le résultat.

e) Ici chaque intervalle (x, y) à temps t est fragmenté indépendamment par l'atome (V_J, X_J), où

$$J := \arg\min\{X_i : X_i > t, V_i \in (x, y)\}.$$

Ainsi, X_J et V_J sont indépendantes, $X_J - t$ suit la loi exponentielle de paramètre $(y - x)$ et V_J est uniformément distribuée dans (x, y), ce qui donne la bonne description avec $\alpha = 1$ et $\beta = 0$.

Exercice 2. Soit $A \subset \mathbb{N}$ un ensemble aléatoire supposé échangeable, au sens où sa loi est invariante par toute permutation de $\mathbb{N}$.

a) Que vaut $\mathbb{P}(A = \{1\})$? Donner la probabilité que A soit fini.

b) Caractériser la loi de A.

c) À présent on se donne (A, B) un couple de sous-ensembles de $\mathbb{N}$, supposé échangeable. Caractériser la loi de (A, B).

Solution de l'exercice 2.

a) Comme $\mathbb{P}(A = \{1\}) = \mathbb{P}(A = \{j\})$ pour tout j et que $\sum_j \mathbb{P}(A = \{j\}) \leq 1$, on a forcément $\mathbb{P}(A = \{1\}) = 0$. Le même raisonnement montre que A est fini avec probabilité 0.

b) Comme (A, A^c) forment une partition échangeable de $\mathbb{N}$, le théorème de de Finetti assure que A a une fréquence asymptotique $p \in [0, 1]$ et que conditionnellement à p la suite $(\mathbb{1}_A(i))$ est une suite de v.a. de Bernoulli indépendantes de paramètre p.

c) De la même manière, $(A \setminus B, B \setminus A, A \cap B, (A \cup B)^c)$ forment une partition échangeable de $\mathbb{N}$. Il existe donc un triplet (p_A, p_B, p_{AB}) aléatoire de réels de $[0, 1]$ de somme ≤ 1 tel que conditionnellement à ce triplet, chaque entier i choisit indépendamment d'appartenir à $(A \setminus B, B \setminus A, A \cap B, {}^c(A \cup B))$ avec probabilités respectives $p_A, p_B, p_{AB}, 1 - p_A - p_B - p_{AB}$.

Exercice 3. Soit $\mathcal{X}$ un ensemble fini, de cardinal n. On appelle $\mathcal{X}$-arbre un arbre (non nécessairement binaire) enraciné à n feuilles dont les feuilles sont étiquetées par $\mathcal{X}$ (de manière bijective). Une collection $\mathscr{H} \subset \mathscr{P}(\mathcal{X})$ de parties de $\mathcal{X}$ est appelée $\mathcal{X}$-*hiérarchie* si : (i) $\mathcal{X} \in \mathscr{H}$; (ii) pour tout $x \in \mathcal{X}$, $\{x\} \in \mathscr{H}$; et

$$\text{(iii) pour tous } A, B \in \mathscr{H},\ \ A \cap B \in \{A, B, \varnothing\}.$$

a) Montrer qu'il existe une bijection naturelle φ entre les $\mathcal{X}$-arbres et les $\mathcal{X}$-hiérarchies. Piste : le cardinal d'une $\mathcal{X}$-hiérarchie est toujours inférieur ou égal à $2n-1$.

b) Montrer que pour tous $\mathcal{X}$-arbres T et T'

$$d(T, T') = \#(\varphi(T)\Delta\varphi(T'))$$

définit une distance (ici Δ désigne la différence symétrique : $C\Delta C' = (C \cup C') \setminus (C \cap C') = (C \setminus C') \cup (C' \setminus C)$). Elle est appelée *distance de Robinson-Foulds.*

c) Expliquer pourquoi si T et T' sont deux $\mathcal{X}$-arbres *binaires* distincts, $d(T, T') \geq 2$.

On se donne maintenant k $\mathcal{X}$-arbres $T_1, \dots, T_k$.

d) On définit

$$\mathscr{H}^{>\frac{1}{2}} = \{A \subset \mathcal{X} : \#\{j : A \in \varphi(T_j)\} > k/2\}.$$

Montrer que $\mathscr{H}^{>\frac{1}{2}}$ est une $\mathcal{X}$-hiérarchie. On appelle $T^{>\frac{1}{2}} = \varphi^{-1}(\mathscr{H}^{>\frac{1}{2}})$ *l'arbre consensus majoritaire.*

e) Montrer que

$$\sum_{i=1}^{k} d(T_i, T^{>\frac{1}{2}}) \leq \frac{k}{2}\#(\cup_{i=1}^{k}\varphi(T_i)) \leq \frac{k}{2}(n+1+k(n-2)).$$

Solution de l'exercice 3.

