# Chapitre 4 - Intervalle de Confiance

Marouane Il Idrissi
il_idrissi.marouane@uqam.ca

**STT 1000 - Automne 2025**

*Département de Mathématiques, Université du Québec à Montréal*

UQÀM

# Intervalle de confiance

Dans le **Chapitre 3**, on cherchait a faire une **estimation ponctuelle** d'un paramètre $\theta$ en se basant sur un échantillon $\mathbf{X} = (X_1, \ldots, X_n)$

Après avoir trouvé **un bon estimateur** $\widehat{\theta}(\mathbf{X})$, et calculé **une estimation** $\widehat{\theta}(\mathbf{x})$ à l'aide d'un **échantillon observé** $\mathbf{x}$, **quelle confiance peut-on avoir en** $\widehat{\theta}(\mathbf{x})$?
On sait que $\widehat{\theta}(\mathbf{x}) \neq \theta$ à moins d'avoir **observé toute la population** (qui est supposée infinie)...

**Intuition:**
Construire un **intervalle**, qui contiendrait **la vraie valeur de** $\theta$ avec **une grande probabilité!**

Par exemple, si on accepte un **risque de se de se tromper de** 5% , est-ce qu'on peut **trouver un intervalle** $[a(\mathbf{X}), b(\mathbf{X})]$ **à partir de l'échantillon** tel que:

$$\mathbb{P}\left(a(\mathbf{X}) \leq \theta \leq b(\mathbf{X})\right) = 0.95$$

☞ **C'est exactement la définition d'un intervalle de confiance!**

# Plan du chapitre

# Sommaire

# Loi du Chi-Carré

Une v.a. de **loi du Chi-Carré** (ou Khi-Carré, ou $\chi^2$) est une v.a. **continue positive**

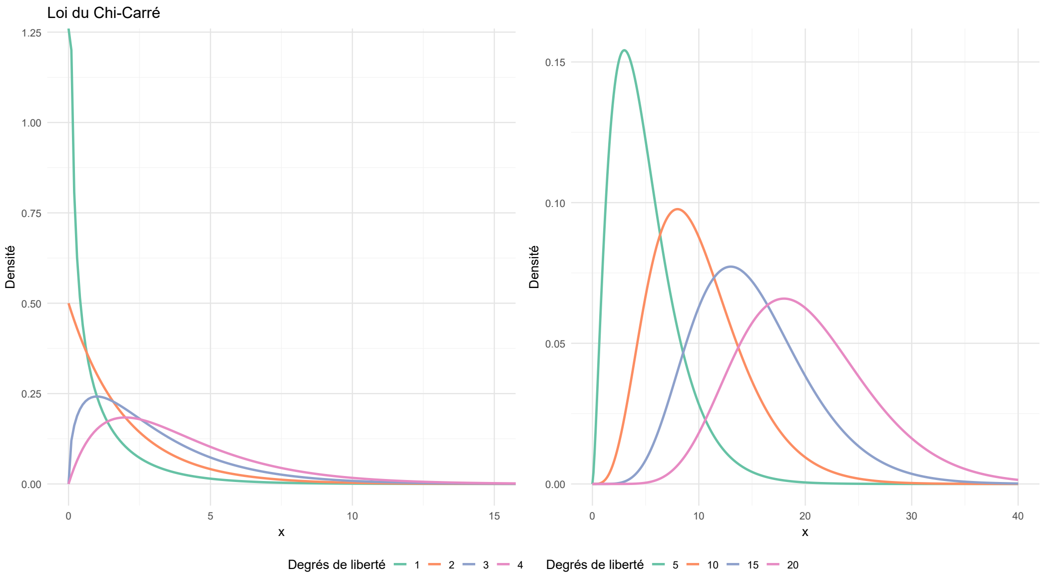
Elle est paramétrée par **un degré de liberté** $k \in \mathbb{N} \setminus \{0\}$

On note $X \sim \chi^2(k)$, et son support est $(0, \infty)$

**Caractéristiques** *(Loi du Chi-Carré)*. ★

- $f_X(t) = \frac{t^{k/2-1} \exp(-t/2)}{2^{k/2}\Gamma(k/2)} \mathbb{1}_{(0,\infty)}(t)$
- $\mathbb{E}[X] = k$
- $\mathbb{V}(X) = 2k$

# Loi du Chi-Carré - Illustration



# Loi du Chi-Carré - Propriétés

**Propriétés** *(Chi-Carré = Somme de carré de Gaussiennes standard).* ★

*Soient $Z_1, \ldots, Z_n$ des **variables aléatoires Gaussiennes standard et indépendantes**. Alors,*

$$V = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$$

☞ Si $Z \sim \mathcal{N}(0,1)$, quelle est la loi de $Z^2$? ✍

☞ En déduire la valeur de $\mathbb{E}\left[Z^4\right]$. ✍

**Propriétés** *(Somme de Chi-Carrés = Chi-Carré).* ★

*Soient $Y_1 \sim \chi^2(n_1)$ et $Y_2 \sim \chi^2(n_2)$ deux v.a.s **indépendantes**, alors*

$$V = Y_1 + Y_2 \sim \chi^2(n_1 + n_2).$$

# Loi de Student

Une v.a. de **loi de Student** est une v.a. **continue**

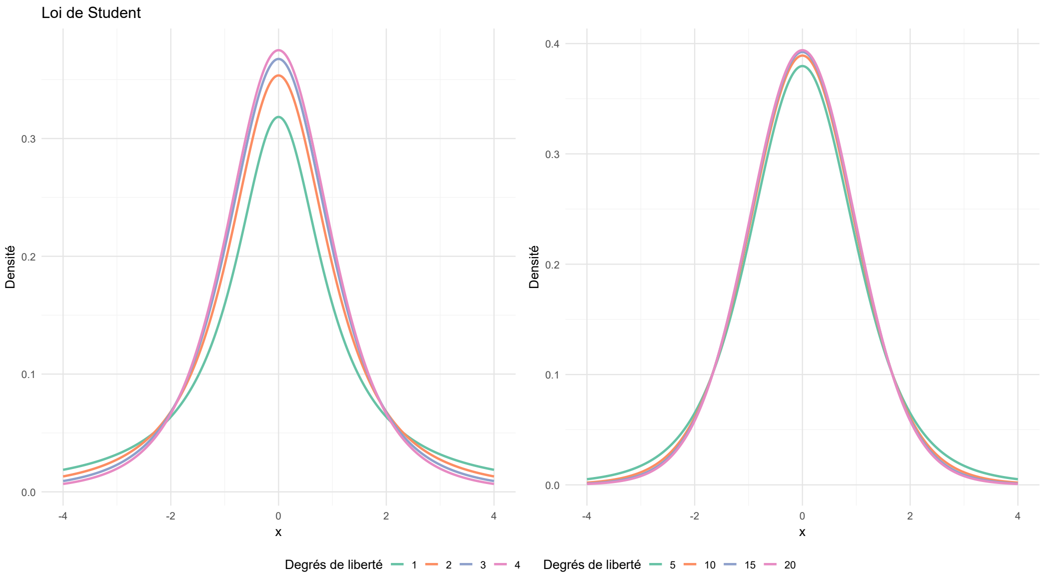
Elle est paramétrée par **un degré de liberté** $k \in \mathbb{N} \setminus \{0\}$

On note $X \sim \mathcal{T}(k)$ (des fois $t(k)$), et son support est $\mathbb{R}$

**Caractéristiques** *(Loi de Student)*. ★

- $f_X(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma(k/2)} \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}$
- $\mathbb{E}[X] = 0$ pour $k \geq 2$, n'existe pas sinon
- $\mathbb{V}(X) = \frac{k}{k-2}$ pour $k > 2$, n'existe pas sinon

# Loi de Student - Illustration



# Loi de Student - Propriétés

**Propriétés** *(Student = Gaussienne divisée par Chi-Carré).* ★

*Soit $Z \sim \mathcal{N}(0,1)$ et $V \sim \chi^2(k)$ deux v.a.s* ***indépendantes****. Alors:*

$$T = Z\sqrt{\frac{k}{V}} \sim \mathcal{T}(k)$$

**Propriétés** *(Comportement asymptotique de la loi de Student).* ★

*Soit $f_T(s;k)$ la densité de la loi de Student, et $\phi(s)$ la densité* ***normale standard****. Alors,*

$$f_T(s;k) \underset{k\to\infty}{\longrightarrow} \phi(s).$$

☞ **Plus le degré de liberté augmente, plus la loi de Student se rapproche d'une $\mathcal{N}(0,1)$.**

# Sommaire

# Intervalles de confiance

L'idée des **intervalles de confiance** (IC) est de **construire un intervalle aléatoire**, qui **contiendrait le paramètre que l'on souhaite estimer avec une grande probabilité**

Un *intervalle aléatoire*, c'est simplement un intervalle $[X, Y]$, ou $X$ et $Y$ sont deux v.a.s.

