14 Théorèmes de convergence en probabilité Statistiques inférentielles

14.1 Estimation

Définition 14.1

Soit X une variable aléatoire admettant une espérance μ et une variance σ^2.

Pour tout entier $n \in \mathbb{N}^*$, on appelle n-échantillon de X tout n-uplet $(X_1, \ldots, X_n)$ de variables aléatoires indépendantes et de même loi *que X.

On appelle alors *estimateur* d'un paramètre θ de X toute suite de variable aléatoire (T_n), chaque T_n étant une fonction de $(X_1, \ldots, X_n)$.

Le but d'un estimateur est de pouvoir, à partir d'une suite d'observations empiriques d'un phénomène aléatoire, calculer une approximation du paramètre recherché.

Exemple : (à connaître)

Soit X une variable aléatoire admettant une espérance μ et une variance σ^2.

Les deux estimateurs suivants sont à connaître :

- la *moyenne empirique*, notée M_n ou $\overline{X_n}$, définie par

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

  est un estimateur de l'espérance μ de X.

- la *variance empirique*, notée S_n^2, définie par

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - M_n^2$$

  est un estimateur de la variance σ^2 de X.

*. On parlera souvent de suite de variables *i.i.d.*, pour « indépendantes et identiquement distribuées ».

14.2 Convergence en loi

Les suites de variables aléatoires peuvent converger sous différents sens. On étudie ici le cas de la convergence *en loi*.

Définition 14.2

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires, et soit X une variable aléatoire.

Pour tout n, on note F_n la fonction de répartition de X_n, et on note F la fonction de répartition de X.

On dit alors que la suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X si pour tout réel t où F est continue, la suite $(F_n(t))_{n\in\mathbb{N}}$ converge vers $F(t)$.

On note alors $X_n \xrightarrow{\mathcal{L}} X$.

Exemple : Soient $U_n \hookrightarrow \mathcal{U}([0,1])$ mutuellement indépendantes. On pose

$$X_n = n(1 - \max(U_1, \dots, U_n)).$$

On commence par chercher la fonction de répartition de X_n : soit $x \in \mathbb{R}$.

$$\begin{aligned} P(X_n \leqslant x) &= P\left(\max(U_1, \dots, U_n) \geqslant 1 - \frac{x}{n}\right) \\ &= 1 - P\left(\max(U_1, \dots, U_n) < 1 - \frac{x}{n}\right) \\ &= 1 - \prod_{i=1}^{n} P\left(U_i < 1 - \frac{x}{n}\right) \\ &= 1 - \left(1 - \frac{x}{n}\right)^n \end{aligned}$$

Or on sait que $(1 - \frac{x}{n})^n \xrightarrow[n\to\infty]{} e^{-x}$, et donc, pour tous n et x

$$F_{X_n}(x) \xrightarrow[n\to\infty]{} 1 - e^{-x}.$$

On reconnaît alors la fonction de répartition d'une loi exponentielle de paramètre 1.

On peut donc dire, en posant $X \hookrightarrow \mathcal{E}(1)$, que

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

Proposition 14.3

Soient (X_n) et X des variables aléatoires à valeurs dans $\mathbb{Z}$. Alors

$$X_n \xrightarrow{\mathcal{L}} X \quad \Leftrightarrow \quad \forall k \in \mathbb{Z},\ \lim_{n\to\infty} P(X_n = k) = P(X = k).$$

Proposition 14.4 – Approximation de loi binomiale par la loi de Poisson

Soit (X_n) une suite de variables aléatoires de lois respectives $\mathcal{B}(n, p_n)$, avec $np_n \to \lambda > 0$.

Alors $X_n \xrightarrow{\mathcal{L}} X$ où $X \hookrightarrow \mathcal{P}(\lambda)$.

Démonstration. Les lois binomiales et de Poisson sont à valeurs entières, et on applique donc le résultat précédent.

On a donc

$$P(X_n = k) = \binom{n}{k}\left(\frac{p_n}{1-p_n}\right)^k (1-p_n)^n.$$

On note que la condition $np_n \to \lambda$ est équivalente à dire que $p_n \sim \frac{\lambda}{n}$, et donc

$$\left(\frac{p_n}{1-p_n}\right)^k \sim p_n^k \sim \frac{\lambda^k}{n^k}.$$

De plus, on a

$$(1-p_n)^n = e^{n\ln(1-p_n)},$$

avec $n\ln(1-p_n) \sim -np_n \sim -\lambda$. Donc $(1-p_n)^n \to e^{-\lambda}$. On peut aussi montrer facilement que $\binom{n}{k} \sim \frac{n^k}{k!}$.