a) À chaque nœud de l'arbre T on peut associer la partie de $\mathcal{X}$ qui est l'ensemble des étiquettes des feuilles qu'il sous-tend. On voit facilement que pour tout arbre T, la collection des parties ainsi associées à ses nœuds est une hiérarchie. En effet soient A et B deux telles parties et ν_A et ν_B les nœuds associés. Si $\nu_A \prec \nu_B$, alors $B \subset A$; bien sûr si $\nu_B \prec \nu_A$, alors $A \subset B$; sinon, $A \cap B = \varnothing$ (sans quoi il existerait une feuille qui descend à la fois de ν_A et de ν_B).
L'inverse de φ peut être définie de la manière suivante. Soit $\mathscr{H}$ une hiérarchie. On construit récursivement $T = \varphi^{-1}(\mathscr{H})$ à l'aide d'une correspondance entre les éléments de $\mathscr{H}$ et les nœuds de T en partant de la racine, qui est associée à $\{\mathcal{X}\}$. Pour tout nœud ν de T déjà construit, associé à $A \in \mathscr{H}$, soit A est un singleton et ν est une feuille, soit A n'est pas un singleton, et chaque fille de ν est associée à chaque $B \in \mathscr{H}$ tel que $B \subset A$ qui est maximal (pour tout $C \in \mathscr{H}$, si $C \subset A$ alors $B \not\subset A$).

b) L'inégalité triangulaire vient du simple fait que de manière générale

$$C\Delta C'' \subset (C\Delta C') \cup (C'\Delta C'').$$

c) Si $T \neq T'$, sans perte de généralité (par symétrie), il existe $A \in \varphi(T) \setminus \varphi(T')$. Soit alors B le plus petit élément de $\varphi(T)$ contenant strictement A. Comme T est binaire, $B = A \sqcup A'$, où $A' \in \varphi(T)$. Si on avait simultanément $B \in \varphi(T')$ et $A' \in \varphi(T')$, alors comme T' est aussi binaire, on aurait $A \in \varphi(T')$. Donc on peut trouver au moins deux éléments dans $\varphi(T') \setminus \varphi(T)$, A d'une part, et d'autre part B ou A'. Donc

$$d(T, T') \geq \#(\varphi(T') \setminus \varphi(T)) \geq 2.$$

d) Comme $\mathcal{X} \in \varphi(T_j)$ pour tous j, on a $\#\{j : \mathcal{X} \in \varphi(T_j)\} = k > k/2$ et donc $\mathcal{X} \in \mathscr{H}^{>\frac{1}{2}}$. De la même manière $\mathscr{H}^{>\frac{1}{2}}$ contient tous les singletons.
Soient maintenant $A, B \in \mathscr{H}^{>\frac{1}{2}}$. Alors $\#\{j : A \in \varphi(T_j)\} > k/2$ et $\#\{j : B \in \varphi(T_j)\} > k/2$. Donc il existe i tel que $A \in \varphi(T_i)$ et $B \in \varphi(T_i)$. Comme $\varphi(T_i)$ est une hiérarchie, $A \cap B \in \{A, B, \varnothing\}$.

e) Pour tout i,

$$d(T_i, T^{>\frac{1}{2}}) = \#(\varphi(T_i)\Delta\mathscr{H}^{>\frac{1}{2}}) = \sum_{A\subset\mathcal{X}} \mathbb{1}(A \in \varphi(T_i)\Delta\mathscr{H}^{>\frac{1}{2}}).$$

Donc

$$\sum_{i=1}^{k} d(T_i, T^{>\frac{1}{2}}) = \sum_{A\subset\mathcal{X}} \sum_{i=1}^{k} \mathbb{1}(A \in \varphi(T_i)\Delta\mathscr{H}^{>\frac{1}{2}}).$$

Or

$$\begin{aligned}
\sum_{i=1}^{k} \mathbb{1}(A \in \varphi(T_i)\Delta\mathscr{H}^{>\frac{1}{2}}) &= \sum_{i=1}^{k} \mathbb{1}(A \in \varphi(T_i) \setminus \mathscr{H}^{>\frac{1}{2}}) + \sum_{i=1}^{k} \mathbb{1}(A \in \mathscr{H}^{>\frac{1}{2}} \setminus \varphi(T_i)) \\
&\leq \frac{k}{2}\mathbb{1}(A \in (\cup_i\varphi(T_i) \setminus \mathscr{H}^{>\frac{1}{2}}) + \frac{k}{2}\mathbb{1}(A \in \mathscr{H}^{>\frac{1}{2}}) \\
&= \frac{k}{2}\mathbb{1}(A \in (\cup_i\varphi(T_i)).
\end{aligned}$$

On obtient la première égalité en sommant sur A. La deuxième égalité est due au fait que chaque $\varphi(T_j)$ contient les mêmes $n+1$ éléments (n feuilles et racine) et additionnellement au plus $n-2$ éléments qui ne sont pas partagés par les autres hiérarchies.