Tests d'Hypothèses

Partons d'un exemple...Une machine fabrique des tiges d'acier. Si la machine est réglée correctement, l'utilisateur obtient une population de tiges de longueurs moyenne m et d'écart-type σ. On désire savoir si cette machine se dérègle. Ainsi, on prélèvera, à intervalles réguliers, des échantillons pour mesurer la longueur effective des tiges.

Nous faisons alors l'hypothèse H_0 dite hypothèse nulle que la machine est bien réglée. On teste alors cette hypothèse: 2 cas se présentent :

- la machine est bien réglée, on accepte H_0.
- la machine est mal réglée, on rejette H_0 et donc on accepte H_1 dite hypothèse alternative.

1. Principes des Tests Statistiques

Définition : un test statistique est une méthode permettant de prendre une **décision** à partir d'informations fournies par un **échantillon**.

Cette décision dépend donc de l'échantillon. Ainsi qu'elle que soit la décision prise, on court deux sortes de risques :

- le risque dit de 1ère espèce noté α, est la probabilité de rejeter l'hypothèse H_0 alors qu'elle est vraie en réalité : $\alpha = p(\text{rejeter } H_0 / H_0 \text{ vraie})$
- le risque dit de 2nde espèce noté β, est la probabilité d'accepter l'hypothèse H_0 alors qu'elle est fausse en réalité : $\beta = p(\text{accepter } H_0 / H_0 \text{ fausse})$.

Un test est bon si on arrive à minimiser α et β.

On considère une population P sur laquelle on veut étudier un paramètre γ inconnue associé à un paramètre c. Sur un échantillon de taille n, on obtient γ_e connu. Sur la base de cette valeur observée γ_e, on se propose de comparer la vraie valeur γ à une valeur γ_0 fixée à priori, constituant la valeur standard.

Soit une population P de grand effectif sur laquelle on étudie un caractère c. La moyenne m de c est inconnue. Sur un échantillon, on a trouvé une moyenne $\overline{x}_e$. On doit tester la moyenne m par rapport à une valeur notée m_0 qui est la valeur standard.

