PROBABILITÉS ET STATISTIQUE
Série 6

Exercice 1. *Complexité d'un algorithme de tri*
La complexité algorithmique temporelle permet de fixer une borne supérieure du nombre d'opérations qui sont nécessaires pour trier un ensemble de n éléments.
On peut montrer que la complexité temporelle "en moyenne" et "dans le pire des cas" d'un algorithme basé sur une fonction de comparaison ne peut pas être meilleure que $n\log(n)$. Les tris qui ne demandent que $n\log(n)$ comparaisons en moyenne sont alors dits optimaux.

1. Le problème du tri consiste, étant donné une suite $u = (u_1, u_2, ..., u_n)$ d'éléments d'un ensemble totalement ordonné (par exemple), à déterminer une permutation $\sigma \in Sym(n)$ telle que : $y = (u_{\sigma(1)}, u_{\sigma(2)}, ..., u_{\sigma(n)})$ soit triée.
   (a) Montrer que la complexité d'un algorithme de tri par comparaison dichotomique ne peut pas être meilleure que $n\log_2(n)$ dans le pire des cas.
   Indication : modéliser l'algorithme par un arbre binaire et en déterminer la hauteur minimale en utilisant Stirling.
   (b) Montrer que la complexité d'un algorithme de tri par comparaison dichotomique ne peut pas être meilleure que $n\log_2(n)$ en moyenne.
   Indication : travailler avec la mesure uniforme sur $Sym(n)$ et raisonner encore avec les arbres binaires.
2. Le principe du *tri fusion* consiste à trier récursivement la moitié gauche d'un tableau, puis la moitié droite, puis d'interclasser les deux moitiés du tableau. L'interclassement a une complexité $O(n)$ où n est la taille du tableau résultat.
   (a) Ecrire une équation de récurrence pour la complexité d'un algorithme de tri fusion.
   (b) On cherche à montrer que le tri fusion est optimal, i.e. qu'il existe une constante a telle que
   $$C_{fusion}(n) \leq an(\log_2(n) + 1)$$
   En faire une démonstration par récurrence, en considérant séparément les cas "n pair" et "n impair".

Exercice 2. Soient X et Y deux variables indépendantes de loi $\mathcal{N}(0,1)$.

1. Montrer que $U := X^2 + Y^2$ suit une loi exponentielle dont on précisera le paramètre.
2. Montrer que le quotient $Z := X/Y$ suit une loi de Cauchy, c'est-à-dire admet la densité suivante sur $\mathbb{R}$:
   $$\phi_Z(x) = \frac{1}{\pi} \times \frac{1}{1+x^2}.$$
3. Que vaut l'espérance de Z ?

Exercice 3.
Si F et f sont respectivement la fonction de répartition et la densité de probabilité d'une variable aléatoire positive, on désigne par taux de panne la fonction ρ défnie pour $x > 0$ par :

$$\rho(x) = \frac{f(x)}{1 - F(x)}$$

Déterminer le taux de panne ρ pour la loi exponentielle de paramètre $\theta > 0$ et pour la loi de Weibull de densité de probabilité $f(x) = \alpha\theta x^{\alpha-1}\exp(-\theta x^{\alpha})$ pour $x > 0$, où α et θ sont deux paramètres positifs.

Exercice 4.

1. Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$. Montrer que

$$\mathbb{E}(X) = \sum_{n\geq 0} \mathbb{P}(X > n)$$

2. Soit maintenant X une variable aléatoire de densité f à valeurs positives. Montrer que

$$\mathbb{E}(X^r) = \int_0^{\infty} r x^{r-1} \mathbb{P}(X > x) dx$$

3. Application : dans le cas discret, calculer la moyenne d'une variable aléatoire géométrique $X \sim geom(p)$, et dans le cas continu, calculer la moyenne d'une variable aléatoire exponentielle $Y \sim exp(\lambda)$.

Exercice 5. *Simulation de variables aléatoires par la méthode d'inversion*
La plupart des langages informatiques donne accès à une fonction retournant des nombres (pseudo) aléatoires suivant une loi uniforme sur l'intervalle $[0, 1]$. Il est alors utile de pouvoir générer à partir de ceux-ci des nombres suivant d'autres lois. Une des méthodes utilisées à cette fin est la méthode d'inversion :
Soit F une fonction de répartition sur $\mathbb{R}$. On désire générer une variable aléatoire de fonction de répartition F. Pour tout $p \in [0, 1]$, on définit l'inverse généralisé de F par

$$G(p) := \inf\{x \in \mathbb{R},\ F(x) \geq p\}.$$

Remarque : on utilise les conventions $\inf \mathbb{R} = -\infty$ et $\inf \emptyset = +\infty$.

1. Justifier l'existence dans $\mathbb{R}$ de $G(p)$ si $p \in]0, 1[$.
2. Montrer que pour tout x dans $\mathbb{R}$: $G(F(x)) \leq x$.
3. Montrer que si $G(p) \in \mathbb{R}$, alors $F(G(p)) \geq p$.
4. Montrer que $G(p) \leq x \Leftrightarrow F(x) \geq p$.
5. Montrer que si U est une variable de loi uniforme sur $[0, 1]$, alors $G \circ U$ admet F pour fonction de répartition ; i.e. la fonction de répartition de la v.a. $G(U)$ est précisément F. Ceci est utilisé en pratique pour simuler des variables aléatoires.
6. Que vaut G lorsque F est bijective sur $]0, 1[$? Comment simuler une variable de loi $\mathcal{E}(\lambda)$ à partir d'une variable de loi uniforme (donner un algorithme) ?