

Chapitre XIII

Estimation par intervalles de confiance

*Un cours complet et accessible
pour les classes préparatoires ECG*

Ce document reprend et développe le cours d'estimation
avec des explications détaillées, des exemples concrets
et des rappels méthodologiques.

Pr. Elakili

prepasmaths360.com

2025–2026

Table des matières

Introduction	3
1 Intervalle de confiance : définition	3
1.1 Cadre général	3
1.2 Vocabulaire et conventions	4
1.3 En pratique	4
2 Construction par l'inégalité de Bienaymé-Tchebychev	4
2.1 Rappel de l'inégalité	4
2.2 Construction de l'intervalle	4
2.3 Exemple : estimation de p pour une loi de Bernoulli	5
3 Intervalle de confiance pour la moyenne d'une loi normale	6
3.1 Cadre et hypothèses	6
3.2 La moyenne empirique et sa loi	7
3.3 Construction de l'intervalle	7
4 Intervalles de confiance asymptotiques	8
4.1 Motivation	8
4.2 Définition	9
4.3 Exemple phare : intervalle asymptotique pour p (loi de Bernoulli)	9
5 Synthèse et comparaison des méthodes	11
6 Formulaire récapitulatif	11

Introduction : pourquoi les intervalles de confiance ?

Lorsqu'on réalise un sondage ou une expérience, on cherche souvent à estimer un paramètre inconnu θ (par exemple, la proportion p de personnes favorables à une mesure).

À partir d'un échantillon $(X_1, X_2, \dots, X_n)$, on peut calculer un **estimateur ponctuel** T_n de θ . Mais donner une seule valeur ne suffit pas : on a besoin de savoir **à quel point cette estimation est fiable**.

L'idée clé

Un **intervalle de confiance** est un intervalle aléatoire $[U_n, V_n]$ qui “capture” le vrai paramètre θ avec une probabilité *contrôlée*. Par exemple, dire “ p est dans $[0,42 ; 0,48]$ avec une confiance de 95 %” est beaucoup plus informatif que de dire simplement “ $\hat{p} = 0,45$ ”.

Plan du chapitre :

1. Définition d'un intervalle de confiance
2. Construction par l'inégalité de Bienaymé-Tchebychev
3. Intervalle de confiance pour la moyenne (loi normale, σ connu)
4. Intervalles de confiance asymptotiques et le théorème central limite

1 Intervalle de confiance : définition

1.1 Cadre général

On se place dans le cadre classique de l'estimation :

- On dispose d'un n -échantillon $(X_1, \dots, X_n)$ de loi dépendant d'un paramètre $\theta \in \Theta$.
- On souhaite estimer une fonction $g(\theta)$ du paramètre.
- On cherche deux statistiques $U_n = U_n(X_1, \dots, X_n)$ et $V_n = V_n(X_1, \dots, X_n)$ telles que $g(\theta)$ tombe dans $[U_n, V_n]$ avec une probabilité *au moins* $1 - \alpha$.

Définition : Intervalle de confiance

Soit $\alpha \in]0, 1[$. Soient U_n et V_n deux estimateurs de $g(\theta)$ tels que, pour tout $\theta \in \Theta$:

$$\mathbb{P}_\theta(g(\theta) \in [U_n, V_n]) \geq 1 - \alpha.$$

On dit que l'intervalle $[U_n, V_n]$ est un **intervalle de confiance** de $g(\theta)$ **au niveau de confiance** au moins $1 - \alpha$ (ou **avec un risque d'au plus α**).

1.2 Vocabulaire et conventions

Bien comprendre les termes

- α est le **risque** (ou **seuil de signification**) : c'est la probabilité maximale de "rater" le vrai paramètre.
- $1 - \alpha$ est le **niveau de confiance** : c'est la probabilité minimale de capturer $g(\theta)$.
- En pratique, on utilise souvent $\alpha = 0,05$ (confiance 95 %) ou $\alpha = 0,01$ (confiance 99 %).