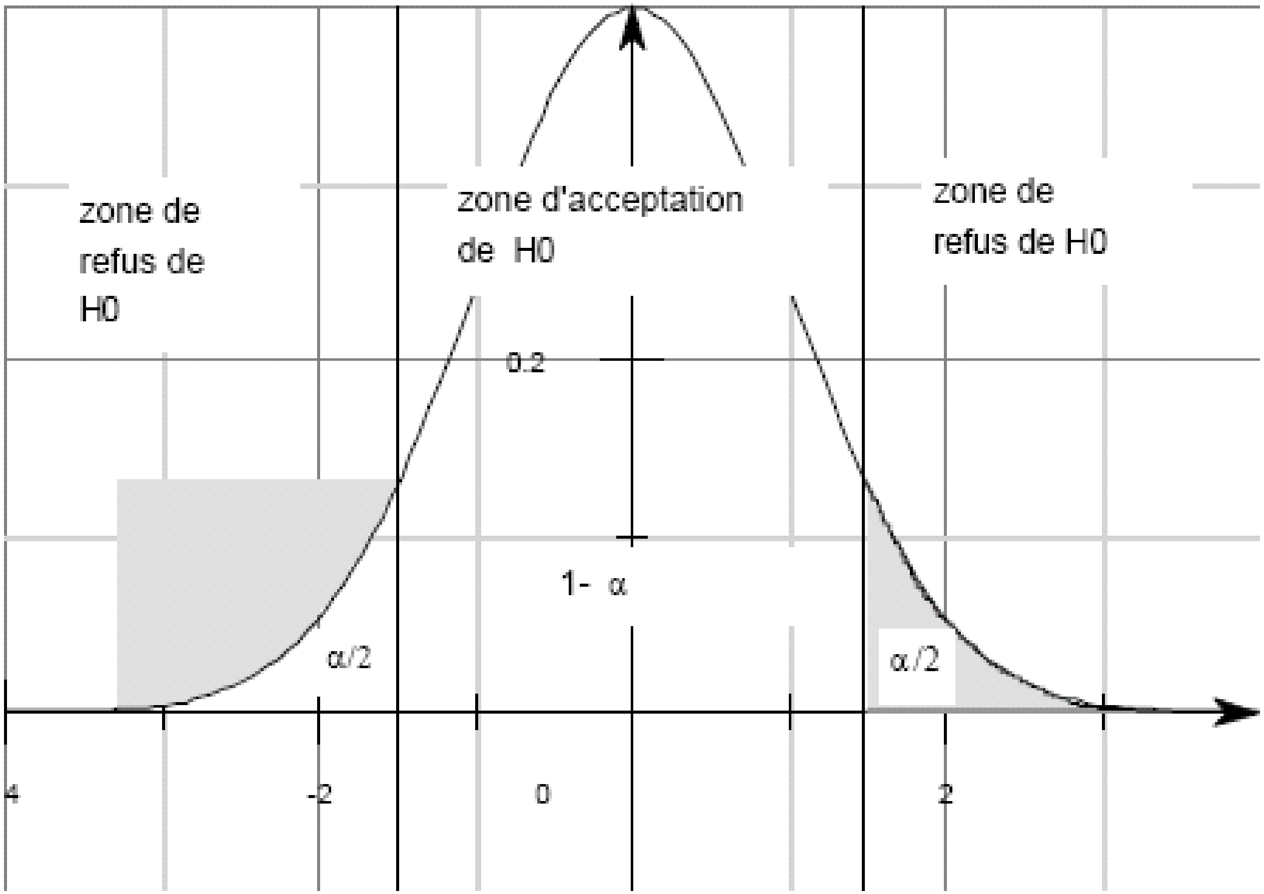
A. Tests Bilatéraux

Soit H_0 : " $m=m_0$ " l'hypothèse nulle et H_1 : " m? m_0 " l'hypothèse alternative
Soit X la variable aléatoire prenant pour valeurs les moyennes des différents échantillons de taille n=30, alors on sait que X suit une $\mathcal{N}(m_0 \,;\, \frac{\sigma}{\sqrt{n}})$, σ étant l'écart-type de la population P.

- On choisit un risque α
- on cherche dans la table de la loi normale centrée réduite $\mathcal{N}(0,1)$, t_α tel que $\pi(t_\alpha) = 1 - \frac{\alpha}{2}$
- soit $\overline{x}_e$ la moyenne de l'échantillon de taille n alors

si $\overline{x}_e \in \left[m_0 - t_\alpha \frac{\sigma}{\sqrt{n}} \,; m_0 + t_\alpha \frac{\sigma}{\sqrt{n}} \right]$, on accepte H_0 avec le risque α

