

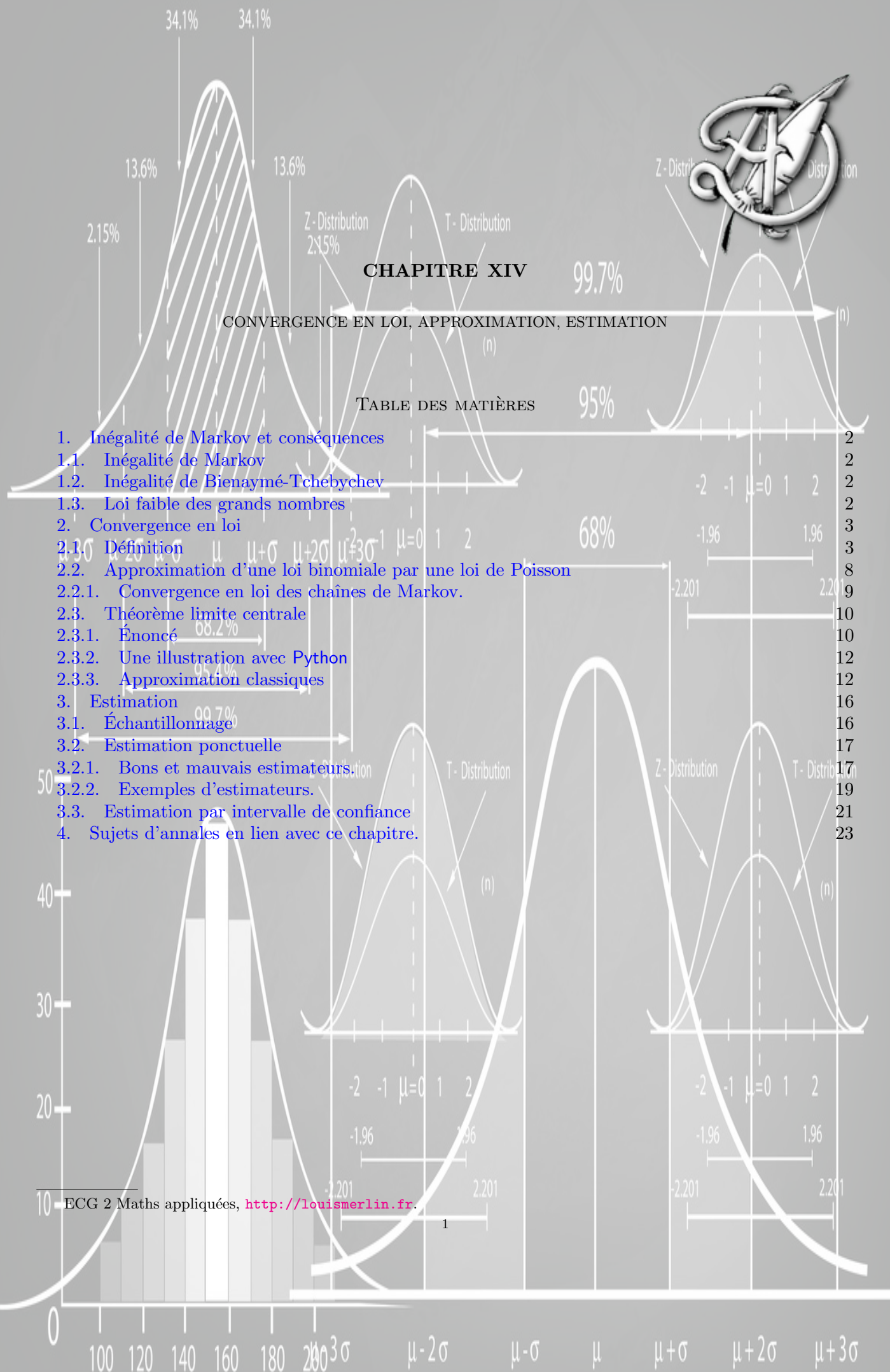


CHAPITRE XIV

CONVERGENCE EN LOI, APPROXIMATION, ESTIMATION

TABLE DES MATIÈRES

- 1. Inégalité de Markov et conséquences
 - 1.1. Inégalité de Markov
 - 1.2. Inégalité de Bienaymé-Tchebychev
 - 1.3. Loi faible des grands nombres
- 2. Convergence en loi
 - 2.1. Définition
 - 2.2. Approximation d'une loi binomiale par une loi de Poisson
 - 2.2.1. Convergence en loi des chaînes de Markov.
 - 2.3. Théorème limite centrale
 - 2.3.1. Énoncé
 - 2.3.2. Une illustration avec Python
 - 2.3.3. Approximation classiques
- 3. Estimation
 - 3.1. Échantillonnage
 - 3.2. Estimation ponctuelle
 - 3.2.1. Bons et mauvais estimateurs.
 - 3.2.2. Exemples d'estimateurs.
 - 3.3. Estimation par intervalle de confiance
- 4. Sujets d'annales en lien avec ce chapitre.



1. INÉGALITÉ DE MARKOV ET CONSÉQUENCES

1.1. **Inégalité de Markov.** L'inégalité de Markov affirme qu'une variable aléatoire positive prend avec petite probabilité des valeurs très grandes. Plus précisément :

Théorème : Inégalité de Markov

Soit X une variable aléatoire réelle **positive** et admettant une espérance. Alors pour tout $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

La preuve de ce résultat est à connaître.

Démonstration. À compléter. □

Exemple 1.1.1. Si X est une variable aléatoire de loi $\mathcal{E}(\lambda)$, alors, pour tout $a > 0$, $P(X \geq a) \leq \frac{1}{a\lambda}$.

Si $a\lambda < 1$, cette majoration ne nous apprend pas grand chose...

En effet, l'inégalité de Markov est souvent utilisée avec de grandes valeurs de a .

1.2. **Inégalité de Bienaymé-Tchebychev.** L'inégalité de Bienaymé-Tchebychev affirme qu'une variable aléatoire qui a un moment d'ordre 2 prend avec petite probabilité des valeurs loin de son espérance. Comme pour l'inégalité de Markov, cette inégalité n'est pas très fine, surtout pour le contrôle des "petits" écarts à la moyenne. Plus précisément,

Théorème : Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire réelle admettant un moment d'ordre 2. Alors pour tout $\varepsilon > 0$,

$$P\left(|X - E(X)| \geq \varepsilon\right) \leq \frac{V(X)}{\varepsilon^2}.$$

La preuve de l'inégalité de Bienaymé-Tchebychev est aussi à connaître.

Démonstration. À compléter. □

Exemple 1.2.1. Si X est une variable aléatoire de loi $\mathcal{E}(\lambda)$, alors, pour tout $\varepsilon > 0$,

$$P\left(\left|X - \frac{1}{\lambda}\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} = \frac{1}{\lambda^2 \varepsilon^2}$$

En particulier, pour $\varepsilon = \frac{2}{\lambda} > 0$,

$$P\left(X \geq \frac{3}{\lambda}\right) \leq P\left(\left|X - \frac{1}{\lambda}\right| \geq \frac{2}{\lambda}\right) \leq \frac{1}{4}$$

car $\left[X \geq \frac{3}{\lambda}\right] \subset \left[\left|X - \frac{1}{\lambda}\right| \geq \frac{2}{\lambda}\right]$.

1.3. **Loi faible des grands nombres.**

Théorème : Loi faible des grands nombres

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles définies sur le même espace probabilisé (Ω, \mathcal{A}, P) , **indépendantes**, admettant chacune la même espérance m et la même variance. Si l'on pose $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, on a, pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} P\left(\left|\bar{X}_n - m\right| \geq \varepsilon\right) = 0.$$

Démonstration. À compléter. □

Remarque 1.3.1. • $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est appelé la **moyenne empirique** des variables aléatoires X_1, X_2, \dots, X_n .

- En particulier, la loi faible des grands nombres est valable pour une suite de variables aléatoires mutuellement indépendantes, de même loi et admettant une espérance et une variance.
- Ce théorème n'est pas quantitatif : on n'a pas d'idée de la "vitesse" à laquelle \bar{X}_n se rapproche de m (mais on aura une idée de la vitesse de convergence dans la suite du cours).

2. CONVERGENCE EN LOI

2.1. Définition.

Définition : Convergence en loi d'une suite de variables aléatoires réelles

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles et X une variable aléatoire réelle, toutes définies sur le même espace probabilisé (Ω, \mathcal{A}, P) . On dit que la suite $(X_n)_{n \in \mathbb{N}^*}$ **converge en loi** vers la variable aléatoire X si, et seulement si, en tout point x où F_X est continue, on a :

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$$

avec F_X (resp. F_{X_n} , $n \in \mathbb{N}^*$) la fonction de répartition de X (resp. X_n , $n \in \mathbb{N}^*$). On note :

$$X_n \xrightarrow{\mathcal{L}} X.$$

Remarque 2.1.1. Dans la convergence en loi, on ne se préoccupe pas directement des variables aléatoires X_n mais uniquement de leur loi.

Il existe d'autres types de convergence qui ne sont pas discutés dans le programme. En particulier, la loi faible des grands nombres est aussi un résultat de convergence de variables aléatoires (mais il ne s'agit pas d'une convergence en loi).

Remarque 2.1.2. La condition qui porte sur les fonctions de répartitions des variables aléatoires :

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$$

porte un nom, on dit que la suite de fonctions $(F_{X_n})_{n \in \mathbb{N}}$ converge *simplement* vers la fonction F_X . Il s'agit d'une façon (parmi d'autres) de définir la convergence d'une suite de fonctions : on fixe un réel x et on s'intéresse à la suite *de réels* $(F_{X_n}(x))_{n \in \mathbb{N}}$. On dit que la suite de fonctions converge simplement si cette suite de réels converge pour tout x .

