PROBABILITÉS ET STATISTIQUE
Série 7

Exercice 1. *Théorie des nombres et entropie*
Le célèbre théorème des nombres premiers démontré indépendamment en 1896 par la Vallée Poussin et Hadamard affirme que lorsque n tend vers l'infini :

$$\pi(n) := \text{Card}\,\{p \text{ premier} \leq n\} \sim \frac{n}{\log(n)}, \quad \text{ou encore} \sum_{\substack{p\leq n \\ p\,\text{premier}}} \frac{\log(p)}{p} \sim \log(n).$$

On se propose ici de montrer par une méthode probabilisite que lorsque n est grand :

$$\log(n) \leq \sum_{\substack{p\leq n \\ p\,\text{premier}}} \frac{\log(p)}{p}.$$

On fixe un nombre entier n (grand). On tire ensuite un nombre $N \in \{1,\ldots,n\}$ selon la loi uniforme et on écrit sa décomposition en facteurs premiers :

$$N = \prod_{p\leq n} p^{X_p}.$$

1. Montrer que $\mathbb{P}(X_p \geq k) = \dfrac{1}{n}\left\lfloor \dfrac{n}{p^k} \right\rfloor \leq p^{-k}$. En déduire que $\mathbb{E}\left[X_p\right] \leq \dfrac{1}{p-1}$.

   À une variable aléatoire Y à valeurs entières, on associe son entropie (au sens de Shannon) :

   $$H(Y) := -\sum_i \mathbb{P}(Y=i)\log \mathbb{P}(Y=i).$$

   De même, pour un vecteur aléatoire de telles variables, on pose

   $$H(Y_1,\ldots,Y_k) := -\sum_{(i_1,\ldots,i_k)} \mathbb{P}((Y_1,\ldots,Y_k)=(i_1,\ldots,i_k))\log \mathbb{P}((Y_1,\ldots,Y_k)=(i_1,\ldots,i_k)).$$

2. Montrer que $\log(n) = H(N) = H(X_p;\, p\leq n)$.

3. Montrer que si Y_1 et Y_2 sont deux variables indépendantes, alors $H(Y_1,Y_2) = H(Y_1)+H(Y_2)$. En fait, on montre plus généralement que $H(Y_1,\ldots,Y_k) \leq H(Y_1)+\ldots+H(Y_k)$ avec égalité si et seulement si les variables sont indépendantes.

4. En utilisant l'approximation $\frac{1}{n}\left\lfloor \frac{n}{p^k}\right\rfloor \approx 1/p^k$, montrer que

   $$H(X_p) \approx \frac{\log(p)}{p-1} - \log\left(1-\frac{1}{p}\right), \quad \text{puis} \quad \sum_{p\leq n} H(X_p) \approx \sum_{p\leq n}\frac{\log(p)}{p}.$$

   En fait on peut montrer qu'on a toujours $\leq$.

5. Déduire l'inégalité recherchée des questions 2, 3 et 4.

Exercice 2. *Variance*
Soit X une v.a. de densité f_X proportionnelle à

$$f_X(x) \propto \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] + \exp\left[-\frac{1}{2}\left(\frac{x+\mu}{\sigma}\right)^2\right].$$

1. Calculer la constante de normalisation de f_X.
2. Calculer la variance de X.

Exercice 3. *Coefficient de corrélation*
Soient X et ε deux variables aléatoires indépendantes des lois respectives $X \sim \mathcal{N}(0,1)$ et $\mathcal{N}(0,\sigma^2)$ où $\sigma > 0$. Soit $\beta \neq 0$ et $Y := \beta X + \varepsilon$.

1. Exprimer le coefficient de corrélation entre X et Y en fonction σ/β.
2. Soit maintenant $U \sim \mathcal{N}(0,1)$ indépendante de X et $Z = UX$. Calculer le coefficient de corrélation entre X et Z. Commenter.

Exercice 4. *Méthode probabiliste*
On appelle graphe complet à n sommets le graphe $G = (V, E)$ où $V = \{1, ..., n\}$ et où chaque sommet est relié par une arête à tous les autres sommets.
Un *tournoi* (d'un ensemble V de n joueurs) est une orientation des arêtes du graphe complet à n sommets. C'est-à-dire que pour une paire (x, y) de sommets, soit $[x, y] \in E$ soit $[y, x] \in E$ (mais pas les deux).
On dira qu'un chemin dans G est admissible s'il relie les n sommets avec $n-1$ arêtes de E (aucun sommet n'est visité deux fois).

1. *Préliminaires...*
   Soit X une v.a. à valeurs dans $\mathbb{N}^*$. Montrer que
   $$\exists j \in X(\Omega) \; tq \quad j \geq \mathbb{E}(X).$$
2. Montrer qu'il existe un tournoi à n joueurs qui possède au moins $n! \cdot 2^{-(n-1)}$ chemins admissibles.
   Indication : travailler dans l'espace probabilisé $(\Omega = \{tournois\}, \mathcal{P}(\Omega), unif_\Omega)$ *en représentant les chemins dans G par des cycles de $Sym(n)$ et utiliser la linéarité de l'espérance.*

Remarque : N.Alon a prouvé dans les années 90 que le nombre maximum de chemins admissibles dans un tournoi à n joueurs est $\frac{n!}{(2-o(1))^n}$. La borne que l'on trouve ici est donc presque optimale !