Chapitre 1

Tests d'hypothèses composites

D'une manière générale on est amené à considérer des tests où les deux hypothèses en présence ne sont pas simples. Les hypothèses en présence sont alors de la forme de la forme : H_0 : "$\theta \in \Theta_0$" et H_1 : "$\theta \in \Theta_1$"

En particulier si $\Theta \subset R$, on distingue

les hypothèses unilatères : "$\theta \leq \theta_0$", "$\theta > \theta_1$"

les hypothèses bilatères : "$\theta \leq \theta_0$ ou $\theta > \theta_1$", "$\theta \neq \theta_0$".

Définition 1 (fonction puissance d'un test)**.** La fonction puissance d'un test Φ de région critique C est la fonction β_Φ définie sur $\Theta_0 \cup \Theta_1$par

$$\beta_\Phi(\theta) = P_\theta(C)$$

Définition 2. Soit Φ un test de H_0 : "$\theta \in \Theta_0$" contre H_1 : "$\theta \in \Theta_1$"

On appelle niveau de signification de Φ le nombre

$$\alpha_\Phi = \sup_{\theta \in \Theta_0} \beta_\Phi(\theta)$$

Définition 3. Soit le problème de test de H_0 : "$\theta \in \Theta_0$" contre H_1 : "$\theta \in \Theta_1$".

Un test Φ est dit niformément le plus puissant au niveau α pour ce problème si

1. $\alpha_\Phi \leq \alpha$
2. pour tout autre test Φ' tel que $\alpha_{\Phi'} \leq \alpha$ on a

$$\beta_{\Phi'}(\theta) \leq \beta_\Phi(\theta) \quad \forall \theta \in \Theta_1$$

1.0.1 Tests d'hypothèses unilatères

Questions :

1) Existe-t-il des tests UMP pour des hypothèses composites

2) Si oui comment les déterminer

Exemple introductif : tests portant sur les paramètres d'une loi normale

Le test le plus puissant au niveau α de "$m = m_0$" contre "$m = m_1$"

avait la région critique $C = \{\widetilde{x}/\overline{x} > m_0 + \frac{S'}{\sqrt{n}} t_{n-1}(1-\alpha)\}$

indépendante de l'alternative "$m = m_1$". Ce test est donc le même quelque soit la valeur de m_1 (pourvu que $m_1 > m_0$)

De même le test de "$\sigma^2 = \sigma_0^2$" contre "$\sigma^2 = \sigma_1^2$" pour la variance avait pour région critique

$C = \{\sum_{i=1}^{n}(x_i - m)^2 \ > \sigma_0^2 \chi_{n-1}^2(1-\alpha)\}$

Il est indépendant de l'alternative "$\sigma^2 = \sigma_1^2$". Il est donc valable quelque soit la valeur de σ_1^2 (pourvu que $\sigma_1^2 > \sigma_0^2$)

Définition 4. La famille de lois $\mathcal{P} = \{(f_\theta)_{\theta\in\Theta}\}$ est dite à rapport de vraisemblance monotone si

1) il existe une statistique $T : \mathcal{X} \longrightarrow R$

$$x \longrightarrow T(x)$$

2) il existe une fonction monotone $g : R \longrightarrow R$

telles que $\forall \ \ \theta_1 > \theta_0$

$$\frac{L(\theta_1, x)}{L(\theta_0, x)} = g(T(x))$$

On parle alors de famille à R.V.M croissant si g est croissante ou de famille à R.V.M décroissant si g est décroissante.

Exemple 1. Considérons la famille des lois de Poisson $\{P(\lambda) \setminus \lambda > 0\}$ et soit $\lambda_1 > \lambda_0$

$$\frac{L(\lambda_1, x)}{L(\lambda_0, x)} = \frac{\dfrac{e^{-\lambda_1}\lambda_1^x}{x!}}{\dfrac{e^{-\lambda_0}\lambda_0^x}{x!}} = e^{-(\lambda_1-\lambda_0)}\left(\frac{\lambda_1}{\lambda_0}\right)^x = e^{-(\lambda_1-\lambda_0)}e^{x\log\left(\frac{\lambda_1}{\lambda_0}\right)} = g(x)$$

g est croissante puisque $\log\left(\frac{\lambda_1}{\lambda_0}\right) > 0$.