On a donc

$$P(X_n = k) \sim \frac{n^k}{k!}\frac{\lambda^k}{n^k}e^{-\lambda} = \frac{\lambda^k}{k!}e^{-\lambda},$$

et on reconnaît la loi d'une variable de Poisson de paramètre λ. □

Dans la pratique, pour n assez grand et certaines conditions, on utilise le résultat d'approximation suivant :

Corollaire 14.5

Si $p \leqslant 0,1$, $n \geqslant 30$ et $np \leqslant 15$, on peut approximer la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(np)$.

14.3 Théorèmes limites

Commençons par rappeler que si X est une variable aléatoire admettant une espérance μ et une variance $\sigma^2 \neq 0$, alors on note X^* la variable aléatoire $\frac{X-\mu}{\sigma}$. Cette variable aléatoire est alors centrée réduite.

On notera toujours $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ la moyenne empirique, où les variables X_i sont indépendantes et de même loi que X.

On a alors

$$\mathbb{E}(M_n) = \mu \quad \text{et} \quad \mathbb{V}(M_n) = \frac{\sigma^2}{n}.$$

La variable centrée réduite associée à M_n est donc M_n^* définie par :

$$M_n^* = \frac{M_n - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

On rappelle alors l'inégalité de Bienaymé-Tchebychev :

Théorème 14.6 – Inégalité de Bienaymé-Tchebytchev

Soit X une variable aléatoire admettant une variance. Alors

$$\forall \varepsilon > 0,\ \mathbb{P}(|X - \mathbb{E}(X)| \geqslant \varepsilon) \leqslant \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

Exemple : On souhaite estimer la valeur de $\frac{\pi}{4}$ à $0,1$ près, avec un risque d'erreur inférieur à 1%. Pour cela, on tire au hasard deux valeurs x et y uniformément dans $[0,1]$. On note que la probabilité que le point de coordonnées (x, y) soit dans le quart de cercle de centre 0 et de rayon 1 est $\frac{\pi}{4}$.

On itère l'expérience : on note X_n la variable aléatoire qui vaut 1 si le n-ième point tiré est dans le cercle, 0 sinon.

On cherche donc n tel que

$$P\left(\left|\overline{X_n} - \frac{\pi}{4}\right| \geqslant 0,1\right) < 0,01.$$

En notant que pour tout $x \in [0,1]$, $x(1-x) \leqslant \frac{1}{4}$, l'inégalité de Bienaymé-Tchebytchev nous donne

$$P\left(\left|\overline{X_n} - \frac{\pi}{4}\right| \geqslant 0,1\right) \leqslant \frac{1}{4n \times 0,1^2} = \frac{25}{n}.$$

Il suffit donc de prendre $n \geqslant 2500$.

On se fixe alors dans la suite une variable aléatoire X admettant une espérance μ et une variance $\sigma^2 \neq 0$.

On en déduit alors directement la loi faible des grands nombres :

Théorème 14.7 – Loi faible des grands nombres

Soit (X_n) une suite de variables aléatoires indépendantes de même loi que X. Alors

$$\forall \varepsilon > 0,\ \mathbb{P}(|M_n - \mu| \geqslant \varepsilon) \xrightarrow[n \to \infty]{} 0.$$

On peut alors raffiner le théorème précédent pour obtenir le célèbre théorème central limite.

Théorème 14.8 – central limite

Soit (X_n) une suite de variables aléatoires indépendantes toutes de même loi que X.

Alors la suite (M_n^*) converge en loi vers une loi normale centrée réduite.

On a donc pour tous réels $a < b$,

$$\mathbb{P}(a < M_n^* < b) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

Nota : Le résultat reste vrai en faisant tendre a ou b vers $\pm\infty$. Par exemple,

$$\mathbb{P}(M_n^* < b) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

Ce théorème nous dit que, en un certain sens, on peut approximer la quantité M_n par une loi normale de paramètres μ et $\frac{\sigma^2}{n}$.

Un inconvénient de ce théorème est qu'il est nécessaire de connaître l'écart-type du phénomène observé. Dans le cas contraire, on peut le remplacer par l'écart-type empirique.

