Haute école de gestion de Genève
Geneva School of Business Administration

Test paramétrique

Dr Sacha Varone

Objectif

- Connaître les principales définitions
- Comprendre le principe d'un test statistique paramétrique

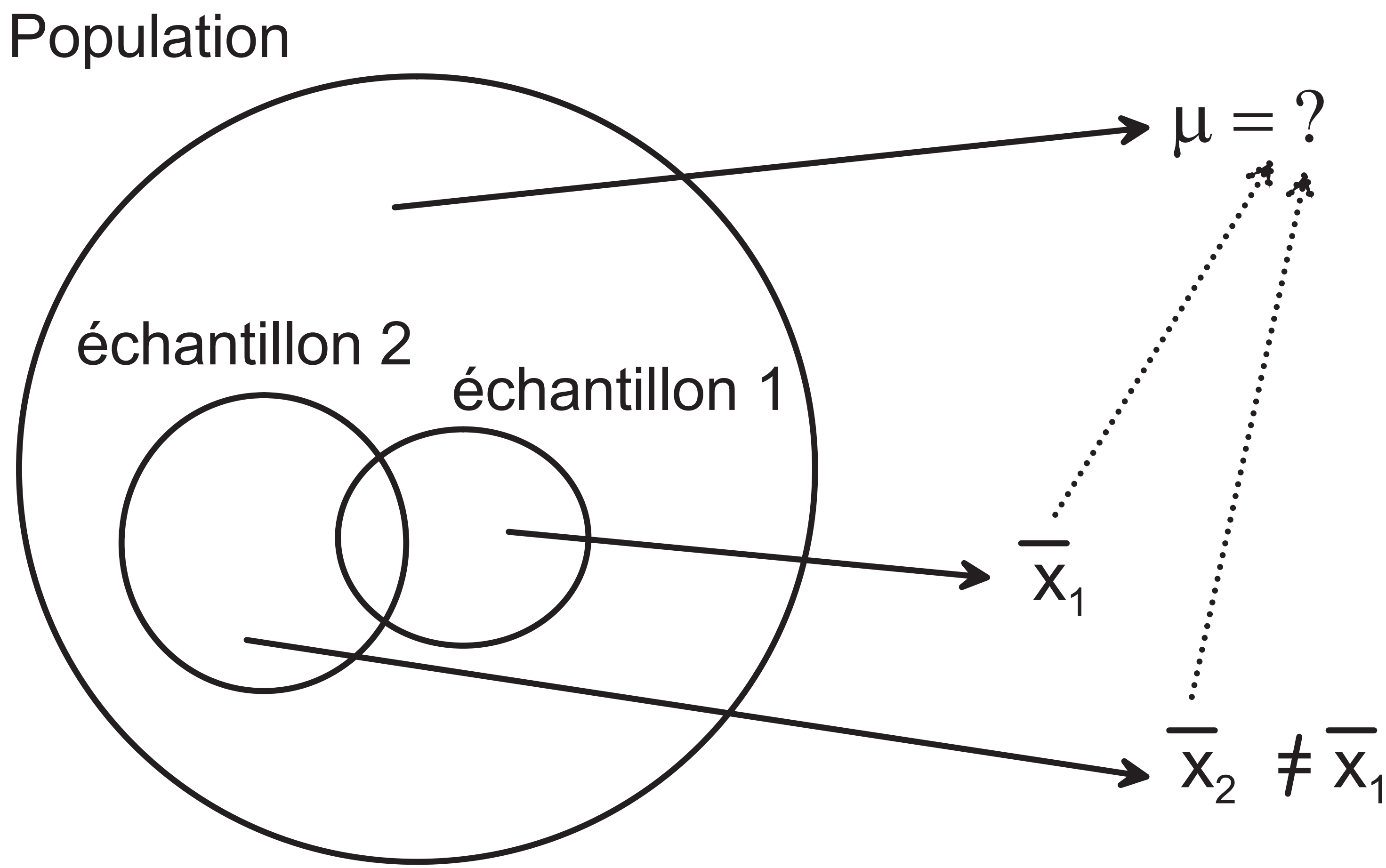
Rappels

Intervalle de confiance

estimé	hypothèse	intervalle
μ	σ^2 connu, distr. normale ou $n \geq 30$	$\mu \in \bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
	σ^2 inconnu, distr. normale	$\mu \in \bar{x} \pm t_{\frac{\alpha}{2},n-1} \frac{s}{\sqrt{n}}$
π	$n\bar{p} \geq 5$ et $n(1-\bar{p}) \geq 5$	$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$
σ^2	μ connu, distr. normale	$\left[\frac{\sum(x_i-\mu)^2}{q_{\frac{\alpha}{2},n}} ; \frac{\sum(x_i-\mu)^2}{q_{1-\frac{\alpha}{2},n}}\right]$
	μ inconnu, distr. normale	$\left[\frac{\sum(x_i-\bar{x})^2}{q_{\frac{\alpha}{2},n-1}} ; \frac{\sum(x_i-\bar{x})^2}{q_{1-\frac{\alpha}{2},n-1}}\right]$

Test d'hypothèse

Illustration



But de l'estimation : quantifier
But de l'inférence : valider/invalider

Type de test

	paramétriques	non paramétriques
Données	distribuées selon une loi particulière	pas de distribution particulière
Puissance	+	-
Risque erreur	-	+
Exemple	moyenne, variance	médiane

Hypothèse nulle/alternative

Hypothèse nulle et alternative

Principe : choisir entre deux hypothèses

Hypothèse nulle H_0 = affirmation testée.
Hypothèse alternative H_1 = ensemble des valeurs non couvertes par l'hypothèse nulle (non H_0).

- H_0 rejeté si contradiction suffisamment évidente.
- Sinon, non rejet de H_0

Exemple