Donc la famille $\{P(\lambda) \setminus \lambda > 0\}$ esr à R.V.M croissant

Exemple 2. Famille des lois exponentielles :

Soit $\theta_1 > \theta_0$

On a $f_\theta(x) = c(\theta)h(x)\exp(a(x)\alpha(\theta)) \implies$

$$\frac{L(\theta_1, x)}{L(\theta_0, x)} = \frac{c(\theta_1)h(x)\exp(a(x)\alpha(\theta_1))}{c(\theta_0)h(x)\exp(a(x)\alpha(\theta_0))} = \frac{c(\theta_1)}{c(\theta_0)}\exp(a(x)(\alpha(\theta_1) - \alpha(\theta_0))$$

$$\frac{L(\theta_1, x)}{L(\theta_0, x)} = g(a(x))$$

Ce rapport est $\begin{cases} \text{croissant si } \alpha \text{ est croissante} \\ \text{décroissant si } \alpha \text{ est décroissante} \end{cases}$

Application : famille des lois $Exp(\lambda)$

$f_\lambda(x) = \lambda e^{-\lambda x} 1_{R_+}(x) = c(\lambda)h(x)\exp(a(x)\alpha(\lambda))$

avec $\alpha(\lambda) = -\lambda$

donc la famille $Exp(\lambda)$ est à R.V.M décroissant.

Théorème 1. *Soit le modèle statistique $(\mathfrak{X}, \mathfrak{B}, (f_\theta)_{\theta\in\Theta})$ où la famille $(f_\theta)_{\theta\in\Theta}$ est à R.V.M croissant alors le test*

$$\Phi(x) = \begin{cases} 1 & si\ T(x) > K \\ \gamma & si\ T(x) = K \\ 0 & si\ T(x) < K \end{cases}$$

de niveau α est UMP au niveau α pour tester "$\theta \leq \theta_0$" contre "$\theta > \theta_0$"

Remarque 1. Si la famille $(f_\theta)_{\theta\in\Theta}$ est R.V.M décroissant alors le test le plus puissant au niveau α est de la forme

$$\Phi(x) = \begin{cases} 1 & \text{si } T(x) < K \\ \gamma & \text{si } T(x) = K \\ 0 & \text{si } T(x) > K \end{cases}$$

K et γ sont déterminés par l'équation $\alpha_\Phi = \alpha$.

Démonstration. La démonstration du théorème est basée sur deux lemmes qui sont utiles en eux-mêmes □

Lemme 2. *Un test est UMP de Θ_0 contre Θ_1 si et seulement s' il est UMP de Θ_0 contre "$\theta = \theta_1$" $\forall\ \theta_1 \in \Theta_1$.*

Lemme 3. *Soit $\Theta_0^{'} \subset \Theta_0$ et Φ un test UMP au niveau α de $\Theta_0^{'}$ contre Θ_1. Si comme test de Θ_0 contre Θ_1, Φ est de niveau α alors Φ est UMP au niveau α de Θ_0 contre Θ_1.*

1.0.2 Tests d'hypothèses bilatères

Les tests de la forme "$\theta = \theta_0$" contre "$\theta \neq \theta_0$" sont parmi les plus importants tests bilatères.

On montre facilement qu'il n'existe pas de test UMP pour ce type de test.

Prenons l'exemple d'un test portant sur la moyenne d'une loi normale à variance connue : "$m = m_0$" contre "$m \neq m_0$",

Si un test UMP existait, alors d'après le lemme 1 :

- il serait UMP de "$m = m_0$" contre "$m > m_0$" et donc aurait pour région crique $C_1 = \{\overline{x}/\overline{x} > m_0 + \frac{\sigma}{\sqrt{n}} q_{1-\alpha}\}$

-il serait UMP de "$m = m_0$" contre "$m > m_0$" et donc aurait pour région crique $C_1 = \{\overline{x}/\overline{x} < m_0 - \frac{\sigma}{\sqrt{n}} q_{1-\alpha}\}$

Ces deux régions critiques sont contradictoires et conduisent à des décisions contraires.

