# Colles série 6 : Les théorèmes limites et les propriétes de plusieurs estimateurs de la moyenne dans trois modèles

**Sujet 1:**

Soit $X$ une variable de loi $\mathcal{N}(\mu, \sigma^2)$ avec $\mu = 1$ et $\sigma^2 = 4$.

1. Simuler un échantillon de taille $n = 1000$ et calculer pour chaque échantillons des k-premières valeurs la moyenne, la moyenne des carrés et la variance empirique.
2. Représenter la suite des $n$ valeurs obtenues ($s_k^2$ pour $k = 1, ..., n$) par un nuage de points en imposant les limites 1 et 7 sur l'axe des ordonnées. Observe-t-on une convergence lorsque $k \to \infty$ ? Si oui vers quelle limite (la faire figurer en rouge sur le graphique) ?

**Sujet 2:**

Soit $X$ une variable de loi $\mathcal{B}(p)$ avec $p = 0.5$.

1. Tirer $N$ (avec $N$ assez grand, au moins $10^3$...) réalisations d'un échantillon de taille $n$ (pour commencer $n$ petit) et les ranger dans une matrice à $N$ lignes et $n$ colonnes. Calculer les moyennes de chacune des $N$ réalisations du $n$-échantillons avec la fonction `rowMeans()` et les affecter à `xbarn`.
2. Calculer la moyenne empirique et la variance empirique corrigée ($s'^2$) de l'échantillon `xbarn` et représenter sa répartition avec un histogramme. Y superposer ensuite la densité d'une loi normale convenablement choisie. Que se passe-t-il pour $n = 5$ ? pour $n = 500$ ? Quel résultat vient d'être illustré ici ?

**pour les sujets 3-4-5-6 : modèle uniforme**

Dans les exercices suivants, on étudie les propriétés de biais de variances de deux estimateurs de $a$ obtenus avec un échantillon de taille $n$ du modèle uniforme sur $[0, a]$. On etudie aussi la loi asymptotique et l'intervalle de confiance symétrique obtenu avec le premier estimateur proposé, $T_1$.

Rappelons que si $X$ suit une uniforme sur $[0, a]$, $f(x) = 1/a$ si $x \in [0, a]$ et 0 sinon. De plus (exo de TD) on sait montrer que $E(X) = a/2$ et $V(X) = a^2/12$. On pose $T_1 = 2\bar{X}_n$ et $T_2 = maxX_i = X_{(n)}$.

**Sujet 3: Biais de $T_1$ et $T_2$.**

1. Choisir trois valeurs pour $a$, $N$ et $n$. Tirer $N$ réplications d'échantillons de taille $n$ de $X$ (uniforme sur $[0, a]$) qui seront stockées dans une matrice à $N$ lignes et $n$ colonnes. Calculer les $N$ réalisations de $T_1$ obtenues et les affecter à un vecteur de taille $N$ qu'on pourra nommer `echT1`. Faire de même pour $T_2$.
2. Proposer une expérience numérique qui montre que $T_1$ est sans biais (i.e. $E(T_1) - a = 0$) alors que $T_2$ a un biais de $-a/(n + 1)$ (i.e. $E(T_2) - a = -a/(n + 1)$)

**Sujet 4: Variances de $T_1$ et $T_2$**

1. Choisir trois valeurs pour $a$, $N$ et $n$. Tirer $N$ réplications d'échantillons de taille $n$ de $X$ (uniforme sur $[0, a]$) qui seront stockées dans une matrice à $N$ lignes et $n$ colonnes. Calculer les $N$ réalisations de $T_1$ obtenues et les affecter à un vecteur de taille $N$ qu'on pourra nommer `echT1`. Faire de même pour $T_2$.

2. Proposer une expérience numérique qui montre que $V(T_1) = a^2/(3n)$ et que $V(T_2) = na^2/(n+2)(n+1)^2$. Lequel de ces deux estimateurs est le moins variable ?

**Sujet 5: Loi asymptotique de $T_1$**

1. Choisir trois valeurs pour $a$, $N$ et $n$. Tirer $N$ réplications d'échantillons de taille $n$ de $X$ (uniforme sur $[0, a]$) qui seront stockées dans une matrice à $N$ lignes et $n$ colonnes. Calculer les $N$ réalisations de $T_1$ obtenues et les affecter à un vecteur de taille $N$ qu'on pourra nommer `echT1`.

2. Montrer graphiquement que pour $n$ assez grand la loi de $T_1$ est approximativement une gaussienne dont on indiquera les paramètres.

**Sujet 6: Intervalle de confiance construit avec $T_1$**

1. Choisir trois valeurs pour $a$, $N$ et $n$. Tirer $N$ réplications d'échantillons de taille $n$ de $X$ (uniforme sur $[0, a]$) qui seront stockées dans une matrice à $N$ lignes et $n$ colonnes. Calculer les $N$ réalisations de $T_1$ obtenues et les affecter à un vecteur de taille $N$ qu'on pourra nommer `echT1`.

2. Montrer numériquement que
   $I(a, \alpha) = [\sqrt{3n}T_1/(\sqrt{3n} + u_{1-\alpha/2}); \sqrt{3n}T_1/(\sqrt{3n} - u_{1-\alpha/2})]$
   pour $n$ grand est un intervalle de confiance pour $a$ de niveau de confiance approximatif $1 - \alpha$.