Supposition : les PME suisses ont en moyenne $\mu = 35$ employés avec une variance égale à 220.
Échantillon aléatoire de taille 20 : $\bar{x} = 27$, $s^2 = 334.7$

2	2	5	6	7	8	12	14	23	26
28	31	40	42	46	47	48	49	52	52

Question : statistiquement admissible que $\mu = 35$?

Exemple

Supposition : les PME suisses ont en moyenne $\mu = 35$ employés avec une variance égale à 220.
Échantillon aléatoire de taille 20 : $\bar{x} = 27$, $s^2 = 334.7$

2	2	5	6	7	8	12	14	23	26
28	31	40	42	46	47	48	49	52	52

Question : statistiquement admissible que $\mu = 35$?

$$\begin{aligned} H_0 &: \mu = 35 \\ H_1 &: \mu \neq 35 \end{aligned}$$

Règles de formulation

1. H_0 et H_1 formulées en termes du paramètre de la population d'intérêt.

2. H_0 = statu quo, condition supposée exister
   H_1 si suffisamment d'évidence de changement.

3. H_0 contient $=$ ou $\geq$ ou $\leq$.

Exemple

Étudiants en emploi $\longrightarrow$ horaire tenant compte de leur situation.
(i.e. travail jusqu'à 25h par semaine, $\approx$ 60%).
La situation réelle a-t-elle changé ?

1. Déterminer le paramètre de la population d'intérêt

Exemple

Étudiants en emploi $\longrightarrow$ horaire tenant compte de leur situation.
(i.e. travail jusqu'à 25h par semaine, $\approx$ 60%).
La situation réelle a-t-elle changé ?

1. Déterminer le paramètre de la population d'intérêt
   Nombre d'heures de travail moyen par semaine des étudiants en emploi.

2. Définir la situation qui est supposée vraie.

Exemple

Étudiants en emploi $\rightarrow$ horaire tenant compte de leur situation. (i.e. travail jusqu'à 25h par semaine, $\approx$ 60%).
La situation réelle a-t-elle changé ?

