Chapitre II : Tests paramétriques à un échantillon

Cours de tests paraméRiques

M. Champion

Université de Paris

2021-2022

II. Introduction

Dans ce chapitre, on s'intéresse aux tests paramétriques à un échantillon. Les tests de comparaison d'échantillons seront étudiés au chapitre III.

Objectif : comparer la caractéristique d'un échantillon à une valeur de référence connue à priori :

- test de comparaison d'une **moyenne** observée à une moyenne théorique,
- test de comparaison d'une **proportion** observée à une proportion théorique,
- test de comparaison d'une **variance** observée à une variance théorique.

II.1 Test de comparaison moyenne observée/théorique

Exemple

Pour apaiser les migraines, les médecins prescrivent généralement le médicament A. Une étude statistique a montré que la durée de disparition de la douleur chez les malades traités avec A est une v.a de loi normale d'espérance $\mu_0 = 30$ mns. Un laboratoire pharmaceutique a conçu un nouveau médicament B et désire tester son efficacité. Pour cela, le médicament a été administré à n cobayes malades et on a mesuré la durée de disparition de la douleur pour chacun d'eux. On suppose que cette durée est aussi une v.a de loi normale d'espérance μ inconnue et de variance σ^2, également inconnue.

Les mesures suivantes ont été obtenues :

```
##  [1] 25 28 20 32 17 24 41 28 25 30 27 24
```

```
## avec pour moyenne  26.75
```

Peut-on mettre le médicament B sur le marché?

II.1 Test de comparaison moyenne observée/théorique

Formalisation

Dans cette partie, on se limite à l'étude des tests sur la **moyenne** d'**un échantillon**. On se donne :

- une v.a X d'espérance $\mu := \mathbb{E}[X]$ et de variance $\sigma^2 := \mathbb{V}[X]$,
- n v.a $(X_1, ..., X_n)$ indépendantes et de même loi que X (un n-échantillon de X),
- n observations $(x_1, ..., x_n)$ des v.a $(X_1, ..., X_n)$.

Exemple

- X est la v.a "durée de disparition de la douleur de la migraine après prise d'un médicament",
- $X \sim \mathcal{N}(\mu, \sigma^2)$,
- $n = 12$ observations de X sont disponibles.

II.1 Test de comparaison moyenne observée/théorique

Hypothèses

Faire un test de comparaison d'une moyenne observée avec une moyenne théorique revient à répondre à la question suivante : la v.a X a t-elle une espérance $\boldsymbol{\mu}$ égale à une valeur $\boldsymbol{\mu_0}$ donnée à l'avance? On tente de répondre à cette question à partir d'une observation de μ.

Il s'agit donc de juger de la pertinence de l'hypothèse :

$$(H_0) : \mu = \mu_0,$$

où μ_0 est l'espérance théorique.

Il s'agit d'un **test de conformité**.

II.1 Test de comparaison moyenne observée/théorique

Hypothèses

Comment définir l'hypothèse alternative (H_1)?

II.1 Test de comparaison moyenne observée/théorique

Hypothèses

Comment définir l'hypothèse alternative (H_1)?

Exemple

- $\mu < \mu_0$: le médicament B est en moyenne plus efficace que le A,
- $\mu > \mu_0$: le médicament B est en moyenne moins efficace que le A,
- $\mu \neq \mu_0$: le médicament B n'a pas la même efficacité que le A.

II.1 Test de comparaison moyenne observée/théorique

Hypothèses

Comment définir l'hypothèse alternative (H_1)?

Exemple

- $\mu < \mu_0$: le médicament B est en moyenne plus efficace que le A,
- $\mu > \mu_0$: le médicament B est en moyenne moins efficace que le A,
- $\mu \neq \mu_0$: le médicament B n'a pas la même efficacité que le A.