Avec les mêmes hypothèses, posons

$$M_n = \frac{1}{n}\sum_{k=1}^n X_k$$
$$S_n^2 = \frac{1}{n}\sum_{k=1}^n (X_k - M_n)^2$$

On a alors

Théorème 14.9 – central limite

Dans le contexte précédent, la suite $\left(\frac{M_n-\mu}{\frac{S_n}{\sqrt{n}}}\right)$ converge en loi vers une loi normale centrée réduite.

On a donc

$$\mathbb{P}\left(a < \frac{M_n-\mu}{\frac{S_n}{\sqrt{n}}} < b\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

Dans le cas particulier de la loi binomiale, on obtient le théorème suivant :

Théorème 14.10 – Moivre-Laplace

Soit $p \in]0,1[$, et soit (X_n) une suite de variables indépendantes telles que $X_n \hookrightarrow \mathcal{B}(n,p)$.

Alors pour tous réels $a < b$,

$$\mathbb{P}\left(a < \frac{X_n - np}{\sqrt{np(1-p)}} < b\right) \xrightarrow[n\to\infty]{} \Phi(b) - \Phi(a).$$

Démonstration. On peut écrire X_n comme la somme de n variables de Bernoulli identiques et indépendantes, de paramètre p : $X_n = \sum_{k=1}^{n} B_k$.

En notant $\overline{B_n} = \frac{1}{n}\sum_{k=1}^{n} B_k$, on a donc $\overline{B_n}^* = \frac{\overline{B_n}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{X_n - np}{\sqrt{np(1-p)}}$.

Le théorème central limite appliqué à la suite (B_n) permet d'affirmer que la suite $(\overline{B_n}^*)$ converge en loi vers une loi normale centrée réduite. □

En pratique, on approche la loi binomiale $\mathcal{B}(n,p)$ par une loi normale $\mathcal{N}(np, \sqrt{npq}^2)$ si $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.

14.4 Test de conformité de la moyenne

En statistique, un *test d'hypothèse* est une méthode qui nous permettra de choisir s'il est raisonnable d'accepter une hypothèse H_0.

L'hypothèse H_0, appelée *hypothèse nulle* est alors faite ; tous les résultats qui suivent supposent donc que cette hypothèse est vraie. L'hypothèse contraire à H_0 est appelée *hypothèse alternative*, notée H_1.

Le test sera alors une méthode qui nous permettra d'accepter ou de rejeter H_0, en fonction d'observations empiriques. Il est cependant statistiquement impossible d'être sûr d'avoir eu raison dans notre conclusion. On distingue alors deux types d'erreurs :

- le risque de première espèce, qui est le risque de rejeter l'hypothèse H_0 sachant qu'elle est vraie ; ce risque est choisi à l'avance, souvent 5% ou 1%
- le risque de seconde espèce, qui est le risque d'accepter l'hypothèse sachant qu'elle est fausse ; il est plus difficile à calculer, et son étude dépasse le cadre de ce cours.

On peut alors utiliser le test suivant pour la valeur de la moyenne :

Proposition 14.11

Soit X une variable aléatoire admettant une espérance μ et une variance $\sigma^2 \neq 0$. Soit $\mu_0 \in \mathbb{R}$.

On suppose $H_0 : \mu = \mu_0$, et on définit l'hypothèse alternative $H_1 : \mu \neq \mu_0$.

Alors

$$\mathbb{P}\left(\left|\frac{M_n - \mu_0}{\frac{S_n}{\sqrt{n}}}\right| > u_{1-\frac{\alpha}{2}}\right) \xrightarrow[n\to\infty]{} \alpha,$$

où $u_{1-\frac{\alpha}{2}}$ est l'unique réel tel que $\Phi\left(u_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$.

En pratique, pour faire un test d'hypothèse de conformité de la moyenne :

- on calcule la moyenne empirique et l'écart type empirique de nos mesures
- on rejette l'hypothèse si la valeur observée de $\frac{M_n - \mu_0}{\frac{S_n}{\sqrt{n}}}$ n'est pas dans l'intervalle $\left[-u_{1-\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}}\right]$

Exemple : Un médicament doit contenir $\mu_0 = 2,5$g de substance active.

En prenant 100 comprimés au hasard, on trouve une moyenne de $M_n = 2,6$g de substance active, avec une variance $S_n^2 = 0.16$.

