Feuille de TP n° 6

Estimation ponctuelle et par intervalle de confiance (correction)

Correction de l'exercice 1 :

```
import numpy as np
import numpy.random as rd

def Estime(a,n):
    S=0
    N=10000
    for k in range(N):
        X=rd.random(n)
        eps=sp.ndtri(1-a/2)*np.sqrt(np.var(X)/n)
        if -eps<=np.mean(X)-1/2<=eps:
            S=S+1
    return S/N
```

Exécuter `Estime(0.05,10000)` m'a donné `0.9502`. Il s'agit bien approximativement de `1-0.05`. C'est attendu puisque la probabilité que m appartienne à cet intervalle est asymptotiquement $1-\alpha$.

Correction de l'exercice 2 :

1) a) Puisque X_1^2 admet un moment d'ordre 2 (puisque X_1 un moment d'ordre 4), la loi faible des grands nombres entraîne que

$$T_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow[n\to+\infty]{\mathbb{P}} \mathbb{E}(X_1^2) = t^2.$$

La loi faible des grands nombres entraîne aussi que $\overline{X}_{!n} \xrightarrow[n\to+\infty]{\mathbb{P}} m$. La fonction carré étant continue sur $\mathbb{R}$, on en déduit que $(\overline{X}_{!n})^2 \xrightarrow[n\to+\infty]{\mathbb{P}} m^2$.

b) Soit $n\in\mathbb{N}^*$. On a

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2\overline{X}_{!n}X_i + (\overline{X}_{!n})^2\right) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{2\overline{X}_{!n}}{n}\sum_{i=1}^{n} X_i + \frac{2(\overline{X}_{!n})^2}{n}\sum_{i=1}^{n} 1 = T_n^2 - 2(\overline{X}_{!n})^2 + (\overline{X}_{!n})^2$$

donc $S_n^2 = T_n^2 - (\overline{X}_{!n})^2$. Puisque, d'après la formule de Koenig-Huygens, $\sigma^2 = t^2 - m^2$, on en déduit que $S_n^2 - \sigma^2 = T_n^2 - t^2 - \left((\overline{X}_n)^2 - m^2\right)$.

2) Pour tout $n\in\mathbb{N}^*$, par inégalité triangulaire,

$$|S_n^2 - \sigma^2| = |(T_n^2 - t^2) - \left((\overline{X}_n)^2 - m^2\right)| \leqslant [T_n^2 - t^2| + |(\overline{X}_n)^2 - m^2|.$$

Si l'événement $\left[\left|T_n^2 - t^2\right| \geqslant \frac{\varepsilon}{2}\right] \cup \left[\left|(\overline{X}_n)^2 - m^2\right| \geqslant \frac{\varepsilon}{2}\right]$ n'est pas réalisé, alors l'événement

$$\left[\left|T_n^2 - t^2\right| < \frac{\varepsilon}{2}\right] \cap \left[\left|(\overline{X}_n)^2 - m^2\right| < \frac{\varepsilon}{2}\right]$$

et donc l'événement $\left[\left|S_n^2 - \sigma^2\right| < \varepsilon\right]$ est réalisé. Par contraposée, on a bien

$$\left[\left|S_n^2 - \sigma^2\right| \geqslant \varepsilon\right] \subset \left[\left|T_n^2 - t^2\right| \geqslant \frac{\varepsilon}{2}\right] \cup \left[\left|(\overline{X}_n)^2 - m^2\right| \geqslant \frac{\varepsilon}{2}\right].$$

3) La question 1 entraîne que

$$\mathbb{P}\left(\left|T_n^2-t^2\right|\geqslant\varepsilon\right)\xrightarrow[n\to+\infty]{}0\qquad\text{et}\qquad\mathbb{P}\left(\left|(\overline{X}_n)^2-m^2\right|\geqslant\frac{\varepsilon}{2}\right)\xrightarrow[n\to+\infty]{}0.$$

La formule de Poincaré entraîne que

$$\begin{aligned}0\leqslant\mathbb{P}\left(\left|S_n^2-\sigma^2\right|\geqslant\varepsilon\right)&\leqslant\mathbb{P}\left(\left[\left|T_n^2-t^2\right|\geqslant\frac{\varepsilon}{2}\right]\cup\left[\left|(\overline{X}_n)^2-m^2\right|\geqslant\frac{\varepsilon}{2}\right]\right)\\&=\mathbb{P}\left(\left|S_n^2-\sigma^2\right|\geqslant\varepsilon\right)+\mathbb{P}\left(\left|(\overline{X}_n)^2-m^2\right|\geqslant\frac{\varepsilon}{2}\right)\\&\qquad-\mathbb{P}\left(\left[\left|T_n^2-t^2\right|\geqslant\frac{\varepsilon}{2}\right]\cap\left[\left|(\overline{X}_n)^2-m^2\right|\geqslant\frac{\varepsilon}{2}\right]\right)\\&\leqslant\mathbb{P}\left(\left|S_n^2-\sigma^2\right|\geqslant\varepsilon\right)+\mathbb{P}\left(\left|(\overline{X}_n)^2-m^2\right|\geqslant\frac{\varepsilon}{2}\right)\xrightarrow[n\to+\infty]{}0.\end{aligned}$$

Par encadrement, $\mathbb{P}\left(\left|S_n^2-\sigma^2\right|\geqslant\varepsilon\right)\xrightarrow[n\to+\infty]{}0$.

Ceci étant valable pour tout $\varepsilon>0$, cela signifie que $S_n^2\xrightarrow[n\to+\infty]{\mathbb{P}}\sigma^2$.

Correction de l'exercice 3 :

1) Par linéarité, $\mathbb{E}(\overline{X}_n)=\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}\mathbb{E}(X_i)=\dfrac{1}{n}\sum_{k=1}^{n}m=m$. Ainsi $\overline{X}_n$ est un estimateur sans biais. Comme les X_i, $i\geqslant 1$, sont indépendants et admettent une variance, la loi faible des grands nombres assure que $\overline{X}_n$ est convergent vers $\mathbb{E}(X_1)=m$.

2) Soit $n\in\mathbb{N}^*$. Par stabilité des loi Normales, comme $X_1,\dots,X_n$ sont indépendantes et de loi Normale, on a $\displaystyle\sum_{k=1}^{n}X_k\hookrightarrow\mathcal{N}(nm,n\sigma^2)$ et donc $\overline{X}_n\hookrightarrow\mathcal{N}\left(m,\dfrac{\sigma^2}{n}\right)$. On en déduit que $\sqrt{n}\left(\dfrac{\overline{X}_n-m}{\sigma}\right)\hookrightarrow\mathcal{N}(0,1)$ et donc

$$\mathbb{P}\left(m\in\left[\overline{X}_n-\frac{u_\alpha\sigma}{\sqrt{n}}\,;\overline{X}_n+\frac{u_\alpha\sigma}{\sqrt{n}}\right]\right)=\mathbb{P}\left(-u_\alpha\leqslant\sqrt{n}\left(\frac{\overline{X}_n-m}{\sigma}\right)\leqslant u_\alpha\right)=\Phi(u_\alpha)-\Phi(-u_\alpha)$$

Ainsi

$$\mathbb{P}\left(m\in\left[\overline{X}_n-\frac{u_\alpha\sigma}{\sqrt{n}}\,;\overline{X}_n+\frac{u_\alpha\sigma}{\sqrt{n}}\right]\right)=2\Phi(u_\alpha)-1=2\left(1-\frac{\alpha}{2}\right)-1.$$

Ainsi $\left[\overline{X}_n-\dfrac{u_\alpha\sigma}{\sqrt{n}}\,;\overline{X}_n+\dfrac{u_\alpha\sigma}{\sqrt{n}}\right]$ est un intervalle de confiance de niveau $1-\alpha$ pour m.