1. Déterminer le paramètre de la population d'intérêt
   Nombre d'heures de travail moyen par semaine des étudiants en emploi.

2. Définir la situation qui est supposée vraie.
   Statu quo $\mu \leq 25$h.

3. Formuler l'hypothèse nulle et l'hypothèse alternative

Exemple

Étudiants en emploi $\rightarrow$ horaire tenant compte de leur situation.
(i.e. travail jusqu'à 25h par semaine, $\approx$ 60%).
La situation réelle a-t-elle changé ?

1. Déterminer le paramètre de la population d'intérêt
   Nombre d'heures de travail moyen par semaine des étudiants en emploi.

2. Définir la situation qui est supposée vraie.
   Statu quo $\mu \leq 25$h.

3. Formuler l'hypothèse nulle et l'hypothèse alternative

$$\begin{aligned} H_0 &: \quad \mu \leq 25 \\ H_1 &: \quad \mu > 25 \end{aligned}$$

Remarque

Une *hypothèse simple* correspond à une valeur spécifique, une situation déterminée (une hypothèse sur une égalité). Une *hypothèse composite* correspond à un ensemble de valeurs, de situations (une hypothèse sur une inégalité).

Région critique

Zone critique

Principe du test : rejet de H_0 si la valeur de la statistique Q_0 est trop différente du paramètre

Deux zones à définir

- A : ensemble des valeurs probables de Q_0 lorsque l'hypothèse H_0 est vraie.
  Il s'agit de la région d'acceptation de H_0.

- R : ensemble des valeurs peu probables de Q_0 lorsque l'hypothèse H_0 est vraie.
  Il s'agit de la région de rejet de H_0 (région critique).

Zone critique

Principe du test : rejet de H_0 si la valeur de la statistique Q_0 est trop différente du paramètre

Deux zones à définir

- A : ensemble des valeurs probables de Q_0 lorsque l'hypothèse H_0 est vraie.
  Il s'agit de la région d'acceptation de H_0.
- R : ensemble des valeurs peu probables de Q_0 lorsque l'hypothèse H_0 est vraie.
  Il s'agit de la région de rejet de H_0 (région critique).

Règle de décision

$$q_0 \notin R \iff \text{Acceptation de } H_0$$

$$q_0 \in R \iff \text{Rejet de } H_0$$

Jeu

Un meneur de jeu, plusieurs joueurs.

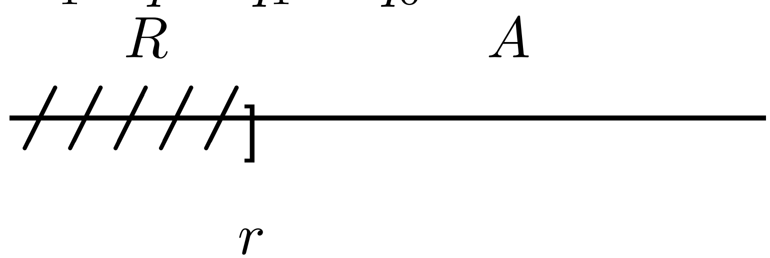
Principe : loterie

1. Le meneur choisit 5 nombres secrets entre 1 et 10.
2. Un joueur annonce son nombre
3. Le meneur de jeu vérifie et annonce au joueur s'il a gagné
4. Retour en 1.

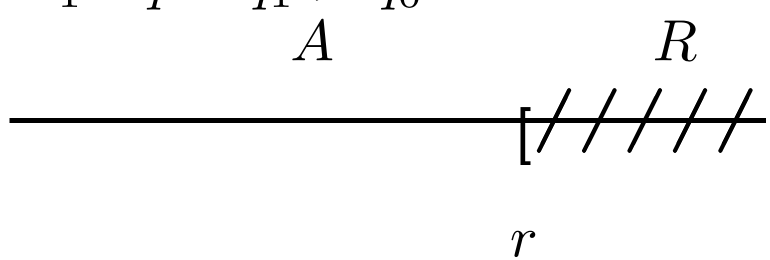
Après quelques itérations, analyser les résultats.