Ici, l'important est de ne pas se tromper si on décide de changer de médicament : il vaut mieux conserver un médicament moins performant plutôt que d'adopter un médicament moins performant que l'ancien. Il faut donc que l'hypothèse alternative à (H_0) soit $\mu < \mu_0$:

$$(H_0) : \mu = \mu_0 \quad \text{contre} \quad (H_1) : \mu < \mu_0.$$

II.1 Test de comparaison moyenne observée/théorique

Hypothèses

De manière générale, on trouve trois différents types de tests :

- test **unilatéral à gauche**, pour lequel l'hypothèse alternative est :

$$(H_1) : \mu < \mu_0,$$

- test **unilatéral à droite**, pour lequel l'hypothèse alternative est:

$$(H_1) : \mu > \mu_0,$$

- test **bilatéral**, pour lequel l'hypothèse alternative est :

$$(H_1) : \mu \neq \mu_0.$$

⚠Le choix du test et de l'hypothèse alternative est très important puisque cela change les calculs et la conclusion.

II.1 Test de comparaison moyenne observée/théorique

Règle de décision

Pour le test de comparaison de moyenne, la variable d'intérêt est définie par :

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

qui est un bon estimateur du paramètre inconnu μ.

Objectif : définir un seuil à partir duquel on choisit (H_1) plutôt que (H_0).

Exemple

Dans l'exemple avec les médicaments, il sera naturel de rejeter (H_0) si on observe un temps moyen d'action du médicament B très inférieur à celui du A, c'est-à-dire à $\mu_0 = 30$ mns. On compare donc $\bar{X}_n$ à μ_0.

II.1 Test de comparaison moyenne observée/théorique

Règle de décision

On rejettera (H_0), suivant les tests :

- **unilatéral à gauche** (H_1) : $\mu < \mu_0$,

$$\text{si } \bar{X}_n - \mu_0 \text{ est loin de 0 dans les négatifs,}$$

- **unilatéral à droite** (H_1) : $\mu > \mu_0$:

$$\text{si } \bar{X}_n - \mu_0 \text{ est loin de 0 dans les positifs,}$$

- **bilatéral** (H_1) : $\mu \neq \mu_0$.

$$\text{si } |\bar{X}_n - \mu_0| \text{ est loin de 0.}$$

II.1 Test de comparaison moyenne observée/théorique

Loi de $\bar{X}_n - \mu_0$

Pour faire le test, il est vital de connaître la loi de $\bar{X}_n - \mu_0$ sous (H_0). Deux cas sont possibles :

- soit on dispose d'un **grand** échantillon et on utilise alors le **TCL** pour approcher $\bar{X}_n - \mu_0$ par une **loi normale**,
- soit on dispose d'un **petit** échantillon qui suit directement une **loi normale**.

Rappel : **Théorème Central Limite**

$$\forall X_1, ..., X_n \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2), \quad \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

II.1 Test de comparaison moyenne observée/théorique

Loi de $\bar{X}_n - \mu_0$ - cas des grands échantillons

Lorsque la taille de l'échantillon est grande, on applique le TCL, qui nous dit que :

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \text{ suit approximativement une loi } \mathcal{N}(0,1).$$

On définit alors la **statistique de test** par :

$$T_n = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}},$$

pour laquelle, sous l'hypothèse (H_0) :

$$T_n \text{ suit approximativement une loi } \mathcal{N}(0,1).$$

⚠ T_n ne peut être utilisé comme statistique de test que si σ^2 est connue, ce qui n'est jamais le cas en pratique!

II.1 Test de comparaison moyenne observée/théorique

Loi de $\bar{X}_n - \mu_0$ - cas des petits échantillons

Lorsque la taille de l'échantillon est petite, on ne peut plus appliquer le TCL. Il faut donc une hypothèse supplémentaire sur la loi des observations :

$$\forall 1 \leq i \leq n, \quad X_i \sim \mathcal{N}(\mu, \sigma^2),$$

ce qui permet de connaître la loi de $\bar{X}_n$ sous (H_0).

On définit alors la **statistique de test** par :

$$T_n = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim_{H_0} \mathcal{N}(0,1).$$

⚠ T_n ne peut être utilisé comme statistique de test que si σ^2 est connue, ce qui n'est jamais le cas en pratique!