Cet intervalle aléatoire est construit à l'aide de **deux statistiques** $\widehat{\theta}_I(\mathbf{X})$ et $\widehat{\theta}_S(\mathbf{X})$

Une statistique, c'est **une fonction d'un échantillon**

**Définition** *(Intervalle de confiance).* ★

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un échantillon i.i.d. de loi populationnelle $f(t, \theta)$. Soit $\alpha \in (0, 1)$ un **niveau de risque**. Un **intervalle de confiance au niveau $1 - \alpha$ pour le paramètre $\theta$** est un intervalle aléatoire de la forme $\mathrm{IC}_\theta(\mathbf{X}) = \left[\widehat{\theta}_I(\mathbf{X}), \widehat{\theta}_S(\mathbf{X})\right]$, tel que:

$$\mathbb{P}\left(\theta \in \mathrm{IC}_\theta(\mathbf{X})\right) = \mathbb{P}\left(\widehat{\theta}_I(\mathbf{X}) \leq \theta \leq \widehat{\theta}_S(\mathbf{X})\right) = 1 - \alpha$$

# Intervalles de confiance

☞ Le **niveau de risque** $\alpha$ contrôle **le risque que** $\theta$ **ne soit pas dans cet intervalle**:

- Pour $\alpha \times 100$ % des **échantillons observés**, $\theta$ ne sera **pas dans cet intervalle**
- pour $(1 - \alpha) \times 100$ % des **échantillons observés**, $\theta$ sera effectivement **dans cet intervalle**

On dit qu'un **intervalle de confiance** pour $\theta$ est **informatif** si: ★

1. $\alpha$ est petit ;
2. La largeur de l'intervalle est petite.

**Par exemple:**

L'intervalle $[0, 1]$ pour le $p$ d'une Bernoulli **n'est pas informatif**:

1. $\alpha = 0$, donc **on est sûr que** $\theta$ est **toujours** dedans ;
2. La largeur de l'intervalle est **trop grande pour être utile**.

☞ Généralement, on **fixe le risque** $\alpha$ à $0.1$, $0.05$ ou $0.01$

Quels sont les **niveaux de confiance associés?** ✍

# Intervalles de confiance - Unilatéral, bilatéral

Il y a **trois types d'IC**: ★

- **Unilatéral à droite**, de la forme $\left[-\infty, \widehat{\theta}_S(\mathbf{X})\right]$, tel que: $\mathbb{P}\left(\theta \leq \widehat{\theta}_S(\mathbf{X})\right) = 1 - \alpha$
- **Unilatéral à gauche**, de la forme $\left[\widehat{\theta}_I(\mathbf{X}), \infty\right]$, tel que: $\mathbb{P}\left(\widehat{\theta}_I(\mathbf{X}) \leq \theta\right) = 1 - \alpha$
- **Bilatéral**, de la forme $\left[\widehat{\theta}_I(\mathbf{X}), \widehat{\theta}_S(\mathbf{X})\right]$, tel que: $\mathbb{P}\left(\widehat{\theta}_I(\mathbf{X}) \leq \theta \leq \widehat{\theta}_S(\mathbf{X})\right) = 1 - \alpha$

☞ Un **intervalle de confiance** bilatéral n'est **pas nécessairement symétrique autour de** $\theta$!

Pourquoi utiliser chaque type d'IC: ★

- **Unilatéral à droite** : **Borne supérieure** pour $\theta$
- **Unilatéral à gauche** : **Borne inférieure** pour $\theta$
- **Bilatéral** : **Encadrer** la valeur précise de $\theta$

# Intervalles de confiance - Recette pour construire un IC exact

Pour construire un IC pour $\theta$ inconnu à partir d'un échantillon $\mathbf{X} = (X_1, \ldots, X_n)$:

1. **Proposer un estimateur** $\widehat{\theta}(\mathbf{X})$ de $\theta$
2. **Déterminer la loi exacte de** $\widehat{\theta}(\mathbf{X})$
3. Construire un **pivot** $\Delta$ entre $\widehat{\theta}(\mathbf{X})$ et $\theta$, par exemple:

$$\Delta = \widehat{\theta}(\mathbf{X}) - \theta \quad \text{ou} \quad \Delta = \widehat{\theta}(\mathbf{X})/\theta.$$

   **Un pivot $\Delta$ est une v.a. dont la loi ne dépend pas de $\theta$!!**

   ☞ Ce dernier point est **crucial** pour la construction d'un IC

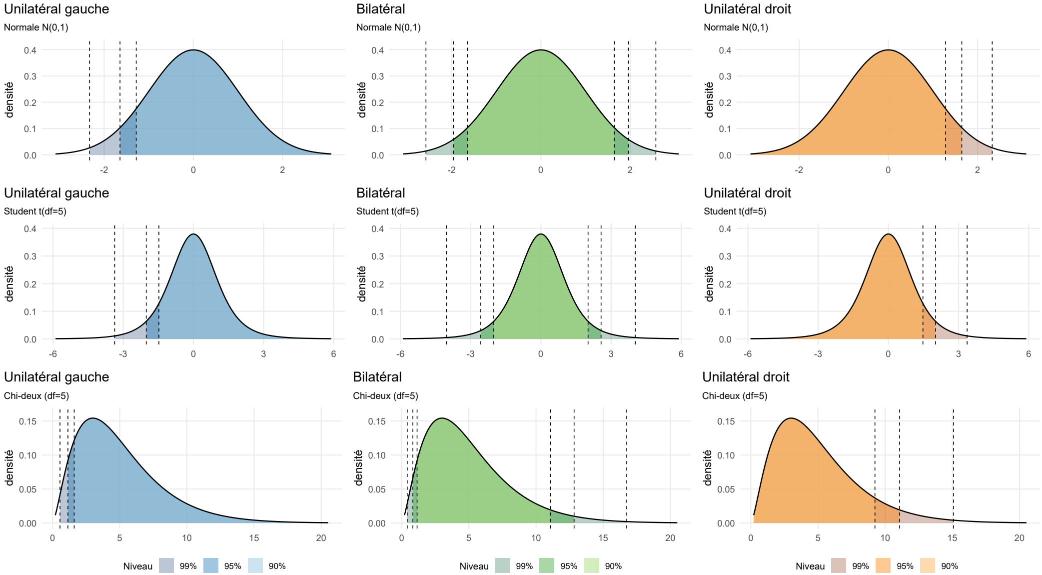
4. Choisir un **niveau de risque** $\alpha$, et donc un **niveau de confiance** $1-\alpha$. Se baser sur la **loi exacte** de $\Delta$ (table de probabilités) pour trouver **deux nombres** $d_I$ et $d_S$ tels que

$$\mathbb{P}\left(d_I \leq \Delta \leq d_s\right) = 1 - \alpha$$

5. Utiliser la définition de $\Delta$, et les quantités $d_I$ et $d_S$ pour **isoler** $\theta$ dans le calcul de probabilité au-dessus, par exemple pour $\Delta = \widehat{\theta}(\mathbf{X}) - \theta$:

$$\mathbb{P}\left(d_I \leq \widehat{\theta}(\mathbf{X}) - \theta \leq d_S\right) = 1 - \alpha, \quad \Longleftrightarrow \quad \mathbb{P}\left(\widehat{\theta}(\mathbf{X}) - d_S \leq \theta \leq \widehat{\theta}(\mathbf{X}) - d_I\right) = 1 - \alpha,$$

# Intervalles de confiance - Illustration, loi du pivot



# Échantillon Gaussien - IC bilatéral pour $\mu$, variance connue

On a $X_1, \dots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu, \sigma_*^2)$, avec $\mu$ **inconnu**, **mais $\sigma_*^2$ est connu**
(Par exemple, on sait que $\sigma_*^2 = 2$).