3)
```
import numpy as np
import numpy.random as rd
import scipy.special as sp
alpha=0.05
u=sp.ndtri(1-alpha/2)
N=10000
def Simul(m,sigma,n):
    eps=u*sigma/np.sqrt(n)
    c=0#compteur
    for k in range(N):
        Xbar=np.mean(rd.normal(m,sigma,n))
        if Xbar-eps<=m<=Xbar+eps:
            c=c+1
    return c/N
```

On observe que, quelles que soient les valeurs de m,σ,n, on obtient une valeur proche de $0,95$.

Correction de l'exercice 4 :

1) Par linéarité, $\mathbb{E}(\overline{X}_n) = \dfrac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_i) = \dfrac{1}{n}\sum_{k=1}^{n} p = p$. Ainsi $\overline{X}_n$ est un estimateur sans biais. Comme les X_i, $i \geqslant 1$, sont indépendants et admettent une variance, la loi faible des grands nombres assure que $\overline{X}_n$ est convergent vers $\mathbb{E}(X_1) = p$.

2) On a $\dfrac{1}{4} - p(1-p) = \dfrac{1-4p+4p^2}{4} = \dfrac{(1-2p)^2}{4} \geqslant 0$ donc $p(1-p) \leqslant \dfrac{1}{4}$.

3) Soit $n \in \mathbb{N}^*$. On a

$$\begin{aligned}
\mathbb{P}\left(p \in \left[\overline{X}_n - \frac{1}{2\sqrt{n\alpha}}\,;\overline{X}_n + \frac{1}{2\sqrt{n\alpha}}\right]\right) &= \mathbb{P}\left(-\frac{1}{2\sqrt{n\alpha}} \leqslant \overline{X}_n - p \leqslant \frac{1}{2\sqrt{n\alpha}}\right) \\
&= \mathbb{P}\left(-\frac{1}{2\sqrt{n\alpha}} \leqslant \overline{X}_n - p \leqslant \frac{1}{2\sqrt{n\alpha}}\right) \\
&= \mathbb{P}\left(\left|\overline{X}_n - p\right| \leqslant \frac{1}{2\sqrt{n\alpha}}\right) \\
&= 1 - \mathbb{P}\left(\left|\overline{X}_n - p\right| > \frac{1}{2\sqrt{n\alpha}}\right).
\end{aligned}$$

Utilisons l'inégalité de Bienaymé-Tchebychev :

$$\begin{aligned}
\mathbb{P}\left(p \in \left[\overline{X}_n - \frac{1}{2\sqrt{n\alpha}}\,;\overline{X}_n + \frac{1}{2\sqrt{n\alpha}}\right]\right) \geqslant 1 - \frac{\mathbb{V}\left(\overline{X}_n\right)}{(\frac{1}{2\sqrt{n\alpha}})^2} &= 1 - 4(n\alpha)\mathbb{V}\left(\overline{X}_n\right) \\
&= 1 - \frac{4\alpha}{n}\mathbb{V}\left(\sum_{k=1}^{n} X_k\right).
\end{aligned}$$

Par indépendance de $X_1, \ldots, X_n$:

$$\begin{aligned}
\mathbb{P}\left(p \in \left[\overline{X}_n - \frac{1}{2\sqrt{n\alpha}}\,;\overline{X}_n + \frac{1}{2\sqrt{n\alpha}}\right]\right) \geqslant 1 - \frac{\mathbb{V}\left(\overline{X}_n\right)}{(\frac{1}{2\sqrt{n\alpha}})^2} &= 1 - 4(n\alpha)\mathbb{V}\left(\overline{X}_n\right) \\
&= 1 - \frac{4\alpha}{n}\sum_{k=1}^{n}\mathbb{V}(X_k) \\
&= 1 - \frac{4\alpha}{n}\sum_{k=1}^{n}\mathbb{V}(X_k)p(1-p) \\
&= 1 - \alpha \times 4p(1-p) \\
&\geqslant 1 - \alpha,
\end{aligned}$$

d'après la question précédente. Ainsi $\left[\overline{X}_n - \dfrac{1}{2\sqrt{n\alpha}}\,;\overline{X}_n + \dfrac{1}{2\sqrt{n\alpha}}\right]$ est un intervalle de confiance de niveau $1-\alpha$ pour p.

4) Soit $n \in \mathbb{N}^*$. On a

$$\mathbb{P}\left(m \in \left[\overline{X}_n - \frac{u_\alpha}{2\sqrt{n}}\,;\overline{X}_n + \frac{u_\alpha}{2\sqrt{n}}\right]\right) = \mathbb{P}\left(-\frac{u_\alpha}{\sqrt{4p(1-p)}} \leqslant \sqrt{n}\frac{\overline{X}_n - p}{\sqrt{p(1-p)}} \leqslant \frac{u_\alpha}{\sqrt{4p(1-p)}}\right).$$

Le Théorème Central Limite entraîne que $\sqrt{n}\dfrac{\overline{X}_n - p}{\sqrt{p(1-p)}} \xrightarrow[n\to+\infty]{\mathscr{L}} N$ avec $N \hookrightarrow \mathcal{N}(0,1)$. Ainsi

$$\begin{aligned}
\mathbb{P}\left(m \in \left[\overline{X}_n - \frac{u_\alpha}{2\sqrt{n}}\,;\overline{X}_n + \frac{u_\alpha}{2\sqrt{n}}\right]\right) \xrightarrow[n\to+\infty]{} \mathbb{P}\left(-\frac{u_\alpha}{\sqrt{4p(1-p)}} \leqslant N \leqslant \frac{u_\alpha}{\sqrt{4p(1-p)}}\right) \\
&= \Phi\left(\frac{u_\alpha}{\sqrt{4p(1-p)}}\right) - \Phi\left(-\frac{u_\alpha}{\sqrt{4p(1-p)}}\right) \\
&= 2\Phi\left(\frac{u_\alpha}{\sqrt{4p(1-p)}}\right) - 1.
\end{aligned}$$

On a $4p(1-p) \leqslant 1$ donc, par stricte croissance de la fonction $t \longmapsto \dfrac{1}{\sqrt{t}}$, $\dfrac{u_\alpha}{\sqrt{4p(1-p)}} \geqslant u_\alpha$. La fonction Φ étant croissante (c'est une fonction de répartition) sur $\mathbb{R}$, on obtient :

$$2\Phi\left(\frac{u_\alpha}{\sqrt{4p(1-p)}}\right) - 1 \geqslant 2\Phi(u_\alpha) - 1 = 2\left(1 - \frac{\alpha}{2}\right) - 1 = 1 - \alpha.$$

Ainsi $\left[\overline{X}_n - \dfrac{u_\alpha}{2\sqrt{n}}\,;\overline{X}_n + \dfrac{u_\alpha}{2\sqrt{n}}\right]$ est un intervalle de confiance asymptotique de niveau $1-\alpha$ pour p.