Il s'agit d'une façon assez peu contraignante (et assez naïve) de faire converger des suites de fonctions : une suite de fonctions est pensée comme une collection de suites de réels $(F_{X_n}(x))_{n \in \mathbb{N}}$. Dans cette version de la convergence de fonctions, on ne tient pas compte des propriétés des fonctions F_{X_n} (leur continuité, dérivabilité, monotonie, convexité,...). Il existe des modes de convergence qui s'intéressent aussi à la préservation de ces propriétés.

Ces considérations sont hors programme, mais on pourra retenir que la convergence en loi est un mode de convergence très souple (et donc très fréquent), ce qui nous permettra dans la suite de ce cours de construire de nombreux exemples de suites de variables aléatoires qui convergent en loi. On retiendra aussi que la continuité des fonctions de répartition n'est pas préservée lors d'un passage à la limite en loi. On pourra alors observer des phénomènes de suites de variables discrètes qui convergent en loi vers des variables à densité et inversement.

Exemple 2.1.3. Pour tout $n \in \mathbb{N}^*$, considérons la variable aléatoire X_n de loi uniforme sur $\left[\frac{-1}{n}, \frac{1}{n}\right]$. La fonction de répartition de X_n est donnée par :

$$F_{X_n}(x) = \begin{cases} 0 & \text{si } x < \frac{-1}{n} \\ \frac{nx+1}{2} & \text{si } \frac{-1}{n} \leq x \leq \frac{1}{n} \\ 1 & \text{si } x > \frac{1}{n}. \end{cases}$$

Si x est un réel strictement négatif, alors pour n assez grand, $x < \frac{-1}{n}$ donc $\lim_{n \rightarrow +\infty} F_{X_n}(x) = 0$. De même, pour $x > 0$, alors pour n assez grand, $x > \frac{1}{n}$ donc $\lim_{n \rightarrow +\infty} F_{X_n}(x) = 1$.

Il en résulte que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers une variable X de loi certaine égale à 0. En effet, la fonction de répartition de X est définie par : $F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0. \end{cases}$

La fonction F_X est discontinue en 0, donc on n'a pas à calculer la limite de $F_{X_n}(x)$ quand $x = 0$.

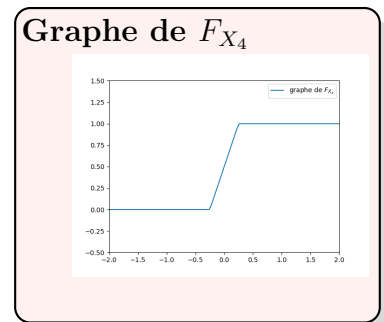
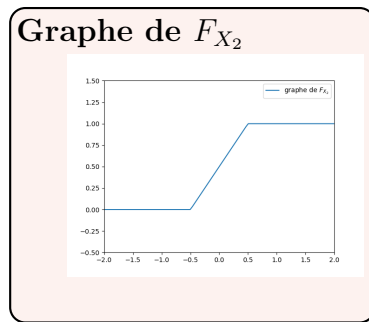
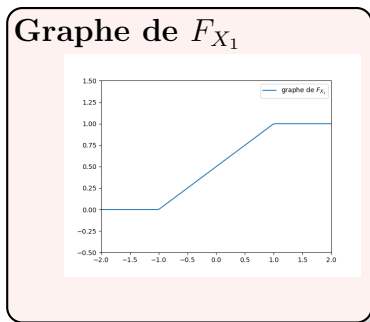
On rappelle une fois pour toutes dans ce chapitre le code Python pour dessiner les graphes de fonctions.

```

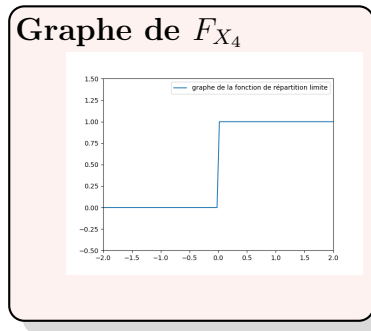
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 def F(x):
5     if (x < -0.25):
6         return 0
7     elif (x > 0.25):
8         return 1
9     else:
10        return (4*x+1)/2
11
12 x = np.linspace(-2,2,100)
13
14 y=[F(t) for t in x]
15
16 plt.plot(x,y, label="graphe de $F_{X_4}$")
17 plt.axis([-2,2,-0.5,1.5])
18 plt.legend()
19 plt.show()

```

Graphes de F_{X_n} :



Graphe de F_X :



Remarque 2.1.4 (Fonction de répartition *empirique*). On peut tracer la fonction de répartition *empirique* d'une variable aléatoire à partir de ses simulations informatiques.

Notons $[a, b]$ le segment sur lequel on souhaite tracer F_X .

On commence par construire le vecteur des abscisses $\mathbf{abs} = [x_1 \ x_2 \ x_3 \ \dots \ x_n]$, où $x_1 = a$ et $x_n = b$.

On doit ensuite construire le vecteur des ordonnées $\mathbf{ordo} = [F_X(x_1) \ F_X(x_2) \ F_X(x_3) \ \dots \ F_X(x_n)]$

L'expression de la fonction F_X étant supposée inconnue, on observe que :

$$\mathbf{ordo} = [P(X \leq x_1) \ P(X \leq x_2) \ P(X \leq x_3) \ \dots \ P(X \leq x_n)]$$

Calculer le vecteur \mathbf{ordo} revient donc à estimer les probabilités $P(X \leq x)$ pour tout $x \in \mathbf{abs}$.

Pour cela on procède comme suit :

- On simule un grand nombre N de fois la variable aléatoire X et on conserve les résultats dans un tableau \mathbf{res} .
- Pour chaque $x \in \mathbf{abs}$, on compte le nombre de valeurs du tableau \mathbf{res} qui sont $\leq x$ et on divise ce nombre par N , ce qui nous donne la proportion de valeurs qui sont $\leq x$ parmi toutes les valeurs du tableau \mathbf{res} .
- La probabilité de l'événement $[X \leq x]$ est approximativement donnée par la proportion de résultats amenant $X \leq x$ lorsque l'on simule un grand nombre de fois X : c'est la proportion que nous avons calculée au point précédent !

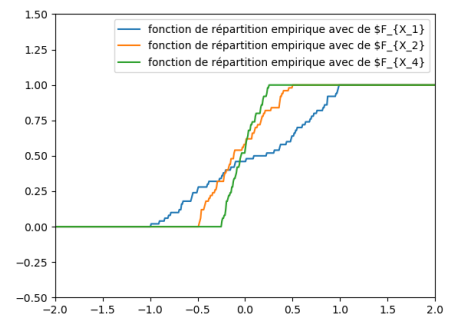
Remarque 2.1.5 (Simulation informatique). On reprend l'exemple précédent.

```

1 import numpy.random as rd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5
6 # Construction de la fonction de repartition
  empirique.
7 N=50 # pour choisir la precision de l approximation
  .
8 X1=[]
9 for k in range (N):
10     simul=rd.uniform(-1,1,1)
11     X1.append(simul)
12
13 X2=[]
14 for k in range (N):
15     simul=rd.uniform(-1/2,1/2,1)
16     X2.append(simul)
17
18 X3=[]
19 for k in range (N):
20     simul=rd.uniform(-1/4,1/4,1)
21     X3.append(simul)
22
23
24 x = np.linspace(-2, 2, 500)
25 y1 = [(sum(t<u for t in X1))/N for u in x]
26 y2 = [(sum(t<u for t in X2))/N for u in x]
27 y3 = [(sum(t<u for t in X3))/N for u in x]
28
29
30 plt.plot(x, y1, label="fonction de repartition
  empirique avec de $F_{X_1}$")
31 plt.plot(x, y2, label="fonction de repartition
  empirique avec de $F_{X_2}$")
32 plt.plot(x, y3, label="fonction de repartition
  empirique avec de $F_{X_4}$")
33 plt.axis([-2, 2, -0.5, 1.5])
34 plt.legend()
35 plt.show()

```

Approximation empirique des fonctions de répartition de lois uniformes.



Théorème : Convergence en loi d'une suite de variables aléatoires réelles discrètes

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles **discrètes** et X une variable aléatoire réelle **discrète**, toutes définies sur le même espace probabilisé (Ω, \mathcal{A}, P) , et prenant leurs valeurs dans \mathbb{Z} . Alors la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers la variable aléatoire X si, et seulement si pour tout $k \in \mathbb{Z}$,

$$\lim_{n \rightarrow +\infty} P(X_n = k) = P(X = k).$$

Bien sûr, cet énoncé est très spécial au cas des variables discrètes car dans le cas d'une variable à densité, chacun des nombres $P(X_n = k)$ et $P(X = k)$ est nul et la condition du théorème est donc toujours satisfaite pour les variables à densités.

Démonstration. Admis.