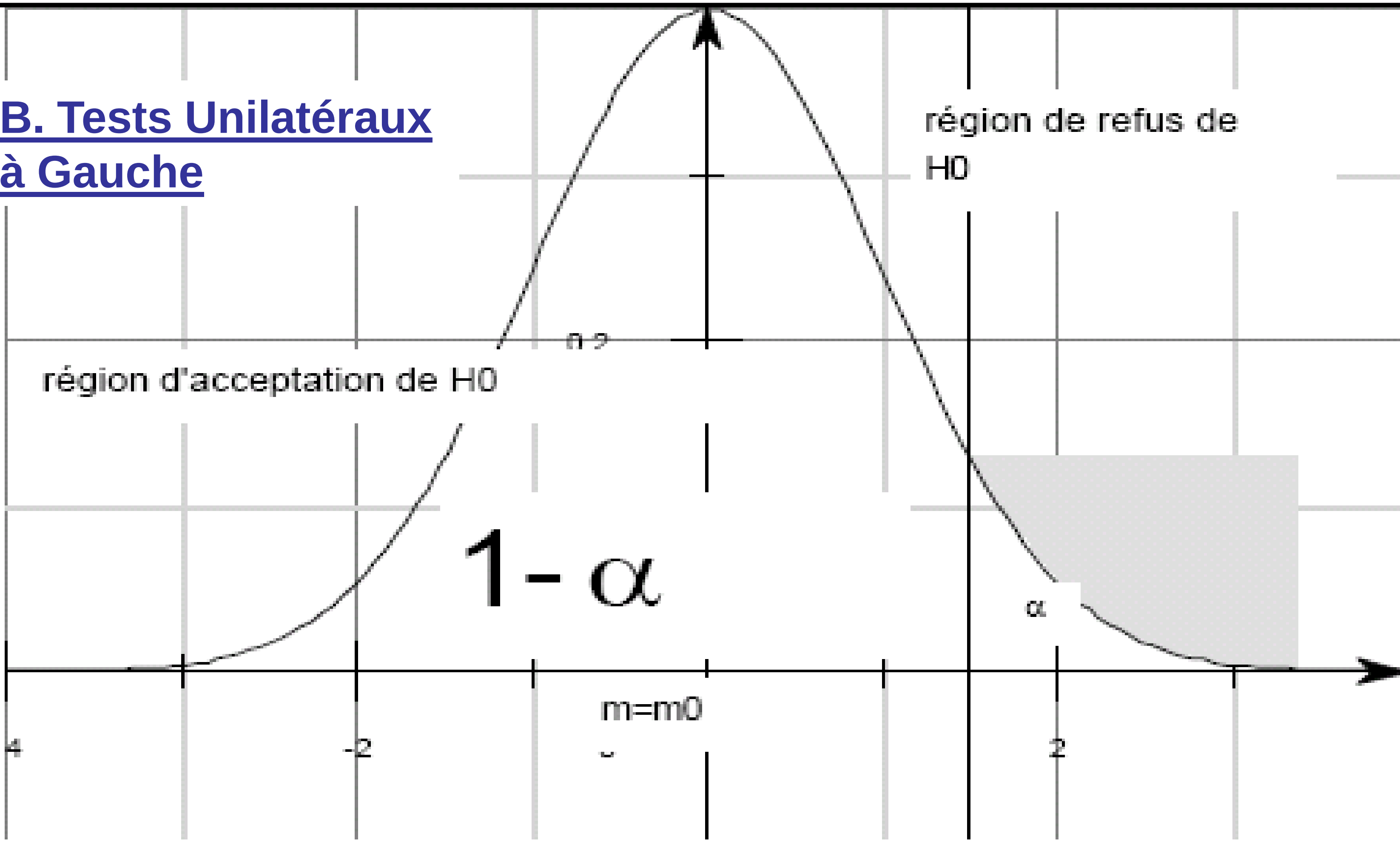
sinon on rejette H_0 et donc on accepte H_1 avec un risque α.



Remarque : dans le cas usuel où $\alpha = 5\%$ alors $t_\alpha = 1{,}96$ et si $\alpha = 1\%$ alors $t_\alpha = 2{,}58$.

Si σ est inconnu (ce qui est souvent le cas) alors on prend son estimateur ponctuel $\sqrt{\frac{n}{n-1}}\sigma_e$ où σ_e est l'écart-type de l'échantillon.