5) Soit $n \in \mathbb{N}^*$. On a

$$\mathbb{P}\left(m \in \left[\overline{X}_n - \frac{u_\alpha S_n}{\sqrt{n}}\,;\overline{X}_n + \frac{u_\alpha S_n}{\sqrt{n}}\right]\right) = \mathbb{P}\left(-u_\alpha \leqslant \sqrt{n}\frac{\overline{X}_n - p}{S_n} \leqslant u_\alpha\right) \xrightarrow[n\to+\infty]{} \mathbb{P}\left(-u_\alpha \leqslant N \leqslant u_\alpha\right)$$

Cette limite vaut $\Phi(u_\alpha) - \Phi(-u_\alpha) = 2\Phi(u_\alpha) - 1 = 2\left(1 - \dfrac{\alpha}{2}\right) - 1 = 1 - \alpha$. Ainsi $\left[\overline{X}_n - \dfrac{u_\alpha S_n}{\sqrt{n}}\,;\overline{X}_n + \dfrac{u_\alpha S_n}{\sqrt{n}}\right]$ est un intervalle de confiance asymptotique de niveau $1-\alpha$ pour p.

6)

```python
import numpy as np
import numpy.random as rd
import scipy.special as sp
alpha=0.05
u=sp.ndtri(1-alpha/2)
N=10000
def Simul(p,n):
    eps1=1/(2*np.sqrt(n*alpha))
    eps2=u/(2*np.sqrt(n))
    eps3=2*eps2
    amp1=2*eps1
    amp2=2*eps2
    amp3=0
    [c1,c2,c3]=[0,0,0]#compteur
    for k in range(N):
        X=rd.binomial(1,p,n)
        Xbar=np.mean(X)
        S=np.sqrt(np.mean((X-Xbar)**2))
        if Xbar-eps1<=p<=Xbar+eps1:
            c1=c1+1
        if Xbar-eps2<=p<=Xbar+eps2:
            c2=c2+1
        if Xbar-eps3*S<=p<=Xbar+eps3*S:
            c3=c3+1
        amp3=amp3+S
    return [c1/N,c2/N,c3/N,amp1,amp2,2*eps3*amp3/N]
```

```
>>> Simul(0.3,10)
[1.0, 0.9871, 0.841, 1.414213562373095, 0.6197950323045616, 0.5254829244142909]
>>> Simul(0.3,100)
[1.0, 0.9611, 0.9492, 0.4472135954999579, 0.1959963984540054, 0.17861416898976343]
 >>> Simul(0.3,1000)
[1.0, 0.9673, 0.9523, 0.1414213562373095, 0.06197950323045616, 0.056788203772554106]
 >>> Simul(0.5,10)
[1.0, 0.9779, 0.8933, 1.414213562373095, 0.6197950323045616, 0.5859655853200887]
 >>> Simul(0.5,100)
[1.0, 0.9468, 0.9468, 0.4472135954999579, 0.1959963984540054, 0.19501186040552754]
 >>> Simul(0.5,1000)
[1.0, 0.9468, 0.9468, 0.1414213562373095, 0.06197950323045616, 0.06194810208346628]
 >>> Simul(0.9,10)
[1.0, 0.9982, 0.6513, 1.414213562373095, 0.6197950323045616, 0.28173511880390817]
```

```
>>> Simul(0.9,100)
[1.0, 0.9985, 0.9304, 0.4472135954999579, 0.1959963984540054, 0.11570966032981762]
>>> Simul(0.9,1000)
[1.0, 0.9992, 0.9527, 0.1414213562373095, 0.06197950323045616, 0.0371294162153281]
```

On remarque que toutes les expériences ont conduit à ce que p appartienne au premier intervalle de confiance. Ce n'est pas étonnant : en moyenne il est de très grand amplitude (plus de 1 pour le cas $n = 10$ et près de $0,5$ pour $n = 100$... alors que $p \in]0\,;1[$). Avec le deuxième intervalle de confiance, on a un niveau réel supérieur à α (y compris pour de petites valeurs de n mais la contrepartie est que l'amplitude moyenne est élevée) mais une amplitude plus faible en moyenne que celle du premier intervalle de confiance. Enfin le troisième intervalle de confiance a un niveau réel plus proche de α (du moins pour $n \geqslant 100$) mais c'est celui des trois qui possède l'amplitude moyenne la plus faible.

Correction du problème :

1) a) La fonction f_θ est continue sur $\mathbb{R}\backslash\{\theta\}$, elle est positive. Par ailleurs elle est nulle en dehors du segment $]0\,;\theta[$ donc $\displaystyle\int_{-\infty}^{+\infty} f_\theta(x)\,dx$ converge et est égale à

$$\int_0^\theta f_\theta(x)\,dx = \int_0^\theta \frac{2x}{\theta^2}\,dx = \left[\frac{x^2}{\theta^2}\right]_0^\theta = 1 - 0 = 1.$$

Ainsi f_θ est bien une densité de probabilité.

b) • Si $t < 0$, alors $F_\theta(t) = \displaystyle\int_{-\infty}^{t} f_\theta(x)\,dx = 0$.

• Si $t \in [0\,;\theta]$, alors $F_\theta(t) = \displaystyle\int_{-\infty}^{0} f_\theta(x)\,dx + \int_0^t f_\theta(x)\,dx = 0 + \int_0^t \frac{2x}{\theta^2}\,dx = \left[\frac{x^2}{\theta^2}\right]_0^t = \frac{t^2}{\theta^2}$.

• Si $t > \theta$, alors $F_\theta(t) = \displaystyle\int_{-\infty}^{0} f_\theta(x)\,dx + \int_0^\theta f_\theta(x)\,dx + \int_\theta^t f_\theta(x)\,dx = 0 + 1 + 0 = 1$.

Ainsi

$$F_\theta : t \longmapsto \begin{cases} 0 & \text{si} \quad t < 0 \\ \dfrac{t^2}{\theta^2} & \text{si} \quad t \in [0\,;\theta] \\ 1 & \text{si} \quad t > \theta \end{cases}$$

2) a) • Si $x < 0$, $\mathbb{P}(\theta\sqrt{U} \leqslant x) = \mathbb{P}(\varnothing) = 0$.

• Si $x \geqslant 0$, par stricte monotonie de la fonction carré sur $\mathbb{R}_+$,

$$\mathbb{P}(\theta\sqrt{U} \leqslant x) = \mathbb{P}(\theta^2 U \leqslant x^2) = \mathbb{P}\left(U \leqslant \frac{x^2}{\theta^2}\right) = F_U\left(\frac{x^2}{\theta^2}\right).$$

Ainsi

$$\mathbb{P}(\theta\sqrt{U} \leqslant x) = \begin{cases} 0 & \text{si} & \dfrac{x^2}{\theta^2} < 0 \\ \dfrac{x^2}{\theta^2} & \text{si} & \dfrac{x^2}{\theta^2} \in [0\,;1] \\ 1 & \text{si} & \dfrac{x^2}{\theta^2} > 1 \end{cases} = \begin{cases} \dfrac{x^2}{\theta^2} & \text{si} & x \in [0\,;\theta] \\ 1 & \text{si} & x > \theta \end{cases}$$

Ainsi $\theta\sqrt{U}$ et X ont même fonction de répartition. Comme la fonction de répartition caractérise la loi, $\theta\sqrt{U}$ et X ont donc la même loi.

b)
```python
import numpy as np
import numpy.random as rd
def Simul_M(n,theta):
    X=theta*np.sqrt(rd.random(n))
    return np.max(X)
```

c)

```python
def Simul_MM(n, theta):
    X=theta*np.sqrt(rd.random(n))
    return [np.max(X[:k]) for k in range(1,n+1)]
```

d)
```python
import matplotlib.pyplot as plt
for k in range(5):
    L=Simul_MM(100,1)
    plt.plot(range(1,101),L)
plt.show()
```

On peut conjecturer que $(M_n)_{n\geqslant 1}$ converge (en probabilité) vers $1=\theta$ et que la convergence est très rapide.