□

Exemple 2.1.6. Pour tout $n \in \mathbb{N}^*$, notons X_n une variable aléatoire discrète de loi uniforme sur l'ensemble $\left\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right\}$. Autrement dit pour tout $k \in \llbracket 1, n \rrbracket$

$$P\left(X = \frac{k}{n}\right) = \frac{1}{n}.$$

On a : $X_n(\Omega) \subset]0; 1]$ donc $F_{X_n}(x) = 0$ si $x \leq 0$ et $F_{X_n}(x) = 1$ si $x > 1$.
Soit $x \in]0; 1]$. On a alors :

$$F_{X_n}(x) = P(X_n \leq x) = \sum_{\substack{k \in \llbracket 1, n \rrbracket \\ \frac{k}{n} \leq x}} P\left(X_n = \frac{k}{n}\right).$$

La somme porte sur les entiers k compris entre 1 et n et tels que $\frac{k}{n} \leq x$, c'est-à-dire les entiers k compris entre 1 et la partie entière de nx :

$$F_{X_n}(x) = \sum_{k=1}^{\lfloor nx \rfloor} \frac{1}{n} = \frac{\lfloor nx \rfloor}{n}.$$

Mais, par définition de la partie entière, $|\lfloor nx \rfloor - nx| \leq 1$ d'où, en divisant par $n > 0$: $\left|\frac{\lfloor nx \rfloor}{n} - x\right| \leq \frac{1}{n}$.

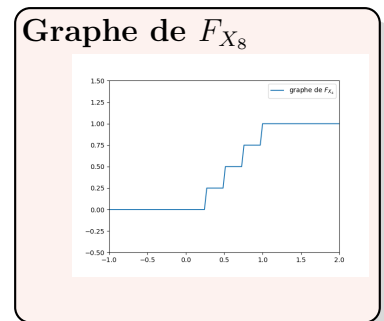
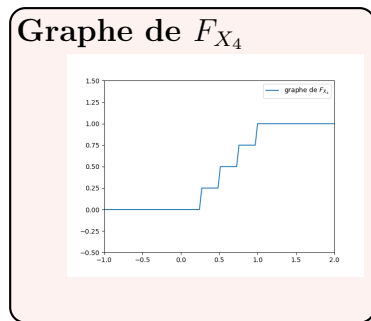
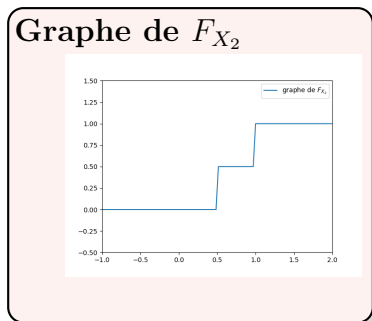
Donc, d'après le théorème des gendarmes, $\lim_{n \rightarrow +\infty} \frac{\lfloor nx \rfloor}{n} = x$.

On obtient donc :

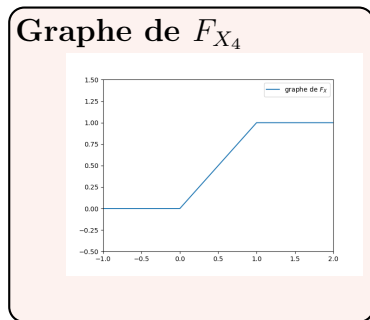
$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } 0 < x \leq 1 \\ 1 & \text{si } x > 1. \end{cases}$$

On reconnaît la fonction de répartition d'une variable aléatoire X de loi uniforme sur l'intervalle $]0; 1]$.
Donc la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers X de loi $\mathcal{U}(]0; 1])$.

Graphes de F_{X_n} :



Graphe de F_X :



Remarque 2.1.7. On remarque sur ces exemples que l'on peut avoir convergence en loi d'une suite de variables aléatoire à densité vers une variable aléatoire discrète, ou bien le contraire.

Exercice 2.1.8. On considère une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires réelles indépendantes suivant une loi uniforme sur $[0; 1]$. Pour $n \geq 1$, on définit $M_n = \max(X_1, X_2, \dots, X_n)$ et $Y_n = n(1 - M_n)$.

1. Déterminer la fonction de répartition de M_n puis celle de Y_n .
2. Montrer que si $x < 0$, $\lim_{n \rightarrow +\infty} F_{Y_n}(x) = 0$.
3. Déterminer, pour $x \geq 0$, $\lim_{n \rightarrow +\infty} F_{Y_n}(x)$.
4. En déduire que Y_n converge en loi vers une variable aléatoire X dont on précisera la loi.

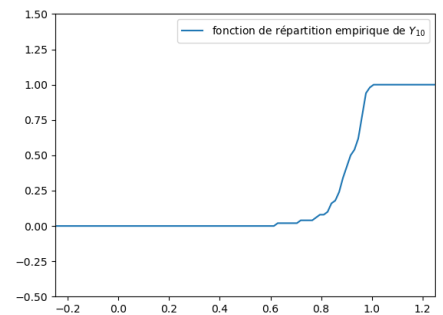
Remarque 2.1.9 (Simulation informatique).

```

1 import numpy.random as rd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5
6 # Construction de la fonction de répartition
  empirique.
7 N=50 # pour choisir la précision de l'approximation
8
9 X=[]
10 for k in range (N):
11     simul=rd.uniform(0,1,10)
12     res = max(simul)
13     X.append(res)
14
15 x = np.linspace(-0.25, 1.25, 100)
16 y = [(sum(t<u for t in X))/N for u in x]
17
18
19 plt.plot(x, y, label="fonction de répartition
  empirique de $Y_{10}$")
20 plt.axis([-0.25, 1.25, -0.5, 1.5])
21 plt.legend()
22 plt.show()

```

Approximation empirique de la fonction de répartition du maximum de 10 lois uniformes.



2.2. Approximation d'une loi binomiale par une loi de Poisson.

Théorème : Convergence en loi d'une suite de variables aléatoires de loi binomiale vers une variable aléatoire de loi de Poisson

Soit λ un réel strictement positif et $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles discrètes telles que, pour tout $n \in \mathbb{N}^*$, X_n suit la loi binomiale de paramètres n et $\frac{\lambda}{n}$: $X_n \leftrightarrow \mathcal{B}\left(n, \frac{\lambda}{n}\right)$.

Alors pour tout $k \in \mathbb{N}$,

$$\lim_{n \rightarrow +\infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Autrement dit, la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire X qui suit la loi de Poisson de paramètre λ .

Beaucoup de questions de concours s'inspirent d'un raisonnement comme dans celui de la preuve.

Démonstration. À compléter. □

Remarque 2.2.1. Dans une urne contenant des boules blanches ou noires, la proportion de boules blanches étant p , on tire n boules avec remise. Si p est proche de 0 et n assez grand, la variable aléatoire égale au nombre de boules blanches tirées suivra approximativement une loi de Poisson de

paramètre np . On dit que la loi de Poisson est la loi des "événements rares". En pratique, on considère que l'approximation est valable si $n \geq 30$, $p \leq 0,1$, et $np \leq 15$ mais les conditions d'application seront toujours précisées en concours. On remplace alors une loi dépendant de deux paramètres par une loi qui ne dépend que d'un réel, et dont on connaît les probabilités données par des tables de loi.

Exemple 2.2.2. Soit X une variable aléatoire de loi binomiale $\mathcal{B}(100; 0,05)$ et on s'intéresse à la probabilité que X prenne la valeur 2.

- **Calcul exact :**

$$P(X = 2) = \binom{100}{2} (0,05)^2 (0,95)^{98} = 4950 \times 0,0025 \times (0,95)^{98} \approx 0,0812.$$

- **Calcul approché :** nous sommes dans les conditions de l'approximation de la loi binomiale par la loi de Poisson. On approche donc la loi $\mathcal{B}(100; 0,05)$ par la loi $\mathcal{P}(5)$ ($100 \times 0,05 = 5$). On obtient :

$$P(X = 2) = e^{-5} \frac{5^2}{2!} = e^{-5} \times 12,5 = 0,0842.$$

Pas mal!

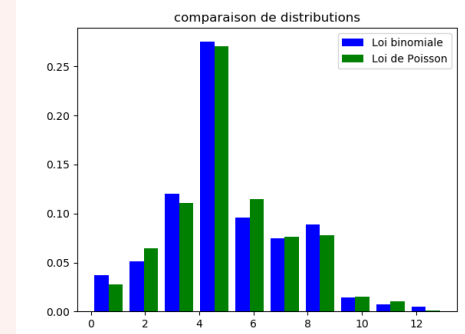
Remarque 2.2.3 (Simulation informatique).

```

1 import numpy.random as rd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5
6 N=1000
7 X = rd.binomial(100,0.05,N)
8 Y = rd.poisson(5,N)
9
10 plt.hist([X, Y], color = ['blue', 'green'], density
11         =True,
12         label = ['Loi binomiale', 'Loi de Poisson'])
13 plt.title('comparaison de distributions')
14 plt.legend()
15 plt.show()

```

Comparaison loi de Poisson / loi binomiale.



2.2.1. *Convergence en loi des chaînes de Markov.* Si $(X_n)_{n \geq 1}$ est une chaîne de Markov de matrice de transition $M = (m_{ij})_{i,j=1,\dots,k}$, on s'intéresse souvent à ses (ou son) état(s) stationnaire(s)

Définition : États stationnaires.

Soit $\mu = (\mu_1, \dots, \mu_k)$ une variables aléatoire à support dans l'ensemble $\{1, 2, \dots, k\}$ des états de la chaîne de Markov. D'après le cours sur les chaînes de Markov, nous savons que les trois conditions suivantes sont équivalentes :

1. Pour tout $n \in \mathbb{N}$, la loi de X_n est donnée par $P(X_n = i) = \mu_i$.

2. Pour tout $j \in \llbracket 1, k \rrbracket$, $\sum_{i=1}^k m_{ij} \mu_i = \mu_j$.

3. Le vecteur colonne $\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}$ est propre pour la matrice ${}^t M$.

Si l'une (donc toutes) de ces propositions se réalise, on dit que μ est un état stationnaire de la chaîne.

Le résultat suivant (déjà vu) affirme qu'il existe toujours un état stationnaire de la chaîne de Markov, et, sous certaines conditions, que cet état stationnaire est unique.

