Sommaire

As so far...

- Un estimateur est une variable aléatoire fonction de l'échantillon, permettant d'inférer la valeur d'un paramètre
  - biais, variance, risque
  - consistance
  - loi : exacte ou approchée par l'asymptotique
- VRAI ou FAUX ?
  - Un estimateur non biaisé est forcément de risque minimum
  - La somme des carrés de n variables aléatoires gaussiennes centrées réduites indépendantes suit une loi du Khi-deux à n ddl
  - Soit X un n-échantillon iid gaussien. Alors,

$$T_n = \frac{\bar{X} - \mathbb{E}(X_1)}{\sum(X_i - \bar{X})^2/(n-1)} \sim \mathcal{T}(n-1);$$

  - Tous les estimateurs consistants ont un comportement asymptotiquement normal

Tests : un exemple

Un constructeur automobile annonce une consommation $\mu_0 = 6.32\ell/100\ km$, avec un écart type $\sigma = 0.21\ell/100\ km$, pour des véhicules d'un type donné. Un organisme indépendant suspecte une sous-estimation de cette consommation et indique que la consommation s'élèverait à $\mu_1 = 6.45\ell/100\ km$.
Sur un 30-échantillon $\bar{x} = 6.43\ell/100\ km$. *Qui a raison ?*

(H_0) conso. conforme au constructeur : $X \sim \mathcal{N}(\mu_0, \sigma^2)$
$\mu = \mu_0 = 6.32$

(H_1) conso. suspectée par l'organisme : $X \sim \mathcal{N}(\mu_1, \sigma^2)$
$\mu = \mu_1 = 6.45$

- Choisir, à partir d'un n-échantillon, entre les deux hypothèses (H_0) et (H_1), en assumant le risque de première espèce α (5%, 10%,..) de choisir (H_1) alors que (H_0) est vrai.

Un exemple (suite)

- Statistique $\bar{X}$, moyenne des consommations de $n = 30$ véhicules
- Loi sous (H_0),

$$\bar{X} \sim \mathcal{N}(\mu_0, \frac{\sigma^2}{n}) \text{ soit } T = \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma} \sim \mathcal{N}(0,1)$$

- Choisir a priori un risque α, calibrant la probabilité de rejet de (H_0) à tort ($\alpha = 5\%$ par exemple)

$$\alpha = \mathbb{P}_{H_0}\Big(\underbrace{T > q^*_{1-\alpha}}_{\mathcal{R}=]q^*_{1-\alpha};\infty[,\ \text{Région de rejet pour } T} \Big)$$

$$= \mathbb{P}_{H_0}\Big(\underbrace{\bar{X} > \mu_0 + q^*_{1-\alpha}\frac{\sigma}{\sqrt{n}}}_{\mathcal{R}=]\mu_0+q^*_{1-\alpha}\frac{\sigma}{\sqrt{n}};\infty[,\ \text{Région de rejet pour } \bar{X}} \Big)$$

avec $q^*_{1-\alpha}$ le quantile d'une loi $\mathcal{N}(0,1)$ d'ordre $1-\alpha$

