Statistique (MATH-F-315, Cours #3)

Thomas Verdebout

Université Libre de Bruxelles

2015

Plan de la partie Statistique du cours

1. Introduction.
2. Théorie de l'estimation.
3. $\boxed{\text{Tests d'hypothèses et intervalles de confiance.}}$
4. Régression.
5. ANOVA.

Problèmes de test

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un échantillon provenant d'un modèle statistique $\mathcal{P} = \{P_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \Theta\}$. Considérons une partition de Θ en

$$\Theta = H_0 \oplus H_1 \quad (\oplus \text{ désigne une réunion disjointe})$$

où H_0 est appelée l'*hypothèse "nulle"* (en anglais, *null hypothesis*) et H_1 est la *"contre-hypothèse"* (en anglais, *alternative*).

Un *problème de test* est un problème de décision dans lequel deux décisions seulement sont possibles:

(i) Le rejet de l'hypothèse nulle (RH_0) après observation de $\mathbf{X}$;

(ii) Le non-rejet de H_0 (NRH_0) après observation de $\mathbf{X}$.

Définition

Un test est une statistique, traditionnellement notée ϕ, à valeurs dans $\{0, 1\}$, qui sert de règle de décision : $\phi(X) = 0$ signifie NRH_0 tandis que $\phi(X) = 1$ signifie RH_0.

Exemple (échantillon de Bernoulli)
Considérons n jets $X_1, \ldots, X_n$ d'une pièce de monnaie : $\mathrm{P}[\text{“face”}] = p$, $\Theta = [0, 1]$. Supposons qu'on mette en doute le fait que cette pièce soit correctement équilibrée, c'est-à -dire l'hypothèse que $p = 1/2$.

Est-il raisonnable, après observation des résultats des n jets, de rejeter cette hypothèse?

Problèmes de test

Le problème de décision ainsi posé est le *problème de test*

$$\begin{cases} H_0 : p = 1/2 \ : & \text{la pièce est correctement équilibrée} \\ H_1 : p \neq 1/2 & \text{la pièce n'est pas correctement équilibrée.} \end{cases}$$

Nous savons qu'un estimateur raisonnable de p est donné par

$$\hat{p} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

Il paraît donc raisonnable de rejeter l'hypothèse si la proportion observée $\hat{p}$ de résultats "face" est "trop petite" ($\hat{p} < 1/2 - a$) ou "trop grande" ($\hat{p} > 1/2 + a$) par rapport à $1/2$. Cette règle de décision est un test de la forme

$$\phi = \phi(\hat{p}) = \begin{cases} 1 & \text{si } \hat{p} \notin [1/2 \pm a] \quad \text{(proportion observée } \hat{p} \text{ "trop" différente de } 1/2\text{)} \\ 0 & \text{si } \hat{p} \in [1/2 \pm a] \quad \text{(proportion observée de } \hat{p} \text{ "proche" de } 1/2\text{).} \end{cases}$$

Mais comment déterminer a ?

Risques de première et de seconde espèce

Soit ϕ un test (pour un problème associé à un paramètre $\boldsymbol{\theta}$, une hypothèse nulle H_0 et une contre-hypothèse H_1). Deux erreurs peuvent être commises :

- l'*erreur de première espèce*, qui consiste à rejeter l'hypothèse nulle correcte
- l'*erreur de seconde espèce*, qui est celle de ne pas rejeter l'hypothèse nulle fausse.

	$\boldsymbol{\theta} \in H_0$	$\boldsymbol{\theta} \in H_1$
$\phi = 0$ (NRH_0)	décision correcte	erreur (risque) de 2nde espèce
$\phi = 1$ (RH_0)	erreur (risque) de 1ère espèce	décision correcte

La probabilité de commettre chacune de ces deux erreurs est appelée *risque*; ce risque est fonction de la valeur (inconnue) de $\boldsymbol{\theta}$.

