Corrigé Examen 12 Mai 2017, 9h00-12h00

Exercice 1. *Variance d'une somme de fonctions indicatrices*

1. On a

$$\begin{aligned}\mathbb{E}\left[S_m^2\right] &= \mathbb{E}\left[\sum_{i=1}^m \mathbb{1}_{A_i} + \sum_{\substack{i,j=1\\ i\sim j}}^m \mathbb{1}_{A_i}\mathbb{1}_{A_j} + \sum_{\substack{i,j=1\\ i\neq j, i\nsim j}}^m \mathbb{1}_{A_i}\mathbb{1}_{A_j}\right]\\ &= \mathbb{E}[S_m] + \Delta_m + \sum_{\substack{i,j=1\\ i\neq j, i\nsim j}}^m \mathbb{P}(A_i)\,\mathbb{P}(A_j)\\ &\leq \mathbb{E}[S_m] + \Delta_m + \mathbb{E}[S_m]^2.\end{aligned}$$

Comme $\mathrm{Var}(S_m) = \mathbb{E}\left[S_m^2\right] - \mathbb{E}\left[S_m\right]^2$, on obtient le résultat.

2. Par méthode du second moment et par la première partie,

$$\mathbb{P}(S_m > 0) \geq 1 - \frac{\mathrm{Var}(S_m)}{\mathbb{E}[S_m]^2} \geq 1 - \frac{(\mathbb{E}[S_m] + \Delta_m)}{\mathbb{E}[S_m]^2} = 1 - o(1).$$

Par ailleurs, soit $\varepsilon > 0$. On a

$$\mathbb{P}\left(\left|\frac{S_m}{\mathbb{E}[S_m]} - 1\right| > \varepsilon\right) \leq \frac{\mathrm{Var}(S_m)}{\mathbb{E}[S_m]^2\varepsilon^2} = o(1).$$

Exercice 2. *Couplage et distance en variation totale*

Solution 1 :

On peut supposer que $\lambda \geq \mu$. Soit $X \sim \mathrm{Poisson}(\mu)$. Indépendamment de X, soit $X' \sim \mathrm{Poisson}(\lambda - \mu)$, défini sur le même espace de probabilité. Posons $Y = X + X'$. Selon l'indication, on a $Y \sim \mathrm{Poisson}(\lambda)$. Donc (X, Y) est un couplage de $(\mathrm{Poisson}(\lambda), \mathrm{Poisson}(\mu))$. En appliquant le Lemme 9 du cours, on obtient

$$d_{VT}\left(\mathrm{Poisson}(\lambda), \mathrm{Poisson}(\mu)\right) \leq \mathbb{P}(X \neq Y) = \mathbb{P}(X' > 0) = 1 - e^{\lambda-\mu} \leq \lambda - \mu.$$

Solution 2 :

Soit $\varepsilon > 0$. Soit $n \in \mathbb{N}$ grand tel que $\max\{\lambda/n, \mu/n\} \leq \varepsilon/2$, et soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d., $\sim \mathrm{Ber}(\lambda/n)$. On peut supposer que l'espace de probabilité contient également des variables aléatoires $Y_1, \ldots, Y_n$ i.i.d., $\sim \mathrm{Ber}(\mu/n)$, tel que (X_i, Y_i), $i = 1, \ldots, n$, est un couplage maximale pour les lois $\mathrm{Ber}(\lambda/n)$ et $\mathrm{Ber}(\mu/n)$, i.e., $d_{VT}(\mathcal{L}(X_i), \mathcal{L}(Y_i) = \mathbb{P}(X_i \neq Y_i)$. Posons $X = \sum_{i=1}^n X_i$ et $Y = \sum_{i=1}^n Y_i$. On a vu en classe que

$$d_{VT}(\mathrm{Poisson}(\lambda), \mathcal{L}(X)) \leq \lambda/n, \quad d_{VT}(\mathrm{Poisson}(\mu), \mathcal{L}(Y)) \leq \mu/n.$$