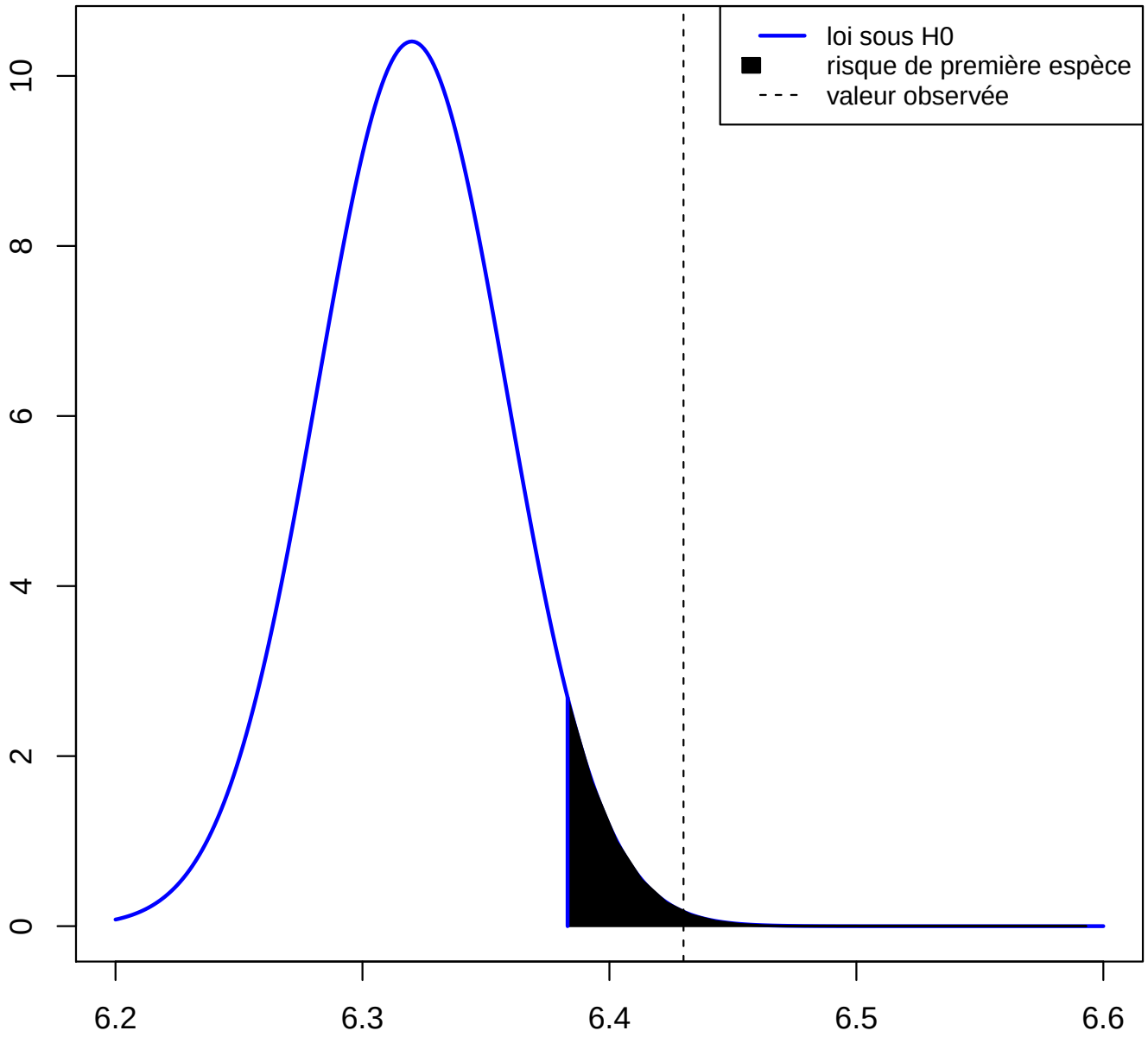
Un exemple (suite)

- Décider :
  - si T est dans la région de rejet, on rejette H_0
  - sinon, on ne rejette pas [conserve/accepte] (H_0) faute de preuves suffisantes
- ici, $\bar{x} = 6.43\ell/100\ km$,

$$t_{obs} = \frac{6.43 - 6.32}{0.21/\sqrt{30}} = 2.86 > 1.64$$

au niveau $\alpha = 5\%$, les données sont significatives pour rejeter (H_0), le constructeur a minimisé la consommation, avec un risque (de première espèce) α.

Représentation graphique



Un exemple (suite)

Un autre cas de figure :

- si la même consommation a été observée sur un échantillon de $n = 9$ véhicules

$$t_{obs} = \frac{6.43 - 6.32}{0.21/\sqrt{9}} = 1.57 < 1.64$$

on ne peut pas rejeter le fait que le constructeur a sous-estimé la consommation, on ne peut rejeter (H_0) qu'on accepte par défaut

  - *avec quelle erreur ?*

Un exemple (suite)

Une autre façon de se tromper :