e) Soit $n \in \mathbb{N}^*$. Soit $x \in \mathbb{R}$. On a, par mutuelle indépendances de $X_1, \ldots, X_n$,

$$F_{M_n}(x) = \mathbb{P}\left(\bigcap_{k=1}^{n} [X_k \leqslant x]\right) = \prod_{k=1}^{n} \mathbb{P}(X_k \leqslant x).$$

Comme $X_1, \ldots, X_n$ ont la même loi, on obtient :

$$F_{M_n}(x) = \prod_{k=1}^{n} F_\theta(x) = (F_\theta(x))^n = \begin{cases} 0 & \text{si } x < 0 \\ \left(\dfrac{x}{\theta}\right)^{2n} & \text{si } x \in [0\,;\theta] \\ 1 & \text{si } x > \theta \end{cases}$$

Il s'agit d'une fonction continue sur $\mathbb{R}$ et de classe C^1 sur $\mathbb{R}\backslash\{\theta\}$ donc M_n est à densité. En dérivant, on trouve qu'une densité de M_n est

$$f_{M_n} : x \longmapsto \begin{cases} 2n\dfrac{x^{2n-1}}{\theta^{2n}} & \text{si } x \in [0\,;\theta] \\ 0 & \text{sinon} \end{cases}$$

f) Puisque $x \longmapsto |x^2| f_{M_n}(x)$ est continue sur $[0\,;\theta]$ et nulle en dehors, l'intégrale $\displaystyle\int_{-\infty}^{+\infty} |x^2| f_{M_n}(x)\,dx$ converge et donc M_n admet un moment d'ordre 2.

- On a
$$\mathbb{E}(M_n) = \int_{-\infty}^{+\infty} x f_{M_n}(x)\,dx = \int_0^\theta 2n\frac{x^{2n}}{\theta^{2n}}\,dx = \frac{2n}{\theta^{2n}}\left[\frac{x^{2n+1}}{2n+1}\right]_0^\theta = \frac{2n\theta}{2n+1}.$$

- Par théorème de transfert, on a
$$\mathbb{E}(M_n^2) = \int_{-\infty}^{+\infty} x^2 f_{M_n}(x)\,dx = \int_0^\theta 2n\frac{x^{2n+1}}{\theta^{2n}}\,dx = \frac{2n}{\theta^{2n}}\left[\frac{x^{2n+2}}{2n+2}\right]_0^\theta = \frac{n\theta^2}{n+1}.$$

g) On a $\mathbb{E}(M_n) = \dfrac{2n}{2n+1}\theta \neq \theta$ mais $\mathbb{E}(M_n) \xrightarrow[n\to+\infty]{} \theta$. Ainsi M_n est un estimateur biaisé de θ mais asymptotiquement sans biais.

h) On a

$$\mathbb{P}(|M_n-\theta| \geqslant \varepsilon) = 1 - \mathbb{P}(|M_n-\theta| < \varepsilon) = 1 - \mathbb{P}(\theta-\varepsilon < M_n < \theta+\varepsilon) = 1 - F_{M_n}(\theta+\varepsilon) + F_{M_n}(\theta-\varepsilon).$$

Comme $\theta+\varepsilon > \theta$, on a $F_{M_n}(\theta+\varepsilon) = 1$ et donc

$$\mathbb{P}(|M_n-\theta| \geqslant \varepsilon) = F_{M_n}(\theta-\varepsilon) = \begin{cases} 0 & \text{si } \theta-\varepsilon < 0 \\ \left(\dfrac{\theta-\varepsilon}{\theta}\right)^{2n} & \text{si } \theta-\varepsilon \in [0\,;\theta] \\ 1 & \text{si } \theta-\varepsilon > \theta \end{cases}$$
$$= \begin{cases} 0 & \text{si } \varepsilon > \theta \\ \left(1-\dfrac{\varepsilon}{\theta}\right)^{2n} & \text{si } \varepsilon \leqslant \theta \end{cases}$$

Dans le cas où $\varepsilon \leqslant \theta$, $0 \leqslant 1 - \dfrac{\varepsilon}{\theta} < 1$ donc $\left(1-\dfrac{\varepsilon}{\theta}\right)^{2n} \xrightarrow[n\to+\infty]{} 0$. Ainsi, pour toutes les valeurs de $\varepsilon > 0$, on a $\mathbb{P}(|M_n-\theta| \geqslant \varepsilon) \xrightarrow[n\to+\infty]{} 0$. Cela signifie que M_n est un estimateur convergent de θ.

i) On a, par linéarité de l'espérance,

$$\mathbb{E}((M_n-\theta)^2)=\mathbb{E}(M_n^2-2\theta M_n+\theta^2)=\mathbb{E}(M_n^2)-2\theta\mathbb{E}(M_n)+\theta^2$$

donc, d'après la question f,

$$\mathbb{E}((M_n-\theta)^2)=\frac{n\theta^2}{n+1}-\frac{4n\theta^2}{2n+1}+\theta^2=\theta^2\,\frac{n(2n+1)-4n(n+1)+(n+1)(2n+1)}{(2n+1)(n+1)}=\frac{\theta^2}{(2n+1)(n+1)}.$$

Par stricte croissante de la fonction carré et par l'inégalité de Markov appliquée à la variable aléatoire positive $(M_n-\theta)^2$, on a, pour tout $\varepsilon>0$,

$$\mathbb{P}(|M_n-\theta|\geqslant\varepsilon)=\mathbb{P}((M_n-\theta)^2\geqslant\varepsilon^2)\leqslant\frac{\mathbb{E}((M_n-\theta)^2)}{\varepsilon^2}=\frac{\theta^2}{(2n+1)(n+1)\varepsilon^2}\xrightarrow[n\to+\infty]{}0.$$

Par théorème d'encadrement, on a $\mathbb{P}(|M_n-\theta|\geqslant\varepsilon)\xrightarrow[n\to+\infty]{}0$. Ainsi $(M_n)_{n\geqslant1}$ converge en probabilité vers θ. On retrouve le fait que M_n est un estimateur convergent de θ.

3) a) Déjà X admet bien une moment d'ordre 2 puisque $x\longmapsto|x^2|f_\theta(x)$ est continue sur $[0\,;\theta]$ et nulle en dehors de $[0\,;\theta]$ si bien que $\displaystyle\int_{-\infty}^{+\infty}|x^2|f_\theta(x)\,dx$.

Par linéarité de l'espérance, $\mathbb{E}(\overline{X}_n)=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\mathbb{E}(X_i)=\dfrac{1}{n}n\mathbb{E}(X)$ donc

$$\mathbb{E}(\overline{X}_n)=\mathbb{E}(X)=\int_{-\infty}^{+\infty}xf_\theta(x)\,dx=\int_0^\theta x\frac{2x}{\theta^2}\,dx=\frac{2}{\theta^2}\left[\frac{\theta^3}{3}\right]_0^\theta=\frac{2\theta}{3}.$$

Par indépendance de $X_1,\dots,X_n$, $\mathbb{V}(\overline{X}_n)=\dfrac{1}{n^2}\displaystyle\sum_{i=1}^{n}\mathbb{V}(X_i)=\dfrac{1}{n^2}n\mathbb{V}(X)=\dfrac{\mathbb{V}(X)}{n}$. Par théorème de transfert :

$$\mathbb{E}(X^2)=\int_{-\infty}^{+\infty}x^2f_\theta(x)\,dx=\int_0^\theta x^2\frac{2x}{\theta^2}\,dx=\frac{2}{\theta^2}\left[\frac{\theta^4}{4}\right]_0^\theta=\frac{\theta^2}{2}.$$