Risques de première et de seconde espèce

Définition

Le risque de première espèce est la probabilité de rejeter l'hypothèse quand celle-ci est correcte (quand $\boldsymbol{\theta} \in H_0$):

$$\mathrm{P}_{\boldsymbol{\theta}}[RH_0] = \mathrm{P}_{\boldsymbol{\theta}}[\phi(\boldsymbol{X}) = 1] = \mathrm{E}_{\boldsymbol{\theta}}[\phi] \qquad \textit{quand } \boldsymbol{\theta} \in H_0.$$

Définition

Le risque de seconde espèce est la probabilité de ne pas rejeter l'hypothèse quand celle-ci est fausse (quand $\boldsymbol{\theta} \in H_1$):

$$\mathrm{P}_{\boldsymbol{\theta}}[NRH_0] = 1 - \mathrm{P}_{\boldsymbol{\theta}}[\phi(\boldsymbol{X}) = 1] = 1 - \mathrm{E}_{\boldsymbol{\theta}}[\phi] \qquad \textit{quand } \boldsymbol{\theta} \in H_1.$$

Le test idéal serait celui qui minimiserait à la fois les risques de première et de seconde espèce.

Risques de première et de seconde espèce

Pour minimiser à la fois les risques de première et de seconde espèce, il faudrait minimiser $\mathrm{E}_{\boldsymbol{\theta}}[\phi]$ pour $\boldsymbol{\theta} \in H_0$ et minimiser $1 - \mathrm{E}_{\boldsymbol{\theta}}[\phi]$ pour $\boldsymbol{\theta} \in H_1$ (ou de façon équivalente maximiser $\mathrm{E}_{\boldsymbol{\theta}}[\phi]$ pour $\boldsymbol{\theta} \in H_1$)

Ces deux objectifs ne peuvent être atteints simultanément. En effet,

- le test $\phi \equiv 0$ (ne jamais rejeter) minimise $\mathrm{E}_{\boldsymbol{\theta}}[\phi]$ pour $\boldsymbol{\theta} \in H_0$
- tandis que le test $\phi \equiv 1$ (toujours rejeter) maximise $\mathrm{E}_{\boldsymbol{\theta}}[\phi]$ pour $\boldsymbol{\theta} \in H_1$.

Le Principe de Neyman

Afin de résoudre ce dilemme, le principe ci-dessous, appelé *Principe de Neyman*, est appliqué. Définissons tout d'abord le concept de puissance.

Définition

La puissance d'un test est donc la probabilité $\mathrm{P}_{\boldsymbol{\theta}}[\phi(\boldsymbol{X}) = 1] = \mathrm{E}_{\boldsymbol{\theta}}[\phi]$, $\boldsymbol{\theta} \in H_1$ pour que celui-ci rejette une hypothèse fausse.

Le Principe de Neyman consiste à

(a) se restreindre aux tests ϕ *de niveau* α *sur* H_0, c'est à dire aux tests satisfaisant la *contrainte de niveau*

$$\mathrm{E}_{\boldsymbol{\theta}}[\phi] \leq \alpha, \qquad \boldsymbol{\theta} \in H_0$$

où α est fixé à l'avance (valeurs usuelles : 0.01; 0.05; 0.01; 0.001).

(b) parmi les test de niveau α, choisir celui (existence ? unicité ?) qui maximise la puissance uniformément en $\boldsymbol{\theta} \in H_1$.

Le Principe de Neyman

Ceci conduit à la définition de la notion de *test à puissance uniformément maximum (PUM*; en anglais, *uniformly most powerful, UMP) dans la classe des tests de niveau* α*.*

Definition

Un test ϕ^* est à PUM dans la classe des test de niveau α (pour H_0 contre H_1) si

(i) ϕ^* est de niveau α : $\mathrm{E}_{\boldsymbol{\theta}}[\phi^*] \leq \alpha \quad$ pour tout $\boldsymbol{\theta} \in H_0$;

(ii) ϕ^* est au moins aussi puissant, en tout $\boldsymbol{\theta} \in H_1$, que tout autre test ϕ satisfaisant à la contrainte de niveau (i) : pour tout ϕ tel que $\mathrm{E}_{\boldsymbol{\theta}}[\phi] \leq \alpha$ en tout $\boldsymbol{\theta} \in H_0$, $\mathrm{E}_{\boldsymbol{\theta}}[\phi] \leq \mathrm{E}_{\boldsymbol{\theta}}[\phi^*]$ pour tout $\boldsymbol{\theta} \in H_1$.

Le problème de l'existence de tests PUM et de leur construction n'est pas toujours simple (et des tests PUM n'existent pas pour tous les problèmes). Dans la suite, nous décrirons les tests les plus usuels.

Echantillon gaussien, variance spécifiée : test d'une moyenne

Soient $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$, de variance σ^2 spécifiée.