B. Tests Unilatéraux à Gauche

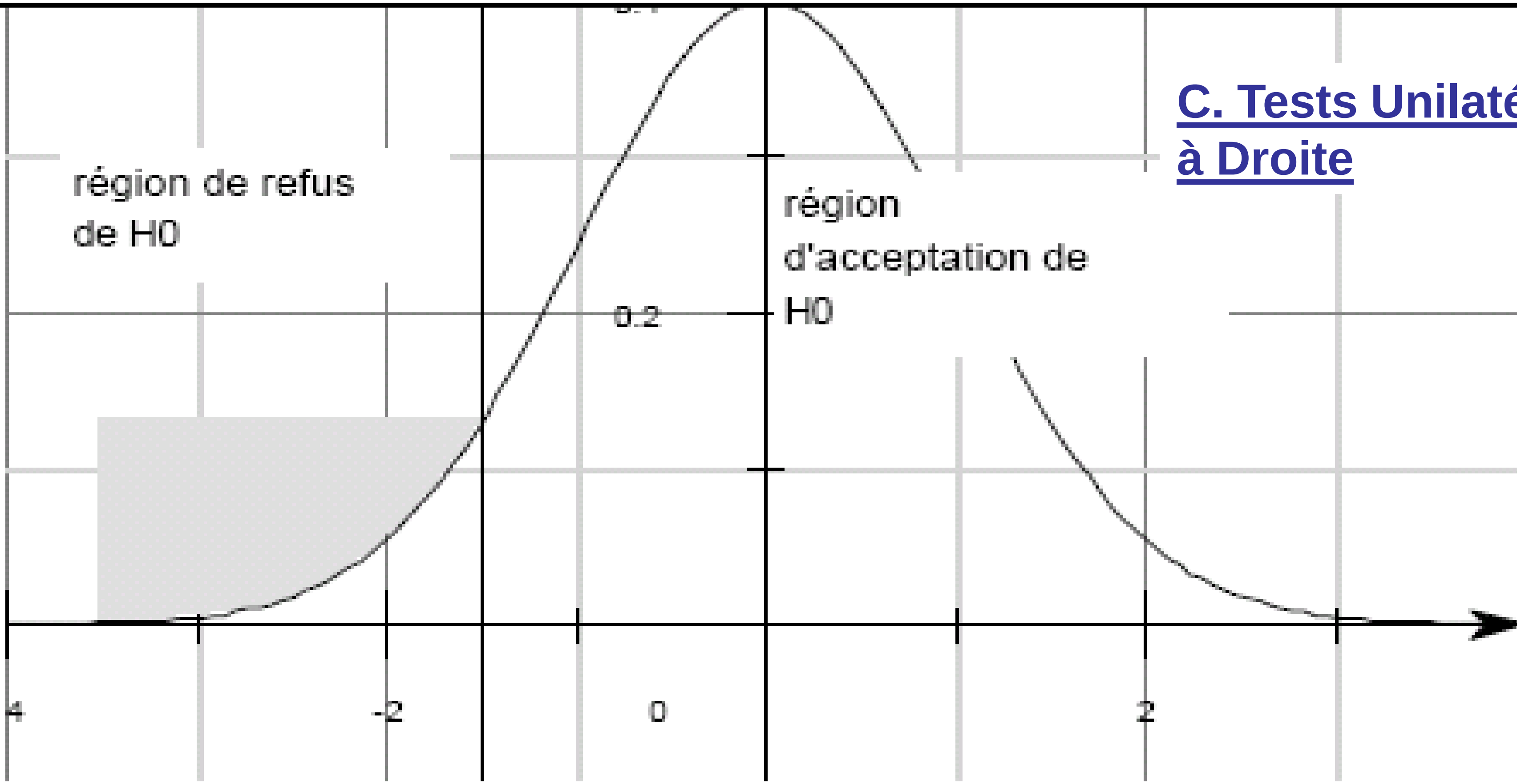


L'hypothèse nulle est H_0 : "$m = m_0$" et l'hypothèse alternative est H_1 : "$m > m_0$"
On peut la retrouver par exemple dans le cas d'un test de dépassement d'une norme.

- On choisit un risque α
- On cherche dans la table de la loi normale centrée réduite $\mathcal{N}(0;1)$ t_α tel que $\pi(t_\alpha) = 1 - \alpha$
- Si $\overline{x}_e \leq m_0 + t_\alpha \dfrac{\sigma}{\sqrt{n}}$ alors on accepte H_0 sinon on refuse H_0 et donc on accepte H_1

Rem : dans les cas usuels : si $\alpha = 5\%$ alors $t_\alpha = 1{,}645$; si $\alpha = 1\%$ alors $t_\alpha = 2{,}33$

C. Tests Unilatéraux à Droite



L'hypothèse nulle est : H_0 : "$m = m_0$" et l'hypothèse alternative est : H_1 : "$m < m_0$"
On la retrouve dans les cas de tests de non égalité d'une norme.