△ Attention : Ce que ça ne veut pas dire

Un intervalle de confiance à 95 % ne signifie *pas* que " $g(\theta)$ a 95 % de chances d'être dans l'intervalle". Le paramètre $g(\theta)$ est **fixe** (mais inconnu). C'est l'**intervalle** $[U_n, V_n]$ qui est **aléatoire** : si on répétait l'expérience un grand nombre de fois, environ 95 % des intervalles construits contiendraient le vrai paramètre.

1.3 En pratique

Concrètement, on part d'un échantillon observé $(x_1, x_2, \dots, x_n)$ et on calcule :

$$u_n = U_n(x_1, \dots, x_n) \quad \text{et} \quad v_n = V_n(x_1, \dots, x_n).$$

L'intervalle *numérique* $[u_n, v_n]$ contient alors $g(\theta)$ avec une probabilité supérieure à $1 - \alpha$.

2 Construction par l'inégalité de Bienaymé-Tchebychev

2.1 Rappel de l'inégalité

Propriété : Inégalité de Bienaymé-Tchebychev

Soit T_n une variable aléatoire admettant une variance. Pour tout $\varepsilon > 0$:

$$\mathbb{P}_\theta(|T_n - \mathbb{E}_\theta(T_n)| \geq \varepsilon) \leq \frac{\mathbb{V}_\theta(T_n)}{\varepsilon^2}.$$

2.2 Construction de l'intervalle

Idée : on part de l'inégalité de Bienaymé-Tchebychev et on la "retourne" pour obtenir un encadrement de $g(\theta)$.

Étape 1 : En passant au complémentaire, on obtient :

$$\mathbb{P}_\theta(|T_n - \mathbb{E}_\theta(T_n)| < \varepsilon) \geq 1 - \frac{\mathbb{V}_\theta(T_n)}{\varepsilon^2}.$$

Étape 2 : Si T_n est un **estimateur sans biais** de $g(\theta)$, c'est-à-dire $\mathbb{E}_\theta(T_n) = g(\theta)$, on obtient :

$$\mathbb{P}_\theta(|T_n - g(\theta)| < \varepsilon) \geq 1 - \frac{\mathbb{V}_\theta(T_n)}{\varepsilon^2}.$$

Étape 3 : Si on peut trouver un entier n tel que $\frac{\mathbb{V}_\theta(T_n)}{\varepsilon^2} \leq \alpha$, alors :

$$\boxed{\mathbb{P}_\theta(T_n - \varepsilon \leq g(\theta) \leq T_n + \varepsilon) \geq 1 - \alpha.}$$

Méthode : Intervalle via Bienaymé-Tchebychev

Si T_n est un estimateur sans biais de $g(\theta)$, l'intervalle de confiance est :

$$[T_n - \varepsilon, T_n + \varepsilon] \quad \text{de longueur } 2\varepsilon,$$

avec ε déterminé par la condition $\frac{\mathbb{V}_\theta(T_n)}{\varepsilon^2} \leq \alpha$.

2.3 Exemple : estimation de p pour une loi de Bernoulli

Exemple : Sondage – Intervalle pour le paramètre p

Contexte. On réalise un sondage auprès de n personnes. Chaque personne répond oui (1) ou non (0) de manière indépendante, avec probabilité p de répondre oui. On note X_i la réponse de la i -ème personne.

Modèle. Chaque X_i suit une loi de Bernoulli $\mathcal{B}(p)$, et les X_i sont indépendantes. On a :

$$\mathbb{E}(X_i) = p \quad \text{et} \quad \mathbb{V}(X_i) = p(1 - p).$$

Estimateur naturel. On utilise la moyenne empirique :

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

C'est un estimateur sans biais de p , et sa variance vaut :

$$\mathbb{E}(\bar{X}_n) = p \quad \text{et} \quad \mathbb{V}(\bar{X}_n) = \frac{p(1 - p)}{n}.$$

Application de Bienaymé-Tchebychev. L'inégalité donne :

$$\mathbb{P}(|\bar{X}_n - p| \leq \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}.$$