On veut construire un IC bilatéral pour $\theta = \mu$.

1. Proposez un estimateur pour $\mu$ ✍
2. Quelle est la loi de cet estimateur? ✍
3. Trouvez un pivot tel quel $\Delta \sim \mathcal{N}(0, 1)$ **(qui ne dépend pas de $\mu$)**✍
4. Quels deux nombres $d_I$ et $d_S$ permettent d'avoir
$$\mathbb{P}\left(d_I \leq \Delta \leq d_S\right) = 1 - \alpha \text{ ?}$$ ✍
5. Proposez un intervalle de confiance $[\widehat{\mu}_I(\mathbf{X}), \widehat{\mu}_S(\mathbf{X})]$ pour $\mu$ au niveau $1 - \alpha$. ✍
6. Supposons que $n = 8$, $\alpha = 0.05$, $\sigma_*^2 = 2$ et que nous ayons comme échantillon observé
$$\mathbf{x} = (45.44, 45.68, 45.65, 44.92, 45.88, 48.61, 43.94, 43.75), \quad \text{et donc} \quad \overline{x}_n = 45.48.$$
Calculez l'intervalle observé $[\widehat{\mu}_I(\mathbf{x}), \widehat{\mu}_S(\mathbf{x})]$ et interprétez-le. ✍

# Échantillon Gaussien - IC unilatéral à droite pour $\mu$, variance connue

On a $X_1, \ldots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu, \sigma_*^2)$, avec $\mu$ **inconnu**, **mais $\sigma_*^2$ est connu**
(Par exemple, on sait que $\sigma_*^2 = 2$).

On veut construire un IC unilatéral à droite pour $\theta = \mu$.

1. **Proposez un estimateur pour $\mu$** ✍
2. **Quelle est la loi de cet estimateur?** ✍
3. **Trouvez un pivot tel quel $\Delta \sim \mathcal{N}(0,1)$ (qui ne dépend pas de $\mu$)**✍
4. Quels nombre $d_I$ permet d'avoir
$$\mathbb{P}(d_I \leq \Delta) = 1 - \alpha \,? \text{ ✍}$$
5. Proposez un intervalle de confiance $(-\infty, \widehat{\mu}_S(\mathbf{X})]$ pour $\mu$ au niveau $1 - \alpha$. ✍
6. Calculez l'intervalle observé $(-\infty, \widehat{\mu}_S(\mathbf{x})]$ basé sur les données précédentes et interprétez-le. ✍

# Échantillon Gaussien - IC unilatéral à gauche pour $\mu$, variance connue

On a $X_1, \ldots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu, \sigma_*^2)$, avec $\mu$ **inconnu**, **mais $\sigma_*^2$ est connu**
(Par exemple, on sait que $\sigma_*^2 = 2$).

On veut construire un IC unilatéral à gauche pour $\theta = \mu$.

1. **Proposez un estimateur pour $\mu$** ✍
2. **Quelle est la loi de cet estimateur?** ✍
3. **Trouvez un pivot tel quel $\Delta \sim \mathcal{N}(0,1)$ (qui ne dépend pas de $\mu$)**✍
4. Quels nombre $d_S$ permet d'avoir

$$\mathbb{P}(\Delta \leq d_S) = 1 - \alpha \, ? \text{ ✍}$$

5. Proposez un intervalle de confiance $[\widehat{\mu}_I(\mathbf{X}), \infty)$ pour $\mu$ au niveau $1 - \alpha$. ✍
6. Calculez l'intervalle observé $[\widehat{\mu}_I(\mathbf{x}), \infty)$ basé sur les données précédentes et interprétez-le. ✍

# Échantillon Gaussien - IC bilatéral pour $\mu$, variance inconnue

On a $X_1, \ldots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu, \sigma_*^2)$, avec $\mu$ et $\sigma^2$ **inconnus**.
On veut construire un IC bilatéral pour $\theta = \mu$.

**Idée:** Remplacer $\sigma^2$ par **un estimateur**!

1. Proposez un estimateur pour $\mu$ et un estimateur pour $\sigma^2$. ✍
2. Quelles sont les lois de ces estimateur? ✍
3. En utilisant le résultat suivant (non démontré):

   > **Lemme** *(Lemme de Basu).* ★
   > *Soit* $\mathbf{X}$ *un* ***échantillon Gaussien i.i.d****. Alors,* $\overline{X}_n \perp\!\!\!\perp S^2$.

   Trouvez un pivot tel quel $\Delta \sim \mathcal{T}(n-1)$. ✍
4. Quels deux nombres $d_I$ et $d_S$ permettent d'avoir

$$\mathbb{P}\left(d_I \leq \Delta \leq d_S\right) = 1 - \alpha \,? \text{ ✍}$$

5. Proposez un intervalle de confiance $[\widehat{\mu}_I(\mathbf{X}), \widehat{\mu}_S(\mathbf{X})]$ pour $\mu$ au niveau $1 - \alpha$. ✍

## Échantillon Gaussien - IC unilatéral à droite pour $\mu$, variance inconnue

On a $X_1, \dots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu, \sigma_*^2)$, avec $\mu$ et $\sigma^2$ **inconnus**.

On veut construire un IC unilatéral à droite pour $\theta = \mu$.

1. **Proposez un estimateur pour $\mu$ et un estimateur pour $\sigma^2$.** ✍
2. **Quelles sont les lois de ces estimateur?** ✍
3. **Trouvez un pivot tel quel $\Delta \sim \mathcal{T}(n-1)$.** ✍
4. Quel nombre $d_l$ permet d'avoir
$$\mathbb{P}\left(d_l \leq \Delta\right) = 1 - \alpha \ ?$$ ✍
5. Proposez un intervalle de confiance $(\infty, \widehat{\mu}_S(\mathbf{X})]$ pour $\mu$ au niveau $1-\alpha$. ✍

# Échantillon Gaussien - IC unilatéral à gauche pour $\mu$, variance inconnue

On a $X_1, \ldots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu, \sigma_*^2)$, avec $\mu$ et $\sigma^2$ **inconnus**.

On veut construire un IC unilatéral à gauche pour $\theta = \mu$.

1. **Proposez un estimateur pour $\mu$ et un estimateur pour $\sigma^2$.** ✍
2. **Quelles sont les lois de ces estimateur?** ✍
3. **Trouvez un pivot tel quel $\Delta \sim \mathcal{T}(n-1)$.** ✍
4. Quel nombre $d_S$ permettent d'avoir

$$\mathbb{P}(\Delta \leq d_S) = 1 - \alpha \text{ ? } ✍$$

5. Proposez un intervalle de confiance $[\widehat{\mu}_I(\mathbf{X}), \infty)$ pour $\mu$ au niveau $1-\alpha$. ✍

# Échantillon Gaussien - IC bilatéral pour $\sigma^2$, moyenne connue

On a $X_1, \ldots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu_*, \sigma^2)$, avec $\mu_*$ **connu**.

On cherche à construire un IC bilatéral pour le paramètre $\theta = \sigma^2$.

1. Proposez un estimateur pour $\sigma^2$.✍
2. Quelle est la loi de cet estimateur? ✍
3. Trouvez un pivot tel quel $\Delta \sim \chi^2(n)$.✍
4. Quels deux nombres $d_I$ et $d_S$ permettent d'avoir

$$\mathbb{P}\left(d_I \leq \Delta \leq d_S\right) = 1 - \alpha \,? \text{ ✍}$$

5. Proposez un intervalle de confiance $\left[\widehat{\sigma^2}_I(\mathbf{X}), \widehat{\sigma^2}_S(\mathbf{X})\right]$ pour $\sigma^2$ au niveau $1 - \alpha$. ✍

# Échantillon Gaussien - IC unilatéral à droite pour $\sigma^2$, moyenne connue

On a $X_1, \ldots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu_*, \sigma^2)$, avec $\mu_*$ **connu**.

On cherche à construire un IC unilatéral à droite pour le paramètre $\theta = \sigma^2$.