D'où

$$\begin{aligned}
&d_{VT}\left(\text{Poisson}(\mu), \text{Poisson}(\lambda)\right) \\
&\leq d_{VT}(\text{Poisson}(\lambda), \mathcal{L}(X)) + d_{VT}(\mathcal{L}(X), \mathcal{L}(Y)) + d_{VT}(\text{Poisson}(\mu), \mathcal{L}(Y)) \\
&\leq (\lambda+\mu)/n + \mathbb{P}(X \neq Y) \leq \varepsilon + n\,\mathbb{P}(X_1 \neq Y_1) \leq \varepsilon + |\lambda - \mu|.
\end{aligned}$$

Solution 3 :
Soient $Z \sim$ Poisson(λ) et $W \sim$Poisson (μ). La méthode de Stein-Chen donne la majoration

$$d_{VT}(W, Z) = \sup_{A \subset \mathbb{N}_0} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| = \sup_{A \subset \mathbb{N}_0} \left|\mathbb{E}[\lambda f_A(W+1) - W f_A(W)]\right|.$$

Par le Lemme 12 du cours, on sait que $\sup_{k \in \mathbb{N}_0} |f_A| \leq 1$. Pour une fonction f bornée par 1,

$$\begin{aligned}
\mathbb{E}[\lambda f(W+1) - W f(W)] &= \left|\lambda \sum_{k=0}^{\infty} \mathbb{P}(W=k) f(k+1) - \sum_{k=0}^{\infty} k\,\mathbb{P}(W=k) f(k)\right| \\
&\quad |(\lambda - \mu) \sum_{k=0}^{\infty} \mathbb{P}(W=k) f(k+1)| \leq |\lambda - \mu|,
\end{aligned}$$

où on a utilisé le fait que $W \sim$ Poisson(μ).

Exercice 3. *Méthode de Stein-Chen*

On considère le graphe $G = G_{n,p}$, avec $p = c/n$, $c > 0$. On considère le nombre de triangles isolés W, que l'on décompose de la manière suivante. Soit $M = \binom{n}{3}$ et $T_1, \ldots, T_M$ une énumération des triangles de K_n. Si on pose $X_i = \mathbb{1}_{\{T_i \text{ présent et isolé dans } G\}}$, alors $W = \sum_{i=1}^{M} X_i$. On pose enfin $\lambda = \mathbb{E}[W]$, et $Z \sim$ Poisson(λ). On va montrer que $d_{VT}(W, Z) \to 0$ quand $n \to \infty$.

1. Question de cours : rappeler l'opérateur de Stein de la loi Poisson(λ), et l'équation caractérisant les fonctions f_A, pour $A \subset \mathbb{N}_0$. Déduire la majoration habituelle de $d_{VT}(W, Z)$. Pour $A \subset \mathbb{N}_0$, et $k \in \mathbb{N}_0$, f_A vérifie l'équation

$$\lambda f_A(k+1) - k f_A(k) = \mathbb{1}_A(k) - \mathbb{P}(Z \in A).$$

En posant $k = W$ et en prenant l'espérance, on obtient $\mathbb{E}[\lambda f_A(W+1) - k f_A(k)] = \mathbb{P}(W \in A) - \mathbb{P}(Z \in A)$. D'où

$$d_{VT}(W, Z) = \sup_{A \subset \mathbb{N}_0} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| = \sup_{A \subset \mathbb{N}_0} \left|\mathbb{E}[\lambda f_A(W+1) - W f_A(W)]\right|.$$

2. L'invariance par réétiquetage permet de dire

$$\begin{aligned}
&\mathbb{E}[\lambda f_A(W+1) - W f_A(k)] \\
&= \mathbb{E}[\lambda f_A(W+1)] - \sum_{i=1}^{M} \mathbb{E}[X_i f_A(W)] \\
&= \lambda\, \mathbb{E}[f_A(W+1)] - M\, \mathbb{E}[X_1 f_A(W)] \\
&= \lambda\, \mathbb{E}[f_A(W+1)] - M\, \mathbb{P}(X_1=1)\, \mathbb{E}[X_1 f_A(W)|X_1=1] - M\, \mathbb{P}(X_1=0)\, \mathbb{E}[X_1 f_A(W)|X_1=0] \\
&= \lambda\, \mathbb{E}[f_A(W+1)] - M\, \mathbb{P}(X_1=1)\, \mathbb{E}[f_A(W)|X_1=1] \\
&= \lambda \left(\mathbb{E}[f_A(W+1)] - \mathbb{E}[f_A(W)|X_1=1]\right).
\end{aligned}$$