La formule de Koenig-Huygens entraîne alors

$$\mathbb{V}(X)=\mathbb{E}(X^2)-\mathbb{E}(X)^2=\frac{\theta^2}{2}-\frac{4\theta^2}{9}=\frac{\theta^2}{18}.$$

Finalement $\mathbb{V}(\overline{X}_n)=\dfrac{\theta^2}{18n}$.

b) On pose $Z_n=\dfrac{3}{2}\overline{X}_n$. Par linéarité de l'espérance, on a alors $\mathbb{E}(Z_n)=\dfrac{3}{2}\mathbb{E}(\overline{X}_n)=\theta$. Ainsi Z_n est un estimateur sans biais de θ.

c)
```
def Simul_Z(n,theta):
    X=theta*np.sqrt(rd.random(n))
    return np.mean(X)*3/2
```

d)
```
def Simul_ZZ(n,theta):
    X=theta*np.sqrt(rd.random(n))
    return (np.cumsum(X)/[k for k in range(1,n+1)])*3/2
```

e)
```
for k in range(5):
    L=Simul_ZZ(100,1)
    plt.plot(range(1,101),L)
plt.show()
```

On peut conjecturer que $(Z_n)_{n\geqslant1}$ converge (en probabilité) vers $1=\theta$. Néanmoins la convergence semble moins rapide que celle de $(M_n)_{n\geqslant1}$.

f) On a, par linéarité de l'espérance,

$$\mathbb{E}((Z_n-\theta)^2)=\mathbb{E}(Z_n^2-2\theta Z_n+\theta^2)=\mathbb{E}(Z_n^2)-2\theta\mathbb{E}(Z_n)+\theta^2=\mathbb{E}(Z_n^2)-\theta^2.$$

La formule de Koenig-Huygens entraîne que

$$\mathbb{E}(Z_n^2)=\mathbb{V}(Z_n)+\mathbb{E}(Z_n)^2=\frac{9}{4}\mathbb{V}(\overline{X}_n)+\theta^2=\frac{9}{4}\frac{\theta^2}{18n}+\theta^2=\frac{\theta^2}{8n}+\theta^2$$

et donc $\mathbb{E}((Z_n-\theta)^2)=\dfrac{\theta^2}{8n}$.

Par stricte croissante de la fonction carré et par l'inégalité de Markov appliquée à la variable aléatoire positive $(Z_n-\theta)^2$, on a, pour tout $\varepsilon>0$,

$$\mathbb{P}(|Z_n-\theta|\geqslant\varepsilon)=\mathbb{P}((Z_n-\theta)^2\geqslant\varepsilon^2)\leqslant\frac{\mathbb{E}((Z_n-\theta)^2)}{\varepsilon^2}=\frac{\theta^2}{8n\varepsilon^2}\xrightarrow[n\to+\infty]{}0.$$

Par théorème d'encadrement, on a $\mathbb{P}(|M_n-\theta|\geqslant\varepsilon)\xrightarrow[n\to+\infty]{}0$. Ainsi $(Z_n)_{n\geqslant1}$ converge en probabilité vers θ. Autrement dit Z_n est un estimateur convergent de θ.

4) a) Pour tout $n\in\mathbb{N}^*$, $\dfrac{\theta^2}{(2n+1)(n+1)}\leqslant\dfrac{\theta^2}{8n}$ si et seulement si $2n^2+3n+1\geqslant 8n$ si et seulement si $2n^2-5n+1\geqslant0$. Un calcul de discriminant classique montre que c'est le cas dès que $n\geqslant3$. Ainsi, pour $n\geqslant3$, le risque quadratique de M_n est inférieur à celui de Z_n. Par ailleurs, lorsque n tend vers $+\infty$, le risque quadratique de M_n est négligeable devant celui de Z_n. Autrement dit, on se trompe beaucoup moins avec M_n qu'avec Z_n. Cela nous pousse à choisir M_n.

b)
```
M=[Simul_M(100,1) for k in range(5000)]
plt.hist(M,bins=50,density=True)
plt.figure()
Z=[Simul_Z(100,1) for k in range(5000)]
plt.hist(Z,bins=50,density=True)
plt.show()
```

On observe que les valeurs prises par Z_n sont beaucoup plus étalées que celles prises par M_n. Ainsi Z_n semble bien moins précis. Cela nous conforme donc dans le choix de M_n.

5) a) Soit $x\in\mathbb{R}$. On a

$$\mathbb{P}(2n(\theta-M_n)\leqslant x)=\mathbb{P}(M_n\geqslant\theta+\frac{x}{2n})=1-F_{M_n}\left(\theta-\frac{x}{2n}\right).$$

Le calcul de F_{M_n} a été fait à la question 2e :

$$\mathbb{P}(2n(\theta-M_n)\leqslant x)=1-\begin{cases}0 & \text{si } \theta-\dfrac{x}{2n}<0\\ \left(\dfrac{\theta-x/(2n)}{\theta}\right)^{2n} & \text{si } \theta-\dfrac{x}{2n}\in[0\,;\theta]\\ 1 & \text{si } \theta-\dfrac{x}{2n}>\theta\end{cases}$$

donc

$$\mathbb{P}(2n(\theta-M_n)\leqslant x)=\begin{cases}1 & \text{si } 2n\theta<x\\ 1-\left(1-\dfrac{x}{2n\theta}\right)^{2n} & \text{si } x\in[0\,;2n\theta]\\ 0 & \text{si } x<0\end{cases}$$

Ainsi :

- Si $x<0$, $\mathbb{P}(2n(\theta-M_n)\leqslant x)=0\xrightarrow[n\to+\infty]{}0$.
- Si $x\geqslant0$ alors, pour n assez grand, $x\leqslant 2n\theta$ et donc $\mathbb{P}(2n(\theta-M_n)\leqslant x)=1-\left(1-\dfrac{x}{2n\theta}\right)^{2n}$. On a

$$ln\left(\left(1-\frac{x}{2n\theta}\right)^{2n}\right)=2n\ln\left(1-\frac{x}{2n\theta}\right)\underset{n\to+\infty}{\sim}2n\left(-\frac{x}{2n\theta}\right)=-\frac{x}{\theta}$$

car $-\dfrac{x}{2n\theta} \xrightarrow[n\to+\infty]{} 0$. Puisque $\exp$ est continue en $-\dfrac{x}{\theta}$, on obtient que $\left(1-\dfrac{x}{2n\theta}\right)^{2n} \xrightarrow[n\to+\infty]{} e^{-x/\theta}$.
Ainsi $\mathbb{P}(2n(\theta - M_n) \leqslant x) \xrightarrow[n\to+\infty]{} 1-e^{-x/\theta}$.

