Feuille d'exercices 1

Exercice 1 (**Queue gaussienne**)
Montrer que pour tout $x > 0$,

$$\max\left(0, \left(\frac{1}{x} - \frac{1}{x^3}\right)\right) e^{-x^2/2} \le \int_x^{+\infty} e^{-t^2/2} dt \le \frac{1}{x} e^{-x^2/2}$$

Exercice 2 (**Inégalité de Hoeffding**)
Soient $X_1, \dots, X_n$ des variables aléatoires réelles indépendantes telles que, pour tout $i \in [\![1, n]\!]$, $a_i \le X_i \le b_i$ avec $a_i < b_i$. Le but de cet exercice est de montrer que, si $Z = \sum_{i=1}^n X_i$, alors pour tout $t \ge 0$,

$$\mathbb{P}\left(Z - \mathbb{E}Z \ge t\right) \le \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

1. Soit Y une variable aléatoire à valeurs dans $[0, 1]$ d'espérance $\mathbb{E}Y = p$. Montrer que pour tout $\lambda \in \mathbb{R}$,
$$\log \mathbb{E}\left[e^{\lambda(Y - \mathbb{E}Y)}\right] \le \varphi(\lambda) \le \frac{\lambda^2}{8},$$
où $\varphi(\lambda) = \log(pe^\lambda + 1 - p) - \lambda p$.
2. En déduire que si X est une variable aléatoire à valeurs dans $[a, b]$ avec $a < b$, alors pour tout $\lambda \in \mathbb{R}$,
$$\log \mathbb{E}\left[e^{\lambda(X - \mathbb{E}X)}\right] \le \frac{\lambda^2(b-a)^2}{8},$$
puis que pour tout $\lambda \in \mathbb{R}$,
$$\log \mathbb{E}\left[e^{\lambda(Z - \mathbb{E}Z)}\right] \le \frac{\lambda^2}{8} \sum_{i=1}^n (b_i - a_i)^2.$$
3. Conclure.

Exercice 3 (**Tirages sans remise**) Soit $n \in \mathbb{N}^*$ et $a_1, \dots, a_n$ des réels. On considère $\eta = (\eta_1, \dots, \eta_n)$ un vecteur i.i.d. de loi uniforme sur $\{1, \dots, n\}$, et $\sigma = (\sigma_1, \dots, \sigma_n)$ une permutation de $\{1, \dots, n\}$ uniformément distribuée. Pour $k \in [\![1, n]\!]$, on note

$$X_k = \sum_{i=1}^k a_{\sigma_i} \qquad \text{et} \qquad Y_k = \sum_{i=1}^k a_{\eta_i}.$$

1. Montrer que l'on peut trouver un couplage (X_k, Y_k) tel que $\mathbb{E}[Y_k \mid X_k] = X_k$.
2. En déduire que pour toute fonction convexe $f : \mathbb{R} \to \mathbb{R}$, on a $\mathbb{E}\left[f(X_k)\right] \le \mathbb{E}\left[f(Y_k)\right]$.
3. Montrer que pour tout $\lambda \in \mathbb{R}$, on a $\log \mathbb{E}\left[e^{\lambda(X_k - \mathbb{E}X_k)}\right] \le \log \mathbb{E}\left[e^{\lambda(Y_k - \mathbb{E}Y_k)}\right]$.

Exercice 4 (**Lemme de Slutsky**) Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de variables aléatoires réelles définies sur le même espace de probabilité. On suppose que $(X_n)_{n\in\mathbb{N}}$ converge en loi vers une variable X et $(Y_n)_{n\in\mathbb{N}}$ converge en loi vers un réel a. Montrer que la suite de couples de variables aléatoires $(X_n, Y_n)_{n\in\mathbb{N}}$ converge en loi vers (X, a).

Exercice 5 (**Delta-méthode**)
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles. On suppose qu'il existe un réel a, une suite de réels positifs $(v_n)_{n\in\mathbb{N}}$ avec $v_n \underset{n\to\infty}{\longrightarrow} +\infty$ et une variable aléatoire X tels que

$$v_n(X_n - a) \xrightarrow[n\to\infty]{\mathcal{L}} X.$$

Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction dérivable en a.

1. Montrer que $v_n(g(X_n) - g(a)) \xrightarrow[n\to\infty]{\mathcal{L}} g'(a)X$.
2. (Stabilisation de la variance.) Soit $\overline{X}_n$ la moyenne empirique associée à un n-échantillon de loi de Poisson de paramètre $\lambda > 0$. Proposer une fonction g telle que $\sqrt{n}\left(g(\overline{X}_n) - g(\lambda)\right)$ converge en loi vers une limite non dégénérée qui ne dépend pas de λ.