II.1 Test de comparaison moyenne observée/théorique

Loi de $\bar{X}_n - \mu_0$ - cas où la variance σ^2 est inconnue

Lorsque σ^2 est **inconnue** (ce qui est toujours le cas en pratique), on ne peut plus utiliser la même statistique de test T_n. On remplace alors σ^2 par un de ses estimateurs :

$$(\hat{\sigma}^2 =) s^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2.$$

Remarque : s^2 est la variance empirique de X, c'est un estimateur sans biais de σ^2.

La **statistique de test** devient alors :

$$T_n = \frac{\bar{X}_n - \mu_0}{s/\sqrt{n}} \sim_{H_0} St(n-1).$$

II.1 Test de comparaison moyenne observée/théorique

Procédure de test unilatéral à gauche - cas où la variance σ^2 est connue

$$(H_0) : \mu = \mu_0 \ \text{ contre } \ (H_1) : \mu < \mu_0.$$

- Statistique de test :

$$T_n = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim_{H_0} \mathcal{N}(0,1).$$

- Seuil à risque α fixé :

$$c_\alpha \ \text{ tel que } \ \mathbb{P}_{H_0}[\text{rejeter } H_0] = \mathbb{P}_{H_0}[T_n < c_\alpha] = \alpha.$$

- Zone de rejet :

$$R_\alpha = \{T_n < c_\alpha\}.$$

- Calcul de la valeur observée t_n de T_n sur les observations.
- Conclusion du test : on rejette (H_0) si $t_n < c_\alpha$.

II.1 Test de comparaison moyenne observée/théorique

Procédure de test unilatéral à gauche - cas où la variance σ^2 est inconnue

$$(H_0) : \mu = \mu_0 \text{ contre } (H_1) : \mu < \mu_0.$$

- Statistique de test :

$$T_n = \frac{\bar{X}_n - \mu_0}{\boldsymbol{s}/\sqrt{n}} \sim_{H_0} St(n-1).$$

- Seuil à risque α fixé :

$$c_\alpha \text{ tel que } \mathbb{P}_{H_0}[\text{rejeter } H_0] = \mathbb{P}_{H_0}[T_n < c_\alpha] = \alpha.$$

- Zone de rejet :

$$R_\alpha = \{T_n < c_\alpha\}.$$

- Calcul de la valeur observée t_n de T_n sur les observations.
- Conclusion du test : on rejette (H_0) si $t_n < c_\alpha$.

II.1 Test de comparaison moyenne observée/théorique

Procédure de test unilatéral à droite - cas où la variance σ^2 est inconnue

$$(H_0) : \mu = \mu_0 \ \text{ contre } \ (H_1) : \mu > \mu_0.$$

- Statistique de test :

$$T_n = \frac{\bar{X}_n - \mu_0}{s/\sqrt{n}} \sim_{H_0} St(n-1).$$

- Seuil à risque α fixé :

$$c_\alpha \ \text{ tel que } \ \mathbb{P}_{H_0}[\text{rejeter } H_0] = \mathbb{P}_{H_0}[T_n > c_\alpha] = \alpha.$$

- Zone de rejet :

$$R_\alpha = \{T_n > c_\alpha\}.$$

- Calcul de la valeur observée t_n de T_n sur les observations.
- Conclusion du test : on rejette (H_0) si $t_n > c_\alpha$.

II.1 Test de comparaison moyenne observée/théorique

Procédure de test bilatéral - cas où la variance σ^2 est inconnue

$$(H_0) : \mu = \mu_0 \ \text{ contre } \ (H_1) : \mu \neq \mu_0.$$

- Statistique de test :

$$T_n = \frac{\bar{X}_n - \mu_0}{s/\sqrt{n}} \sim_{H_0} St(n-1).$$

- Seuil à risque α fixé :

$$c_\alpha \ \text{ tel que } \ \mathbb{P}_{H_0}[\text{rejeter } H_0] = \mathbb{P}_{H_0}[|T_n| > c_\alpha] = \alpha.$$

- Zone de rejet :

$$R_\alpha = \{|T_n| > c_\alpha\}.$$

- Calcul de la valeur observée t_n de T_n sur les observations.
- Conclusion du test : on rejette (H_0) si $|t_n| > c_\alpha$.

