# TP 13: Intervalles de confiance.

28 février 2025

## 1 Introduction.

Les estimations ponctuelles faites dans le TP 10 n'apportent pas d'information sur la précision des résultats. Plus précisément, elles ne tiennent pas compte des erreurs dues aux fluctuations d'échantillonnage. Pour évaluer la confiance que l'on peut avoir en une estimation, il est nécessaire de déterminer un intervalle contenant, avec une certaine probabilité fixée au préalable, la vraie valeur du paramètre : c'est l'estimation par intervalle de confiance.
Soit $p$ la probabilité d'un événement que l'on cherche à estimer. On considère $X_1, \cdots, X_n$ des variables aléatoires indépendantes suivant la même loi de Bernoulli de paramètre $p$. On sait qu'alors $\bar{X_n} = \dfrac{1}{n} \sum_{i=1}^{n} X_i$ est un estimateur sans biais de $p$.
On reprend l'exemple du TP 10. On lance successivement une pièce honnête 10 fois. On cherche à estimer la probabilité $p$ d'avoir au moins une fois une succession de trois résultats identiques. On considère la fonction `simul` qui simule cette expérience et renvoie 1 si cet événement se réalise et 0 sinon.

```
import numpy.random as rd
def simul():
   U = [rd.randint(1,3) for i in range(10)]
   x = 0
   for i in range(2,10):
       if U[i]==U[i-1]==U[i-2] :
           x = 1
   return x
```

## Exercice 1 : Intervalle de confiance par Bienaymé-Tchebychev

1. Montrer que pour tout réel $x \in [0,1]$ on a : $x(1-x) \leqslant \dfrac{1}{4}$.
2. Montrer que pour tout réel $t > 0$ on a : $\mathbb{P}(|\bar{X_n} - p| > t) \leq \dfrac{1}{4nt^2}$.
3. Montrer que si l'on prend $t = \dfrac{1}{2\sqrt{n\alpha}}$ alors $[\bar{X_n} - t, \bar{X_n} + t]$ est un intervalle de confiance pour $p$ avec un niveau de confiance supérieur à $1-\alpha$.
4. Pour $1 - \alpha = 0.95$, écrire les instructions qui permettent d'afficher l'intervalle de confiance ci-dessus pour $n = 1000$. Le réel $\lambda$ étant donné par l'utilisateur.
5. On a vu lors du précédent TP que $p \simeq 0.82$. On considère les instructions :

```python
n = 1000
p = 0.82
X = np.zeros(n)
M = np.zeros(n)
t = np.zeros(n)
for i in range(1,n):
    X[i] = simul()
    M[i] = np.mean(X[:i])
    t[i] = 2.24/np.sqrt(i)
plt.close()
plt.plot(range(n),M,color='blue')
plt.plot(range(n),M-t,color='red')
plt.plot(range(n),M+t,color='red')
plt.plot(range(n),p*np.ones(n),color='black')
plt.show()
```

Expliquer le fonctionnement de ce programme et commenter le graphique obtenu.

6. **Vérification.** On fait $N = 1000$ fois l'estimation de $p$ faite à la question 4 et pour chaque estimation on regarde si $p$ appartient à notre intervalle de confiance. Le pourcentage de réussite devrait être environ de 95%.

   Compléter les instructions suivantes. Faire un commentaire.

```python
N = 1000
n = 1000
X = np.zeros([N,N])
for i in range(N):
    for j in range(N):
        X[i][j] = simul()
M=np.mean(X, axis=0)
k = 0
t = ....
for i in range(N):
    if .....:
        k = k+1
print("confiance empirique : ",k/N*100,"%")
```

# Exercice 2 : Intervalle de confiance asymptotique par TCL.

1. Montrer que $Y_n = \dfrac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}}$ converge en loi vers une loi normale centrée réduite.
2. Montrer que si l'on pose $t = \dfrac{\Phi^{-1}(1-\frac{\alpha}{2})}{2\sqrt{n}}$ alors $\left[\bar{X}_n - t, \bar{X}_n + t\right]$ est un intervalle de confiance asymptotique à un niveau de confiance supérieur à $1-\alpha$.
3. La fonction `norm.ppf()` de la bibliothèque `scipy.stats` importée par `from scipy.stats import norm` renvoie les valeurs approchées de $\Phi^{-1}$ où $\Phi$ est la fonction de répartition de la loi normale centrée réduite.

Calculer la valeur de $\Phi^{-1}(1-\frac{\alpha}{2})$ pour $1-\alpha=95\%$.

Ecrire les instructions qui permettent d'afficher l'intervalle de confiance ci-dessus.

4. Modifier les instructions de l'exercice 1 afin d'obtenir le niveau empirique de notre intervalle de confiance. Faire un commentaire.

## Exercice 3 : Intervalle de confiance par TCL+Slutsky.

On admet que $Z_n = \dfrac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}$ converge en loi vers une loi normale centrée réduite.

1. En déduire que $\left[\bar{X}_n - t, \bar{X}_n + t\right]$ est un intervalle de confiance asymptotique à un niveau de confiance $1-\alpha$ pour
$$t = \frac{\Phi^{-1}(1-\frac{\alpha}{2})\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}}$$
2. Reprendre les questions 3 et 4 de l'exercice 2 dans le cas de cet intervalle de confiance asymptotique.

## Exercice 4 : Estimation du paramètre d'une loi de Poisson.

On considère $X_1, \cdots, X_n$ des variables aléatoires indépendantes suivant la même loi de Poisson de paramètre $\lambda$.

On a vu au précédent TP que $Y_n = \dfrac{1}{n}\sum_{i=1}^n X_i$ est un estimateur sans biais de $\lambda$ et $V_n = \dfrac{1}{n-1}\sum_{i=1}^n (X_i - Y_n)^2$ est un estimateur sans biais de $\lambda$.

1. Ecrire un programme qui simule avec `rd.poisson()` les variables aléatoires $X_1, \cdots, X_n$ et qui mémorise dans `M` la variable aléatoire $Y_n$ et dans `V` la variable aléatoire $V_n$.
2. Montrer que les intervalles $[Y_n - t_1; Y_n + t_1]$ et $[Y_n - t_2; Y_n + t_2]$ pour :
$$t_1 = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\sqrt{Y_n}}{\sqrt{n}} \text{ et } t_2 = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\sqrt{V_n}}{\sqrt{n}}$$
sont des intervalles de confiance asymptotiques de $\lambda$ au niveau de confiance $1-\alpha$.
3. Ecrire un programme qui pour un entier $n$ donné par l'utilisateur, affiche ces deux intervalles de confiance pour $1-\alpha = 0.95$. On rappelle que la variance empirique débiaisée se calcule en Python avec l'instruction `np.var` en précisant l'option `ddof=1`.
4. Compléter le programme suivant afin qu'il calcule les confiances empiriques pour $N = 1000$ expériences.

```python
n = 1000
N = 1000
l = 3
X = rd.poisson(l, size=[N,n])
M = np.mean(X, axis=0)
V = np.var(X,ddof=1, axis=0)
k=0
for i in range(N):
    if ....:
        k = k+1
print("niveau empirique 1 : ",(k/N*100))
k = 0
for i in range(N):
    if ....:
        k = k+1
print("niveau empirique 2 : ",(k/N*100))
```