Problème de test (unilatéral): $H_0 : \ \mu \leq \mu_0$ et $H_1 : \ \mu > \mu_0$

Statistique de test : $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ ou $Z := \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Loi sous $\mu = \mu_0$: $\bar{X} \sim \mathcal{N}(\mu_0, \sigma^2/n) \Longrightarrow Z \sim \mathcal{N}(0,1)$

Règle de comportement : $\phi = 1$ (RH_0) si

$$Z > z_{1-\alpha} \quad \text{ou, de façon équivalente, si} \quad \bar{X} > \mu_0 + z_{1-\alpha}\frac{\sigma_0}{\sqrt{n}}$$

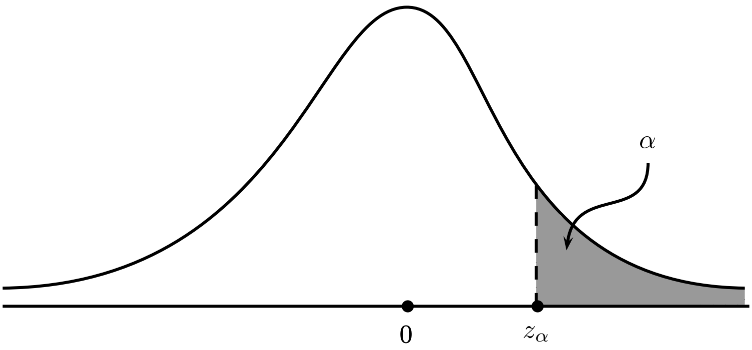
où $z_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la variable normale centrée réduite.

Le contenu intuitif de cette règle de comportement est clair : on rejette H_0 si $\bar{X}$ est "trop grand" par rapport à μ_0, la magnitude du "trop grand" étant calibrée par la condition de niveau.

Echantillon gaussien, variance spécifiée : test d'une moyenne

La condition de niveau est satisfaite. En effet,

- en $\mu = \mu_0$, $\mathrm{E}_{\mu_0}(\phi) = \mathrm{P}_{\mu_0}(\phi = 1) = \mathrm{P}_{\mu_0}(RH_0) = \mathrm{P}[Z > z_{1-\alpha}] = \alpha$;

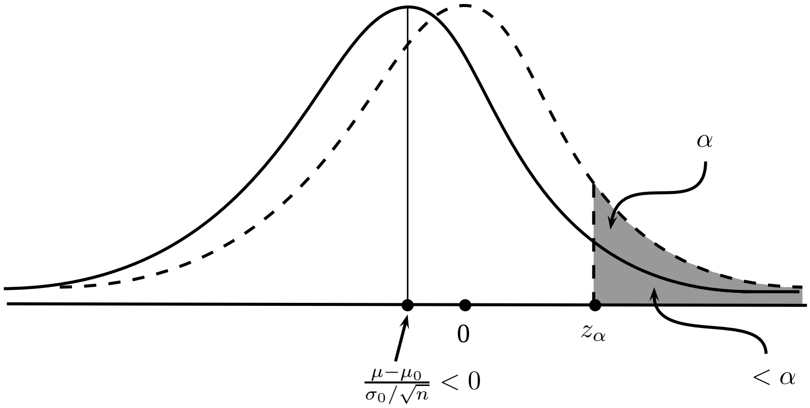


En gris : Probabilité de rejet sous $\mu = \mu_0$.

Echantillon gaussien, variance spécifiée : test d'une moyenne

- en $\mu < \mu_0$, $\mathrm{E}_\mu(\phi) = \mathrm{P}_\mu(\phi = 1) = \mathrm{P}_\mu(RH_0) < \alpha$. En effet,

$$Z = \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} = \underbrace{\frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}}}_{\mathcal{N}(0,1)} + \underbrace{\frac{\mu - \mu_0}{\sigma_0/\sqrt{n}}}_{<0} \sim \mathcal{N}\left(\underbrace{\frac{\mu - \mu_0}{\sigma_0/\sqrt{n}}}_{<0}, 1\right)$$



Echantillon gaussien, variance spécifiée : test d'une moyenne

Ou de manière analytique:

$$\mathrm{P}_\mu(\phi = 1) = \mathrm{P}_\mu(RH_0) = \mathrm{P}\left[\mathcal{N}\left(\frac{\mu - \mu_0}{\sigma_0/\sqrt{n}}, 1\right) > z_{1-\alpha}\right]$$