D'où

$$d_{VT}(W,Z) \leq \lambda \sup_{A\subset\mathbb{N}_0} \left|\mathbb{E}[f_A(W+1)] - \mathbb{E}[f_A(W)|X_1=1]\right|.$$

3. On a $W = \sum_{i=1}^M X_i$. Donc par construction la loi de sachant $\{X_1 = 1\}$ est la même que la loi de $\sum_{i=1}^M Y_i$. D'où

$$\begin{aligned}
\lambda \left|\mathbb{E}[f_A(W+1)] - \mathbb{E}[f_A(W)|X_1=1]\right| &= \lambda \left|\mathbb{E}\left[f_A\left(\sum_{i=1}^M X_i + 1\right) - f_A\left(\sum_{i=1}^M Y_i\right)\right]\right| \\
&\leq \lambda(1\wedge\lambda^{-1})\, \mathbb{E}\left[\left|\sum_{i=1}^M X_i + 1 - \sum_{i=1}^M Y_i\right|\right] \\
&= \lambda(1\wedge\lambda^{-1})\, \mathbb{E}\left[\left|\sum_{i=2}^M X_i + X_1 + 1 - Y_1 - \sum_{i=1}^M Y_i\right|\right] \\
&\leq (\lambda\wedge 1)\left(\mathbb{E}[X_1] + \sum_{i=2}^M \mathbb{E}[|X_i - Y_i|]\right).
\end{aligned}$$

On a utilisé le fait que $Y_1 = 1$ p.s. En prenant le sup, on déduit

$$d_{VT}(W,Z) \leq (\lambda\wedge 1)\left(\mathbb{E}[X_1] + \sum_{i=2}^M \mathbb{E}[|X_i - Y_i|]\right).$$

4. Quand $i \neq 1$ et $T_i \cap T_1 \neq \emptyset$, alors au moins une des arêtes de T_i relie T_1 à $K_n \setminus T_1$. Cette arête n'est jamais présente dans G' par construction. Donc $Y_i = 0$.

   Quand $T_i \cap T_1 = \emptyset$, on sait que les arêtes de T_i sont inchangées entre G et G', et les arêtes incidentes à T_i ne peuvent que disparaître. Donc si $X_i \neq Y_i$, le seul cas possible est que T_i est présent dans G mais pas isolé, et que les arêtes incidentes à T_i dans G ne sont plus dans G'. Ce sont donc des arêtes incidentes à T_1. Donc

$$\mathbb{P}(X_i \neq Y_i) \leq \mathbb{P}(T_i \text{ présent dans } G \text{ et il existe une arête entre } T_1 \text{ et } T_i).$$

Cette dernière probabilité vaut

$$\mathbb{P}\left(\bigcup_{e:T_1\leftrightarrow T_i}\{T_i\subset G,\, e\in G\}\right)\leq 9p^4.$$

On déduit de tout ça que

$$d_{VT}(W,Z)\leq(\lambda\wedge 1)\left(\mathbb{E}[X_1]+\left(3(n-3)+\frac{3(n-3)(n-4)}{2}\right)\mathbb{E}[X_1]+\binom{n-3}{3}\times 9p^4\right).$$

Dans notre régime $p=O(1/n)$ donc $\mathbb{E}[X_1]=O(1/n^3)$ et $\lambda=O(1)$. D'où

$$d_{VT}(W,Z)\leq O(1)\left(\mathbb{E}[X_1]+O(n^2)O(1/n^3)+O(n^3)O(1/n^4)\right)=O(1/n).$$

5. On a $d_{VT}(W,Z)\to 0$. Or $Z\sim\text{Poisson}(\lambda)$ et $\lambda\to e^{-c\frac{c^3}{6}}$. Donc par l'exercice 2, $d_{VT}(Z,\text{Poisson}(e^{-c\frac{c^3}{6}}))\leq|\lambda-e^{-c\frac{c^3}{6}}|\to 0$, d'où $d_{VT}(W,\text{Poisson}(e^{-c\frac{c^3}{6}}))\to 0$ par inégalité triangulaire.