On utilise un test de conformité à la moyenne, avec un risque de 5%.

On choisit alors $u_{1-\frac{\alpha}{2}} = 1.96$, et on a alors

$$\frac{M_n - \mu_0}{\frac{S_n}{\sqrt{n}}} = 2.5 \notin [-1.96, 1.96].$$

On peut donc rejeter notre hypothèse.

14.5 Exercices

Exercice 1

Soit (X_n) une suite de variables aléatoires identiques et indépendantes, admettant une espérance μ et une variance σ^2.

1. Montrer que $\mathbb{E}(S_n^2) \neq \sigma^2$.
2. On appelle alors *estimateur corrigé de la variance* la suite

$$S_n'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n)^2.$$

   Montrer que $\mathbb{E}(S_n'^2) = \sigma^2$.

Exercice 2

Soient $n \in \mathbb{N}^*$ et $a \in \mathbb{R}$. On définit la fonction

$$f_n : \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \frac{an}{\pi(1+n^2x^2)} \end{array}.$$

1. Calculer la valeur de a pour que f_n soit une densité.
2. Soient X_n des variables aléatoires de densité f_n. Montrer que X_n n'admet ni espérance, ni variance.
3. Étudier la convergence en loi de la suite (X_n).

Réponse de l'exercice

1. Les fonctions f_n sont continues et positives sur $\mathbb{R}$. Reste à calculer ; soient $b < c$:

$$\begin{aligned} \int_b^c f_n(x)\,\mathrm{d}x & \\ &= \frac{a}{\pi}\int_b^c \frac{n}{1+n^2x^2}\,\mathrm{d}x \\ &= \frac{a}{\pi}\left(\arctan(nc) - \arctan(nb)\right) \\ &\to a \end{aligned}$$

   Il faut et il suffit de prendre $a = 1$.
2. Au voisinage de $+\infty$, on a $xf_n(x) \sim \frac{1}{nx}$, dont l'intégrale diverge.
   Par théorème de comparaison des intégrales de fonctions positives, X_n n'admet donc pas d'espérance, et donc pas de variance.

3. On note que pour tout x :
$$F_n(x) = \frac{1}{\pi}\left(\arctan(nx) + \frac{\pi}{2}\right).$$
Ainsi, quand n tend vers l'infini, on a donc
$$F_n(x) \to \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{sinon} \end{cases}.$$
Ainsi, la suite (X_n) converge en loi vers la variable certaine égale à 0.

Exercice 3

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes. Pour tout $n \in \mathbb{N}^*$,
$$\mathbb{P}(X_n = 1) = p \text{ et } \mathbb{P}(X_n = -1) = q = 1 - p.$$

1. On cherche la loi de $Y_n = \prod_{k=1}^{n} X_k$.
   a) Quel est l'univers image de Y_n ?
   b) On pose $\alpha = \mathbb{P}(Y_n = 1)$. Exprimer l'espérance de Y_n en fonction de α.
   c) En déduire alors la loi de Y_n.
2. Montrer alors que la suite (Y_n) converge en loi.

Réponse de l'exercice

Exercice 4

Une compagnie aérienne, disposant de 300 places dans un avion, souhaite optimiser au mieux son remplissage. On estime qu'une personne ayant réservé oublie de se présenter à l'aéroport avec une probabilité de 10%. Les choix des passagers de venir ou non sont supposés indépendants.

Pour un entier n représentant le nombre de réservations faites, on note S_n le nombre de personnes qui se présentent à l'aéroport.

1. Calculer $\mathbb{E}(S_n)$ et $\mathbb{V}(S_n)$.
2. Déterminer le nombre maximal de réservations que peut proposer la compagnie pour que le risque de ne pas pouvoir embarquer tout le monde soit inférieur à 5%.

Réponse de l'exercice

1. La variable S_n suit une loi binomiale $\mathcal{B}(n, 0.9)$. On a donc $\mathbb{E}(S_n) = 0.9n$ et $\mathbb{V}(S_n) = 0.09n$.

2. On note que $S_n \leqslant 300$ si et seulement si $\frac{\frac{1}{n}S_n - 0.9}{\frac{\sqrt{0.09}}{\sqrt{n}}} \leqslant \frac{\frac{1}{n}300 - 0.9}{\frac{\sqrt{0.09}}{\sqrt{n}}}$.