- On choisit un risque α
- On cherche dans la table de la loi normale centrée réduite $\mathbf{N}(0;1)$ t_α tel que $\pi(t_\alpha) = 1 - \alpha$
- Si $\overline{x}_e > m_0 - t_\alpha \dfrac{\sigma}{\sqrt{n}}$ on accepte H_0, sinon on rejette H_0 et donc on accepte H_1

Dans ces deux cas, il est très fréquent qu'on ne connaisse pas σ ; on a alors $\dfrac{\sigma}{\sqrt{n}} = \dfrac{\sigma_e}{\sqrt{n-1}}$

Tous les tests que l'on vient de voir restent valables ; il suffit de remplacer m par p (proportion inconnue dans la population P), $\bar{x}_e$ par f_e (proportion effective sur l'échantillon) et

$$\frac{\sigma}{\sqrt{n}} \text{ par } \sqrt{\frac{f_e\left(1-f_e\right)}{n-1}}$$

Un test bilatéral ne peut être utilisé à la place d'un test unilatéral dans la mesure où la recherche de t_α ne se fait pas de la même façon. D'autre part le test bilatéral est un test « draconien ». C'est un test de non dépassement d'un côté ou de l'autre d'une norme. On peut imaginer par exemple la fabrication de pièces usinées. Les tests unilatéraux ne limitent eux que d'un seul côté.

	Population P1		Population P2	
Caractères Etudiés	C		C	
Moyenne Ecart-type	m_1 σ_1	inconnus	m_2 σ_2	inconnus
	Echantillon e_1		Echantillon e_2	
Taille Moyenne Ecart-type	$n_1=30$ $\bar{x}_{e_1}$ σ_{e_1}	connus	$n_2=30$ $\bar{x}_{e_2}$ σ_{e_2}	connus

Règle du test de comparaison de 2 moyennes

L'hypothèse nulle est : H_0 : "$m_1 = m_2$" et l'hypothèse alternative est H_1 : "$m_1 ? m_2$"

- On choisit un risque α
- On cherche dans la table de la loi normale centrée réduite $N(0;1)$ t_α tel que $\pi(t_\alpha) = 1- \alpha/2$
- Si $\dfrac{\bar{x}_{e_1} - \bar{x}_{e_2}}{\sqrt{\dfrac{\sigma_{e_1}^2}{n_1 - 1} + \dfrac{\sigma_{e_2}^2}{n_2 - 1}}} \in \left[-t_\alpha ; t_\alpha\right]$ on accepte H_0 sinon on rejette H_0 et donc on accepte H_1
- Si H_0 est acceptée, on dit que la différence m_1-m_2 n'est pas significative au risque α

	Population P_1	Population P_2
Caractère étudié Pourcentage	C p_1 inconnu	C p_2 inconnu
	Echantillon e_1	Echantillon e_2
Taille Pourcentage	$n_1=30$ f_1 connu	$n_2=30$ f_2 connu

L'hypothèse nulle est H_0 : "$p_1=p_2$" et l'hypothèse alternative H_1 : "$p_1 ? p_2$"

- On choisit un risque α
- On cherche dans la table de la loi normale centrée réduite $\mathcal{N}(0;1)$ t_α tel que $\pi(t_\alpha) = 1- \alpha/2$
- Soit $f = \dfrac{n_1f_1 + n_2f_2}{n_1 + n_2}$ alors si $\dfrac{f_1 - f_2}{\sqrt{f(1-f)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \in [-t_\alpha ; t_\alpha]$ on accepte H_0 sinon on rejette H_0

  et on accepte donc H_1.
- Si H_0 est acceptée, on dit que $p_1 - p_2$ n'est pas significative au risque α.