II.1 Test de comparaison moyenne observée/théorique

Hypothèses composites

Les hypothèses **composites** (par opposition à l'hypothèse **simple** $(H_0) : \mu = \mu_0$) sont définies par :

$$(H_0) : \mu \leq \mu_0 \;\text{ contre }\; (H_1) : \mu > \mu_0$$

ou encore

$$(H_0) : \mu \geq \mu_0 \;\text{ contre }\; (H_1) : \mu < \mu_0.$$

Ici, la construction du test est basée sur le choix d'un c_α où le risque α dépend de μ. En effet, sous (H_0), μ peut prendre toutes les valeurs réelles telles que $\mu \leq \mu_0$.

On peut cependant se sortir de cette situation en considérant que α correspond au risque maximum que l'on accepte de prendre en rejetant (H_0) à tort :

$$\alpha = \sup_{\mu \in (H_0)} \alpha_\mu.$$

On construit donc finalement la zone de rejet en utilisant la valeur μ_0 seule qui correspond au risque le plus élevé.

II.1 Test de comparaison moyenne observée/théorique

p-valeurs

Le **degré de significativité** ou **p-valeur** du test est le plus petit seuil pour lequel on rejette (H_0). Il est défini, suivant les tests, par :

- **unilatéral à gauche**
$$\text{p-valeur} = \mathbb{P}_{H_0}[T_n < t_n],$$
- **unilatéral à droite**
$$\text{p-valeur} = \mathbb{P}_{H_0}[T_n > t_n],$$
- **bilatéral**
$$\text{p-valeur} = \mathbb{P}_{H_0}[|T_n| > t_n].$$

La **p-valeur** correspond à la probabilité d'avoir observé une valeur aussi extrême pour T_n que celle observée t_n.

II.1 Test de comparaison moyenne observée/théorique

Bilan

On teste l'hypothèse (H_0) : $\mu = \mu_0$. Suivant les cas :

	Statistique de test	Loi de T_n sous (H_0) Cas gaussien n petit	Cas non gaussien n grand	Hypothèse (H_1)	Région de rejet
Cas σ^2 connue	$T_n = \frac{\bar{X}_n - \mu_0}{\sigma / \sqrt{n}}$	$T_n \sim_{H_0} \mathcal{N}(0,1)$	$T_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$	$\mu > \mu_0$	$T_n > c_\alpha$
Cas σ^2 inconnue	$T_n = \frac{\bar{X}_n - \mu_0}{s / \sqrt{n}}$	$T_n \sim_{H_0} St(n-1)$	$T_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$	$\mu < \mu_0$	$T_n < c_\alpha$

où c_α est un quantile dépendant de la loi de T_n au niveau α fixé.

II.2 Test de comparaison de proportion observée/théorique

Exemple

Un an avant les élections municipales, un maire d'un petit village de 300 habitants se demande quelles seront les intentions de vote aux prochaines élections. Il charge un élu du conseil municipal d'aller sonder un échantillon représentatif de la population de 50 habitants en leur demandant pour qui ils comptent voter. Aux dernières élections, ce maire avait obtenu 65% des votes, il espère faire mieux l'année prochaine!

Les résultats font apparaître 69% d'intentions positives de vote en faveur du maire. Que peut-on en conclure?

II.2 Test de comparaison de proportion observée/théorique

Exemple

Un an avant les élections municipales, un maire d'un petit village de 300 habitants se demande quelles seront les intentions de vote aux prochaines élections. Il charge un élu du conseil municipal d'aller sonder un échantillon représentatif de la population de 50 habitants en leur demandant pour qui ils comptent voter. Aux dernières élections, ce maire avait obtenu 65% des votes, il espère faire mieux l'année prochaine!

Les résultats font apparaître 69% d'intentions positives de vote en faveur du maire. Que peut-on en conclure?

Test de comparaison d'une proportion observée à une proportion théorique

II.2 Test de comparaison de proportion observée/théorique

Formalisation

Dans cette partie, on se limite à l'étude des tests sur les **proportions** d'**un échantillon**. On se donne :

- une v.a X suivant une loi de Bernoulli de paramètre p : $X \sim \mathcal{B}(p)$,

$$\mathbb{P}(X = 1) = p \text{ et } \mathbb{P}(X = 0) = 1 - p.$$

- n v.a $(X_1, ..., X_n)$ indépendantes et de même loi que X (un n-échantillon de X),
- n observations $(x_1, ..., x_n)$ des v.a $(X_1, ..., X_n)$.