Exercice 6 (**Delta-méthode - ordre 2**)
Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d., centrées, de variance 1, et $\overline{X}_n = \sum_{i=1}^n X_i/n$.

1. Étudier la convergence p.s. de $\overline{X}_n$ et en loi de $\sqrt{n}\overline{X}_n$.
2. Montrer que $\sqrt{n}(\cos(\overline{X}_n) - 1)$ converge en loi et identifier la loi limite. Pourquoi la limite est-elle dégénérée ?
3. En reprenant la démonstration de la Delta méthode, déterminer une suite de réels positifs $(v_n)_{n\in\mathbb{N}}$ tendant vers l'infini et une v.a. réelle Z non constante telles que $v_n(\cos(\overline{X}_n) - 1) \xrightarrow{\mathcal{L}} Z$.

Exercice 7 (**Propriétés des moyenne et médiane empiriques**)
Soit $X_1, \dots, X_n$ un échantillon de fonction de répartition F avec $\mathbb{E}(X_1^2) < \infty$.

1. (a) Montrer que $\mathbb{E}(X_1) = \operatorname{argmin}_{t\in\mathbb{R}} \mathbb{E}[(X_1 - t)^2]$.
   (b) Déterminer $\operatorname{argmin}_{t\in\mathbb{R}} \sum_{i=1}^n (X_i - t)^2/n$.
2. On note $x_{1/2} = F^{-1}(1/2)$ la médiane de la loi de X_1.
   (a) Montrer que si F est continue sur $\mathbb{R}$ et strictement croissante sur un voisinage de $x_{1/2}$, alors $x_{1/2} = \operatorname{argmin}_{t\in\mathbb{R}} \mathbb{E}[|X_1 - t|]$.
   (b) Déterminer, selon la parité de n, $\operatorname{argmin}_{t\in\mathbb{R}} \sum_{i=1}^n |X_i - t|/n$.

Exercice 8 (**Lien médiane-espérance**)
Soit X une variable aléatoire dont la médiane est notée m, l'espérance μ et l'écart-type σ. Montrer que

$$|m - \mu| \le \sigma.$$

Exercice 9 (**Loi uniforme**)

1. Soit $(X_1, \dots, X_n)$ un échantillon i.i.d. de loi uniforme sur $[0, 1]$. On considère le ré-arrangement croissant de ces variables : $X_{(1)} \le \dots \le X_{(n)}$.
   (a) Donner une densité de la loi de la variable $R_n = X_{(n)} - X_{(1)}$.
   (b) Étudier les différents modes de convergence de R_n quand $n \to \infty$.
   (c) Étudier le comportement en loi de $n(1 - R_n)$ quand $n \to \infty$.
2. Soit $X_1, \dots, X_n$ un échantillon de loi $\mathcal{U}([0, \theta])$, on veut estimer $\theta > 0$.
   (a) Déterminer un estimateur de θ par la méthode des moments. On le notera $\hat{\theta}_n$.
   (b) À quelle vitesse cet estimateur converge-t-il vers θ ? Entre $\hat{\theta}_n$ et $X_{(n)} = \max_{1\le i\le n} X_i$, lequel choisiriez-vous ?
   (c) Donner un intervalle de confiance non asymptotique de niveau $1 - \alpha$ pour θ.

Exercice 10 Au cours de la seconde guerre mondiale, l'armée alliée notait les numéros de série $X_1, \dots, X_n$ de tous les tanks nazis capturés ou détruits, afin d'obtenir un estimateur du nombre total N de tanks produits. On propose un modèle de tirage uniforme sans remise dans $\{1, \dots, N\}$. Dans ce cadre, les observations ne sont pas indépendantes, mais elles demeurent échangeables.

1. Étudier la loi de $X_{(n)} = \max X_i$ et calculer l'espérance de $\overline{X}_n$. En déduire deux estimateurs non biaisés de N.
2. Proposer deux intervalles de confiance de niveau α. On pourra utiliser le fait que l'inégalité de Hoeffding s'applique également aux tirages sans remise.

Note : Selon Ruggles et Broodie (1947, JASA), la méthode statistique a fourni comme estimation une production moyenne de 246 tanks/mois entre juin 1940 et septembre 1942. Des méthodes d'espionnage traditionnelles donnaient une estimation de 1400 tanks/mois. Les chiffres officiels du ministère nazi des Armements ont montré après la guerre que la production moyenne était de 245 tanks/mois.