Théorème : Existence et unicité des états stationnaires (hors programme)

1. Il existe toujours au moins un état stationnaire.
2. Si la chaîne de Markov est irréductible, c'est-à-dire si pour tout couple (i, j) d'états de la chaîne, il existe un chemin allant de i à j et un chemin allant de j à i , alors l'état stationnaire est unique.

Ce résultat n'est pas à connaître dans cette version générale. En revanche, on demande souvent de trouver des états stationnaires. L'une des façons possibles consiste à vérifier le critère suivant :

Théorème : Convergence en loi d'une chaîne de Markov vers un état stationnaire.

Si une chaîne de Markov converge en loi, on note

$$\mu_1 = \lim_{n \rightarrow +\infty} P(X_n = 1) \quad \mu_2 = \lim_{n \rightarrow +\infty} P(X_n = 2) \quad , \dots \quad , \quad \mu_k = \lim_{n \rightarrow +\infty} P(X_n = k).$$

Alors $\mu = (\mu_1, \mu_2, \dots, \mu_k)$ est un état stationnaire de la chaîne de Markov.

Remarque 2.2.4. Ce théorème est très semblable au théorème de convergence des trajectoires de systèmes différentiels : si une trajectoire converge, c'est vers un point d'équilibre du système.

► **Pour s'entraîner : exo 1.**

2.3. Théorème limite centrale.

2.3.1. *Énoncé.*

Théorème : Théorème limite centrale (version moyenne empirique)

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur le même espace probabilisé. On suppose les variables X_n **indépendantes** et **de même loi**, ayant chacune une espérance m et une variance σ^2 non nulle. On note :

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \bar{X}_n^* = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}$$

(\bar{X}_n^* est la variable centrée réduite associée à \bar{X}_n .)

Alors la suite $(\bar{X}_n^*)_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire de loi normale centrée réduite.

On a donc, pour tout réel x :

$$\lim_{n \rightarrow +\infty} P(\bar{X}_n^* \leq x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt,$$

et, plus généralement, pour tous a et b tels que $-\infty \leq a < b \leq +\infty$:

$$\lim_{n \rightarrow +\infty} P(a \leq \bar{X}_n^* \leq b) = \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt.$$

► **Pour s'entraîner : exo 2.**

Démonstration. Admis. □

En notant $S_n = X_1 + X_2 + \dots + X_n$, on a $S_n = n\bar{X}_n$. On calcule : $E(S_n) = nm$ par linéarité et $V(S_n) = n\sigma^2$ puisque les variables X_i sont deux à deux indépendantes.

On remarque alors :

$$S_n^* = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{n\bar{X}_n - nm}{\sigma\sqrt{n}} = \frac{n(\bar{X}_n - m)}{\sqrt{n}\sigma} = \sqrt{n} \frac{\bar{X}_n - m}{\sigma} = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} = \bar{X}_n^*.$$

Cela signifie qu'avec les mêmes hypothèses, le théorème central limite peut aussi s'énoncer ainsi :

Corollaire : Théorème limite central (version somme)

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur le même espace probabilisé. On suppose les variables X_n **indépendantes** et **de même loi**, ayant chacune une espérance m et une variance σ^2 non nulle. On note :

$$S_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i \quad \text{et} \quad S_n^* = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - nm}{\sigma\sqrt{n}}$$

(S_n^* est la variable centrée réduite associée à S_n .)

Alors la suite $(S_n^*)_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire de loi normale centrée réduite.

On a donc, pour tout réel x :

$$\lim_{n \rightarrow +\infty} P(S_n^* \leq x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt,$$

et, plus généralement, pour tous a et b tels que $-\infty \leq a < b \leq +\infty$:

$$\lim_{n \rightarrow +\infty} P(a \leq S_n^* \leq b) = \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt.$$

On a de plus :

$$\bar{X}_n = \frac{\sigma}{\sqrt{n}} \bar{X}_n^* + m \quad \text{et} \quad S_n = \sigma\sqrt{n} S_n^* + nm.$$

Donc la variable aléatoire \bar{X}_n suit approximativement la loi $\mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$ et la variable aléatoire S_n suit approximativement la loi $\mathcal{N}(nm, n\sigma^2)$. On ne peut pas parler de convergence en loi ici puisque les lois "limites" dépendent de n .

Ce théorème est remarquable puisqu'avec peu d'hypothèses, il fournit un résultat très fort et met en évidence le rôle central joué par la loi normale en probabilités et statistiques.

► **Pour s'entraîner : exo 4 et 5.**

Méthode : Utilisation du théorème central limite

Considérons une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires **indépendantes** et **de même loi**, admettant une **espérance** m et une **variance** σ^2 .

- Si $S_n = X_1 + X_2 + \dots + X_n$, alors $E(S_n) = nm$ et $V(S_n) = n\sigma^2$, donc on peut approcher la loi de S_n par une loi normale de paramètres nm et $n\sigma^2$.
- Si $\bar{X}_n = \frac{1}{n} S_n = \frac{X_1 + X_2 + \dots + X_n}{n}$, alors $E(S_n) = m$ et $V(S_n) = \frac{\sigma^2}{n}$, donc on peut approcher la loi de \bar{X}_n par une loi normale de paramètres m et $\frac{\sigma^2}{n}$.

Exemple 2.3.1. 1. Soit $(U_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires telle que, pour tout $n \in \mathbb{N}^*$, $U_n \hookrightarrow \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$. Montrer que

$$\sqrt{\frac{12}{n}} \sum_{k=1}^n U_k \xrightarrow{\mathcal{L}} U$$

où $U \hookrightarrow \mathcal{N}(0, 1)$.

2. Une montre fait une erreur d'au plus une demi-minute par jour (en plus ou en moins). Déterminer la probabilité que l'erreur commise au bout d'une année soit inférieure ou égale à un quart d'heure.

On donne $\frac{15\sqrt{12}}{\sqrt{365}} \approx 2,72$.

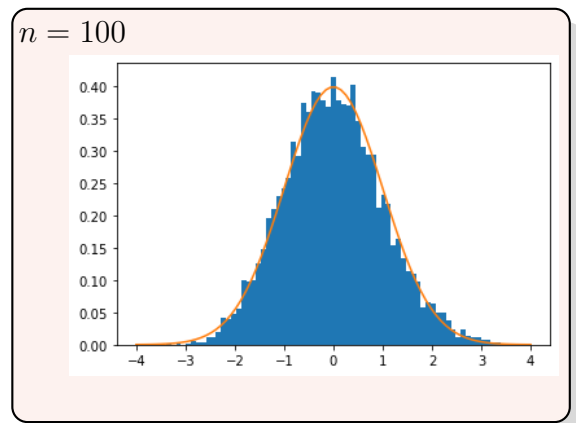
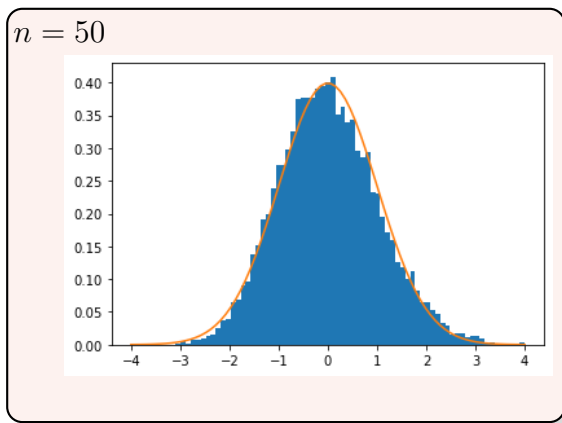
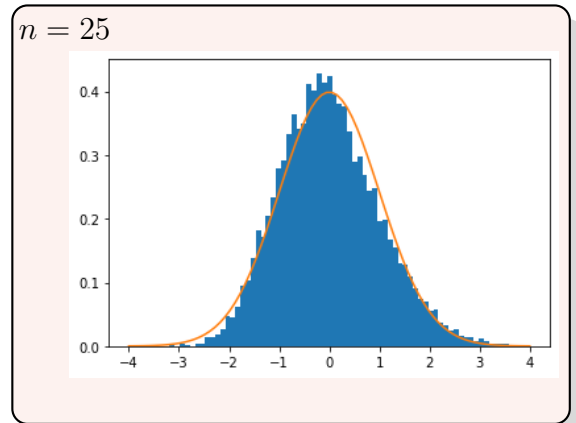
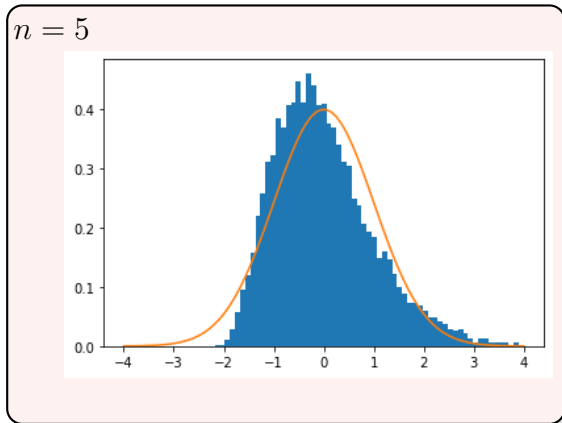
2.3.2. Une illustration avec Python.