- erreur de seconde espèce : ne pas rejeter (H_0) alors que (H_1) est vraie
- Sous (H_1), $\bar{X} \sim \mathcal{N}(\mu_1, \frac{\sigma^2}{n})$ et le risque de seconde espèce est

$$\begin{aligned} \beta &= \mathbb{P}_{H_1}\left(\bar{X} < \mu_0 + q^*_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right) \\ &= \mathrm{F}^*\left(\sqrt{n}\frac{\mu_0 - \mu_1}{\sigma} + q^*_{1-\alpha}\right) \end{aligned}$$

- App.Num : $n = 9$, $\beta \simeq 0.41$

la puissance $\pi = 1 - \beta$ n'est pas très grande

Procédure de test

Définition

- *Un test est une procédure de décision qui permet de trancher, au vu des résultats d'un échantillon, entre deux hypothèses l'hypothèse nulle (H_0) et une hypothèse alternative (H_1), dont une seule est vraie.*
- *La région critique ou région de rejet $\mathcal{R}$ est l'ensemble des valeurs de la variable de décision T qui conduisent à écarter (H_0) au profit de (H_1).*
- *La région d'acceptation par défaut du test est $\overline{\mathcal{R}}$.*

Construire un test (rappel)

Choisir entre deux hypothèses (H_0) et (H_1), en calibrant le risque de choisir (H_1) à tort.

1. Définir le modèle
2. Définir les hypothèses nulle (H_0) et alternative (H_1)
3. Choisir une statistique de test $T(X)$, calculer sa loi sous (H_0)
4. Définir la règle de décision en calibrant la région de rejet $\mathcal{R}_\alpha$ de (H_0) suivant le risque de première espèce α

$$\mathbb{P}_{(H_0)}(T(X) \in \mathcal{R}_\alpha) = \alpha.$$

5. Calcul de la statistique observée et décision
6. Calcul de la puissance

$$\mathbb{P}_{(H_1)}(T(X) \in \mathcal{R}_\alpha) = \pi = 1 - \beta.$$

La décision

La décision du test, à partir de la valeur observée t de la statistique de test T est :

- si $t \in \mathcal{R}$, on rejette (H_0) au risque α : l'erreur commise est de risque $\alpha = \mathbb{P}_{(H_0)}(T \in \mathcal{R})$.
- si $t \notin \mathcal{R}$, on ne rejette pas (conserve) (H_0) : les données ne sont pas significatives pour accepter (H_1). L'erreur de seconde espèce commise est de risque $\beta = \mathbb{P}_{(H_1)}(T \notin \mathcal{R})$, en général inconnu.