   Le théorème central limite nous affirme, les variables admettant une variance non nulle, que

$$\mathbb{P}\left(\frac{\frac{1}{n}S_n - 0.9}{\frac{\sqrt{0.09}}{\sqrt{n}}} \leqslant \frac{\frac{1}{n}300 - 0.9}{\frac{\sqrt{0.09}}{\sqrt{n}}}\right) \approx \Phi\left(\frac{\frac{1}{n}300 - 0.9}{\frac{\sqrt{0.09}}{\sqrt{n}}}\right).$$

   On cherche donc n tel que $\Phi\left(\frac{\frac{1}{n}300-0.9}{\frac{\sqrt{0.09}}{\sqrt{n}}}\right) \geqslant 0.95 \approx \Phi(1,65)$.

   Par croissance de Φ, on a alors

$$\begin{aligned}
&\Phi\left(\frac{\frac{1}{n}300 - 0.9}{\frac{\sqrt{0.09}}{\sqrt{n}}}\right) \geqslant 0.95 \approx \Phi(1,65)\\
\Leftrightarrow &\frac{\frac{1}{n}300 - 0.9}{\frac{\sqrt{0.09}}{\sqrt{n}}} \geqslant 1.65\\
\Leftrightarrow &n \leqslant 323
\end{aligned}$$

Exercice 5

Montrer que

$$\lim_{n\to\infty} e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} = \frac{1}{2}.$$

On pourra utiliser des variables aléatoires X_i indépendantes, suivant toutes une loi de Poisson $\mathcal{P}(1)$.

Réponse de l'exercice

Par stabilité des lois de Poisson par somme, on a donc $S_n \hookrightarrow \mathcal{P}(n)$. Le théorème central limite nous affirme alors

$$\mathbb{P}\left(\frac{\frac{1}{n}S_n - 1}{\frac{1}{\sqrt{n}}} \leqslant 0\right) \to \Phi(0) = \frac{1}{2}.$$

D'autre part, on a

$$\mathbb{P}\left(\frac{\frac{1}{n}S_n - 1}{\frac{1}{\sqrt{n}}} \leqslant 0\right) = \mathbb{P}(S_n \leqslant n) = \sum_{k=0}^{n}\mathbb{P}(S_n) = k = e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}.$$

Exercice 6

On souhaite estimer un paramètre $p \in]0;1[$. On note : $q = 1 - p$.

Soit un entier $n \geqslant 1$ fixé. On considère $X_1, \ldots, X_n$ des variables aléatoires, indépendantes, suivant une même loi de Bernoulli de paramètre p et définies sur un même espace probabilisé.

On note : $\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k$.

1. a) Justifier que $p(1-p) \leqslant \frac{1}{4}$.

   b) En utilisant l'inégalité de Bienaymé-Tchebychev, montrer que p est dans l'intervalle $\left[\overline{X_n} - \sqrt{\frac{5}{n}}, \overline{X_n} + \sqrt{\frac{5}{n}}\right]$ avec probabilité supérieure à $0,95$.

2. Écrire une fonction Python `test(n,p,a,b)` qui prend en arguments un entier n, une probabilité p, deux flottants a et b, simule une réalisation de $\overline{X_n}$ et retourne 1 si $\overline{X_n}$ appartient à $[a, b]$ et 0 sinon.

On cherche par la suite un intervalle auquel p appartient avec probabilité supérieure à $0,95$, d'amplitude plus petite.