1. **Proposez un estimateur pour $\sigma^2$.** ✍
2. **Quelle est la loi de cet estimateur?** ✍
3. **Trouvez un pivot tel quel $\Delta \sim \chi^2(n)$.** ✍
4. Quel nombre $d_I$ permettent d'avoir

$$\mathbb{P}(\Delta \geq d_I) = 1 - \alpha \text{ ?}$$ ✍

5. Proposez un intervalle de confiance $\left[\widehat{\sigma^2}_I(\mathbf{X}), \infty\right)$ pour $\mu$ au niveau $1 - \alpha$. ✍

# Échantillon Gaussien - IC unilatéral à gauche pour $\sigma^2$, moyenne connue

On a $X_1, \ldots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu_*, \sigma^2)$, avec $\mu_*$ **connu**.

On cherche à construire un IC unilatéral à gauche pour le paramètre $\theta = \sigma^2$.

1. **Proposez un estimateur pour $\sigma^2$.** ✍
2. **Quelle est la loi de cet estimateur?** ✍
3. **Trouvez un pivot tel quel $\Delta \sim \chi^2(n)$.** ✍
4. Quel nombre $d_I$ permettent d'avoir
$$\mathbb{P}(\Delta \leq d_S) = 1 - \alpha \, ? \text{ ✍}$$
5. Proposez un intervalle de confiance $\left(\infty, \widehat{\sigma^2}_S(\mathbf{X})\right]$ pour $\mu$ au niveau $1 - \alpha$. ✍

# Échantillon Gaussien - IC bilatéral pour $\sigma^2$, moyenne inconnue

On a $X_1, \ldots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu, \sigma^2)$, avec $\mu$ et $\sigma^2$ **inconnus**.

On cherche à construire un IC bilatéral pour le paramètre $\theta = \sigma^2$.

**Idée:** Remplacer $\mu$ par **un estimateur!**

1. Proposez un estimateur pour $\mu$ et $\sigma^2$. ✍
2. Quelle est la loi de cet estimateur? ✍
3. Trouvez un pivot tel quel $\Delta \sim \chi^2(n-1)$. ✍
4. Quels deux nombres $d_I$ et $d_S$ permettent d'avoir

$$\mathbb{P}\left(d_I \leq \Delta \leq d_S\right) = 1 - \alpha \; ? \text{ ✍}$$

5. Proposez un intervalle de confiance $\left[\widehat{\sigma^2}_I(\mathbf{X}), \widehat{\sigma^2}_S(\mathbf{X})\right]$ pour $\sigma^2$ au niveau $1 - \alpha$. ✍

# Échantillon Gaussien - IC unilatéral à droite pour $\sigma^2$, moyenne inconnue

On a $X_1, \ldots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu, \sigma^2)$, avec $\mu$ et $\sigma^2$ **inconnus**.

On cherche à construire un IC bilatéral pour le paramètre $\theta = \sigma^2$.

**Idée:** Remplacer $\mu$ par **un estimateur!**

1. **Proposez un estimateur pour $\mu$ et un estimateur pour $\sigma^2$.** ✍
2. **Quelle est la loi de cet estimateur?** ✍
3. **Trouvez un pivot tel quel $\Delta \sim \chi^2(n-1)$.** ✍
4. Quel nombre $d_I$ permettent d'avoir

$$\mathbb{P}(\Delta \leq d_S) = 1 - \alpha \text{ ?}$$ ✍

5. Proposez un intervalle de confiance $\left[\widehat{\sigma^2}_I(\mathbf{X}), \infty\right)$ pour $\sigma^2$ au niveau $1 - \alpha$. ✍

# Échantillon Gaussien - IC unilatéral à gauche pour $\sigma^2$, moyenne inconnue

On a $X_1, \ldots, X_n$ i.i.d. de loi populationnelle $\mathcal{N}(\mu, \sigma^2)$, avec $\mu$ et $\sigma^2$ **inconnus**.

On cherche à construire un IC bilatéral pour le paramètre $\theta = \sigma^2$.

**Idée:** Remplacer $\mu$ par **un estimateur!**

1. **Proposez un estimateur pour $\mu$ et un estimateur pour $\sigma^2$.** ✍
2. **Quelle est la loi de cet estimateur?** ✍
3. **Trouvez un pivot tel quel $\Delta \sim \chi^2(n-1)$.** ✍
4. Quel nombre $d_I$ permettent d'avoir
$$\mathbb{P}\left(d_I \leq \Delta\right) = 1 - \alpha \ ? \ ✍$$
5. Proposez un intervalle de confiance $\left(\infty, \widehat{\sigma^2}_S(\mathbf{X})\right]$ pour $\sigma^2$ au niveau $1 - \alpha$. ✍

# Échantillon Gaussien - Récapitulatif des ICs

| Paramètre et type | Situation | $\widehat{\theta}(\mathbf{X})$ | Pivot $\Delta$ | IC exact |
|---|---|---|---|---|
| Bilatéral pour $\mu$ | $\sigma^2$ connu | $\overline{X}_n$ | $\frac{\overline{X}_n-\mu}{\sigma} \sim \mathcal{N}(0,1)$ | $\left[\overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\ ;\ \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$ |
| Uni. droite pour $\mu$ | $\sigma^2$ connu | $\overline{X}_n$ | $\frac{\overline{X}_n-\mu}{\sigma} \sim \mathcal{N}(0,1)$ | $\left(-\infty\ ;\ \overline{X}_n + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right]$ |
| Uni. gauche pour $\mu$ | $\sigma^2$ connu | $\overline{X}_n$ | $\frac{\overline{X}_n-\mu}{\sigma} \sim \mathcal{N}(0,1)$ | $\left[\overline{X}_n - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\ ;\ \infty\right)$ |
| Bilatéral pour $\mu$ | $\sigma^2$ inconnu | $\overline{X}_n,\ S^2$ | $\frac{\overline{X}_n-\mu}{S/\sqrt{n}} \sim \mathcal{T}(n-1)$ | $\left[\overline{X}_n - t_{1-\alpha/2,n-1}\frac{S}{\sqrt{n}}\ ;\ \overline{X}_n + t_{1-\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$ |
| Uni. droite pour $\mu$ | $\sigma^2$ inconnu | $\overline{X}_n,\ S^2$ | $\frac{\overline{X}_n-\mu}{S/\sqrt{n}} \sim \mathcal{T}(n-1)$ | $\left(-\infty\ ;\ \overline{X}_n + t_{1-\alpha,n-1}\frac{S}{\sqrt{n}}\right]$ |
| Uni. gauche pour $\mu$ | $\sigma^2$ inconnu | $\overline{X}_n,\ S^2$ | $\frac{\overline{X}_n-\mu}{S/\sqrt{n}} \sim \mathcal{T}(n-1)$ | $\left[\overline{X}_n - t_{1-\alpha,n-1}\frac{S}{\sqrt{n}}\ ;\ \infty\right]$ |
| Bilatéral pour $\sigma^2$ | $\mu$ connu | $S_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2$ | $\frac{nS_0^2}{\sigma^2} \sim \chi^2(n)$ | $\left[\frac{nS_0^2}{c_{1-\alpha/2,n}}\ ;\ \frac{nS_0^2}{c_{\alpha/2,n}}\right]$ |
| Uni. droite pour $\sigma^2$ | $\mu$ connu | $S_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2$ | $\frac{nS_0^2}{\sigma^2} \sim \chi^2(n)$ | $\left(0\ ;\ \frac{nS_0^2}{c_{\alpha,n}}\right]$ |
| Uni. gauche pour $\sigma^2$ | $\mu$ connu | $S_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2$ | $\frac{nS_0^2}{\sigma^2} \sim \chi^2(n)$ | $\left[\frac{nS_0^2}{c_{1-\alpha,n}}\ ;\ \infty\right)$ |
| Bilatéral pour $\sigma^2$ | $\mu$ inconnu | $\overline{X}_n,\ S^2$ | $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ | $\left[\frac{(n-1)S^2}{c_{1-\alpha/2,n-1}}\ ;\ \frac{(n-1)S_0^2}{c_{\alpha/2,n-1}}\right]$ |
| Uni. droite pour $\sigma^2$ | $\mu$ inconnu | $\overline{X}_n,\ S^2$ | $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ | $\left(0\ ;\ \frac{(n-1)S^2}{c_{\alpha,n-1}}\right]$ |
| Uni. gauche pour $\sigma^2$ | $\mu$ inconnu | $\overline{X}_n,\ S^2$ | $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ | $\left[\frac{(n-1)S^2}{c_{1-\alpha,n-1}}\ ;\ \infty\right)$ |