L'astuce clé. Le problème est que $p(1-p)$ dépend de p , qui est inconnu ! On utilise alors le fait que :

$$0 \leq p(1-p) \leq \frac{1}{4} \quad \text{pour tout } p \in [0, 1],$$

ce qui donne la majoration :

$$\mathbb{P}(|\bar{X}_n - p| \leq \varepsilon) \geq 1 - \frac{1}{4n\varepsilon^2}.$$

Choix de ε . On veut que le membre de droite soit égal à $1 - \alpha$, donc :

$$\frac{1}{4n\varepsilon^2} = \alpha \quad \Longleftrightarrow \quad \varepsilon = \frac{1}{2\sqrt{n\alpha}}.$$

Conclusion. L'intervalle de confiance de p au niveau $1 - \alpha$ est :

$$\left[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}} \right].$$

Application numérique. Pour $\alpha = 0,05$ et $n = 100$:

$$\varepsilon = \frac{1}{2\sqrt{100 \times 0,05}} = \frac{1}{2\sqrt{5}} \approx 0,22.$$

On obtient l'intervalle $[\bar{X}_n - 0,22; \bar{X}_n + 0,22]$ avec un niveau de confiance de 95 %.

Remarque : cet intervalle est assez large (amplitude $2 \times 0,22 = 0,44$, soit une marge de $\pm 0,22$ autour de $\bar{X}_n$) ! C'est le prix à payer pour n'utiliser que l'inégalité de Bienaymé-Tchebychev, qui est très générale mais peu précise. On verra que le théorème central limite permet de faire beaucoup mieux.

3 Intervalle de confiance pour la moyenne d'une loi normale

3.1 Cadre et hypothèses

On se place maintenant dans un cadre plus précis : on connaît la loi des observations.

Hypothèses :

- $(X_1, X_2, \dots, X_n)$ est un n -échantillon issu d'une loi normale $\mathcal{N}(\mu, \sigma^2)$.
- L'écart type σ est **connu**.

— L'espérance μ est **inconnue** : c'est elle que l'on cherche à estimer.

3.2 La moyenne empirique et sa loi

La moyenne empirique est un estimateur **sans biais et convergent** de μ :

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Grâce à la **stabilité de la loi normale par combinaison linéaire**, on sait que :

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

En centrant et réduisant, on obtient la variable pivot :

$$Y_n = \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

3.3 Construction de l'intervalle

Soit $t_\alpha > 0$ le réel tel que $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$, où Φ est la fonction de répartition de $\mathcal{N}(0, 1)$.

Rappel : quantile de la loi normale

Le réel t_α vérifie :

$$\mathbb{P}(-t_\alpha \leq Y_n \leq t_\alpha) = \Phi(t_\alpha) - \Phi(-t_\alpha) = 2\Phi(t_\alpha) - 1 = 1 - \alpha.$$

Valeurs courantes :

Risque α	0,05	0,01	0,10
Confiance $1 - \alpha$	95 %	99 %	90 %
t_α	$\approx 1,96$	$\approx 2,58$	$\approx 1,645$

Dérivation pas à pas. On part de :

$$\mathbb{P}(-t_\alpha \leq Y_n \leq t_\alpha) = 1 - \alpha.$$

En remplaçant $Y_n = \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma}$ et en isolant μ :

$$\begin{aligned} -t_\alpha \leq \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma} \leq t_\alpha &\iff -t_\alpha \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq t_\alpha \frac{\sigma}{\sqrt{n}} \\ &\iff \bar{X}_n - t_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_\alpha \frac{\sigma}{\sqrt{n}}. \end{aligned}$$

Propriété : Intervalle de confiance – Loi normale, σ connu

Sous les hypothèses ci-dessus, l'intervalle de confiance de μ au niveau $1 - \alpha$ est :

$$\left[\bar{X}_n - t_\alpha \frac{\sigma}{\sqrt{n}}, \bar{X}_n + t_\alpha \frac{\sigma}{\sqrt{n}} \right]$$

où $t_\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

Exemple : Application numérique

On mesure la taille de $n = 36$ étudiants. On suppose que la taille suit une loi $\mathcal{N}(\mu, \sigma^2)$ avec $\sigma = 8$ cm (connu). On observe $\bar{x}_{36} = 172$ cm.