On simule S_n^* où $S_n = \sum_{i=1}^n X_i$ avec $X_i \hookrightarrow \mathcal{E}(1)$. On applique à S_n^* le théorème centrale limite et on obtient donc que $S_n^* \xrightarrow{\mathcal{L}} Z$ avec $Z \hookrightarrow \mathcal{N}(0, 1)$.

```

1 import numpy.random as rd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 def S_star(n) :
6     S=[rd.exponential(1) for k in range(n)]
7     return (np.sum(S)-n)/np.sqrt(n)
8
9 n = 5 # puis n=25, n=50 et n=100
10 U=[S_star(n) for k in range(10000)]
11 X=np.linspace(-4,4,80)
12 Y=[np.exp(-(x**2/2))/np.sqrt(2*np.pi) for x in X]
13 plt.hist(U,X, density=True)
14 plt.plot(X,Y)
15 plt.show()

```



2.3.3. Approximation classiques.

Proposition : Approximation d'une loi binomiale par une loi normale

Soit $p \in]0; 1[$, $q = 1 - p$ et $(S_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles définies sur le même espace probabilisé et telle que S_n suit la loi binomiale de paramètres n et p : $S_n \hookrightarrow \mathcal{B}(n, p)$.

Alors la suite de terme général $S_n^* = \frac{S_n - np}{\sqrt{npq}}$ converge en loi vers une variable aléatoire de loi normale centrée réduite.

Démonstration. À compléter. □

Remarque 2.3.2 (Correction de continuité). Dans la pratique, dès que $n \geq 20$ et p proche de 0,5 (mais ce sera indiqué en concours), on considère qu'on peut approcher $\mathcal{B}(n, p)$ par $\mathcal{N}(np, npq)$. On approche ici une loi discrète par une loi continue. On doit alors utiliser la **correction de continuité** pour calculer $P(X_n = k)$ en l'approchant par $P(k - 0,5 < X < k + 0,5)$ où X suit une loi normale. En effet, si on calculait $P(X = k)$, on trouverait 0 ce qui ne donnerait pas une approximation correcte.

Méthode : Approximation d'une loi binomiale $\mathcal{B}(n, p)$ par une loi normale $\mathcal{N}(np, npq)$

Lorsque n est grand, on peut approcher une loi binomiale X de paramètres n et p par une loi normale N de paramètres np et npq (ayant même espérance et même variance que X).

Pour approcher $P(X = k)$, on utilise la correction de continuité :

$$\begin{aligned} P(X = k) &= P(k - 0,5 < X < k + 0,5) \\ &= P(k - 0,5 - np < X - np < k + 0,5 - np) \\ &= P\left(\frac{k - 0,5 - np}{\sqrt{npq}} < \frac{X - np}{\sqrt{npq}} < \frac{k + 0,5 - np}{\sqrt{npq}}\right) \\ &= \Phi\left(\frac{k + 0,5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k - 0,5 - np}{\sqrt{npq}}\right) \end{aligned}$$

avec $X^* = \frac{X - np}{\sqrt{npq}} \hookrightarrow \mathcal{N}(0, 1)$ donc on lit les probabilités sur la table de la loi normale centrée réduite.

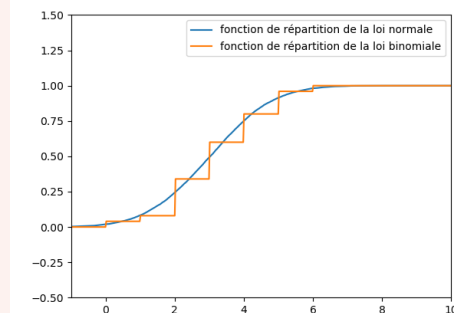
Remarque 2.3.3 (Simulation informatique).

```

1 import numpy.random as rd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5
6 N=10000
7 n=10
8 p=0.3
9 X=[]
10 for k in range (N):
11     simul=rd.normal(n*p,(n*p*(1-p))*(1/2),1)
12     X.append(simul)
13
14 T=50
15 Y = rd.binomial(n, p, T)
16
17 x = np.linspace(-1, 10, 500)
18 y1 = [(sum(t<u for t in X))/N for u in x]
19 y2 = [(sum(t<u for t in Y))/T for u in x]
20
21 plt.plot(x, y1, label="fonction de repartition de
22     la loi normale")
23 plt.plot(x, y2, label="fonction de repartition de
24     la loi binomiale")
25 plt.axis([-1, 10, -0.5, 1.5])
26 plt.legend()
27 plt.show()

```

Comparaison loi binomiale / loi normale.



Exemple 2.3.4. Soit X une variable binomiale de paramètres 900 et 0,5. On cherche à calculer la probabilité : $P(420 \leq X \leq 480)$.

Le calcul avec les formules exactes est long et nécessite l'utilisation d'un ordinateur, pour une valeur de toute façon approchée (un ordinateur ne fait pas de calculs exacts avec des nombres réels). Mais nous sommes dans les conditions d'approximation de la loi binomiale par la loi normale. On approche la loi $\mathcal{B}(900; 0,5)$ par la loi $\mathcal{N}(450; 225)$ et on considère que la variable $\frac{X - 450}{15}$ suit approximativement la loi normale centrée réduite. On obtient :

$$P(420 \leq X \leq 480) = P\left(-2 \leq \frac{X - 450}{15} \leq 2\right) \approx 2\Phi(2) - 1 \approx 2 \times 0,9772 - 1 \approx 0,9544.$$

(On retrouve l'intervalle de confiance à 95%.)

Exercice 2.3.5. 2 000 personnes sont reliées à un même central téléphonique. À une heure de pointe, chacune est censée téléphoner 3 minutes dans l'heure. Combien doit-on construire de lignes pour que la probabilité de saturation soit inférieure ou égale à 1 pour cent ?

(On pourra définir une variable aléatoire X égale au nombre de personnes téléphonant 1 minutes dans cette heure).

Exercice 2.3.6. On dispose de 200 boules identiques qu'on répartit au hasard dans 10 boîtes numérotées de 1 à 10. Quelle est la probabilité que la boîte n°1 contienne 20 boules ?

Proposition : Approximation d'une loi de Poisson par une loi normale

Soit λ un réel strictement positif et $(S_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles définies sur le même espace probabilisé et telle que S_n suit la loi de Poisson de paramètres $n\lambda$: $S_n \hookrightarrow \mathcal{P}(n\lambda)$.

Alors la suite de terme général $S_n^* = \frac{S_n - n\lambda}{\sqrt{n\lambda}}$ converge en loi vers une variable aléatoire de loi normale centrée réduite.

Démonstration. À compléter

□

Remarque 2.3.7. Dans la pratique, dès que $\lambda \geq 15$ (mais ce sera indiqué en concours), on considère qu'on peut approcher la loi $\mathcal{P}(\lambda)$ par $\mathcal{N}(\lambda, \lambda)$.

Méthode : Approximation d'une loi de Poisson $\mathcal{P}(\lambda)$ par une loi normale $\mathcal{N}(\lambda, \lambda)$

Lorsque λ est grand, on peut approcher une loi de Poisson X de paramètre λ par une loi normale N de paramètres λ et λ (ayant même espérance et même variance que X).

Pour approcher $P(X = k)$, on utilise la correction de continuité :

$$\begin{aligned} P(X = k) &= P(k - 0,5 < X < k + 0,5) \\ &= P(k - 0,5 - n\lambda < X - n\lambda < k + 0,5 - n\lambda) \\ &= P\left(\frac{k - 0,5 - n\lambda}{\sqrt{n\lambda}} < \frac{X - n\lambda}{\sqrt{n\lambda}} < \frac{k + 0,5 - n\lambda}{\sqrt{n\lambda}}\right) \\ &= \Phi\left(\frac{k + 0,5 - n\lambda}{\sqrt{n\lambda}}\right) - \Phi\left(\frac{k - 0,5 - n\lambda}{\sqrt{n\lambda}}\right) \end{aligned}$$

avec $X^* = \frac{X - n\lambda}{\sqrt{n\lambda}} \hookrightarrow \mathcal{N}(0, 1)$ donc on lit les probabilités sur la table de la loi normale centrée réduite.

Remarque 2.3.8 (Simulation informatique).

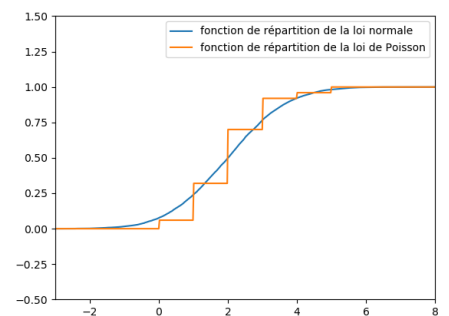
Comparons les fonctions de répartition :

```

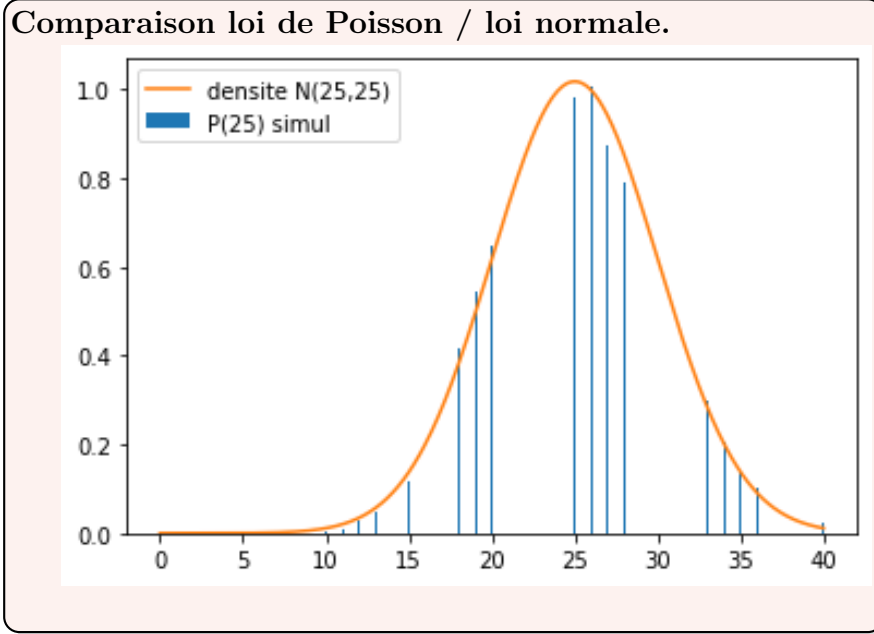
1 import numpy.random as rd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5
6 N = 10000
7 lam = 2
8 X=[]
9 for k in range (N):
10     simul=rd.normal(lam,(lam)**(1/2),1)
11     X.append(simul)
12
13 T=50
14 Y = rd.poisson(lam, T)
15
16 x = np.linspace(-3, 8, 500)
17 y1 = [(sum(t<u for t in X))/N for u in x]
18 y2 = [(sum(t<u for t in Y))/T for u in x]
19
20 plt.plot(x, y1, label="fonction de répartition de
    la loi normale")
21 plt.plot(x, y2, label="fonction de répartition de
    la loi de Poisson")
22 plt.axis([-3, 8, -0.5, 1.5])
23 plt.legend()
24 plt.show()

```

Comparaison loi de Poisson / loi normale.