Forme de la région critique

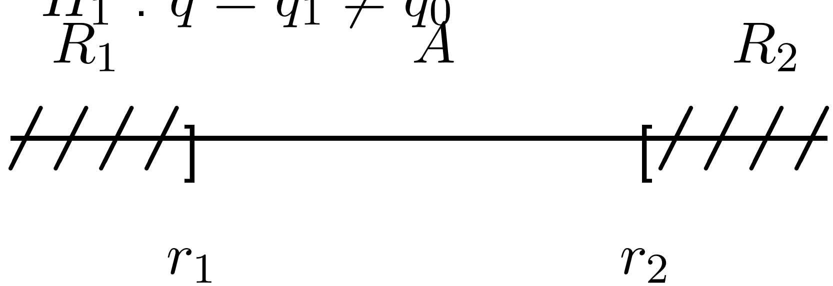
- test unilatéral à gauche
  $H_1 : q = q_1 < q_0$



- test unilatéral à droite
  $H_1 : q = q_1 > q_0$



- test bilatéral
  $H_1 : q = q_1 \neq q_0$



Valeur critique

La *valeur critique* r, aussi appelée seuil critique, est la valeur d'une statistique correspondant à un certain niveau de signification.

But : déterminer la frontière pour le rejet de H_0
Choix : de façon à limiter le risque d'erreur.

source : "The Cartoon Guide to Statistics", L. Gonick & W. Smith

Risques

Type de risque

Risque de première espèce α (risque de type I)
Rejet de H_0 alors qu'elle est en fait vraie.

$$\alpha = P(Q_0 \in R \mid H_0)$$

Type de risque

Risque de première espèce α (risque de type I)
Rejet de H_0 alors qu'elle est en fait vraie.

$$\alpha = P(Q_0 \in R \mid H_0)$$

Risque de deuxième espèce β (risque de type II)
Accepter H_0 alors qu'elle est en fait fausse.

$$\beta = P(Q_0 \notin R \mid H_1)$$

Type de risque

Risque de première espèce α (risque de type I)
Rejet de H_0 alors qu'elle est en fait vraie.

$$\alpha = P(Q_0 \in R \mid H_0)$$

Risque de deuxième espèce β (risque de type II)
Accepter H_0 alors qu'elle est en fait fausse.