# Intervalle de confiance - Illustration

**Exemple** *(Contrôleur de paquets cafés)*. Vous êtes contrôleur qualité dans une compagnie qui vend du café. Pour assurer la conformité des paquets de cafés, ils doivent en moyenne peser **500g**, avec une variabilité acceptée de **3**g². On suppose que les masses (en g) suivent une loi Normale. Vous prélevez un échantillon de taille 10 :

$$498.7,\ 501.2,\ 499.8,\ 500.5,\ 497.9,\ 502.1,\ 499.3,\ 500.8,\ 501.0,\ 498.9$$

On observe $\overline{x}_n = 500.2$, $s^2 = 1.73$, $s = 1.32$, $s_0^2 = 1.56$

☞ Quelle est la loi de la population? ✍

☞ Calculez les ICs bilatéraux aux niveaux 90%, 95% et 99% pour $\mu$ et $\sigma^2$, dans toutes les configurations possibles. ✍

☞ Peut-on conclure que la production de café est conforme à la qualité attendue? ✍

# Intervalle de confiance - Interprétation

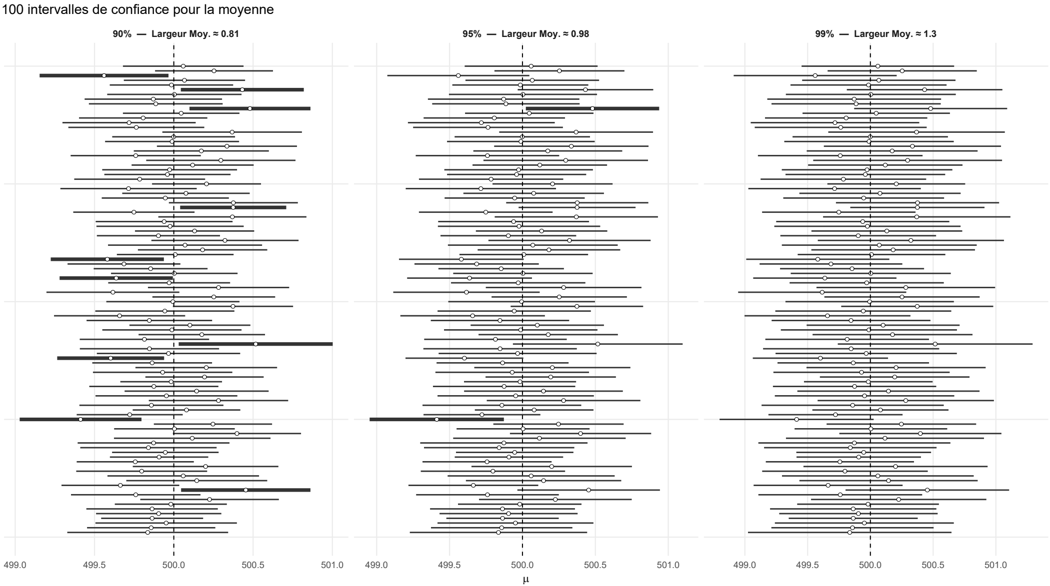
Le niveau de confiance **ne garantie pas que $\theta$ soit à l'intérieur de l'intervalle!**

Interprétation d'un **intervalle de confiance** $\mathrm{IC}_\theta(\mathbf{X})$ au niveau $(1-\alpha)$:

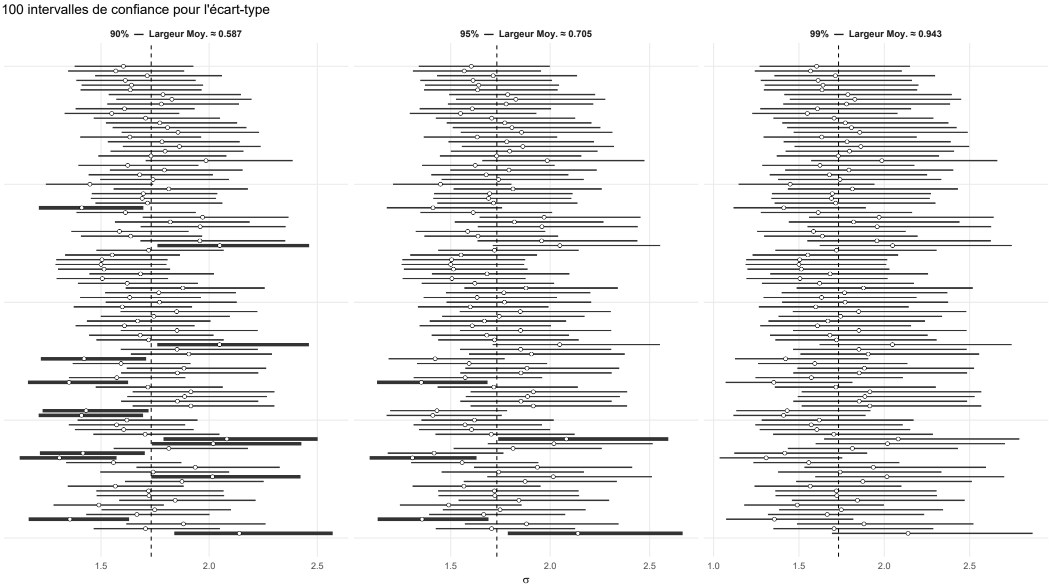
- On suppose que l'on peut simuler **beaucoup d'échantillons <u>observés</u> de même taille** $n$
- A partir de chacun échantillon, on **calcule un IC au niveau** $1-\alpha\%$
- Alors, $\theta$ **sera dans approximativement** $1-\alpha\%$ **de ces échantillons**
- **Si on augmente le nombre d'échantillons simulés, alors la proportion d'IC qui contiennent $\theta$ va tendre vers $1-\alpha\%$**

☞ **Pour un seul échantillon observé et un seul IC, on ne sait pas si $\theta$ appartient effectivement à cet échantillon**

# Intervalle de confiance - Interprétation



# Intervalle de confiance - Interprétation



# Sommaire

# Intervalle de confiance asymptotiques

Jusqu'à présent, on a **supposé que les échantillons étaient Gaussiens**
☞ C'était bien pratique pour **déterminer la loi du pivot**

Mais c'est un cadre de travail très limitant
On ne peut pas modéliser **TOUS** les phénomènes par les v.a. Gaussiennes

Par exemple, si on s'intéresse **un échantillon de durées de vie d'ampoules**
☞ C'est un phénomène qui se modélise généralement avec une loi exponentielle.

**Idée:** **Utiliser le Théorème de la Limite Centrale** pour approcher la loi du pivot
C'est la base des **ICs asymptotiques**

# Théorème de la Limite Centrale

**Théorème** *(Théorème de la limite centrale (TLC)).* ★

*Soient $X_1, \ldots, X_n$ des v.a.s* ***i) indépendantes**, **ii) ayant toutes la même loi iii) avec*** $\mathbb{E}[X_i] = \mu \in \mathbb{R}$***, et*** $\mathbb{V}(X_i) = \sigma^2 < \infty$.