Comparons maintenant la densité de la loi normale avec l'histogramme des valeurs de la loi de Poisson.



Exercice 2.3.9. Soit X une variable aléatoire de loi de Poisson de paramètre $\lambda = 16$. Évaluer $P(X = 16)$ à l'aide de la table de la loi normale centrée réduite, puis $P(X \leq 16)$ et $P(X \geq 16)$ avec et sans correction de continuité.

► Pour s'entraîner : exo 3.

3. ESTIMATION

Les statisticiens connaissent, en général, le type de loi qui décrit un phénomène, de par l'observation qu'ils en ont fait, mais souvent ils ne connaissent pas tous les paramètres de cette loi. Ils doivent donc les estimer : c'est l'objectif de la *statistique inférentielle*.

On considère un phénomène aléatoire et une variable aléatoire X qui lui est associée, dont le type de loi est connu et dépend d'un paramètre réel noté θ (par exemple, le paramètre λ d'une loi exponentielle, l'étendue $b - a$ d'une loi uniforme sur $[a, b]$, le paramètre p d'une variable de Bernoulli, l'espérance m d'une loi normale, etc.) L'objectif est de donner une *estimation* de la vraie valeur de ce paramètre, ou d'une fonction simple $g(\theta)$ de ce paramètre.

Il y a deux types d'estimation : l'estimation ponctuelle, qui vise à trouver une valeur (approchée) du paramètre θ , et l'estimation par intervalle de confiance, qui cherche à déterminer un intervalle qui contient, avec une probabilité donnée, la valeur du paramètre qu'on cherche à évaluer.

Dans toute la suite, n est un entier naturel supérieur ou égal à 2.

3.1. Échantillonnage.

Définition : Échantillonnage

Soit X une variable aléatoire réelle définie sur un espace probabilisé (Ω, \mathcal{A}, P) . On appelle *n -échantillon* de la loi de X tout n -uplet (X_1, X_2, \dots, X_n) de variables aléatoires réelles définies sur le même espace probabilisé (Ω, \mathcal{A}, P) , **mutuellement indépendantes**, et **de même loi que X** .

Une **réalisation** ou un **échantillon observé** de cet échantillon est un n -uplet (x_1, x_2, \dots, x_n) de réels tels que, pour tout $k \in \llbracket 1, n \rrbracket$, x_k est la valeur prise par X_k .

Remarque 3.1.1. Il ne faut pas confondre l'échantillon (X_1, X_2, \dots, X_n) constitué de variables aléatoires, avec sa réalisation (x_1, x_2, \dots, x_n) constituée de réels. En pratique, pour réaliser un échantillon, on effectue n épreuves identiques et indépendantes pour lesquelles, après la k^e épreuve, la variable aléatoire X_k a pris la valeur x_k .

3.2. Estimation ponctuelle. On suppose que les variables aléatoires constituant l'échantillon

$$(X_1, X_2, \dots, X_n)$$

sont définies sur un espace probabilisable (Ω, \mathcal{A}) , muni d'une probabilité P_θ dépendant de θ (qu'on notera simplement P pour plus de commodité), et on rappelle que ces variables aléatoires sont indépendantes.

Définition : Estimateur

Si (X_1, X_2, \dots, X_n) est un n -échantillon de la loi de X (dépendant d'un paramètre θ), on appelle **estimateur** de θ (ou d'une fonction $g(\theta)$ de θ) toute variable aléatoire $T_n = \varphi(X_1, X_2, \dots, X_n)$ fonction de X_1, X_2, \dots, X_n , et indépendante de θ .

Remarque 3.2.1. La réalisation $\varphi(x_1, x_2, \dots, x_n)$ de l'estimateur T_n est l'estimation de θ (ou $g(\theta)$), c'est-à-dire que $\varphi(x_1, x_2, \dots, x_n)$ **est la valeur que le statisticien accordera à θ (ou $g(\theta)$)**.

Exemple 3.2.2. Si (X_1, X_2, \dots, X_n) est un n -échantillon de la loi de X ,

$$T_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

$$Y_n = \min(X_1, X_2, \dots, X_n)$$

et

$$Z_n = \max(X_1, X_2, \dots, X_n)$$

sont des estimateurs de θ . Comment alors choisir un "bon" estimateur ? Tous ne sont pas forcément pertinents.

3.2.1. Bons et mauvais estimateurs. Imaginons que l'on cherche à estimer la note d'un étudiant au concours (vu comme une variable aléatoire d'espérance inconnue). On tire alors 20 fois à Pile ou Face une pièce équilibrée et on compte le nombre de Pile.

On peut alors fabriquer un estimateur de la note de l'étudiant en déclarant que le nombre de Pile obtenu sera sa note au concours. Bien évidemment cette stratégie est absurde et il n'y a aucune raison que la valeur que prend l'estimateur soit une estimation fiable de la note obtenue. Cette stratégie consiste donc en l'utilisation d'un **mauvais** estimateur.

De manière plus pertinente, on peut aussi, pour se faire une idée de la note de l'étudiant au concours, regarder ses notes de DS pendant l'année. Ces notes de DS sont vues comme la réalisation de la même variable aléatoire, qui modélise la performance de l'étudiant. On fait alors la moyenne des notes pendant l'année et on estime la note finale au concours par cette moyenne de notes. Formellement, on dit qu'on a estimé la moyenne théorique par la moyenne empirique. Plusieurs expériences aléatoires étudiées au cours de l'année peuvent nous faire penser que cet stratégie nous permet de construire un **bon** estimateur.

Pour étudier les qualités d'un estimateur, nous allons utiliser des idées classiques en probabilités : un bon estimateur doit prendre en moyenne une valeur proche de la quantité que l'on cherche à estimer. C'est la moindre des choses.

Mais aussi, puisque dans la pratique nous n'avons accès qu'à une seule réalisation de l'estimateur (*un sondage, une seule année pendant laquelle nous étudions les performances de l'étudiant,...*), nous voulons minimiser la probabilité que l'estimateur se réalise trop loin de sa valeur moyenne, donc trop loin de la valeur à estimer.

Ces critères sont faciles à comprendre puisqu'ils s'expriment facilement avec les concepts habituels des variables aléatoires : un bon estimateur est un estimateur dont l'espérance est le paramètre à estimer et dont la variance est petite. Un estimateur peut d'ailleurs gagner en pertinence lorsque la taille de l'échantillon augmente. Dans la pratique, nous observerons très fréquemment que la variance d'un estimateur diminue lorsque la taille de l'échantillon augmente.

Remarque 3.2.3. Dans la suite de ce paragraphe, nous allons utiliser une terminologie qui est devenue hors programme. Par exemple la différence entre l'espérance de l'estimateur et le paramètre à estimer s'appelle le biais. Bizarrement, même si l'étude de la qualité des estimateurs est restée au programme, les concepts qui servent à décrire ces qualités ont disparu du programme, il faudra donc se contenter de l'espérance et de la variance des estimateurs.

Définition : Biais d'un estimateur (hors programme)

Si T_n est un estimateur de θ admettant une espérance, on appelle **biais** de T_n le nombre réel :

$$b_\theta(T_n) = E(T_n) - \theta.$$

On dit qu'un estimateur est **sans biais** lorsque $b_\theta(T_n) = 0$ i.e. $E(T_n) = \theta$.

Ainsi un estimateur sans biais est exactement un estimateur qui prend en moyenne la valeur que l'on cherche à estimer.

Remarque 3.2.4. Si on cherche à estimer $g(\theta)$, on peut aussi écrire $b_\theta(T_n) = E(T_n) - g(\theta)$ et l'estimateur est sans biais si, et seulement si $E(T_n) = g(\theta)$. Ou alors, si on cherche à estimer θ mais que l'estimateur a un biais exprimé par $E(T_n) = g(\theta)$, alors on peut estimer $g(\theta)$ avec T_n et essayer d'en déduire θ .

Définition : Risque quadratique d'un estimateur (hors programme)

Si T_n est un estimateur de θ admettant une variance, on appelle **risque quadratique** de T_n le nombre réel :

$$r_\theta(T_n) = E((T_n - \theta)^2).$$

Théorème : Lien entre biais et risque quadratique (hors programme)

Soit T_n un estimateur de θ admettant une variance, b son biais et r son risque quadratique. Alors :

$$r = b^2 + V(T_n).$$

Ainsi, si l'estimateur T_n est sans biais, on a : $r = V(T_n)$.