Exemple

- X est la v.a "intention de vote aux prochaines municipales",
- $X = 1$ si le vote est en faveur du maire, 0 sinon,
- $n = 50$ observations de X sont disponibles.

II.2 Test de comparaison de proportion observée/théorique

Lien avec le test de comparaison de moyenne

Ce cadre est un cas particulier du test de comparaison de moyenne avec un n-échantillon $(X_1, ..., X_n)$ de X de :

- loi $\mathcal{B}(p)$,
- espérance $\mu := \mathbb{E}(X) = p$,
- variance $\sigma^2 := \mathbb{V}(X) = p(1-p)$.

La variable d'intérêt est définie par :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

qui est un bon estimateur du paramètre inconnu p.

II.2 Test de comparaison de proportion observée/théorique

Hypothèses

Pour le test de comparaison de proportion correspondant à l'exemple précédent, les hypothèses sont définies de la manière suivante :

$$(H_0) : p = p_0 \;\text{ contre }\; (H_1) : p > p_0.$$

La **statistique de test** est définie par :

$$T_n = \frac{\bar{X}_n - p_0}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - p_0)}{\sqrt{p_0(1-p_0)}},$$

dont on connaît la loi sous (H_0) sous des hypothèses peu restrictives.

Remarque : Ici, on se ramène à un cas de test de comparaison de moyenne observée/théorique pour lequel la variance σ^2 est connue puisque celle-ci peut être directement calculée à partir de p_0.

II.2 Test de comparaison de proportion observée/théorique

Loi de $\bar{X}_n - p_0$ - cas des grands échantillons

La variance σ^2 de X étant connue, lorsque la taille de l'échantillon est grande, le TCL appliqué à T_n nous donne :

$$T_n = \frac{\bar{X}_n - p_0}{\sigma/\sqrt{n}} \text{ suit approximativement une loi } \mathcal{N}(0,1).$$

Remarque : La taille de l'échantillon doit être suffisamment grande au sens où :

$$n > 30, \quad np_0 \geq 5 \text{ et } n(1-p_0) \geq 5.$$

II.2 Test de comparaison de proportion observée/théorique

Procédure de test unilatéral à droite

$$(H_0) : p = p_0 \quad \text{contre} \quad (H_1) : p > p_0.$$

- Statistique de test :

$$T_n = \frac{\sqrt{n}(\bar{X}_n - p_0)}{\sqrt{p_0(1-p_0)}} \sim_{H_0} \mathcal{N}(0,1).$$

- Seuil à risque α fixé :

$$c_\alpha \quad \text{tel que} \quad \mathbb{P}_{H_0}[\text{rejeter } H_0] = \mathbb{P}_{H_0}[T_n > c_\alpha] = \alpha.$$

- Zone de rejet : $R_\alpha = \{T_n > c_\alpha\}$.
- Calcul de la proportion observée p_n puis de la réalisation t_n de T_n :

$$t_n = \frac{\sqrt{n}(p_n - p_0)}{\sqrt{p_0(1-p_0)}}.$$

- Conclusion du test : on rejette (H_0) si $t_n > c_\alpha$.

II.2 Test de comparaison proportion observée/théorique

Bilan

On teste l'hypothèse (H_0) : $\mu = \mu_0$. Suivant les cas :

	Statistique de test	Loi de T_n sous (H_0) Cas gaussien n petit	Cas non gaussien n grand	Hypothèse (H_1)	Région de rejet
Cas σ^2 connue	$T_n = \frac{\bar{X}_n - \mu_0}{\sigma / \sqrt{n}}$	$T_n \sim_{H_0} \mathcal{N}(0,1)$	$T_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$	$\mu > \mu_0$	$T_n > c_\alpha$
Cas σ^2 inconnue	$T_n = \frac{\bar{X}_n - \mu_0}{s / \sqrt{n}}$	$T_n \sim_{H_0} St(n-1)$	$T_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$	$\mu < \mu_0$	$T_n < c_\alpha$
Cas particulier $X \sim Ber(p)$	$T_n = \frac{\bar{X}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	???	$T_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$	$\mu \neq \mu_0$	$\lvert T_n \rvert > c_\alpha$

où c_α est un quantile dépendant de la loi de T_n au niveau α fixé.