3. On fixe un réel strictement positif t quelconque et ε un réel strictement positif quelconque.
   a) Établir l'égalité : $\mathbb{P}\left(\overline{X_n} - p \geqslant \varepsilon\right) = \mathbb{P}\left(e^{nt\overline{X_n}} \geq e^{nt(p+\varepsilon)}\right)$.
   b) En utilisant l'inégalité de Markov, établir l'inégalité suivante :
   $$\mathbb{P}\left(\overline{X_n} - p \geqslant \varepsilon\right) \leqslant e^{n\left(\ln(pe^t+q)-t(p+\varepsilon)\right)}.$$
   c) On admet l'inégalité : $\ln(pe^t + q) - tp \leqslant \frac{t^2}{8}$. Ainsi, on a l'inégalité suivante :
   $$\forall t \in \mathbb{R}_+,\ \mathbb{P}\left(\overline{X_n} - p \geqslant \varepsilon\right) \leqslant e^{n\left(\frac{t^2}{8}-t\varepsilon\right)}.$$
   En déduire l'inégalité : $\mathbb{P}\left(\overline{X_n} - p \geqslant \varepsilon\right) \leqslant e^{-2n\varepsilon^2}$.
4. On pose $\overline{Y_n} = \frac{1}{n}\sum\limits_{k=1}^{n}(1 - X_k)$. Établir l'inégalité : $\mathbb{P}\left(\overline{Y_n} - q \geqslant \varepsilon\right) \leqslant e^{-2n\varepsilon^2}$.
5. Déduire des questions 3.3 et 4. l'inégalité :
   $$\forall \varepsilon \in \mathbb{R}_+^*,\ \mathbb{P}\left(\left|\overline{X_n} - p\right| \geqslant \varepsilon\right) \leqslant 2e^{-2n\varepsilon^2}.$$
6. Comment choisir ε pour répondre au problème ? L'amplitude de l'intervalle de confiance est-elle plus réduite que celle obtenue à la question 1.b) ?

Réponse de l'exercice

1. a) Il suffit d'étudier la fonction $x \mapsto x(1 - x)$.
   b) On note que $\overline{X_n}$ admet bien un moment d'ordre 2 et donc d'après l'inégalité de Biemaymé-Tchebytchev, on a pour tout $\varepsilon > 0$
   $$\mathbb{P}\left(\left|\overline{X_n} - p\right| \geqslant \varepsilon\right) \leqslant \frac{p(1-p)}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2}.$$
   On en déduit, en posant $\varepsilon = \sqrt{\frac{5}{n}}$, que
   $$\mathbb{P}\left(p \in \left[\overline{X_n} - \sqrt{\frac{5}{n}}, \overline{X_n} + \sqrt{\frac{5}{n}}\right]\right) = 0.95.$$
2. On propose le code suivant :

```python
import random as rd

def bernoulli(p):
if rd.random() <p:
return 1
else:
return 0

def test(n,p,a,b):
X = [bernoulli(p) for _ in range(n)]
X = sum(X)/n
return (a<=X and X<=b)

```

3. a) Il suffit d'utiliser la croissance de la fonction exponentielle et la positivité de t.

   b) La variable $\overline{X_n}$ admet bien un espérance, et on a donc

$$\mathbb{P}(\overline{X_n} - p \geqslant \varepsilon) \leqslant e^{-n(t+p)\varepsilon}\mathbb{E}\left(e^{tn\overline{X_n}}\right).$$

   Or on a

$$\mathbb{E}\left(e^{tn\overline{X_n}}\right) = \mathbb{E}\left(e^{tX_1}\right)^n = (q + pe^t)^2,$$

   et on retrouve bien le résultat voulu.

   c) On note qu'en étudiant la fonction $t \mapsto \frac{1}{8}t^2 - t\varepsilon$, elle atteint son minimum en $t = 4\varepsilon$, et donc en prenant $t = 4\varepsilon$, on a bien le résultat voulu.

4. On remarque que les variables Y_k suivent une loi de Bernoulli de paramètre q, et il suffit donc d'appliquer l'inégalité précédente.

5. On a

$$\begin{aligned}[|\overline{X_n} - p| \geqslant \varepsilon] &= [\overline{X_n} - p \geqslant \varepsilon] \cup [\overline{X_n} - p \leqslant -\varepsilon] \\ &= [\overline{X_n} - p \geqslant \varepsilon] \cup [\overline{Y_n} - q \geqslant \varepsilon]\end{aligned}$$

   On a alors, par σ-additivité

$$\mathbb{P}\left(|\overline{X_n} - p| \geqslant \varepsilon\right) \leqslant 2e^{-2n\varepsilon^2}.$$

6. Il suffit donc d'avoir $2e^{-2n\varepsilon^2} \leqslant 0.05$, et o n peut donc prendre

$$\varepsilon = \sqrt{\frac{-\ln(0.025)}{2n}},$$

   On a alors approximativement l'intervalle de confiance $\left[\overline{X_n} - \varepsilon, \overline{X_n} + \varepsilon\right] \approx \left[\overline{X_n} - \sqrt{\frac{1.8}{n}}, \overline{X_n} + \sqrt{\frac{1.8}{n}}\right]$, qui est donc un peu meilleur.