Pour un niveau de confiance de 95 % ($\alpha = 0,05$, $t_\alpha \approx 1,96$) :

$$\varepsilon = t_\alpha \frac{\sigma}{\sqrt{n}} = 1,96 \times \frac{8}{\sqrt{36}} = 1,96 \times \frac{8}{6} \approx 2,61 \text{ cm.}$$

L'intervalle de confiance est $[172 - 2,61; 172 + 2,61] = [169,39; 174,61]$.

Interprétation : on est confiant à 95 % que la taille moyenne des étudiants se situe entre 169,4 cm et 174,6 cm.

Observations importantes

- **Quand n augmente**, l'intervalle **rétrécit** (sa longueur est proportionnelle à $\frac{1}{\sqrt{n}}$). Plus on a de données, plus on est précis.
- **Quand on augmente la confiance** (diminuer α), l'intervalle **s'élargit** (t_α augmente). Plus on veut être sûr, moins on est précis.
- Il y a donc un **compromis** entre la confiance et la précision.

4 Intervalles de confiance asymptotiques

4.1 Motivation

Dans l'exemple précédent, on a eu la chance de connaître la loi *exacte* de $\bar{X}_n$. En pratique, on ne connaît pas toujours la loi exacte de l'estimateur. C'est là qu'intervient le **théorème**

central limite (TCL), qui donne une *approximation* de la loi pour n grand.

Le prix à payer : le résultat n'est qu'asymptotique, c'est-à-dire valable "quand n est assez grand".

4.2 Définition

Définition : Intervalle de confiance asymptotique

Soit $\alpha \in]0, 1[$. Soient U_n et V_n deux estimateurs de $g(\theta)$ tels que, pour tout $\theta \in \Theta$:

$$\mathbb{P}_\theta(g(\theta) \in [U_n, V_n]) \geq 1 - \alpha_n \quad \text{avec} \quad \alpha_n \xrightarrow{n \rightarrow +\infty} \alpha.$$

On dit que $[U_n, V_n]$ est un **intervalle de confiance asymptotique** de $g(\theta)$ avec un risque d'au plus α (ou au niveau de confiance au moins $1 - \alpha$).

Différence avec l'IC exact

Pour un intervalle de confiance **exact**, la probabilité $\mathbb{P}(g(\theta) \in [U_n, V_n]) \geq 1 - \alpha$ est vraie *pour tout* n .

Pour un IC **asymptotique**, la probabilité est $\geq 1 - \alpha_n$ avec $\alpha_n \rightarrow \alpha$: la garantie n'est parfaite que *quand* n est grand.

4.3 Exemple phare : intervalle asymptotique pour p (loi de Bernoulli)

Exemple : Intervalle asymptotique pour le paramètre p

On reprend le même modèle de sondage que précédemment : $(X_1, \dots, X_n)$ i.i.d. de loi $\mathcal{B}(p)$.

Étape 1 – Loi de la somme. Par indépendance :

$$n\bar{X}_n = \sum_{i=1}^n X_i \sim \mathcal{B}(n, p) \quad (\text{loi binomiale}).$$

Étape 2 – Application du TCL. Le théorème central limite donne :

$$X_n^* = \frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\sqrt{\mathbb{V}(\bar{X}_n)}} = \sqrt{n} \cdot \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Z \sim \mathcal{N}(0, 1).$$

Étape 3 – Encadrement. Soit $t_\alpha > 0$ tel que $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$. On a :

$$\mathbb{P}\left(-t_\alpha \leq \sqrt{n} \cdot \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \leq t_\alpha\right) \xrightarrow{n \rightarrow +\infty} \Phi(t_\alpha) - \Phi(-t_\alpha) = 1 - \alpha.$$