On en déduit que $F_{2n(\theta-M_n)}(x) \xrightarrow[n\to+\infty]{} (1-e^{-x/\theta})\mathbb{1}_{\mathbb{R}_+}(x)$. On reconnaît une fonction de réparation d'une variable aléatoire de loi $\mathcal{E}(1/\theta)$ à la limite. Par conséquent $\big(2n(\theta-M_n)\big)_{n\geqslant 1}$ converge en loi vers une variable aléatoire de loi $\mathcal{E}(1/\theta)$

b) Remarquons que $\mathbb{P}(M_n \leqslant \theta) = 1$. Ensuite, on a
Pour tout $x \in \mathbb{R}_+^*$, on a $\mathbb{P}(2n(\theta - M_n) \leqslant x) \xrightarrow[n\to+\infty]{} 1-e^{-x/\theta}$. Si on prend $x = -\theta\ln(\alpha)$, on obtient $\mathbb{P}(2n(\theta - M_n) \leqslant x) \xrightarrow[n\to+\infty]{} 1-e^{-x} = 1-\alpha$. Par ailleurs

$$[2n(\theta - M_n) \leqslant x] = [2n(\theta - M_n) \leqslant -\theta\ln(\alpha)] = \left[2n\left(\frac{M_n}{\theta}-1\right) \geqslant \ln(\alpha)\right] = \left[\frac{M_n}{\theta} \geqslant 1+\frac{\ln(\alpha)}{2n}\right].$$

Ainsi, pour tout $n\in\mathbb{N}^*$, tel que $1+\dfrac{\ln(\alpha)}{2n} > 0$ (c'est le cas pour n assez grand),

$$[2n(\theta - M_n) \leqslant x] = \left[\frac{\theta}{M_n} \leqslant \frac{1}{1+\frac{\ln(\alpha)}{2n}}\right].$$

Finalement

$$\mathbb{P}\left(M_n \leqslant \theta \leqslant \frac{M_n}{1+\frac{\ln(\alpha)}{2n}}\right) \xrightarrow[n\to+\infty]{} 1-\alpha.$$

Ainsi $\left[M_n ; \dfrac{M_n}{1+\frac{\ln(\alpha)}{2n}}\right]$ est un intervalle de confiance asymptotique de niveau $1-\alpha$ pour θ.

6) a) Puisque $(X_i)_{i\geqslant 1}$ est une suite de variables aléatoires indépendantes et de même loi admettant pour espérance $\dfrac{2\theta}{3}$ et pour variance $\dfrac{\theta^2}{18}$ (cf. question 3f), le théorème Central Limite entraîne que

$$\sqrt{n}\frac{\overline{X}_n - \frac{2\theta}{3}}{\sqrt{\frac{\theta^2}{18}}} \xrightarrow[n\to+\infty]{\mathscr{L}} N$$

avec $N \hookrightarrow \mathcal{N}(0,1)$. On a

$$\sqrt{n}\frac{\overline{X}_n - \frac{2\theta}{3}}{\sqrt{\frac{\theta^2}{18}}} = \sqrt{n}\frac{2}{3}\frac{Z_n-\theta}{\frac{\theta}{3\sqrt{2}}} = \frac{2\sqrt{2}}{\theta}\sqrt{n}\,(Z_n-\theta)\,.$$

Puisque $x \longmapsto \dfrac{\theta}{2\sqrt{2}}x$ est continue, on en déduit que $\left(\sqrt{n}\,(Z_n-\theta)\,\right)_{n\geqslant 1}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}\left(0,\dfrac{\theta^2}{8}\right)$.

b) Reprenons un peu plus haut :

$$2\sqrt{2n}\left(\frac{Z_n}{\theta}-1\right) = \frac{2\sqrt{2}}{\theta}\sqrt{n}\,(Z_n-\theta) \xrightarrow[n\to+\infty]{\mathscr{L}} N.$$

Grand classique : si on note $u_{1-\alpha/2} = \Phi^{-1}\left(1-\dfrac{\alpha}{2}\right)$,

$$\mathbb{P}(|N| \leqslant u_{1-\alpha/2}) = \mathbb{P}(-u_{1-\alpha/2} \leqslant N \leqslant u_{1-\alpha/2}) = \Phi\left(u_{1-\alpha/2}\right) - \Phi\left(-u_{1-\alpha/2}\right) = 2\Phi\left(u_{1-\alpha/2}\right) - 1$$

donc, par définition, $\mathbb{P}(|N| \leqslant u_{1-\alpha/2}) = 2\left(1-\dfrac{\alpha}{2}\right) - 1 = 1-\alpha$. Ainsi

$$\mathbb{P}\left(\left|2\sqrt{2n}\left(\frac{Z_n}{\theta}-1\right)\right| \leqslant u_{1-\alpha/2}\right) \xrightarrow[n\to+\infty]{} 1-\alpha$$

donc

$$\mathbb{P}\left(1-\frac{u_{1-\alpha/2}}{2\sqrt{2n}} \leqslant \frac{Z_n}{\theta} \leqslant 1+\frac{u_{1-\alpha/2}}{2\sqrt{2n}}\right) \xrightarrow[n\to+\infty]{} 1-\alpha$$

donc, pour tout $n \in \mathbb{N}^*$ tel que $1-\dfrac{u_{1-\alpha/2}}{2\sqrt{2n}}>0$ (c'est le cas pour n assez grand),

$$\mathbb{P}\left(\frac{1}{1+\frac{u_{1-\alpha/2}}{2\sqrt{2n}}} \leqslant \frac{\theta}{Z_n} \leqslant \frac{1}{1-\frac{u_{1-\alpha/2}}{2\sqrt{2n}}}\right) \xrightarrow[n\to+\infty]{} 1-\alpha$$

Ainsi $\left[\dfrac{Z_n}{1+\frac{u_{1-\alpha/2}}{2\sqrt{2n}}}; \dfrac{Z_n}{1-\frac{u_{1-\alpha/2}}{2\sqrt{2n}}}\right]$ est un intervalle de confiance asymptotique de niveau $1-\alpha$ pour θ.

7) a) Examinons l'amplitude des intervalles de confiance asymptotiques.

- Pour celui utilisant M_n, l'amplitude est

$$a(M_n)=\frac{M_n}{1+\frac{\ln(\alpha)}{2n}}-M_n = M_n\frac{-\frac{\ln(\alpha)}{2n}}{1+\frac{\ln(\alpha)}{2n}} \underset{n\to+\infty}{\sim} \frac{-M_n\ln(\alpha)}{2n}.$$

  La suite $(M_n)_{n\geqslant 1}$ converge en probabilité vers θ. On peut donc en déduire que, approximativement l'amplitude de ce second intervalle de confiance asymptotique est de $\dfrac{-\theta\ln(\alpha)}{2n}$.

- Pour celui utilisant Z_n, l'amplitude est

$$\begin{aligned} a(Z_n) &= \frac{Z_n}{1-\frac{u_{1-\alpha/2}}{2\sqrt{2n}}}-\frac{Z_n}{1+\frac{u_{1-\alpha/2}}{2\sqrt{2n}}} = Z_n\left(\frac{1}{1-\frac{u_{1-\alpha/2}}{2\sqrt{2n}}}-\frac{1}{1+\frac{u_{1-\alpha/2}}{2\sqrt{2n}}}\right) \\ &= Z_n\frac{\frac{u_{1-\alpha/2}}{\sqrt{2n}}}{\left(1-\frac{u_{1-\alpha/2}}{2\sqrt{2n}}\right)\left(1+\frac{u_{1-\alpha/2}}{2\sqrt{2n}}\right)} \\ &\underset{n\to+\infty}{\sim} \frac{Z_n u_{1-\alpha/2}}{\sqrt{2n}} \end{aligned}$$

  La suite $(Z_n)_{n\geqslant 1}$ converge en probabilité vers θ. On peut donc en déduire que, approximativement l'amplitude de ce second intervalle de confiance asymptotique est de $\dfrac{\theta u_{1-\alpha/2}}{\sqrt{2n}}$.