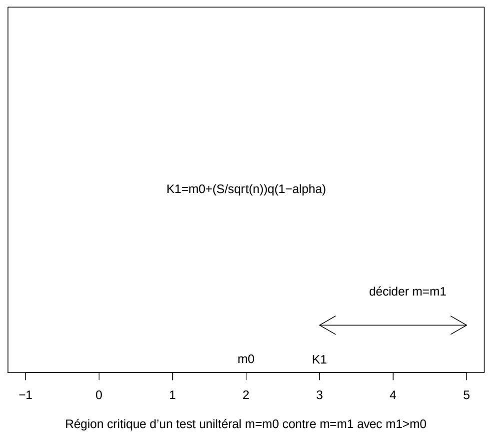

FIGURE 1.1 – Région critique du test unilatéral 1

Une solution est de considérer un test ayant pour région critique un compromis entre les deux régions précédentes de la forme

$$C = \{\overline{x}/\overline{x} < K_1 \ ou \ \overline{x} > K_2\}$$

Définition 5. Un test Φ de H_0 : "$\theta \in \Theta_0$" contre H_1 : "$\theta \in \Theta_1$" est dit sans biais si

$$\alpha_\Phi \leq \beta_\Phi(\theta) \quad \forall \theta \in \Theta_1$$

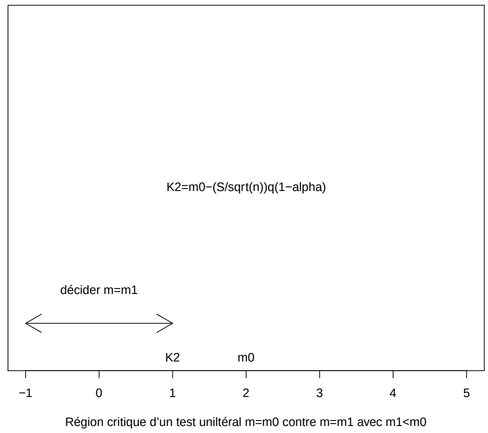

FIGURE 1.2 – Région critique du test unilatéral 2

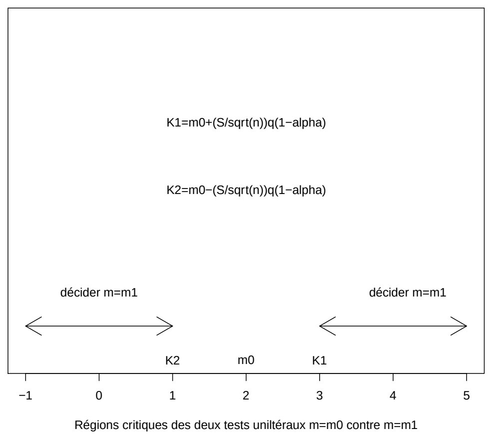

FIGURE 1.3 – Région critique d'un test blatéral

Théorème 4. *Soit* $(\mathfrak{X}, \mathfrak{B}, \{f_\theta(x) = c(\theta)h(x)\exp(a(x)\alpha(\theta)\ \theta \in \Theta\})^{(n)}$ *un modèle statistique exponentiel où la fonction* $\theta \longrightarrow \alpha(\theta)$ *est strictement monotone. Soit* θ_0, θ_1 *2 réels tels que* $\theta_0 < \theta_1$*. Alors* $\forall \alpha \in [0,1]$ *il existe un test UMPB au niveau* α *de "*$\theta_0 \leq \theta \leq \theta_1$*" contre "*$\theta < \theta_0$ *ou* $\theta > \theta_1$*"*

de la forme $\Phi(x) = \begin{cases} 1 & si \sum\limits_{i=1}^{n} a(x_i) > K_1 \ ou \sum\limits_{i=1}^{n} a(x_i) > K_2 \\ \gamma_1, \gamma_2 & si \sum\limits_{i=1}^{n} a(x_i) = K_1 ou \sum\limits_{i=1}^{n} a(x_i) = K_2 \\ 0 & si\ K_1 < \sum\limits_{i=1}^{n} a(x_i) < K_2 \end{cases}$ $K_1, K_2, \gamma_1, \gamma_2$ *étant déterminés par les équations :*

$$\beta_\Phi(\theta_0) = \alpha\ ,\ \beta_\Phi(\theta_1) = \alpha$$

Corollaire 5. *Si* $\theta_0 = \theta_1$*, il existe un test UMPB de "*$\theta = \theta_0$*" contre "*$\theta \neq \theta_0$ *" de la forme précédente,* $K_1, K_2, \gamma_1, \gamma_2$ *étant déterminés par les équations :*

$$\beta_\Phi(\theta_0) = \alpha\ ,\ \beta'_\Phi(\theta_0) = 0$$

Exemple 3. Test sur la moyenne d'une loi normale à variance connue : "$m = m_0$" contre "$m \neq m_0$ "