$$\beta = P(Q_0 \notin R \mid H_1)$$

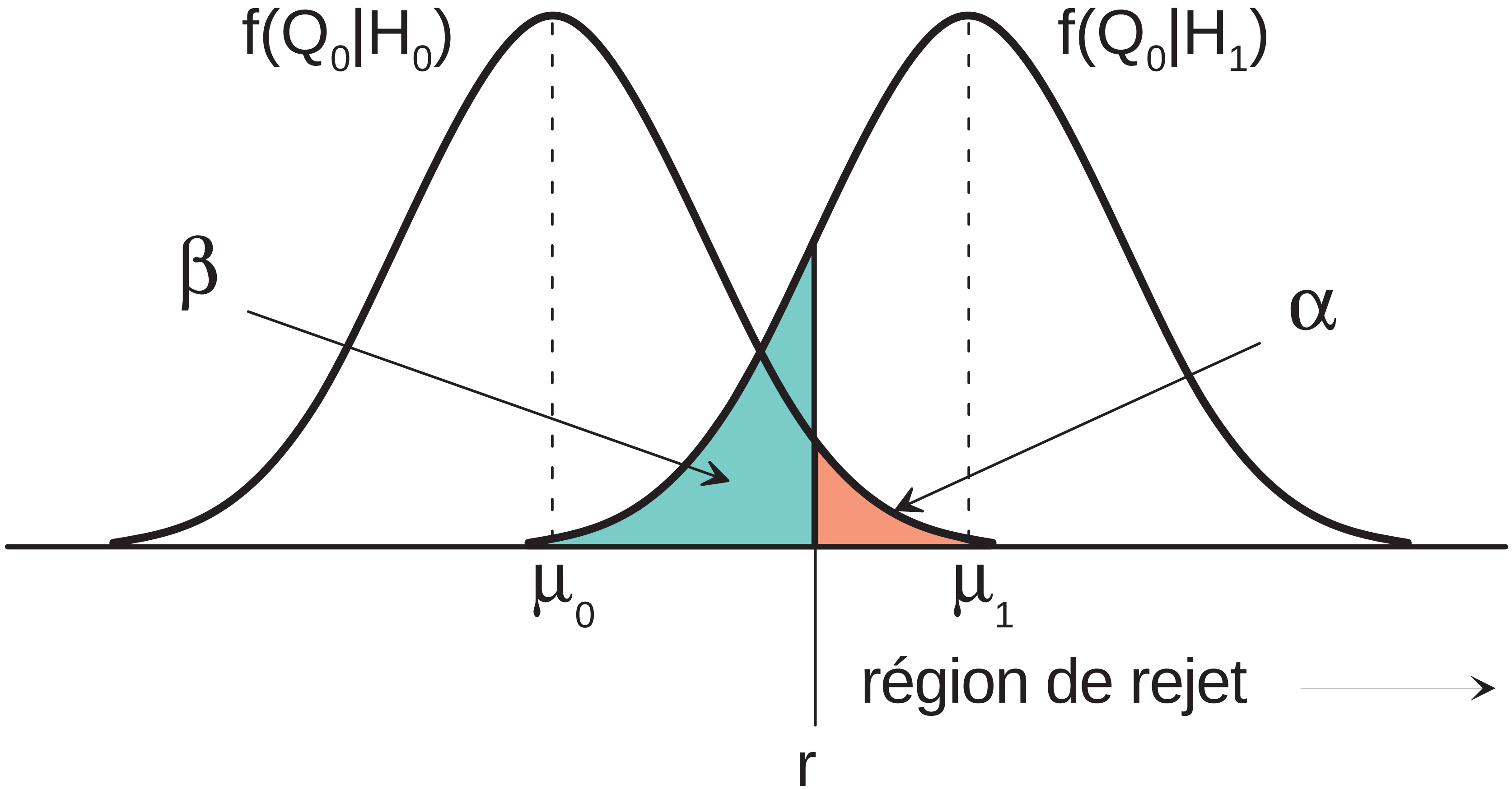
Remarque : α est aussi appelé *niveau de signification*.