*Alors la fonction (de n) suivante admet comme limite*

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) \underset{n \to \infty}{\longrightarrow} Z \sim \mathcal{N}(0, 1)$$

**Théorème** *(Loi asymptotique du pivot pour $\mu$).* ★

*Dans le cadre du TLC, on a que:*

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{S}\right) \underset{n \to \infty}{\longrightarrow} Z \sim \mathcal{N}(0, 1)$$

☞ **Quelle est la différence entre ces deux résultats?** ✍

# Intervalles de confiance asymptotique pour $\mu$ - Recette

Pour construire un IC **asymptotique** pour $\mathbb{E}[X_1] = \mu$ à partir d'un échantillon $\mathbf{X} = (X_1, \ldots, X_n)$:

1. **Utiliser l'estimateur** $\overline{X}_n$ pour $\mu$
2. Utiliser le **pivot**:

$$\Delta = \sqrt{n}\left(\frac{\overline{X}_n - \mu}{S}\right)$$

3. **Vérifier que l'on est dans un régime asymptotique:** $n > 30$
   ☞ **Si (et seulement si) c'est le cas, supposer l'approximation suivante:**

$$\Delta = \sqrt{n}\left(\frac{\overline{X}_n - \mu}{S}\right) \sim \mathcal{N}(0, 1)$$

4. Choisir un **niveau de risque** $\alpha$, et donc un **niveau de confiance** $1 - \alpha$. Se baser sur la **loi asymptotique** de $\Delta$ (table de probabilités) qui donne, pour un IC bilatéral:

$$\mathbb{P}\left(-z_{1-\alpha/2} \leq \sqrt{n}\left(\frac{\overline{X}_n - \mu}{S}\right) \leq z_{1-\alpha/2}\right) = 1 - \alpha$$

5. Utiliser la définition de $\Delta$ pour **isoler** $\mu$ dans le calcul de probabilité au-dessus.

# IC pour l'espérance - Échantillon quelconque

On a $X_1, \ldots, X_n$ i.i.d. de **loi populationnelle inconnue**, avec:

- $\mu = \mathbb{E}[X_1] < \infty$
- $\mathbb{E}\left[X_1^2\right] < \infty$
- $n > 30$

1. Quelle est la loi asymptotique du pivot $\Delta = \sqrt{n}\left(\frac{\overline{X}_n - \mu}{S}\right)$? ✍
2. Sommes-nous dans un régime asymptotique? ✍
3. Quelle approximation pouvons-nous accepter? ✍
4. Sous cette approximation, quels nombres $d_I$ et $d_S$ permettent d'avoir:

   $$\mathbb{P}\left(d_I \leq \Delta \leq d_S\right) = 1 - \alpha?$$

   ✍
5. Proposez un IC asymptotique bilatéral $[\widehat{\mu}_S(\mathbf{X}, \widehat{\mu}_S(\mathbf{X})]$ pour $\mu$ au niveau $1 - \alpha$. ✍
6. Proposez des ICs asymptotiques unilatéraux pour $\mu$ au niveau $1 - \alpha$. ✍

# ICs asymptotiques pour l'espérance - Formules générales

| Paramètre et type | Condition | $\widehat{\theta}(\mathbf{X})$ | Pivot $\Delta$ | Loi asympto. | IC asympto |
|---|---|---|---|---|---|
| Bilatéral pour $\mu$ | $n > 30$ + TLC | $\overline{X}_n$ | $\sqrt{n}\left(\frac{\overline{X}_n - \mu}{S}\right)$ | $\mathcal{N}(0,1)$ | $\left[\overline{X}_n - z_{1-\alpha/2}\frac{S}{\sqrt{n}}\ ;\ \overline{X}_n + z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$ |
| Uni. droite pour $\mu$ | $n > 30$ + TLC | $\overline{X}_n$ | $\sqrt{n}\left(\frac{\overline{X}_n - \mu}{S}\right)$ | $\mathcal{N}(0,1)$ | $\left(-\infty\ ;\ \overline{X}_n + z_{1-\alpha}\frac{S}{\sqrt{n}}\right]$ |
| Uni. gauche pour $\mu$ | $n > 30$ + TLC | $\overline{X}_n$ | $\sqrt{n}\left(\frac{\overline{X}_n - \mu}{S}\right)$ | $\mathcal{N}(0,1)$ | $\left[\overline{X}_n - z_{1-\alpha}\frac{S}{\sqrt{n}}\ ;\ \infty\right)$ |

# Intervalle de confiance asymptotique pour $\mu$ - Illustration

**Exemple** *(Contrôleur d'ampoules)*. Vous êtes contrôleur qualité dans une compagnie qui produit des ampoules. Pour assurer la qualité des produits, elles doivent en moyenne durer **20000h**. On ne connaît pas la loi de la durée de vie de ces ampoules, mais on sait que son espérance et sa variance existent. Vous prélevez un échantillon de taille 50 avec:

$$\overline{x}_n = 20026.94, \quad s = 164.49$$

☞ Sommes-nous dans un régime asymptotique? Si oui, quel est la loi asymptotique du pivot? ✍

☞ Calculez les ICs asymptotiques bilatéraux et unilatéraux aux niveaux 90%, 95% et 99% pour $\mu$. ✍

☞ Peut-on conclure que la durée de vie des ampoules est conforme? ✍

# Sommaire

# Deux échantillons Gaussiens

Jusqu'à présent, nous n'avions accès **qu'à un seul échantillon i.i.d.**
Ça nous a permis d'inférer de l'information sur les paramètres de la **loi de la population**

Comment faire si l'on a **deux populations différentes** que l'on souhaite comparer?

**Exemple** *(Méthodes de manufacture)*. Vous êtes en charge d'une usine de construction de vélo. Vous voulez comparer **deux méthodes de fabrication**. On modélise le **temps de fabrication par la méthode A** par une Gaussienne d'espérance $\mu_X$, et le **temps de fabrication par la méthode B** par une Gaussienne d'espérance $\mu_Y$.

On cherche **une estimation par IC** du gain de productivité moyen $\mu_X - \mu_Y$.

Objectif de cette section:
**Construire des outils pour répondre à ce genre de problématique dans le cas Gaussien**

# Deux configurations

On a accès à **deux échantillons** issus de **deux populations différentes**:

- $\mathbf{X} = (X_1, \dots, X_{n_X})$ i.i.d. de taille $n_X$ et de loi populationnelle $\mathcal{N}(\mu_X, \sigma_X^2)$
- $\mathbf{Y} = (Y_1, \dots, Y_{n_Y})$ i.i.d. de taille $n_Y$ et de loi populationnelle $\mathcal{N}(\mu_Y, \sigma_Y^2)$

On dit que:

- $\mathbf{X}$ et $\mathbf{Y}$ sont **indépendants** si, pour $i = 1, \dots, n_X$ et $j = 1, \dots, n_Y$: $X_i \perp\!\!\!\perp Y_j$ ★
- $\mathbf{X}$ et $\mathbf{Y}$ sont **appariés** si **les individus forment des paires**: ★
  - Les deux échantillons ont la même taille: $n_X = n_Y = n$
  - $X_i$ et $Y_i$ proviennent du **même individu** et forment une **paire** $(X_i, Y_i)$
  - Les paires sont **indépendantes entre elles**: $(X_i, Y_i) \perp\!\!\!\perp (X_j, Y_j)$ pour $i \neq j$.

**Exemple** *(Échantillons appariés)*. On veut mesurer **l'effet d'un médicament sur la fièvre**:

- $X_i$: température corporelle de l'individu $i$ **avant la prise du médicament**
- $Y_i$: température corporelle du même individu $i$ **4h après la prise du médicament**

# Différence de moyennes - Échantillons indépendants, variances connues

On a $\mathbf{X} = (X_1, \ldots, X_{n_X})$ i.i.d. de loi populationnelle $\mathcal{N}(\mu_X, \sigma_X^2)$ et $\mathbf{Y} = (Y_1, \ldots, Y_{n_Y})$ i.i.d. de loi populationnelle, $\mathcal{N}(\mu_Y \sigma_Y^2)$ où $\sigma_X^2$ et $\sigma_Y^2$ **sont supposés connus**, et $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$.

On cherche à construire des ICs pour $\mu_X - \mu_Y$.

1. Proposez un estimateur pour $\mu_X$ et pour $\mu_Y$.✍
2. Quelle est la loi de la différence de ces estimateurs? ✍
3. Trouvez un pivot tel quel $\Delta \sim \mathcal{N}(0, 1)$.✍
4. En déduire les intervalles de confiance unilatéraux et bilatéraux au niveau $1 - \alpha$. ✍

# Différence de moyennes - Échantillons indépendants, variances égales, inconnues

On a $\mathbf{X} = (X_1, \dots, X_{n_X})$ i.i.d. de loi populationnelle $\mathcal{N}(\mu_X, \sigma^2)$ et $\mathbf{Y} = (Y_1, \dots, Y_{n_Y})$ i.i.d. de loi populationnelle, $\mathcal{N}(\mu_Y \sigma^2)$ où $\sigma^2$ est **inconnu** et $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$.

**Très important: On suppose que les lois populationnelles des deux échantillons ont la même variance!**

On cherche à construire des ICs pour $\mu_X - \mu_Y$.