La loi normale vérifie les hypothèses précédentes, il existe donc un test UMPB pour ce problème de la forme :

$$\Phi(x) = \begin{cases} 1 & \text{si } \sum\limits_{i=1}^{n} x_i > K_1 \text{ ou } \sum\limits_{i=1}^{n} x_i > K_2 \\ 0 & \qquad \text{si } K_1 \leq \sum\limits_{i=1}^{n} x_i \leq K_2 \end{cases}$$

Sa région critique est : $C = \{\tilde{x} / \sum\limits_{i=1}^{n} x_i < K_1 \text{ ou } \sum\limits_{i=1}^{n} x_i > K_2\} = \{\overline{x} / \overline{x} < K'_1 \text{ ou } \overline{x} > K'_2\}$

K_1, K_2 sontdéterminés par les équations :

$$\beta_\Phi(m_0) = \alpha \ , \ \beta'_\Phi(m_0) = 0$$

$\beta_\Phi(m_0) = P_{m_0}(C) = P_{m_0}(\sum\limits_{i=1}^{n} x_i < K_1 \ ou \sum\limits_{i=1}^{n} x_i > K'_2) = \alpha$

$\beta_\Phi(m_0) = P_{m_0}(\sum\limits_{i=1}^{n} x_i < K_1) + P(\sum\limits_{i=1}^{n} x_i > K'_2)$

$\beta_\Phi(m_0) = F_{N(nm_0, n\sigma^2)}(K'_1) + 1 - F_{N(nm_0, n\sigma^2)}(K'_2)$

$\beta_\Phi(m_0) = F_{N(0,1)}\left(\dfrac{K'_1 - nm_0}{\sqrt{n}\sigma}\right) + 1 - F_{N(0,1)}\left(\dfrac{K'_2 - nm_0}{\sqrt{n}\sigma}\right)$

$\beta_\Phi(m_0) = \Phi\left(\dfrac{K'_1 - nm_0}{\sqrt{n}\sigma}\right) + 1 - \Phi\left(\dfrac{K'_2 - nm_0}{\sqrt{n}\sigma}\right) = \alpha$

$\Longrightarrow \Phi\left(\dfrac{K'_2 - nm_0}{\sqrt{n}\sigma}\right) - \Phi\left(\dfrac{K'_1 - nm_0}{\sqrt{n}\sigma}\right) = 1 - \alpha$

$\beta'_\Phi(m_0) = -\dfrac{\sqrt{n}}{\sigma}\left(\phi\left(\dfrac{K'_2 - nm_0}{\sqrt{n}\sigma}\right) - \phi\left(\dfrac{K'_1 - nm_0}{\sqrt{n}\sigma}\right)\right) = 0$

avec $\phi = \Phi'$: densité de la loi $N(0,1)$

On a donc à résoudre le système d'équations

$\Phi\left(\dfrac{K'_2 - nm_0}{\sqrt{n}\sigma}\right) - \Phi\left(\dfrac{K'_1 - nm_0}{\sqrt{n}\sigma}\right) = 1 - \alpha \qquad (1)$

$\phi\left(\dfrac{K'_2 - nm_0}{\sqrt{n}\sigma}\right) - \phi\left(\dfrac{K'_1 - nm_0}{\sqrt{n}\sigma}\right) = 0 \qquad (2)$

Solution :

(1) et (2) $\iff \dfrac{K'_2 - nm_0}{\sqrt{n}\sigma} = q_{1-\frac{\alpha}{2}}$ et $\dfrac{K'_1 - nm_0}{\sqrt{n}\sigma} = q_{\frac{\alpha}{\acute{e}}} = -q_{1-\frac{\alpha}{2}}$

donc $K'_2 = nm_0 + \sqrt{n}\sigma q_{1-\frac{\alpha}{2}}$ et $K'_1 = nm_0 - \sqrt{n}\sigma q_{1-\frac{\alpha}{2}}$

Le test UMPB de "$m = m_0$" contre "$m \neq m_0$ " a la région critique :

$$C = \{\tilde{x} / \sum_{i=1}^{n} x_i < nm_0 - \sqrt{n}\sigma q_{1-\frac{\alpha}{2}} \ ou \ \sum_{i=1}^{n} x_i > nm_0 + \sqrt{n}\sigma q_{1-\frac{\alpha}{2}}\}$$

ce qui est équivalent à :

$$C = \{\tilde{x} / |\overline{x}| > m_0 + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \ \}$$