II.3 Test de comparaison de variance observée/théorique

Exemple

Un manufactureur vient d'installer un machine toute neuve qui remplit des bouteilles de jus de fruit. Lors des premiers réglages de la machine, les techniciens du contrôle de qualité souhaitent vérifier que l'écart-type des contenances ne dépasse pas $\sigma_0 = 5$mL. Ils mesurent donc la contenance (en mL) de 30 bouteilles de jus de fruit au hasard. Les résultats suivants ont été obtenus :

```
##  [1] 205 195 203 204 199 200 204 203 188 203 202 201 199 199 204
## [20] 199 202 201 197 203 201 200 198 199 203 201
```

```
## pour un écart-type de 3.501
```

Que peuvent-ils en conclure?

II.3 Test de comparaison de variance observée/théorique

Exemple

Un manufactureur vient d'installer un machine toute neuve qui remplit des bouteilles de jus de fruit. Lors des premiers réglages de la machine, les techniciens du contrôle de qualité souhaitent vérifier que l'écart-type des contenances ne dépasse pas $\sigma_0 = 5$mL. Ils mesurent donc la contenance (en mL) de 30 bouteilles de jus de fruit au hasard. Les résultats suivants ont été obtenus :

```
##  [1] 205 195 203 204 199 200 204 203 188 203 202 201 199 199 204
## [20] 199 202 201 197 203 201 200 198 199 203 201
```

```
## pour un écart-type de 3.501
```

Que peuvent-ils en conclure? *Test de comparaison d'une variance observée à une variance théorique*

II.3 Test de comparaison de variance observée/théorique

Formalisation

Dans cette partie, on se limite à l'étude des tests sur la **variance** d'**un échantillon**. On se donne :

- une v.a X d'espérance $\mathbb{E}[X] := \mu$ et de variance $\mathbb{V}[X] := \sigma^2$,
- n v.a $(X_1, ..., X_n)$ indépendantes et de même loi que X (un n-échantillon de X),
- n observations $(x_1, ..., x_n)$ des v.a $(X_1, ..., X_n)$.

Exemple

- X est la v.a "contenance de bouteilles de jus de fruits' ',
- $n = 30$ observations de X sont disponibles.

II.3 Test de comparaison de variance observée/théorique

Hypothèses

Pour le test de comparaison de variance correspondant à l'exemple précédent, les hypothèses sont définies de la manière suivante :

$$(H_0) : \sigma^2 = \sigma_0^2 \ \text{ contre } \ (H_1) : \sigma^2 \neq \sigma_0^2.$$

Ce test est basé sur l'estimation de la variance σ^2 par l'estimateur :

$$\tilde{s}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2.$$

Si μ est **inconnu**, on le remplace par son estimateur et l'estimateur de la variance devient :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2.$$

II.3 Test de comparaison de variance observée/théorique

Cas où μ est connu - grands échantillons de loi quelconque

Notons $Z_i := (X_i - \mu)^2$, de telle sorte que $\tilde{s}^2 = \bar{Z}_n$. Les v.a Z_i sont i.i.d, d'espérance σ^2 et de variance finie, sous l'hypothèse que $\mathbb{E}(X_i^4)$ existe et est finie.

La **statistique de test** est alors définie comme :

$$T_n = \frac{\sqrt{n}(\tilde{s}^2 - \sigma_0^2)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)^2}},$$

dont on connaît la loi sous (H_0). En effet, si on dispose de suffisamment d'échantillons, le TCL appliqué aux variables Z_i permet de montrer que

$$T_n \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Remarque : tester une égalité sur la variance de X_i revient à tester une égalité sur la moyenne de Z_i.