Démonstration. À compléter. □

Remarque 3.2.5. Si un estimateur est sans biais, c'est-à-dire s'il prend en moyenne la valeur du paramètre à estimer, la variance est donc le paramètre qu'il faut chercher à minimiser, pour être sûr que l'estimateur se concentre autour de la valeur à estimer.

Dans la pratique, il se peut qu'un estimateur ait du biais mais que ce biais diminue à mesure que la taille de l'échantillon augmente. Ainsi l'estimateur T_n a une espérance qui tend vers le paramètre à estimer lorsque n tend vers l'infini. Avec la terminologie dorénavant hors programme, on rend compte de cette situation avec la définition suivante.

Définition : Estimateurs asymptotiquement sans biais (hors programme)

On dit qu'une suite $(T_n)_{n \in \mathbb{N}^*}$ d'estimateur de θ est **asymptotiquement sans biais** si, et seulement si, $\lim_{n \rightarrow +\infty} E(T_n) = \theta$.

Plus simplement, on dit que T_n est asymptotiquement sans biais.

La dernière définition sert précisément à dire qu'un estimateur ne prend pas souvent des valeurs trop éloignées du paramètre à estimer, à condition de pouvoir faire grandir la taille de l'échantillon.

Définition : Estimateur convergent

Une suite $(T_n)_{n \in \mathbb{N}^*}$ d'estimateurs de θ est dite **convergente** lorsque pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P\left(|T_n - \theta| > \varepsilon\right) = 0$$

(i.e. lorsque la probabilité que l'écart entre T_n et θ dépasse ε tend vers 0, pour tout $\varepsilon > 0$.)

Par abus de langage, on dit aussi que T_n est convergent.

Remarque 3.2.6. En fait, il s'agit d'une autre notion de convergence pour une suite de variables aléatoires, dite convergence en probabilités (la même que celle de la loi faible des grands nombres). Dans ce cours, la convergence en probabilité n'est utilisée que pour les estimateurs, il n'y a donc pas de confusion avec la convergence en loi du paragraphe précédent.

Théorème : Condition suffisante de convergence d'un estimateur

Soit T_n un estimateur possédant une variance et tel que $E(T_n) = \theta$. Alors T_n est un estimateur convergent de θ .

Remarque 3.2.7. Bien sûr, on aurait pu formuler ce théorème avec les notions de biais et de risque quadratique. Mais j'ai choisi d'énoncer ce résultat important uniquement avec les mots du programme.

Démonstration. À compléter. □

Méthode : Comment choisir un estimateur ?

Soit T_n et T'_n deux estimateurs du même paramètre θ .

1. Parmi deux estimateurs qui ont la même espérance θ , on choisira celui des deux qui a la plus petite variance.
2. Parmi deux estimateurs qui ont la même variance, on choisira celui des deux qui a l'espérance la plus proche de θ .
3. La situation se complique un peu lorsque les deux estimateurs n'ont ni la même espérance, ni la même variance et le statisticien fait son choix selon les critères qui semblent le plus pertinent. En général, on considère que si l'un des deux a pour espérance θ et que l'autre a une espérance qui converge vers θ , alors il n'y a pas de grosse différence de qualité et on préfère alors dans ce cas celui des deux qui a la plus petite variance.

3.2.2. Exemples d'estimateurs.

Définition : Moyenne empirique

Si (X_1, X_2, \dots, X_n) est un échantillon de la loi de X , on appelle **moyenne empirique** associée à l'échantillon la variable aléatoire :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Théorème : Moyenne empirique/moyenne théorique.

Soit (X_1, X_2, \dots, X_n) est un échantillon de la loi de X (admettant une espérance et une variance) et $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ la moyenne empirique associée. Alors \bar{X}_n est un estimateur de $\theta = E(X)$. Il a pour espérance la valeur θ et il est convergent.

Démonstration. À compléter. □

Remarque 3.2.8. Si (x_1, x_2, \dots, x_n) est un échantillon observé, une estimation de $E(X)$ est la moyenne observée : $\frac{1}{n}(x_1 + x_1 + \dots + x_n)$.

Exercice 3.2.9. On considère une variable aléatoire X suivant une loi uniforme sur $[0; a]$, où a est un réel inconnu. On dispose d'un n -échantillon de X qu'on note (X_1, X_2, \dots, X_n) . On voudrait faire une estimation de a .

1. Quelle est l'espérance de X ?
2. On pose $T_n = 2\bar{X}_n$ où \bar{X}_n désigne la moyenne empirique de (X_1, X_2, \dots, X_n) .
 - a. Montrer que T_n est un estimateur de a d'espérance a .
 - b. Montrer que T_n est un estimateur convergent de a .
3. On pose $U_n = \max(X_1, X_2, \dots, X_n)$.
 - a. Déterminer la loi de U_n , son espérance et sa variance.
 - b. Montrer que U_n est un estimateur asymptotiquement sans biais et convergent de a .
4. On pose $V_n = \frac{n+1}{n}U_n$.
 - a. Déterminer son espérance et sa variance.
 - b. Montrer que V_n est un estimateur convergent de a .
5. Parmi les trois estimateurs proposés, lequel choisiriez-vous pour de grandes valeurs de n ?

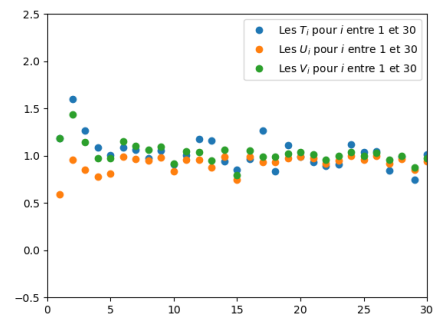
Remarque 3.2.10 (Simulation informatique).

```

1 import numpy.random as rd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5
6 n=1000
7 a=1
8 N=30
9 T=[]
10 U=[]
11 V=[]
12 for k in range (1,N+1):
13     simul = rd.uniform(0,a,k)
14     T.append(2*sum(simul)/k)
15     U.append(max(simul))
16     V.append((k+1)/k*max(simul))
17
18 x = np.linspace(1, 30, 30)
19
20 plt.plot(x, T, "o", label="Les $T_{i}$ pour $i$
21     entre 1 et 30")
22 plt.plot(x, U, "o", label="Les $U_{i}$ pour $i$
23     entre 1 et 30")
24 plt.plot(x, V, "o", label="Les $V_{i}$ pour $i$
25     entre 1 et 30")
26 plt.axis([-0, 30, -0.5, 2.5])
27 plt.legend()
28 plt.show()

```

Estimateurs.



Dans l'exercice précédent, nous avons vu apparaître le maximum des valeurs prises par l'échantillon. C'est en général un (très bon) estimateur du maximum du support de la loi.

Proposition : Minimum et Maximum d'un échantillon

Si (X_1, X_2, \dots, X_n) est un échantillon de la loi de X (qui admet une espérance dont le support est l'intervalle $[a, b]$), on note

$$M_n = \max(X_1, \dots, X_n)$$

et

$$m_n = \min(X_1, \dots, X_n).$$

Alors M_n est un estimateur de b dont l'espérance tend vers b et m_n est un estimateur de a dont l'espérance tend vers a .

Démonstration. Il suffit de reprendre l'exercice précédent. \square

Attention, les deux estimateurs précédents sont biaisés et ne sont seulement que asymptotiquement sans biais. On les utilise pourtant souvent dans la pratique car il est fréquent que leur variance soit petite.

► **Pour s'entraîner : exo 7.**

Définition : Maximum de vraisemblance

Soit (X_1, \dots, X_n) un n -échantillon d'une loi discrète et commune $X \hookrightarrow \mathcal{L}(\theta)$ et soit $(x_1, \dots, x_n) \in X(\Omega)$. La fonction L_n de θ définie par

$$L_n : \theta \mapsto \prod P_\theta(X_i = x_i)$$

s'appelle la **vraisemblance** de la loi \mathcal{L} .

En notant, θ^* la valeur où L_n est maximale (c'est-à-dire telle que, pour tout θ , $L_n(\theta) \leq L_n(\theta^*)$), l'estimateur du **maximum de vraisemblance** est l'estimateur défini par

$$\hat{\theta}_n = \varphi(X_1, \dots, X_n).$$

L'idée qui motive la définition précédente est simple mais très maline : on sait que l'événement

$$(X_1 = x_1) \cap (X_2 = x_2) \cap \dots \cap (X_n = x_n)$$

s'est réalisé et on estime θ par la valeur qui donne à cet événement la plus grande probabilité de se réaliser. On choisit donc de considérer que si un événement s'est réalisé, c'est que c'était le plus probable qu'il se réalise !

Théorème : Maximum de vraisemblance de la loi $\mathcal{B}(p)$

Soit (X_1, X_2, \dots, X_n) un n -échantillon de la loi $\mathcal{B}(p)$. On cherche ici à étudier le paramètre $\theta = p$. Alors le maximum de vraisemblance $\hat{\theta}_n$ est égal à la moyenne empirique :

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Exercice 3.2.11. À l'aide d'un n -échantillon d'une loi de Poisson de paramètre λ , fabriquer l'estimateur du maximum de vraisemblance pour estimer le paramètre λ . Même chose pour estimer le paramètre p d'une loi géométrique de paramètre p .