Ainsi $a(M_n) \underset{+\infty}{=} \mathrm{o}\left(a(Z_n)\right)$. C'est le premier (celui avec M_n) intervalle de confiance asymptotique qui est meilleur. Cela confirme l'observation faite sur les courbes de la question 2d où les courbes de la suite $(M_n)_{n\geqslant 1}$ semblent être bien plus proches de θ que celles de $(Z_n)_{n\geqslant 1}$

b)

```
n=100
N=10000
theta=2
alpha=0.05
X=theta*np.sqrt(rd.random([n,N]))#Une matrice nxN de réalisations de X
M=np.max(X,0)#N réalisations de M_n
Z=3/2*np.mean(X,0)#N réalisations de Z_n

#Définition des bornes des intervalles de confiance :
aM=M; bM=M/(1+np.log(alpha)/(2*n))
import scipy.special as sp
u=sp.ndtri(1-alpha/2)#on a Phi(u)=1-alpha/2
aZ=Z/(1+u/(2*np.sqrt(2*n))); bZ=Z/(1-u/(2*np.sqrt(2*n)))

#"Incertitude" : Largeur moyenne des deux intervalles de confiances :
LM=np.mean(bM-aM)
LZ=np.mean(bZ-aZ)

#"Niveau réel" : Proportion d'intervalles de confiances contenant theta :
pM=(np.sum(theta<=bM)-np.sum(theta<aM))/N
pZ=(np.sum(theta<=bZ)-np.sum(theta<aZ))/N
```

```
print("Estimation ponctuelle de theta par l'estimateur M_n : "+str(M[0]))
print("Intervalles de confiance obtenus avec l'estimateur M_n :")
print('Largeur moyenne : '+str(LM)+ ' et niveau réel : '+str(pM*100))
print("Estimation ponctuelle de theta par l'estimateur Z_n : "+str(Z[0]))
print("Intervalles de confiance obtenus avec l'estimateur Z_n :")
print('Largeur moyenne : '+str(LZ)+' et niveau réel : '+str(pZ*100))
```

La commande `np.sum(theta<=bM)` compte le nombre de coordonnées qui sont `True` dans le vecteurs de booléens `theta<=bM`. Il s'agit du nombre de coordonnées de `bM` qui sont supérieures ou égales à θ. Autrement dit c'est le nombre de fois où θ est inférieur à la borne supérieure de l'intervalle de confiance asymptotique associé à M_n lors des N réalisations. De même `np.sum(theta<aM)` compte le nombre de fois où θ est inférieur strictement à la borne inférieure de l'intervalle de confiance asymptotique associé à M_n lors des N réalisations. Par conséquent `np.sum(theta<=bM)-np.sum(theta<aM)` est le nombre de fois où θ est dans l'intervalle de confiance asymptotique associé à M_n lors des N réalisations. En divisant par N, on obtient donc la proportion des N intervalles de confiances asymptotiques qui contiennent θ. Même chose pour l'autre intervalle de confiance asymptotique.

c) Voici ce que l'on obtient :

```
Estimation ponctuelle de theta par l'estimateur M_n : 1.993859224172605
Intervalles de confiance obtenus avec l'estimateur M_n :
Largeur moyenne : 0.030260507104987913 et niveau réel : 95.09
Estimation ponctuelle de theta par l'estimateur Z_n : 1.9770532367979352
Intervalles de confiance obtenus avec l'estimateur Z_n :
Largeur moyenne : 0.2785593651339157 et niveau réel : 95.30999999999999
```

On constate que les deux estimateurs ont founi un bonne approximation de $\theta = 2$ et que les deux intervalles de confiance asymptotiques ont un niveau réel proche de 95%. En revanche la largeur moyenne du premier est bien plus petite que celle du deuxième. Il est donc plus resserré et cela conforte le choix du premier pour le meilleur.

8) a) Les variables aléatoires $\mathbb{1}_{X_i<\theta/\sqrt{2}}$, $i \in [\![1\,;n]\!]$, sont indépendantes (par théorème des coalitions) et suivent une loi de Bernoulli de paramètre

$$\mathbb{P}\left(X < \frac{\theta}{\sqrt{2}}\right) = F_\theta\left(\frac{\theta}{\sqrt{2}}\right) = \frac{(\theta/\sqrt{2})^2}{\theta^2} = \frac{1}{2}.$$

Par conséquent leur somme D_n suit une loi binomiale de paramètres n et $\dfrac{1}{2}$.

b) Soit $i \in [\![1\,;n]\!]$. L'événement $\left[Y_i \geqslant \dfrac{\theta}{\sqrt{2}}\right]$ est réalisé si et seulement si les variables aléatoires $Y_i, Y_{i+1}, \ldots, Y_n$ prennent des valeurs supérieures ou égales à $\dfrac{\theta}{\sqrt{2}}$ (puisque, pour tout $\omega \in \Omega$, $Y_i(\omega) \leqslant Y_{i+1}(\omega) \leqslant \cdots \leqslant Y_n(\omega)$) si et seulement si il y a strictement moins de i variables aléatoires parmi $Y_1, \ldots, Y_n$ qui prennent des valeurs strictement inférieures à $\dfrac{\theta}{\sqrt{2}}$ si et seulement si il y a strictement moins de i variables aléatoires parmi $X_1, \ldots, X_n$ qui prennent des valeurs strictement inférieures à $\dfrac{\theta}{\sqrt{2}}$ si et seulement si le nombre de variables aléatoires parmi les $X_1, \ldots, X_n$ qui sont strictement inférieures à $\dfrac{\theta}{\sqrt{2}}$ est strictement inférieur à i si et seulement si $[D_n < i]$ est réalisé. On a donc $\left[Y_i \geqslant \dfrac{\theta}{\sqrt{2}}\right] = [D_n < i]$.

c) Soit $\ell \in [\![0\,;n]\!]$. On a

$$\mathbb{P}(D_n \geqslant n-\ell) = \sum_{i=n-\ell}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^i \left(1-\frac{1}{2}\right)^{n-i} = \frac{1}{2^n} \sum_{i=n-\ell}^{n} \binom{n}{i}.$$

Faisons le changement d'indice $j = n - i$:

$$\mathbb{P}(D_n \geqslant n - \ell) = \frac{1}{2^n} \sum_{j=0}^{\ell} \binom{n}{n-j} = \frac{1}{2^n} \sum_{j=0}^{\ell} \binom{n}{j}.$$

De plus

$$\mathbb{P}(D_n \leqslant \ell) = \sum_{i=0}^{\ell} \binom{n}{i} \left(\frac{1}{2}\right)^i \left(1 - \frac{1}{2}\right)^{n-i} = \frac{1}{2^n} \sum_{i=0}^{\ell} \binom{n}{i}$$

donc $\mathbb{P}(D_n \geqslant n - \ell) = \mathbb{P}(D_n \leqslant \ell)$ (les indices i et j sont muets dans les deux sommes).

En prenant $\ell = k$, on a $n - \ell = 2k + 1 - k = k + 1$ donc $\mathbb{P}(D_n \geqslant k + 1) = \mathbb{P}(D_n \leqslant k)$. Mais on a $[D_n \geqslant k + 1] \cup [D_n \leqslant k] = \Omega$ puisque D_n prend des valeurs entières donc

$$1 = \mathbb{P}(D_n \geqslant k + 1) + \mathbb{P}(D_n \leqslant k) = \mathbb{P}(D_n \geqslant k + 1) + \mathbb{P}(D_n \geqslant k + 1) = 2\mathbb{P}(D_n \geqslant k + 1)$$

donc $\mathbb{P}(D_n \geqslant k + 1) = \dfrac{1}{2}$.

d) On a $\mathbb{P}(D_n \geqslant n + 1) = 0 \leqslant \dfrac{\alpha}{2}$ donc $n + 1 \in E_{n,\alpha}$. Ainsi l'ensemble $E_{n,\alpha}$ est une partie non vide de $\mathbb{N}$. Il admet donc un minimum.

e) Si $i \in [0\,; k]$, alors

$$\mathbb{P}(D_n \geqslant i) = \mathbb{P}(D_n \geqslant k + 1) + \mathbb{P}(i \leqslant D_n \leqslant k) = \frac{1}{2} + \mathbb{P}(i \leqslant D_n \leqslant k) > \frac{1}{2}.$$

Comme $\alpha < 1$, on a $\dfrac{\alpha}{2} < \dfrac{1}{2}$ et donc $P(D_n \geqslant i) > \dfrac{\alpha}{2}$. Ainsi $i \notin E_{n,\alpha}$. On vient de montrer que $[0\,; k] \subset \overline{E_{n,\alpha}}$. Par contraposée, $E_{n,\alpha} \subset \mathbb{N} \setminus [0\,; k]$. Le minimum j_α de $E_{n,\alpha}$ est donc supérieur ou égal à $k + 1$.