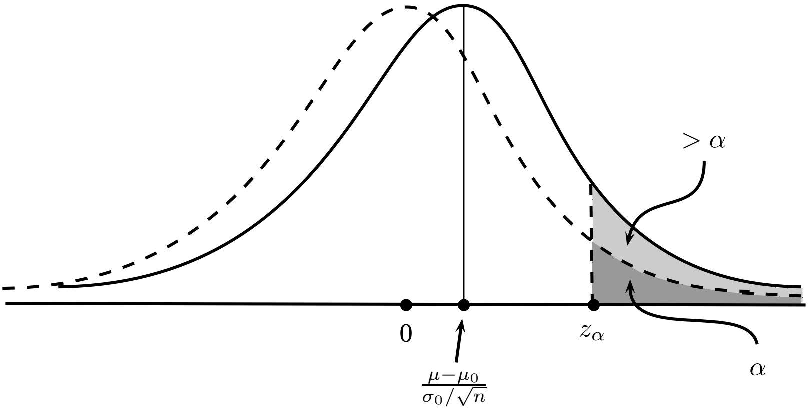
$$= \mathrm{P}\left[\mathcal{N}(0,1) > \underbrace{z_{1-\alpha} - \underbrace{\frac{\mu - \mu_0}{\sigma_0/\sqrt{n}}}_{<0}}_{>z_{1-\alpha}}\right] < \alpha.$$

Donc le test satisfait à la condition de niveau.

Echantillon gaussien, variance spécifiée : test d'une moyenne

De même, en $\mu > \mu_0$,

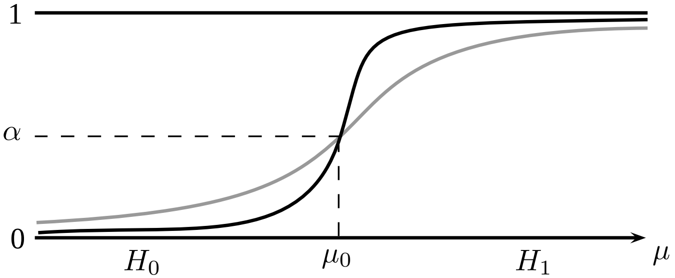
$$Z = \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} = \underbrace{\frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}}}_{\mathcal{N}(0,1)} + \underbrace{\frac{\mu - \mu_0}{\sigma_0/\sqrt{n}}}_{>0} \sim \mathcal{N}\Bigg(\underbrace{\frac{\mu - \mu_0}{\sigma_0/\sqrt{n}}}_{>0}, 1\Bigg)$$



Echantillon gaussien, variance spécifiée : test d'une moyenne

$$
\begin{aligned}
\mathrm{P}_\mu(\phi = 1) = \mathrm{P}_\mu(RH_0) &= \mathrm{P}\left[\mathcal{N}\left(\frac{\mu-\mu_0}{\sigma_0/\sqrt{n}}, 1\right) > z_{1-\alpha}\right] \\
&= \mathrm{P}\left[\mathcal{N}(0,1) > z_{1-\alpha} - \underbrace{\frac{\mu-\mu_0}{\sigma_0/\sqrt{n}}}_{<0}\right] > \alpha \\
&\longrightarrow 1 \text{ pour } \mu \to \infty \text{ ou pour } n \to \infty.
\end{aligned}
$$

Le graphe de la fonction $\mu \mapsto \mathrm{E}_\mu(\phi)$ pour ce test (noir) est de la forme



En gris : la puissance d'un autre test de même niveau, uniformément moins puissant

Echantillon gaussien, variance spécifiée : test d'une moyenne

La règle de décision pour ce test est

$$\begin{aligned} RH_0 \text{ si } Z > z_{1-\alpha} \quad &\Longleftrightarrow F_{\mathcal{N}(0,1)}(Z) > F_{\mathcal{N}(0,1)}(z_{1-\alpha}) \\ &\Longleftrightarrow 1 - \Phi(Z) < \alpha \end{aligned}$$

où $F_{\mathcal{N}(0,1)}$ est la fonction de répartition (notée Φ) de la variable $\mathcal{N}(0,1)$, ce qui implique $\Phi(z_{1-\alpha}) = 1 - \alpha$.

