TD Théorème central limite, variables gaussiennes.

__Exercice 1 :__ Application du TCL : intervalle de confiance asymptotique. Dans une population, on veut estimer la proportion p de fumeurs. On fait un sondage sur n personnes; on note $\bar{X}_n$ la proportion de fumeurs dans l'échantillon. On cherche un moyen de construire un intervalle $I_n = [\bar{X}_n - \varepsilon_n, \bar{X}_n + \varepsilon_n]$ tel que pour au moins 95% des réalisations de l'échantillon, on ait $p \in I_n$.

a. En utilisant l'inégalité de Tchebychev, évaluer la taille de l'échantillon à sonder si on veut $\varepsilon_n \leq 0.01$).

b. Donner une approximation de ε_n en utilisant le TCL.
Indication : $F(1.96) - F(-1.96) = 0.95$ si F est la fonction de répartition de la loi normale centrée réduite.

c. On veut toujours $\varepsilon_n \leq 0.01$; évaluer la taille approximative de l'échantillon à sonder en utilisant **b.**.

__Exercice 2 :__ Applications du TCL (Feller pp 241-243; Foata-Fuchs pp 240-241).
Formule de Stirling. Soient $X_1, \ldots, X_n$ des v.a.i.i.d. suivant une loi exponentielle de paramètre 1. On note S_n leur somme.
a. Calculer en utilisant le TCL la limite quand n tend vers l'infini de

$$P\left(\sqrt{n}\left(\frac{S_n}{n} - 1\right) \in [\alpha,\ \beta]\right) \ ,$$

pour α et β deux réels.
b. En utilisant le fait que S_n suit une loi $\gamma(n, 1)$, calculer directement la même quantité que ci-dessus.
c. Montrer la formule de Stirling en utilisant **a.** et **b.**.

Feller donne deux autres applications à l'étude des permutations, qui nécessitent l'emploi d'un TCL pour des va indépendantes, mais pas identiquement distribuées.

__Exercice 3 :__ De Moivre-Laplace (Feller T1 3e edition pp 179-181).
Soient X_i une suite de v.a. de Bernoulli indépendantes de paramètre 1/2; on

note $S_n = X_1 + \ldots + X_n$. Pour simplifier, on ne considère que des n pairs; on note $n = 2\nu$. On note, pour $-\nu \leq k \leq \nu$, $a(\nu, k) = P(S_n = \nu + k)$ (k est la déviation par rapport à la moyenne).

a. Montrer que

$$a(\nu, k) = \frac{(2\nu)!}{(\nu + k)!(\nu - k)!} 2^{-2\nu} \ .$$

b. On suppose maintenant que $0 \leq k \leq K_\nu$, avec $K_\nu^3/\nu^2 \to 0$ quand ν tend vers l'infini. Montrer que

$$a(\nu, k) = a(\nu, 0) e^{-k^2/\nu} \left(1 + O(\frac{K_\nu^3}{\nu^2}) \right)$$

c. On note $h = 2/\sqrt{n}$ et ϕ la densité de la loi normale centrée réduite. En utilisant la formule de Stirling $n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n}$, montrer que

$$\frac{a(\nu, k)}{h\phi(kh)} \to 1 \ .$$

Remarque : Il faut ensuite approximer une somme discrète par une intégrale pour obtenir le théorème de De Moivre-Laplace pour $p = 1/2$. Pour $p \neq 1/2$, c'est très similaire, mais plus calculatoire. Autre remarque, on peut aussi utiliser ce moyen pour calculer la constante $\sqrt{2\pi}$ dans la formule de Stirling.

Exercice 4 : Application du TCL : test du χ^2 (Ouvrard T2 pp 326-330).

Soient (X_n) une suite de v.a.i.i.d. réelles, de loi μ. Soient $A_1, \ldots, A_k$ un ensemble d'intervalles disjoints, tels que $\cup_{i=1}^k A_i = \mathbb{R}$ et $\mu(A_i) = p_i > 0$ pour tout i. On définit pour tout $i = 1, \ldots, k$ les v.a.

$$N_i^n = \sum_{j=1}^n \mathbb{I}_{A_i}(X_j) \ ,$$

ainsi que la v.a.

$$\chi_{k,n}^2 = \sum_{i=1}^k \frac{(N_i^n - np_i)^2}{np_i} \ .$$

N_i^n compte le nombre de X_j qui sont dans A_i. On veut montrer que $\chi_{k,n}^2$ converge en loi vers χ_{k-1}^2, la loi du chi-deux à $k-1$ degrés de liberté.

a. On définit pour tout j la v.a. à valeurs dans $\mathbb{R}^k$

$$Y_j = \begin{pmatrix} \mathbb{I}_{A_1}(X_j) \\ \vdots \\ \mathbb{I}_{A_k}(X_j) \end{pmatrix}$$

Calculer l'espérance $\vec{p}$ et la matrice de covariance C de Y_j. Montrer que $C = diag(p_1, \ldots, p_k) - \vec{p} \cdot \vec{p}^t$

b. On définit la matrice diagonale M par $M_{ii} = 1/p_i$ pour $i = 1, \ldots, k$, et la v.a. Z_n

$$Z_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (Y_j - \vec{p}) \ .$$

Montrer que

$$\chi^2_{k,n} = \|M^{1/2} Z_n\|^2 \ .$$

c. Montrer que $M^{1/2} C M^{1/2} = I - uu^t$, avec u un vecteur unitaire. En déduire qu'il existe une matrice orthogonale O telle que

$$O M^{1/2} C M^{1/2} O^t = I - \vec{e_1} \vec{e_1}^t = \begin{pmatrix} 0 & 0 \\ 0 & I_{\mathbb{R}^{k-1}} \end{pmatrix} \ .$$

d. Montrer que si U_n, suite de v.a. de $\mathbb{R}^k$, converge en loi vers $\mathcal{N}(0, B)$, et si A est une matrice $k \times k$, alors AU_n converge en loi vers $\mathcal{N}(0, ABA^t)$ (on pourra utiliser le théorème de Lévy).

e. Étudier la convergence en loi de la v.a. $OM^{1/2} Z_n$ et conclure.

Rappel : Si U est une va dans $\mathbb{R}^d$ qui suit une loi gaussienne $N(0, I_d)$, alors $\|U\|^2$ suit une loi du χ^2 à d degrés de liberté.