Remarque 2. Si la variance est inconnue on a le test de Student de région critique

$$C = \{\tilde{x} / |\overline{x}| > m_0 + \frac{S}{\sqrt{n}} t_{n-1}(1 - \frac{\alpha}{2}) \ \}$$

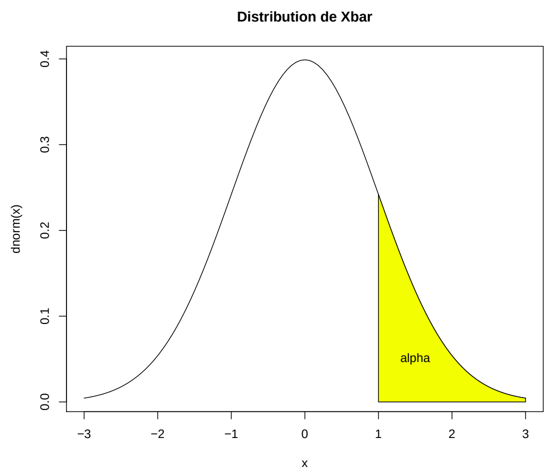

FIGURE 1.4 – Région critique du test unilatéral 1

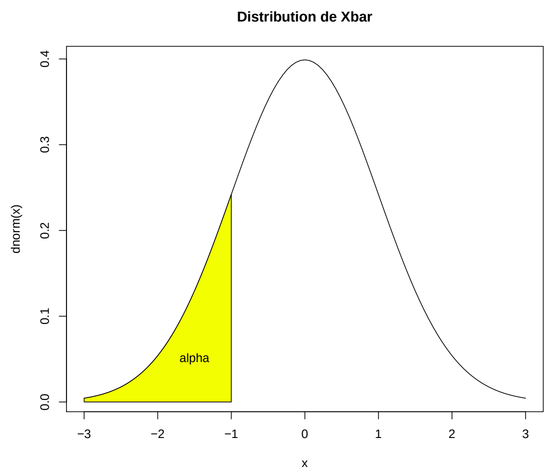

FIGURE 1.5 – Région critique du test unilatéral 2

Exemple 4. Application : Test sur la variance d'une loi normale à moyenne connue : "$\sigma^2 = \sigma_0^2$" contre "$\sigma^2 \neq \sigma_0^2$"

La loi normale vérifie les hypothèses précédentes, il existe donc un test UMPB pour ce problème de la forme :

$$\Phi(x) = \begin{cases} 1 \quad \text{si } \sum_{i=1}^{n}(x_i - m)^2 > K_1 \text{ ou } \sum_{i=1}^{n}(x_i - m)^2 > K_2 \\ 0 \qquad\qquad \text{si } K_1 \leq \sum_{i=1}^{n}(x_i - m)^2 \leq K_2 \end{cases}$$

Sa région critique est : $C = \{\widetilde{x} / \sum_{i=1}^{n}(x_i - m)^2 < K_1 \text{ ou } \sum_{i=1}^{n}(x_i - m)^2 > K_2\}$

K_1, K_2 sontdéterminés par les équations :

$$\beta_\Phi(\sigma_0^2) = \alpha \ , \ \beta'_\Phi(\sigma_0^2) = 0$$

$$\beta_\Phi(\sigma_0^2) = P_{\sigma_0^2}(C) = P_{\sigma_0^2}(\sum_{i=1}^{n}(x_i - m)^2 < K_1 \ ou \sum_{i=1}^{n}(x_i - m)^2 > K_2) = \alpha$$