Definition

$1 - \Phi(Z)$ est appelée *p-valeur* du test (c'est une statistique, donc une quantité aléatoire, fonction de $(X_1, \ldots, X_n)$).

La règle de décision peut donc s'énoncer aussi sous la forme

$$RH_0 \text{ si } p\text{-valeur}(X_1, \ldots, X_n) < \alpha.$$

Echantillon gaussien, variance non spécifiée : test d'une moyenne

Considérons à présent le cas où $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$ où $\left(\mu, \sigma^2\right) \in \mathbb{R} \times \mathbb{R}_+$.

Problème de test :

$$\begin{cases} H_0 \ : \ \mu \leq \mu_0 & (\sigma^2 \text{ non spécifié}) \\ H_1 \ : \ \mu > \mu_0 & (\sigma^2 \text{ non spécifié}), \end{cases}$$

c'est-à -dire, sous forme de partition de Θ,

$$\begin{cases} H_0 = (-\infty, \mu_0] \times (0, \infty) \\ H_1 = (\mu_0, \infty) \times (0, \infty); \end{cases}$$

σ^2 est appelé *paramètre de nuisance*; μ est appelé *paramètre d'intérêt*.

Grâce au lemme de Fisher:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \qquad et \qquad \frac{ns^2}{\sigma^2} \sim \chi^2_{n-1},$$

il découle donc du lemme de Fisher que

$$\frac{\frac{\sqrt{n}\,(\bar{X} - \mu)}{\sigma}}{\sqrt{\left(\frac{ns^2}{\sigma^2}\right) / (n-1)}} = \frac{(\bar{X} - \mu)}{s/\sqrt{n-1}} = \frac{(\bar{X} - \mu)}{S/\sqrt{n}} \sim t_{n-1}.$$

Echantillon gaussien, variance non spécifiée : test d'une moyenne

Le *test de Student (unilatéral) pour* un *échantillon ((one-sided) one-sample Student test)* est donné par

Statistique de test : $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n-1}}$

Loi sous $\mu = \mu_0$: $T \sim t_{n-1}$

Règle de comportement : $\phi = 1$ (RH_0) si

$$T > t_{n-1;1-\alpha} \quad \text{ou, de façon équivalente, si} \quad \bar{X} > \mu_0 + t_{n-1;1-\alpha} \frac{S}{\sqrt{n}}$$

où $t_{n-1;1-\alpha}$ est le quantile d'ordre $(1-\alpha)$ de la variable de Student à $n-1$ degrés de liberté.

Le contenu intuitif de cette règle de comportement est exactement le même que dans l'exemple précédent.

Echantillon gaussien, variance non spécifiée : test d'une moyenne

On peut montrer que ce test

- satisfait à la condition de niveau
- est à puissance uniformément maximale dans la classe des *tests sans biais de niveau* α, c'est-à -dire la classe des test ϕ satisfaisant

$$\begin{cases} \mathrm{E}_\theta[\phi] \leq \alpha & \theta \in H_0 \\ \mathrm{E}_\theta[\phi] \geq \alpha & \theta \in H_1. \end{cases}$$

Remarque 1 : Ici encore, la règle de comportement peut être décrite en faisant intervenir la notion de p-valeur; celle-ci se définit comme $F_{t_{n-1}}(T)$, où $F_{t_{n-1}}$ est la fonction de répartition de la variable de Student à $n-1$ degrés de liberté.

Remarque 2 : Pour n grand ($n \geq 30$), $T \sim t_{n-1} \approx \mathcal{N}(0,1)$, et les quantiles $t_{n-1;1-\alpha}$ peuvent être remplacés par les quantiles gaussiens $z_{1-\alpha}$; grâce au théorème central-limite, l'hypothèse gaussienne peut donc être abandonnée.

Echantillon gaussien, variance non spécifiée : test d'une moyenne

Remarque 3 : Le problème unilatéral symétrique du précédent

$$\begin{cases} H_0 & : \ \mu \geq \mu_0 \\ H_1 & : \ \mu < \mu_0 \end{cases}$$

se traite de la même façon, mais le rejet se fait pour les "petites" valeurs de $\bar{X}$ ou de T.