Remarque 3.2.12. On peut aussi définir la vraisemblance d'une loi à densité. En notant f_θ une densité d'une variable aléatoire X échantillonnée et de loi inconnue $\mathcal{L}(\theta)$, et (x_1, x_2, \dots, x_n) un n -uplet de valeurs de $X(\Omega)$, il s'agit de la fonction

$$L_n : \theta \mapsto \prod_{i=1}^n f_\theta(x_i)$$

L'estimateur du maximum de vraisemblance est alors définie comme précédemment.

Exercice 3.2.13. On considère un n -échantillon de la loi $\mathcal{U}([0, \theta])$ et on cherche à estimer θ . Fabriquer l'estimateur du maximum de vraisemblance pour estimer θ .

► **Pour s'entraîner : exo 8.**

3.3. Estimation par intervalle de confiance. Dans cette partie, U_n et V_n sont deux estimateurs de θ tels que $P(U_n \leq V_n) = 1$ pour tout $n \in \mathbb{N}^*$ (U_n prend presque-sûrement des valeurs toujours inférieures ou égales à V_n).

Définition : Intervalle de confiance

Soit α un réel élément de $[0; 1]$. L'intervalle $[U_n; V_n]$ est appelé **intervalle de confiance** de θ **au niveau de confiance** $1 - \alpha$ si, et seulement si :

$$P(U_n \leq \theta \leq V_n) \geq 1 - \alpha.$$

Autrement dit, on a une probabilité supérieure ou égale à $1 - \alpha$ que θ soit dans l'intervalle $[U_n; V_n]$.

Le réel α est appelé le **seuil de risque**.

Définition : Intervalle de confiance asymptotique

Soit α un réel élément de $[0; 1]$. L'intervalle $[U_n; V_n]$ est appelé **intervalle de confiance asymptotique** de θ **au niveau de confiance** α si, et seulement s'il existe une suite de réels $(\alpha_n)_{n \in \mathbb{N}^*}$ à valeurs dans $[0; 1]$ et de limite α telle que pour tout $n \geq 1$,

$$P(U_n \leq \theta \leq V_n) \geq 1 - \alpha_n.$$

Remarque 3.3.1. En particulier, $[U_n; V_n]$ est un intervalle de confiance asymptotique de θ au niveau de confiance $1 - \alpha$ lorsque : $\lim_{n \rightarrow +\infty} P(U_n \leq \theta \leq V_n) \geq 1 - \alpha$.

Méthode : Intervalle de confiance grâce à l'inégalité de Bienaymé-Tchebychev

Pour trouver un intervalle de confiance $[U_n; V_n]$ au niveau de confiance $1 - \alpha$ en utilisant l'inégalité de Bienaymé-Tchebychev, on peut procéder de la façon suivante :

1. On montre que l'estimateur T_n de θ dont on dispose est sans biais.
2. On écrit l'inégalité appliquée à T_n , pour ε quelconque. On obtient : $P\left(\left|T_n - \theta\right| \geq \varepsilon\right) \leq \dots$
3. On considère l'événement contraire pour faire apparaître $P\left(\left|T_n - \theta\right| < \varepsilon\right) \geq \dots$
4. On choisit ε pour que le membre de droite soit égal à $1 - \alpha$.
5. On transforme l'écriture avec la valeur absolue pour avoir $P(U_n \leq \theta \leq V_n) \geq 1 - \alpha$.

Exercice 3.3.2. Au 2ème tour des élections présidentielles, les citoyens ont le choix entre Jean-Michel Peste et Jean-Pierre Choléra. Un institut de sondage interroge 100 personnes, dont 60 déclarent voter pour M. Peste. On souhaite estimer la proportion finale de gens qui voteront pour M. Peste. On modélise le choix par un échantillon $(X_1, X_2, \dots, X_{100})$ de variables aléatoires indépendantes de même loi de Bernoulli de paramètre p ($X_i = 1$ si la i^e personne vote pour M. Peste, $X_i = 0$ sinon, et p représente la probabilité de voter pour M. Peste, ou la proportion de gens qui vont voter pour M. Peste, $p = E(X)$, et $q = 1 - p$).

On considère la moyenne empirique $\bar{X} = \frac{1}{100} \sum_{i=1}^{100} X_i$. L'une des observations donne $\bar{x} = 0,6$.

1. Déterminer l'espérance et la variance de \bar{X} .
2. Écrire l'inégalité de Bienaymé-Tchebychev pour la variable \bar{X} .
3. Majorer $p(1 - p)$ puis en déduire un intervalle de confiance pour p au niveau de confiance 99%.

Méthode : Intervalle de confiance grâce à au théorème central limite

Pour trouver un intervalle de confiance asymptotique $[U_n; V_n]$ de l'espérance m d'une variable aléatoire X au niveau de confiance $1 - \alpha$ en appliquant le théorème central limite, on peut procéder de la façon suivante :

1. On considère la moyenne empirique \bar{X}_n ou la somme S_n d'un n -échantillon de la loi de X , et on détermine la variable centrée réduite $\bar{X}_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$. \bar{X}_n^* fait apparaître le paramètre m que l'on veut estimer.

2. On applique le théorème central limite (en vérifiant ses hypothèses : une suite de variables indépendantes et de même loi, admettant une espérance et une variance, et la variable étudiée est une somme ou une moyenne empirique de variables de cette suite), et on utilise les propriétés de Φ : pour tout $x > 0$,

$$\lim_{n \rightarrow +\infty} P(-x \leq \bar{X}_n^* \leq x) = 2\Phi(x) - 1.$$

3. On considère le réel x tel que $2\Phi(x) - 1 = 1 - \alpha$ i.e. $\Phi(x) = 1 - \frac{\alpha}{2}$ (Φ est bijective) et on vérifie que :

$$\lim_{n \rightarrow +\infty} P(-x \leq \bar{X}_n^* \leq x) = 1 - \alpha$$

4. On utilise l'expression de \bar{X}_n^* pour transformer l'écriture de $P(-x \leq \bar{X}_n^* \leq x)$ et faire apparaître une expression du type $P(U_n \leq m \leq V_n)$.

Exercice 3.3.3. On reprend les données de l'exercice 3.3.2.

1. On note \bar{X}^* la moyenne empirique centrée réduite. Par quelle loi peut-on approcher celle de \bar{X}^* ? On considère dans la suite que \bar{X}^* suit cette loi.
2. Déterminer un réel t tel que $P(-t \leq \bar{X}^* \leq t) \geq 0,99$.
3. En déduire que $P\left(\bar{X} - t\sqrt{\frac{pq}{n}} \leq p \leq \bar{X} + t\sqrt{\frac{pq}{n}}\right) \geq 0,99$.
4. Majorer $p(1-p)$ puis en déduire que, lorsque $n = 100$, $\left[\bar{X} - \frac{t}{20}; \bar{X} + \frac{t}{20}\right]$ est un intervalle de confiance pour p au niveau de confiance 99%.

► **Pour s'entraîner : exo 6 et 9.**

4. SUJETS D'ANNALES EN LIEN AVEC CE CHAPITRE.

Remarque 4.0.1. 1. Nous traiterons certains des sujets suivants en exercices, en travaux dirigés, en colles ou en devoir. Pour les autres, il existe des corrigés que l'on trouve facilement sur Internet. Ces corrigés sont parfois très rapides, n'hésitez pas à venir m'en parler si vous pensez qu'une question mérite des explications supplémentaires.

2. Les sujets de concours sont souvent pensés pour faire appel à plusieurs parties du programme. Dans la liste qui suit figurent les exercices pour lequel il est *nécessaire* de connaître les résultats de ce chapitre. Mais parfois *ce n'est pas suffisant* car d'autres parties du cours sont aussi impliquées. J'indique ces situations avec le symbole *.
3. Cette liste n'est pas exhaustive.

1. ECRICOME

- 1990 Problème.
- 1991 Problème.
- 1992 Problème.
- 1993 Problème.
- 1995 Exercice 1 (mais quelques fonctions sont maintenant hors programme).
- 2018 Exercice 3.
- 2020 Exercice 3.

2. EDHEC

- 2011 Exercice 3 (pas vraiment dans le thème mais on vérifie la loi des grands nombres à la main pour une suite de variables non indépendantes).
- 2012 Exercice 3 et Problème.
- 2013 Problème.
- 2014 Exercice 3.
- 2018 Exercice 3.
- 2019 Exercice 3.
- 2020 Exercice 2.
- 2021 Exercice 2 (on vérifie le théorème central limite dans un cas particulier).
- 2022 Problème.

3. EML

- 2006 Exercice 3.
- 2012 Exercice 3.
- 2016 Exercice 3.
- 2017 Exercice 3.
- 2020 Exercice 3.

4. ESCP

- 1984 épreuve II Problème 2.
- 1988 épreuve III Exercice 3.
- 1998 épreuve III Exercice 3.
- 2002 épreuve II (à la fin).
- 2003 épreuve III Exercice.

5. ESC

- 2006 Exercice 3.
- 2007 Exercice 3.
- 2009 Exercice 4.

6. ESSEC

- 1989 épreuve II Partie II.
- 2004 épreuve II Partie III.
- 2004 épreuve III Exercice 2.
- 2005 épreuve II tout le sujet.
- 2006 épreuve II Partie I.
- 2008 épreuve II.
- 2013 épreuve II Partie II.
- 2014 épreuve II Partie I.
- 2016 épreuve I Partie II (avec aussi de la convergence en probabilité) et Partie III.
- 2017 épreuve II tout le sujet est une analyse statistique.
- 2020 épreuve II.
- 2021 épreuve II avec de la convergence en probabilité.
- 2022 épreuve II tout le sujet.

7. HEC

- 2010 Problème.
- 2012 Problème.
- 2013 Problème à la fin.
- 2014 Problème.
- 2016 Problème.
- 2017 Problème.