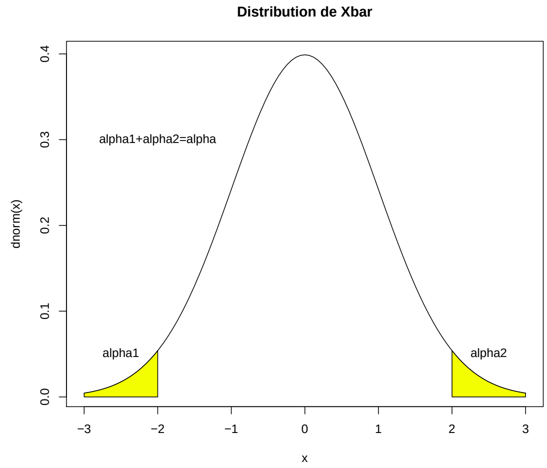

FIGURE 1.6 – Région critique d'un test blatéral

$\beta_\Phi(\sigma_0^2) = P_{\sigma_0^2}(\sum_{i=1}^n (x_i - m)^2 < K_1) + P_{\sigma_0^2}(\sum_{i=1}^n (x_i - m)^2 > K_2)$
$\beta_\Phi(\sigma_0^2) = F_{\mathcal{X}_n^2}(\dfrac{K_1}{\sigma_0^2}) + 1 - F_{\mathcal{X}_n^2}(\dfrac{K_2}{\sigma_0^2})) = \alpha$
$\beta'_\Phi(\sigma_0^2) = \dfrac{-1}{\sigma_0^4}\left(K_1 f_{\mathcal{X}_n^2}(\dfrac{K_1}{\sigma_0^2}) - K_2 f_{\mathcal{X}_n^2}(\dfrac{K_2}{\sigma_0^2})\right) = 0$
On a donc à résoudre le système d'équations
$F_{\mathcal{X}_n^2}(\dfrac{K_2}{\sigma_0^2}) - F_{\mathcal{X}_n^2}(\dfrac{K_1}{\sigma_0^2}) = 1 - \alpha \qquad (1)$
$K'_1 f_{\mathcal{X}_n^2}(\dfrac{K_1}{\sigma_0^2}) - K'_2 f_{\mathcal{X}_n^2}(\dfrac{K_2}{\sigma_0^2}) = 0 \qquad (2)$
Ce système ne peut être résolu analytiquement et ne peut l'être que numériquement.
Pour simplifier on prend
$F_{\mathcal{X}_n^2}(\dfrac{K_1}{\sigma_0^2}) = \dfrac{\alpha}{2}$ et $F_{\mathcal{X}_n^2}(\dfrac{K_2}{\sigma_0^2}) = 1 - \dfrac{\alpha}{2}$
i.e : $K_1 = \sigma_0^2 \mathcal{X}_n^2(\dfrac{\alpha}{2})$ et $K_2 = \sigma_0^2 \mathcal{X}_n^2(1 - \dfrac{\alpha}{2})$
ce qui donne le test de région critique

$$C = \{\widetilde{x} / \sum_{i=1}^n (x_i - m)^2 < \sigma_0^2 \mathcal{X}_n^2(\frac{\alpha}{2}) \ ou \ \sum_{i=1}^n (x_i - m)^2 > \sigma_0^2 \mathcal{X}_n^2(1 - \frac{\alpha}{2})\}$$

Remarque 3. Dans le cas où la moyenne est inconnue on a le test de région critique

$$C = \{\widetilde{x} / \sum_{i=1}^n (x_i - \overline{x})^2 < \sigma_0^2 \mathcal{X}_{n-1}^2(\frac{\alpha}{2}) \ ou \ \sum_{i=1}^n (x_i - \overline{x})^2 > \sigma_0^2 \mathcal{X}_{n-1}^2(1 - \frac{\alpha}{2})\}$$

1.0.3 Comparaison de deux populations normales

Il s'agit de tester l'égalité de deux distributions gaussiennes. Les lois normales étant caractérisées par leurs moyennes et variances cela revient à tester l'égalité de leurs moyennes d'une part et de leurs variances d'autre part.