1. **Proposez un estimateur pour $\mu_X$ et pour $\mu_Y$.**
2. **Quelle est la loi de la différence de ces estimateurs?**
3. En utilisant le pivot
$$\Delta = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_p \sqrt{1/n_X + 1/n_Y}} \sim \mathcal{T}(n_X + n_Y - 2),$$
où
$$S_p^2 = \frac{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}{n_X + n_Y - 2},$$
déduire les intervalles de confiance unilatéraux et bilatéraux au niveau $1 - \alpha$. ✍

# Différence de moyennes - Échantillons appariés

On a $\mathbf{X} = (X_1, \ldots, X_n)$ i.i.d. de loi populationnelle $\mathcal{N}(\mu_X, \sigma_X^2)$ et $\mathbf{Y} = (Y_1, \ldots, Y_n)$ i.i.d. de loi populationnelle, $\mathcal{N}(\mu_Y, \sigma_Y^2)$ où $\sigma^2$ est **inconnu** et pour tout $i \neq j$, $(X_i, Y_i) \perp\!\!\!\perp (X_j, Y_j)$.

**Très important: Les deux échantillons sont appariés et donc pas indépendants. Leur taille est donc nécessairement la même.**

On cherche à construire des ICs pour $\mu_X - \mu_Y$

1. Pour une paire $(X_i, Y_i)$, déterminer la loi de $D_i = X_i - Y_i$. ✍
2. En déduire la loi de $\overline{D}_n$. ✍
3. Proposez un estimateur pour la variance de $D_i$. Quelle est sa loi? ✍
4. Trouvez un pivot tel quel $\Delta \sim \mathcal{T}(n-1)$.
5. En déduire les intervalles de confiance unilatéraux et bilatéraux au niveau $1 - \alpha$. ✍

# Différence de deux moyennes - Récapitulatif des ICs

Variance "poolée": $S_p^2 = \frac{(n_X-1)S_X^2+(n_Y-1)S_Y^2}{n_X+n_Y-2}$

| Ech. | Type | Situation | Estimateurs | Pivot $\Delta$ |
|---|---|---|---|---|
| Indép. | Bilat. | V.C | $\overline{X}, \overline{Y}$ | $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_X^2/n_X+\sigma_Y^2/n_Y}} \sim \mathcal{N}(0,1)$ |
| Indép. | Droite | V.C | $\overline{X}, \overline{Y}$ | $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_X^2/n_X+\sigma_Y^2/n_Y}} \sim \mathcal{N}(0,1)$ |
| Indép. | Gauche | V.C | $\overline{X}, \overline{Y}$ | $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_X^2/n_X+\sigma_Y^2/n_Y}} \sim \mathcal{N}(0,1)$ |
| Indép. | Bilat. | V.E.I | $\overline{X}, \overline{Y}, S_p^2$ | $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_p\sqrt{1/n_X+1/n_Y}} \sim \mathcal{T}(n_X+n_Y-2)$ |
| Indép. | Droite | V.E.I | $\overline{X}, \overline{Y}, S_p^2$ | $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_p\sqrt{1/n_X+1/n_Y}} \sim \mathcal{T}(n_X+n_Y-2)$ |
| Indép. | Gauche | V.E.I | $\overline{X}, \overline{Y}, S_p^2$ | $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_p\sqrt{1/n_X+1/n_Y}} \sim \mathcal{T}(n_X+n_Y-2)$ |
| Appariés | Bilat. | V.I | $\overline{D}_n, S_D^2$ | $\frac{\overline{D}_n-\mu_D}{S_D/\sqrt{n}} \sim \mathcal{T}(n-1)$ |
| Appariés | Droite | V.I | $\overline{D}_n, S_D^2$ | $\frac{\overline{D}_n-\mu_D}{S_D/\sqrt{n}} \sim \mathcal{T}(n-1)$ |
| Appariés | Gauche | V.I | $\overline{D}_n, S_D^2$ | $\frac{\overline{D}_n-\mu_D}{S_D/\sqrt{n}} \sim \mathcal{T}(n-1)$ |

V.C = Variances connues; V.E.I = Variances égales inconnues ; V.I = Variances inconnues

# Différence de deux moyennes - Récapitulatif des ICs

| Ech. | Type | Situation | Estimateurs | IC exact |
|---|---|---|---|---|
| Indép. | Bilat. | V.C | $\overline{X}, \overline{Y}$ | $\left[(\overline{X}-\overline{Y}) \pm z_{1-\alpha/2}\sqrt{\sigma_X^2/n_X+\sigma_Y^2/n_Y}\right]$ |
| Indép. | Droite | V.C | $\overline{X}, \overline{Y}$ | $\left(-\infty\,;\,(\overline{X}-\overline{Y}) + z_{1-\alpha}\sqrt{\sigma_X^2/n_X+\sigma_Y^2/n_Y}\right]$ |
| Indép. | Gauche | V.C | $\overline{X}, \overline{Y}$ | $\left[(\overline{X}-\overline{Y}) - z_{1-\alpha}\sqrt{\sigma_X^2/n_X+\sigma_Y^2/n_Y}\,;\,\infty\right)$ |
| Indép. | Bilat. | V.E.I | $\overline{X}, \overline{Y}, S_p^2$ | $\left[(\overline{X}-\overline{Y}) \pm t_{1-\alpha/2,n_X+n_Y-2}\, S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\right]$ |
| Indép. | Droite | V.E.I | $\overline{X}, \overline{Y}, S_p^2$ | $\left(-\infty\,;\,(\overline{X}-\overline{Y}) + t_{1-\alpha,n_X+n_Y-2}\, S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\right]$ |
| Indép. | Gauche | V.E.I | $\overline{X}, \overline{Y}, S_p^2$ | $\left[(\overline{X}-\overline{Y}) - t_{1-\alpha,n_X+n_Y-2}\, S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\,;\,\infty\right)$ |
| Appariés | Bilat. | V.I | $\overline{D}_n, S_D^2$ | $\left[\overline{D}_n \pm t_{1-\alpha/2,n-1}\frac{S_D}{\sqrt{n}}\right]$ |
| Appariés | Droite | V.I | $\overline{D}_n, S_D^2$ | $\left(-\infty\,;\,\overline{D}_n + t_{1-\alpha,n-1}\frac{S_D}{\sqrt{n}}\right]$ |
| Appariés | Gauche | V.I | $\overline{D}_n, S_D^2$ | $\left[\overline{D}_n - t_{1-\alpha,n-1}\frac{S_D}{\sqrt{n}}\,;\,\infty\right)$ |

V.C = Variances connues; V.E.I = Variances égales inconnues ; V.I = Variances inconnues

# Loi Fisher-Snedecor

Une v.a. de **loi de Fisher-Snedecor** (ou simplement Fisher) est une v.a. **continue positive**

Elle est paramétrée par **deux degrés de libertés** $\nu_1, \nu_2 \in \mathbb{N} \setminus \{0\}$

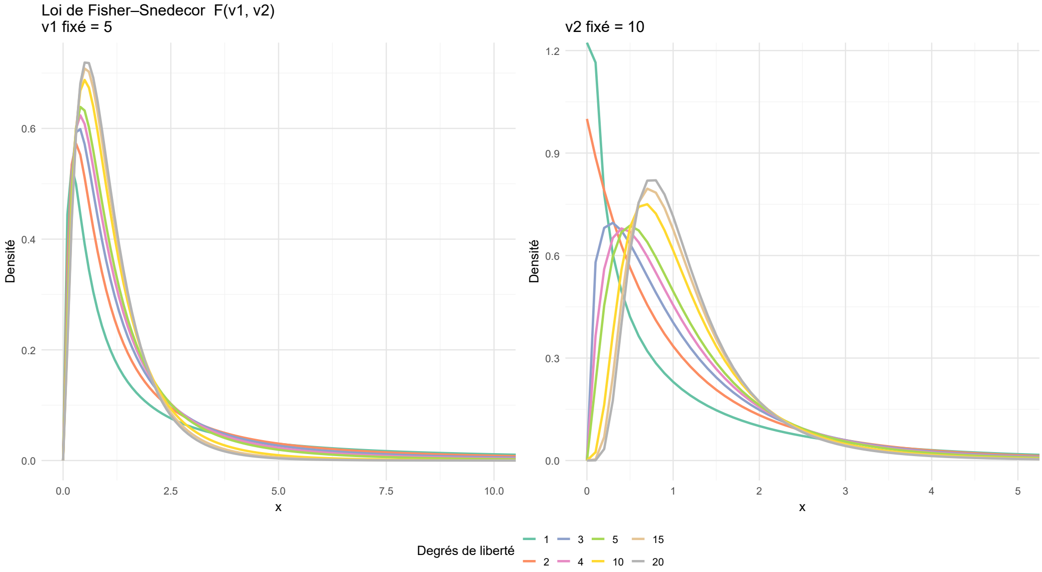
On note $X \sim \mathcal{F}(\nu_1, \nu_2)$, et son support est $[0, \infty)$ et sa densité est donnée par:

$$f_X(t) = \frac{\left(\frac{\nu_1 t}{\nu_1 t + \nu_2}\right)^{\nu_1/2}\left(1 - \frac{\nu_1 t}{\nu_1 t + \nu_2}\right)^{\nu_2/2}}{tB(\nu_1/2, \nu_2/2)} \mathbb{1}_{[0,\infty)}(t), \quad \text{où} \quad B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