Tableau des risques

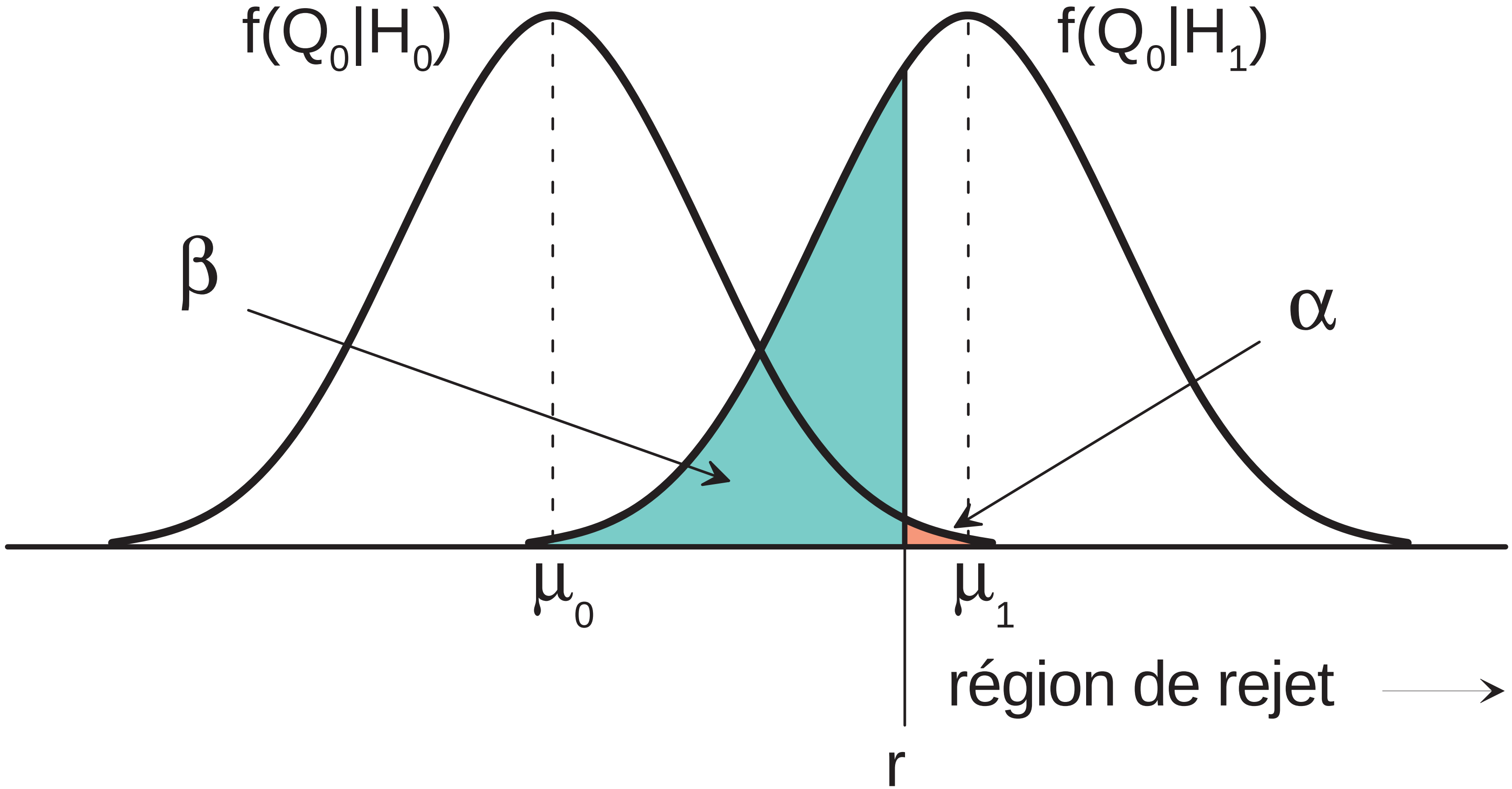
	État de la nature	
	H_0 vraie	H_1 vraie
H_0 accepté	correct	β
H_1 accepté	α	correct