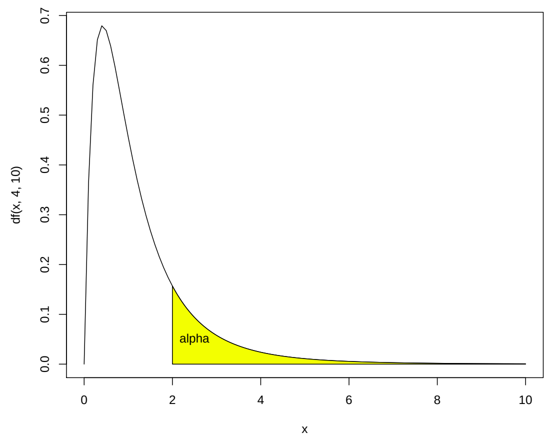

Figure 1.7 – Région critique 1

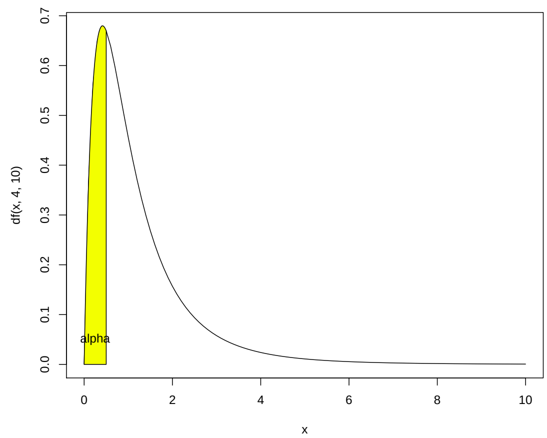

Figure 1.8 – Région critique 2

Soit deux variables indépendantes : $X \sim N(m_1, \sigma_1^2)$ et $Y \sim N(m_2, \sigma_2^2)$. On dispose d'un n-échantillon de X et d'un p-échantillon de Y.

Problèmes :

- Tester "$m_1 = m_2$" contre "$m_1 \neq m_2$"
- Tester "$\sigma_1^2 = \sigma_2^2$" contre "$\sigma_1^2 \neq \sigma_2^2$"

Test "$m_1 = m_2$" contre "$m_1 \neq m_2$" $\iff$ Test "$m_1 - m_2 = 0$" contre "$m_1 - m_2 \neq 0$"

On a : $\overline{X} - \overline{Y} \sim N(m_1 + m_2, \dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{p})$

Si les variances sont connues

$$\frac{\overline{X} - \overline{Y} - (m_1 + m_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{p}}} \sim N(0,1)$$

On a le test de région critique

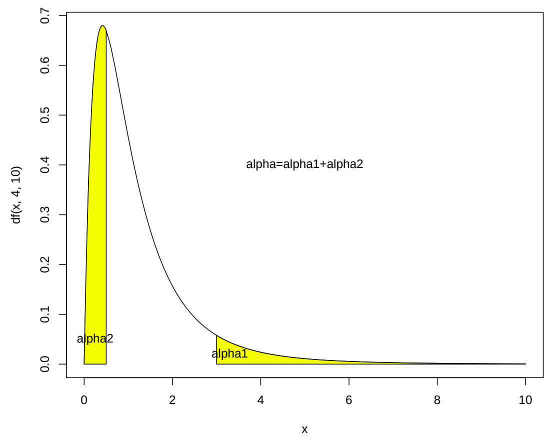

FIGURE 1.9 – Région critique 3

$$C = \{\widetilde{x}, \widetilde{y} / \ |\overline{x} - \overline{y}| > m_1 - m_2 + \left(\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{p}}\right) q_{1-\frac{\alpha}{2}} \ \}$$

Si les variances sont inconnues mais égales, alors elles sont estimées par $\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{p}}$
On a le test de région critique

$$C = \{\widetilde{x}, \widetilde{y} / \ |\overline{x} - \overline{y}| > m_1 - m_2 + \left(\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{p}}\right) t_{n+p-2}(1 - \frac{\alpha}{2}) \ \ \}$$

Test "$\sigma_1^2 = \sigma_2^2$" contre "$\sigma_1^2 \neq \sigma_2^2$" $\Longleftrightarrow$ Test "$\frac{\sigma_1^2}{\sigma_2^2} = 1$" contre "$\frac{\sigma_1^2}{\sigma_2^2} \neq 1$"

On a $\frac{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}{\sigma_1^2} \sim$ et $\frac{\sum\limits_{i=1}^{p} (y_i - \overline{y})^2}{\sigma_2^2} \sim$

$\Longrightarrow \ \frac{\sigma_2^2}{\sigma_1^2} \frac{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}}{\frac{\sum_{i=1}^{p}(y_i-\overline{y})^2}{p}} = \frac{\sigma_2^2}{\sigma_1^2} \frac{S_1^2}{S_2^2} \sim F_{(n-1,p-1)}$

D'ou le test de région critique
$C = \{\widetilde{x}, \widetilde{y} / \ \frac{S_1^2}{S_2^2} < F_{(n-1,p-1)}(\frac{\alpha}{2})$ ou $\frac{S_1^2}{S_2^2} > F_{(n-1,p-1)}(1 - \frac{\alpha}{2})\}$