En isolant p , on obtient l'intervalle :

$$\left[\bar{X}_n - t_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, \bar{X}_n + t_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right].$$

Étape 4 – Éliminer p . Le problème est que les bornes dépendent encore de p (qu'on ne connaît pas). On utilise encore $p(1-p) \leq \frac{1}{4}$:

$$\left[\bar{X}_n - t_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, \bar{X}_n + t_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right] \subset \left[\bar{X}_n - \frac{t_\alpha}{2\sqrt{n}}, \bar{X}_n + \frac{t_\alpha}{2\sqrt{n}} \right].$$

Par croissance de la probabilité, on en déduit :

$$\mathbb{P}\left(\bar{X}_n - \frac{t_\alpha}{2\sqrt{n}} \leq p \leq \bar{X}_n + \frac{t_\alpha}{2\sqrt{n}}\right) \geq 1 - \alpha_n \quad \text{avec} \quad \alpha_n \rightarrow \alpha.$$

Conclusion. L'intervalle de confiance asymptotique de p au niveau $1 - \alpha$ est :

$$\left[\bar{X}_n - \frac{t_\alpha}{2\sqrt{n}}, \bar{X}_n + \frac{t_\alpha}{2\sqrt{n}} \right].$$

Application numérique. Pour $\alpha = 0,05$ ($t_\alpha \approx 1,96$) et $n = 1000$:

$$\varepsilon = \frac{1,96}{2\sqrt{1000}} = \frac{1,96}{63,25} \approx 0,031.$$

L'intervalle est $[\bar{X}_n - 0,031; \bar{X}_n + 0,031]$, soit une marge de $\pm 3,1$ points de pourcentage. C'est **bien plus précis** que l'intervalle par Bienaymé-Tchebychev !

5 Synthèse et comparaison des méthodes

	Bienaymé-Tchebychev	TCL (asymptotique)
Hypothèses	Très faibles (existence de la variance)	n assez grand
IC pour p (Bernoulli)	$\bar{X}_n \pm \frac{1}{2\sqrt{n\alpha}}$	$\bar{X}_n \pm \frac{t_\alpha}{2\sqrt{n}}$
Pour $n = 100$, $\alpha = 0,05$	$\pm 0,22$	$\pm 0,098$
Précision	Faible (intervalle large)	Bonne (intervalle étroit)
Validité	Exacte pour tout n	Approximative (bonne dès $n \geq 30$)

Morale

L'inégalité de Bienaymé-Tchebychev est un outil **universel** mais **peu précis**. Dès que n est assez grand, on préfère utiliser le TCL pour construire des intervalles **asymptotiques** beaucoup plus étroits. En pratique, c'est le TCL qui est systématiquement utilisé dans les sondages.

6 Formulaire récapitulatif

Propriété : Formules à retenir

1. Loi de Bernoulli $\mathcal{B}(p)$, méthode de Bienaymé-Tchebychev :

$$\text{IC de } p : \left[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}} \right] \quad (\text{exact, niveau } 1 - \alpha).$$

2. Loi normale $\mathcal{N}(\mu, \sigma^2)$, σ connu :

$$\text{IC de } \mu : \left[\bar{X}_n - t_\alpha \frac{\sigma}{\sqrt{n}}, \bar{X}_n + t_\alpha \frac{\sigma}{\sqrt{n}} \right] \quad (\text{exact, niveau } 1 - \alpha).$$

3. Loi de Bernoulli $\mathcal{B}(p)$, méthode du TCL :

$$\text{IC de } p : \left[\bar{X}_n - \frac{t_\alpha}{2\sqrt{n}}, \bar{X}_n + \frac{t_\alpha}{2\sqrt{n}} \right] \quad (\text{asymptotique, niveau } 1 - \alpha).$$

Quantiles usuels de $\mathcal{N}(0, 1)$: $t_{0,05} \approx 1,96$; $t_{0,01} \approx 2,58$; $t_{0,10} \approx 1,645$.