h e g

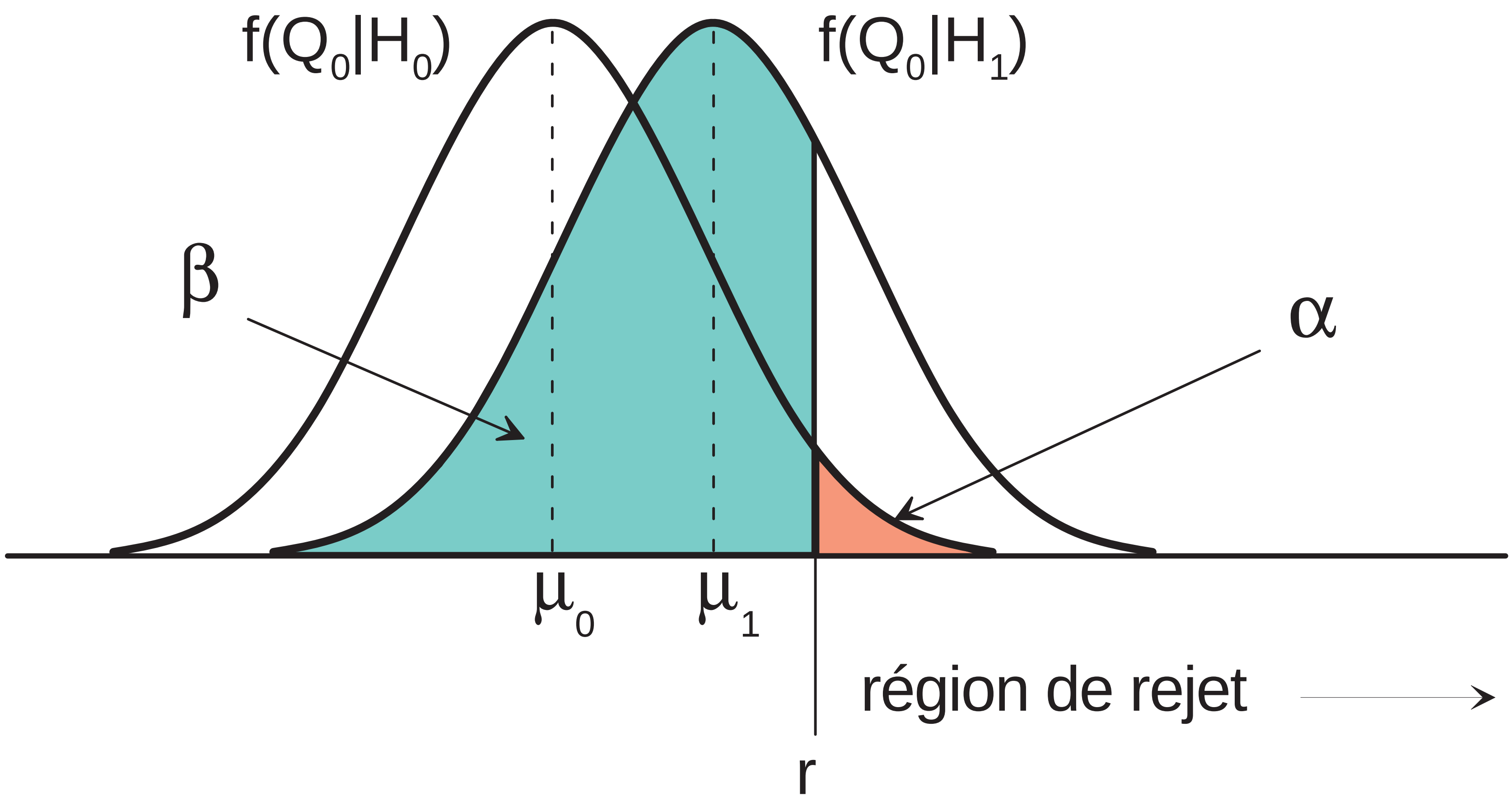
Risques α et β



α petit, donc β grand



H_1 peu différent de $H_0 \Rightarrow \beta$ grand



Risque total

Le *risque total d'erreur* est défini par la relation

$$\alpha \underbrace{P(H_0)}_{\text{inconnu}} + \beta \underbrace{P(H_1)}_{\text{inconnu}}$$

En pratique

En pratique, on détermine le seuil critique r pour un α choisi arbitrairement petit (en général $5\ \%$ ou $10\ \%$).

Test

h e g

Procédure de test

1. Spécifier la valeur de la population d'intérêt.
2. Formuler l'hypothèse nulle H_0 et l'hypothèse alternative H_1
3. Choisir le niveau de signification α
4. Déterminer la région critique.
5. Calculer la statistique associée à l'échantillon.
6. Rejeter H_0 si la statistique appartient à la région critique. Ne pas rejeter H_0 dans le cas contraire.
7. Énoncer une conclusion.