II.3 Test de comparaison de variance observée/théorique

Cas où μ est connu - petits échantillons de loi gaussienne

Lorsque la taille de l'échantillon est petite, on ne peut plus appliquer le TCL. Il faut donc une hypothèse supplémentaire sur la loi des observations :

$$\forall 1 \leq i \leq n, \quad X_i \sim \mathcal{N}(\mu, \sigma^2).$$

On définit alors la **statistique de test** par :

$$T_n = \frac{n\tilde{s}^2}{\sigma_0^2},$$

qui suit une loi du chi-deux à n degrés de liberté sous (H_0) :

$$T_n \sim_{H_0} \chi_2(n).$$

II.3 Test de comparaison de variance observée/théorique

Cas où μ est inconnu - petits échantillons de loi gaussienne

Lorsque μ est inconnue, l'estimateur de la variance σ^2 devient :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2.$$

Dans le cas où l'on dispose de petits échantillons de loi gaussienne, la statistique de test devient :

$$T_n = \frac{(n-1)s^2}{\sigma_0^2},$$

qui suit une loi $\chi_2(n-1)$ sous (H_0) :

$$T_n \sim_{H_0} \chi_2(n-1).$$

II.3 Test de comparaison de variance observée/théorique

Procédure de test

$$(H_0) : \sigma^2 = \sigma_0^2 \;\text{ contre }\; (H_1) : \sigma^2 \neq \sigma_0^2.$$

- Statistique de test : T_n à définir suivant les cas (μ connu, inconnu, petits ou grands échantillons).
- Seuil à risque α fixé :

$$c_\alpha \;\text{ tel que }\; \mathbb{P}_{H_0}[\text{rejeter } H_0] = \mathbb{P}_{H_0}[|T_n| > c_\alpha] = \alpha.$$

- Zone de rejet :

$$R_\alpha = \{|T_n| > c_\alpha\}.$$

- Calcul de la valeur observée t_n de T_n sur les observations.
- Conclusion du test : on rejette (H_0) si $|t_n| > c_\alpha$.

II.3 Test de comparaison de variance observée/théorique

Bilan

On teste l'hypothèse (H_0) : $\sigma = \sigma_0$. Suivant les cas :

	Cas gaussien n quelconque Statistique de test	 Loi de T_n	Cas non gaussien n grand Statistique de test	 Loi de T_n
Cas μ connu	$T_n = \frac{n\tilde{s}^2}{\sigma_0^2}$	$T_n \sim_{H_0} \chi_2(n)$	$T_n = \frac{\sqrt{n}(\tilde{s}^2 - \sigma_0^2)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)^2}}$	$T_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$
Cas μ inconnu	$T_n = \frac{(n-1)s^2}{\sigma_0^2}$	$T_n \sim_{H_0} \chi_2(n-1)$	???	???

Les régions de rejet sont définies de la même manière que précédemment en utilisant les quantiles c_α correspondant à la loi de T_n au niveau α fixé.

II.4 Applications sous R

Test de comparaison moyenne observée/théorique unilatéral à gauche

```r
x <- c(25,28,20,32,17,24,41,28,25,30,27,24)
t.test(x,mu=30, alternative="less")
```

```
##
##  One Sample t-test
##
## data:  x
## t = -1.8526, df = 11, p-value = 0.04547
## alternative hypothesis: true mean is less than 30
## 95 percent confidence interval:
##      -Inf 29.90056
## sample estimates:
## mean of x
##     26.75
```

Le test est significatif au niveau 5%!

II.4 Applications sous R

Test de comparaison proportion observée/théorique bilatéral

```
x <- c(0,1,0,1,1,1,1,0,1,0,0,0,1,1,0)
prop.test(sum(x),p=0.25,n=15)
```

```
## Warning in prop.test(sum(x), p = 0.25, n = 15): Chi-squared appro
## incorrect
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  sum(x) out of 15, null probability 0.25
## X-squared = 5, df = 1, p-value = 0.02535
## alternative hypothesis: true p is not equal to 0.25
## 95 percent confidence interval:
##  0.2742328 0.7772383
## sample estimates:
##         p
## 0.5333333
```