Exercice 4. *Opérateur de Stein pour la loi géometrique*

1. Soit $\mathcal{A}f(k)=(1-p)f(k+1)-f(k)+pf(0)$. Si $Z\sim\text{Geom}(p)$, on a pour $f:\mathbb{N}_0\to\mathbb{R}$ bornée,

$$\begin{aligned}\mathbb{E}\left[(1-p)f(Z+1)\right]&=(1-p)\sum_{m=0}^{\infty}f(m+1)(1-p)^m p\\&=\sum_{m=0}^{\infty}(1-p)^m pf(m)-pf(0)=\mathbb{E}\left[f(Z)-pf(0)\right],\end{aligned}$$

d'où $\mathbb{E}[\mathcal{A}f(Z)]=0$.

2. Soit $f_j(k)=\mathbb{1}_{\{k=j\}}$, $j\in\mathbb{N}_0$. On a $0=\mathbb{E}[\mathcal{A}f_0(W)]=-\mathbb{P}(W=0)+p$, et donc $\mathbb{P}(W=0)=p$. Pour $j\geq 1$,

$$0=\mathbb{E}[\mathcal{A}f_j(W)]=-\mathbb{P}(W=j)+(1-p)\,\mathbb{P}(W=j-1),$$

d'où, par récurrence, $\mathbb{P}(W=j)=(1-p)^j\,\mathbb{P}(W=0)=(1-p)^j p$. Autrement dit, $W\sim\text{Geom}(p)$.

Exercice 5.

1. Pour un graphe G sur $\{1,\dots,n\}$ ayant m arêtes,

$$\mathbb{P}(G_{n,1/2}=G)=\left(\frac{1}{2}\right)^m\left(\frac{1}{2}\right)^{\binom{n}{2}-m}=\frac{1}{2^{\binom{n}{2}}}.$$

2. Posons $X_{K_{k_n}}$ le nombre de copies de K_{k_n} dans $G_{n,1/2}$, on a par l'inégalité de Markov

$$\mathbb{P}(X_{K_{k_n}}>0)\leq\mathbb{E}[X_{K_{k_n}}]=g(n,k_n)=o(1).$$

3. Soit $C_1, \ldots, C_{m(n)}$ une énumération des copies de K_{k_n} dans K_n. On pose $S_n = \sum_{i=1}^m \mathbb{1}_{\{C_i \subset G_{n,1/2}\}}$, et

$$\Delta_n = \sum_{i,j \in \{1,\ldots,m\}:\, i \sim j} \mathbb{P}\left(\{C_i \subset G_{n,1/2}\} \cap \{C_j \subset G_{n,1/2}\}\right),$$

où $i \sim j$ si $i \neq j$ et $\{C_i \subset G_{n,1/2}\}$ et $\{C_j \subset G_{n,1/2}\}$ ne sont pas indépendants, i.e. $C_i \neq C_j$ et $\#(C_i \cap C_j) \geq 2$. En utilisant l'Exercice 1, il suffit de montrer que $\Delta_n = o(\mathbb{E}[S_n]^2)$. On a par symétrie

$$\begin{aligned}
\Delta_n &= \sum_{i \sim j} \mathbb{P}\left(C_i \subset G_{n,1/2}\right) \mathbb{P}\left(C_j \subset G_{n,1/2} \,|\, C_i \subset G_{n,1/2}\right) \\
&= \sum_{j : j \sim 1} \mathbb{P}\left(C_j \subset G_{n,1/2} \,|\, C_1 \subset G_{n,1/2}\right) \sum_i \mathbb{P}\left(C_i \subset G_{n,1/2}\right) \\
&= \mathbb{E}[S_n] \sum_{j : j \sim 1} \mathbb{P}\left(C_j \subset G_{n,1/2} \,|\, C_1 \subset G_{n,1/2}\right).
\end{aligned}$$