Règle de comportement : $\phi = 1$ (RH_0) si

$$T < t_{n-1;\alpha} \quad \text{ou, de façon équivalente, si} \quad \bar{X} < \mu_0 + t_{n-1;\alpha} \frac{S}{\sqrt{n}}.$$

Remarquons, que la symétrie par rapport à zéro de la fonction de densité de la distribution de Student implique que $t_{n-1;\alpha} = -t_{n-1;1-\alpha}$ ce qui implique RH_0 si $\bar{X} < \mu_0 - t_{n-1;1-\alpha} \frac{S}{\sqrt{n}}$

Ici, la p-valeur doit êre calculée comme une "probabilité à gauche" : p-valeur$(X_1, \ldots, X_n) = F_{t_{n-1}}(T)$. La règle de comportement consiste alors à rejeter l'hypothèse nulle si la p-valeur est inférieure à α.

Remarque 4 : Choix du test unilatéral

Deux problèmes unilatéraux a priori sont possibles :

$$\begin{cases} H_0 & : \quad \mu \leq \mu_0 \\ H_1 & : \quad \mu > \mu_0 \end{cases} \qquad \text{et} \qquad \begin{cases} H_0 & : \quad \mu \geq \mu_0 \\ H_1 & : \quad \mu < \mu_0 \end{cases} .$$

Le choix du problème unilatéral doit se faire sur base des objectifs à atteindre.

Dans un test d'hypothèse, *seul le rejet de l'hypothèse nulle est concluant* car le risque correspondant (risque de première espèce) est contrôlé (au plus égal à α).

Dans le cas d'un non-rejet, le risque de seconde espèce n'est pas contrôlé : il est aussi petit que possible, mais sa valeur, inconnue, peut être proche de un. Le non-rejet est donc une suspension du jugement : les observations ne permettent pas de rejeter l'hypothèse, mais cela n'implique aucune certitude.

Le non-rejet de H_0 ne doit en aucune façon être interprétée comme une "acceptation de H_0".

Echantillon gaussien, variance non spécifiée : test d'une moyenne

Remarque 5 : Problèmes de test bilatéraux

Considérons des hypothèses de la forme

$$\begin{cases} H_0 & : \ \mu = \mu_0 \\ H_1 & : \ \mu \neq \mu_0 \end{cases}$$

(σ^2 non spécifié). On utilise le *test de Student bilatéral (two-sided)*, version bilatérale des tests précédents.

Règle de comportement : $\phi = 1$ (RH_0) si

$T \notin [-t_{n-1;1-\alpha/2}; t_{n-1;1-\alpha/2}]$ ou, de façon équivalente, si $\bar{X} \notin \left[\mu_0 \pm t_{n-1;1-\alpha/2} \dfrac{S}{\sqrt{n}}\right]$.

La notion de p-valeur est un peu plus délicate que dans le cas unilatéral.

Supposons que $T > 0$: on rejette l'hypothèse si $T > t_{n-1;1-\alpha/2}$, c'est-à -dire $(1 - F_{t_{n-1}}(T)) < \frac{\alpha}{2}$, ou encore $2(1 - F_{t_{n-1}}(T)) = 2(1 - F_{t_{n-1}}(|T|)) < \alpha$.

On peut donc définir la p-valeur, dans ce test bilatéral, comme

$$2(1 - F_{t_{n-1}}(|T|),$$

Echantillon gaussien, variance non spécifiée : test d'une moyenne

Le graphe de la fonction $\mu \mapsto \mathrm{E}_\mu(\phi) = \mathrm{P}_\mu(RH_0)$ a la forme typique suivante :

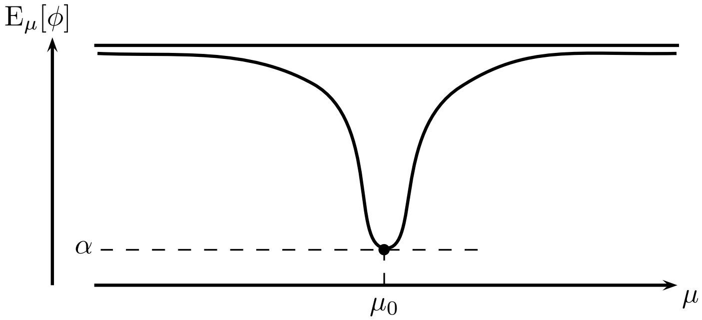


Figure: La fonction $\mu \mapsto \mathrm{E}_\mu(\phi)$ pour le test de Student bilatéral.