**Caractéristiques** *(Loi de Fisher-Snedecor)*. ★

- $\mathbb{E}[X] = \frac{\nu_2}{\nu_2 - 2}$ pour $\nu_2 > 2$
- $\mathbb{V}(X) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ pour $\nu_2 > 4$

# Loi Fisher-Snedecor - Illustration



# Loi Fisher-Snedecor - Propriétés

**Propriétés** *(Fisher = Ratio de deux Chi-Carrés indépendantes).* ★

*Pour* $U_1 \sim \chi^2(\nu_1)$, $U_2 \sim \chi^2(\nu_2)$, *et* $U_1 \perp\!\!\!\perp U_2$:

$$\frac{U_1/\nu_1}{U_2/\nu_2} \sim \mathcal{F}(\nu_1, \nu_2)$$

**Propriétés** *(Transformation et lien avec d'autres lois).* ★

- *Si* $X \sim \mathcal{F}(\nu_1, \nu_2)$, *alors* $\frac{1}{X} \sim \mathcal{F}(\nu_2, \nu_1)$
- *Si* $T \sim \mathcal{T}(\nu)$, *alors* $T^2 \sim \mathcal{F}(1, \nu)$

**Propriétés** *(Quantile d'ordre de la loi de Fisher).* ★

*Soit* $f_{\alpha,\nu_1,\nu_2}$ *le quantile d'ordre* $\alpha$ *de* $X \sim \mathcal{F}(\nu_1, \nu_2)$. *Alors, pour tout* $\alpha \in (0, 1)$:

$$f_{1-\alpha,\nu_2,\nu_1} = \frac{1}{f_{\alpha,\nu_1,\nu_2}}$$

# Ratio de variances - Échantillons indépendants

On a $\mathbf{X} = (X_1, \dots, X_{n_X})$ i.i.d. de loi populationnelle $\mathcal{N}(\mu_X, \sigma_X^2)$ et $\mathbf{Y} = (Y_1, \dots, Y_{n_Y})$ i.i.d. de loi populationnelle, $\mathcal{N}(\mu_Y \sigma_Y^2)$ où **aucun des paramètres n'est connu** et $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$.

On cherche à construire des ICs pour $\frac{\sigma_X^2}{\sigma_Y^2}$.

1. Proposez un estimateur pour $\sigma_X^2$ et pour $\sigma_Y^2$. Quelles sont leurs lois? ✍
2. Trouvez un pivot tel quel $\Delta \sim \mathcal{F}(n_X - 1, n_Y - 1)$.✍
3. En déduire les intervalles de confiance unilatéraux et bilatéraux au niveau $1 - \alpha$. ✍

# Ratio de variances — Récapitulatif des ICs (échantillons indépendants)

| Éch. | Type | Situation | Estimateurs | Pivot $\Delta$ | IC exact |
|---|---|---|---|---|---|
| Indép. | Bilatéral | Moys. inconnues | $S_X^2,\ S_Y^2$ | $\dfrac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \sim F_{n-1,\,m-1}$ | $\left[\dfrac{S_X^2/S_Y^2}{F_{1-\alpha/2,n_X-1,n_Y-1}}\,;\,\dfrac{S_X^2/S_Y^2}{F_{\alpha/2,n_X-1,n_Y-1}}\right]$ |
| Indép. | Uni. droite | Moys. inconnues | $S_X^2,\ S_Y^2$ | $\dfrac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \sim F_{n-1,\,m-1}$ | $\left(0\,;\,\dfrac{S_X^2/S_Y^2}{F_{1-\alpha,n_X-1,n_Y-1}}\right]$ |
| Indép. | Uni. gauche | Moys. inconnues | $S_X^2,\ S_Y^2$ | $\dfrac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \sim F_{n-1,\,m-1}$ | $\left[\dfrac{S_X^2/S_Y^2}{F_{\alpha,n_X-1,n_Y-1}}\,;\,\infty\right)$ |

# Sommaire

# Précision d'un intervalle de confiance

Dans plusieurs domaines, il est crucial de déterminer **combien d'observations** sont nécessaires pour que l'**intervalle de confiance soit suffisamment précis**

Par ex: Essais cliniques, sondages...

On doit d'abord définir la notion de **précision** d'un IC **bilatéral**

**Définition** *(Précision d'un intervalle de confiance).* ★

La **précision** d'un IC bilatéral $\left[\widehat{\theta}_S(\mathbf{X}), \widehat{\theta}_I(\mathbf{X})\right]$ correspond à sa **demi-largeur**:

$$\delta = \frac{\widehat{\theta}_S(\mathbf{X}) - \widehat{\theta}_I(\mathbf{X})}{2} \quad \text{(pour un IC bilatéral).}$$

☞ **Plus $\delta$ est petit, plus l'IC est précis**

# Précision d'un IC - Exemple Gaussien

**Exemple — moyenne d'une population gaussienne (variance connue)** :

$$\left[\overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\ ;\ \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

La **précision** est alors :

$$\delta = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

**Remarques importantes :** ★

- $\delta$ **diminue quand $n$ augmente**: plus d'observations = intervalle plus serré
- $\delta$ **augmente avec $1-\alpha$**: plus le niveau de confiance est grand, plus l'intervalle est large ;
- $\delta$ **augmente avec $\sigma$**: plus le phénomène est incertain, moins on est précis.

**Objectif :** fixer une précision $\delta$ désirée et en déduire la taille minimale d'échantillon $n$ nécessaire.

Pour l'exemple précédent, à partir de la relation:

$$\delta = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

on obtient: ★

$$\boxed{n^* = \left(\frac{z_{1-\alpha/2}\,\sigma}{\delta}\right)^2}$$

☞ L'arrondit à l'entier supérieur de $n^*$ assure d'avoir une précision d'au moins $\delta$ pour cet IC

## Calcul de taille échantillonnale - Exemple

**Exemple :**

- On veut estimer la moyenne $\mu$ d'une population avec un IC à 95% de confiance
- On connaît $\sigma = 3$
- On souhaite une précision $d = 0.5$.

Alors :

$$n = \left(\frac{1.96 \times 3}{0.5}\right)^2 = 138.3 \Rightarrow \boxed{n = 139.}$$

Ainsi, pour une précision donnée:

- Doubler la précision (diviser $d$ par 2) $\Rightarrow$ **multiplier $n$ par 4**;
- Augmenter le niveau de confiance $\Rightarrow$ **Augmente la taille nécessaire**.

# Calcul de taille échantillonnale - Illustration

**Exemple** *(Sondage)*.

Durant des élections municipales une personne candidate souhaite estimer **par un IC bi-latéral asymptotique à 95%** la proportion $p$ de citoyens qui voteront pour elle. Elle souhaite une précision de 3 points de pourcentages ($\delta = 0.03$).

Combien de personnes doit-il interroger au minimum?

☞ Comment modéliser l'échantillon? Quelle est sa loi populationnelle? ✍

☞ Donnez un estimateur de $p$. ✍

☞ Proposez un pivot asymptotique pour cet estimateur. En déduire un IC asymptotique bilatéral. ✍

☞ Quelle est la précision de cet IC? Quelle valeur de $p$ la maximise? ✍

☞ Dans le pire cas, quelle taille d'échantillon permettrait d'avoir un IC avec la précision demandée? ✍