Si $j \sim 1$, on a $2 \leq \#(C_j \cap C_1) \leq k_n - 1$, et donc

$$\sum_{j : j \sim 1} \mathbb{P}\left(C_j \subset G_{n,1/2} \,|\, C_1 \subset G_{n,1/2}\right) = \sum_{i=2}^{k_n - 1} \binom{k_n}{i} \binom{n - k_n}{k_n - i} 2^{-\left[\binom{k_n}{2} - \binom{i}{2}\right]}.$$

En utilisant l'indication, on arrive à

$$\frac{1}{\mathbb{E}[S_n]} \sum_{j : j \sim 1} \mathbb{P}\left(C_j \subset G_{n,1/2} \,|\, C_1 \subset G_{n,1/2}\right) = \sum_{i=2}^{k_n - 1} \frac{\binom{k_n}{i}\binom{n-k_n}{k_n - i}}{\binom{n}{k_n}} 2^{\binom{i}{2}} = o(1).$$

On a montré que $\Delta_n = o(\mathbb{E}[S_n]^2)$.

4. Notons que

$$g(n,k) = \binom{k}{n} 2^{-\binom{k}{n}} \leq n^k 2^{-\binom{k}{n}} = 2^{k(\log_2 n - \frac{k-1}{2})}.$$

En remplacent k par $\lfloor (1+\varepsilon) 2 \log_2(n) \rfloor$ on déduit que $g(n, \lfloor (1+\varepsilon) 2 \log_2(n) \rfloor) \to 0$ quand $n \to \infty$. D'autre part, pour $k^2 = o(n)$, $\delta > 0$ et n grand,

$$g(n,k) \geq \frac{1-\delta}{k!} 2^{k(\log_2 n - \frac{k-1}{2})},$$

ce qui montre que $g(n, \lfloor (1-\varepsilon) 2 \log_2(n) \rfloor) \to \infty$ quand $n \to \infty$. Comme

$$\frac{g(n,k+1)}{g(n,k)} = \frac{n-k}{k+1} 2^{-k}, \tag{1}$$

on va que $k \mapsto g(n,k)$ est montone décroissante pour $k \geq 2\log_2 n$. Tout cela montre que $k_0(n)$ est contenu dans

$$[\lfloor (1-\varepsilon) 2 \log_2(n) \rfloor), \lfloor (1+\varepsilon) 2\log_2(n) \rfloor)]$$

à partir d'un certain rang, d'où le résultat.

5. Par (1), on obtient

$$\frac{g(n, k_0(n)-1)}{g(n, k_0(n))} \to \infty \quad \text{quand } n \to \infty.$$

En utilisant $g(n, k_0(n)) \geq 1$ et 3., on en déduit que le nombre de clique pour $G_{n,1/2}$ est au moins $k_0(n) - 1$ avec grande probabilité. D'autre part, on a aussi

$$\frac{g(n, k_0(n)+1)}{g(n, k_0(n)+2)} \to \infty,$$

ce qui montre que le nombre de clique pour $G_{n,1/2}$ est au maximum $k_0(n)+1$ avec grande probabilité (en utilisant le fait que $g(n, k_0(n)+1) \leq 1$ et donc nécessairement $g(n, k_0(n)+2) \to 0$).

6. Le graphe $G'_{n,1/2}$ a la même loi que $G_{n,1/2}$. Donc avec grande probabilité il ne contient pas de clique de taille $k_0(n)+1$. Si on considère un coloriage optimal de $G_{n,1/2}$ et le plus grand ensemble de sommets de même couleur V', alors $|V'| \geq n/\chi(G_{n,1/2})$. Par définition V' est un stable de $G_{n,1/2}$ donc une clique de $G'_{n,1/2}$. On en déduit qu'avec grande probabilité $n/\chi(G_{n,1/2}) \leq k_0(n)+1$, d'où le résultat.