On a donc $n - j_\alpha + 1 \leqslant n - (k + 1) + 1 = n - k = k + 1$ (on rappelle que $n = 2k + 1$) donc $n - j_\alpha + 1 \leqslant j_\alpha$. Ainsi, presque sûrement, $Y_{n-j_\alpha+1} \leqslant Y_{j_\alpha}$.

f) On a

$$\mathbb{P}\left(Y_{n-j_\alpha+1} < \frac{\theta}{\sqrt{2}} \leqslant Y_{j_\alpha}\right) = \mathbb{P}\left(Y_{j_\alpha} \geqslant \frac{\theta}{\sqrt{2}}\right) - \mathbb{P}\left(Y_{n-j_\alpha+1} \geqslant \frac{\theta}{\sqrt{2}}\right).$$

D'après la question 8b,

$$\mathbb{P}\left(Y_{n-j_\alpha+1} < \frac{\theta}{\sqrt{2}} \leqslant Y_{j_\alpha}\right) = \mathbb{P}(D_n < j_\alpha) - \mathbb{P}(D_n < n - j_\alpha + 1).$$

Comme j_α est le minimum de $E_{n,\alpha}$, il lui appartient donc $\mathbb{P}(D_n \geqslant j_\alpha) \leqslant \dfrac{\alpha}{2}$ donc $1 - \mathbb{P}(D_n < j_\alpha) \leqslant \dfrac{\alpha}{2}$ et donc $\mathbb{P}(D_n < j_\alpha) \geqslant 1 - \dfrac{\alpha}{2}$.

Ensuite la question 8c appliquée à $\ell = n - j_\alpha$ entraîne que

$$\mathbb{P}(D_n < n - j_\alpha + 1) = \mathbb{P}(D_n \leqslant n - j_\alpha) = \mathbb{P}(D_n \geqslant j_\alpha) \leqslant \frac{\alpha}{2}.$$

On en déduit que

$$\mathbb{P}\left(Y_{n-j_\alpha+1} < \frac{\theta}{\sqrt{2}} \leqslant Y_{j_\alpha}\right) = \mathbb{P}(D_n < j_\alpha) - \mathbb{P}(D_n < n - j_\alpha + 1) \geqslant 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.$$

g) On en déduit que

$$\mathbb{P}\left(\sqrt{2}Y_{n-j_\alpha+1} < \theta \leqslant \sqrt{2}Y_{j_\alpha}\right) \geqslant= 1 - \alpha.$$

On en déduit que $\left[\sqrt{2}Y_{n-j_\alpha+1}; \sqrt{2}Y_{j_\alpha}\right]$ est un intervalle de confiance pour θ au niveau de confiance $1 - \alpha$.

On a $n - j_\alpha + 1 < k + 1 \leqslant j\alpha$ d'après la question 8e donc, presque sûrement, $Y_{n-j_\alpha+1} \leqslant Y_{k+1} \leqslant Y_{j_\alpha}$. On multiplie par $\sqrt{2}$ et on obtient $(\sqrt{2}Y_{n-j_\alpha+1} \leqslant T_n \leqslant (\sqrt{2}Y_{j_\alpha}$. Autrement dit, presque sûrement, l'estimateur T_n prend ses valeurs dans l'intervalle de confiance $\left[\sqrt{2}Y_{n-j_\alpha+1}; \sqrt{2}Y_{j_\alpha}\right]$.

h) En exécutant cette fonction, i prend initialement la valeur $n+1$, $s = p = \dfrac{1}{2^n} = \mathbb{P}(D_n = n)$. Ensuite i, p, s prennent successivement les valeurs suivantes :

- $i = n$, $p = \dfrac{n}{2^n} = \dbinom{n}{1}\dfrac{1}{2^n} = \mathbb{P}(D_n = n-1)$ et $s = \mathbb{P}(D_n \geqslant n-1)$,
- $i = n-1$, $p = \dfrac{n(n-1)}{2 \times 2^n} = \dbinom{n}{2}\dfrac{1}{2^n} = \mathbb{P}(D_n = n-2)$ et $s = \mathbb{P}(D_n \geqslant n-2)$,
- $i = n-2$, $p = \dfrac{n(n-1)(n-2)}{2 \times 3 \times 2^n} = \dbinom{n}{3}\dfrac{1}{2^n} = \mathbb{P}(D_n = n-3)$ et $s = \mathbb{P}(D_n \geqslant n-3)$,

etc. jusqu'à ce que $s = \mathbb{P}(D_n \geqslant i-1) > \dfrac{\alpha}{2}$. Dans ce cas i prend bien la valeur minimale pour laquelle $\mathbb{P}(D_n \geqslant i) \leqslant \dfrac{\alpha}{2}$.

i)
```
k=50
n=2*k+1
N=10000
theta=2
alpha=0.05
X=theta*np.sqrt(rd.random([n,N]))#Une matrice de taille nxN de ré
    alisations de X
Y=np.sort(X,0)#Tri colonne par colonne
j=minE(n,alpha)#Calcul de j_alpha

#Définition des bornes des intervalles de confiance :
aT=Y[n-j,:]*np.sqrt(2); bT=Y[j-1,:]*np.sqrt(2)

#"Incertitude" : Largeur moyenne des deux intervalles de confiances :
LT=np.mean(bT-aT)

#"Niveau réel" : Proportion d'intervalles de confiances contenant theta :
pT=(np.sum(theta<=bT)-np.sum(theta<aT))/N

print("Estimation ponctuelle de theta par l'estimateur T_n : "
        +str(Y[k,0]*np.sqrt(2)))
print("Intervalles de confiance obtenus avec l'estimateur T_n :")
print('Largeur moyenne : '+str(LT)+' et niveau réel : '+str(pT*100))
```

On obtient :

```
Estimation ponctuelle de theta par l'estimateur T_n : 2.006887919422553
Intervalles de confiance obtenus avec l'estimateur T_n :
Largeur moyenne : 0.3940772344382862 et niveau réel : 95.26
```

Pour une valeur de n comparable, cet intervalle de confiance fait moins bien que celui avec Z_n (puisque sa largeur moyenne est supérieure) avec un niveau réel proche d niveau théorique.

Si on recommencer l'exécution avec $n = 5$ au lieu de $n = 50$, on obtient :

```
Estimation ponctuelle de theta par l'estimateur T_n : 1.9454015983877693
Intervalles de confiance obtenus avec l'estimateur T_n :
1.474829179978698 et niveau réel : 98.89
```

On constate que le niveau réel se trouve augmenté. L'estimation de $2 = \theta$ semble bonne mais, en baissant la valeur de n, on agrandit l'amplitude de l'intervalle de confiance en moyenne. On a donc une estimation qui nécessite moins d'observations mais au prix de perdre en précision. Moralité : c'est quand même mieux lorsque l'on a un grand échantillon de données !