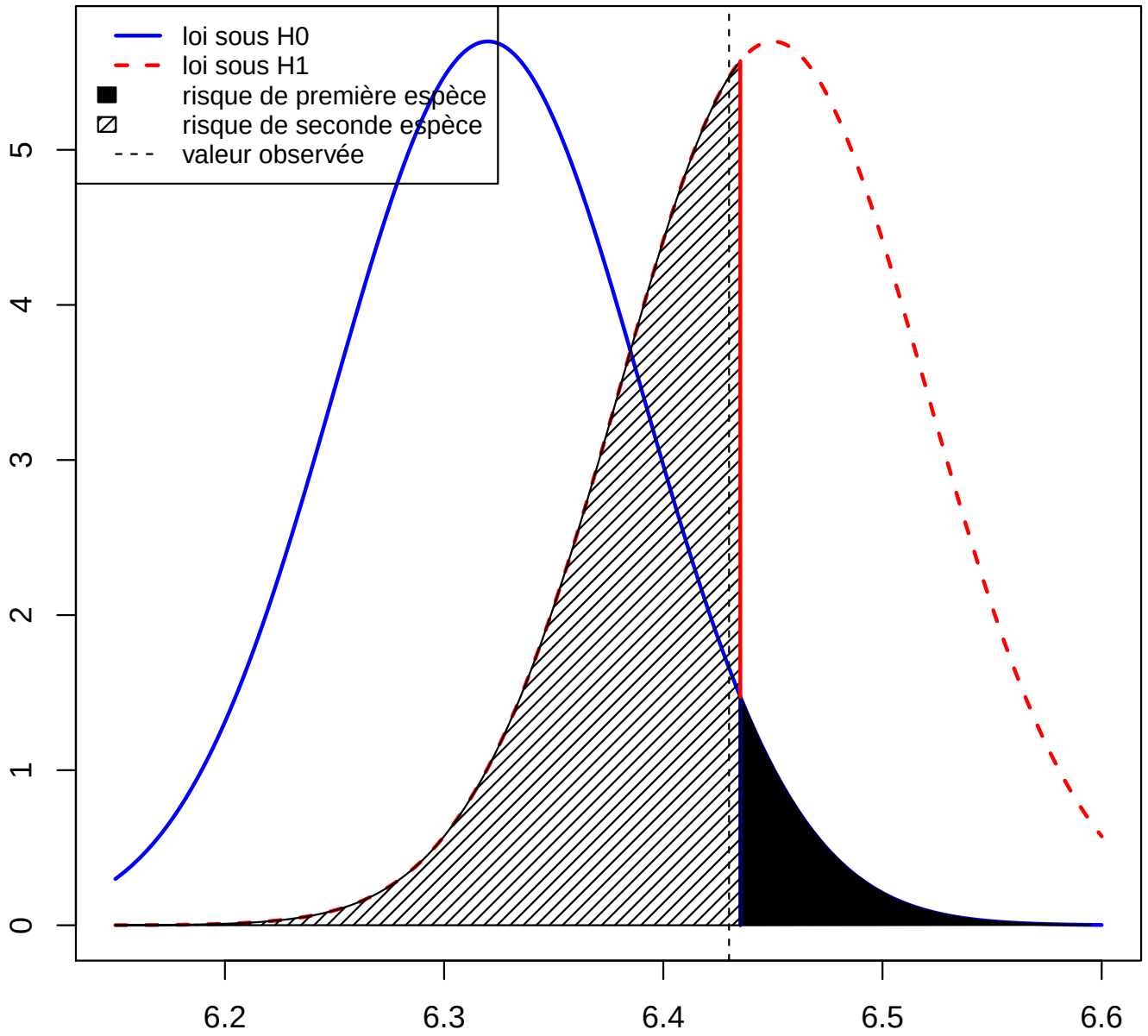
La décision dépend de l'échantillon

Procédure de test

A l'issue du test, les quatre situations suivantes sont possibles

	Choix (H_0)	Choix (H_1)
(H_0) vraie	$1-\alpha$ bonne décision	$\alpha = \mathbb{P}_{H_0}(T \in \mathcal{R})$ risque première espèce mauvaise décision
(H_1) vraie	$\beta = \mathbb{P}_{H_1}(T \notin \mathcal{R})$ risque seconde espèce mauvaise décision	$\pi = 1-\beta$ puissance bonne décision

Représentation graphique



Dissymétrie de la situation de test

- Le risque n'est contrôlé que pour le rejet de (H_0)
  - $\hookrightarrow$ La véritable décision est celle qui rejette (H_0).
  - $\hookrightarrow$ (H_0) et (H_1) ne sont pas interchangeables.
- Il faut connaître la loi de la statistique de test sous (H_0)
- Il faut que cette loi soit différente sous (H_1)
- Entre deux tests de même risque de 1ère espèce α, il faut choisir le plus puissant

Optimalité

Le cadre de la théorie de Neyman-Pearson permet de construire des tests les plus puissants parmi les tests de niveau fixé

Définition

Un test est uniformément plus puissant (UPP) si, quelle que soit la valeur de θ, sa puissance $\pi(\theta)$ est supérieure à la puissance de tout autre test de niveau α.

Théorème (Neyman-Pearson, deux hypothèses simples)

Soit $L(\theta; x)$ la vraisemblance des observations. La région critique optimale du test de $\theta = \theta_0$ vs $\theta = \theta_1$ au niveau α est définie par

$$\mathcal{R}^{\alpha}_{opt} = \{x \in \mathbb{R}^n; \frac{L(\theta_1; x)}{L(\theta_0; x)} > k_\alpha\}; \quad \mathbb{P}(\{X \in \mathcal{R}^{\alpha}_{opt}\}) = \alpha$$

NP : preuve

Supposons que k_α existe. Soit $\mathcal{R}_\alpha$ une autre région de rejet de niveau α.

$$\begin{aligned}\mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha) &= \mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}) + \mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha \cap \mathcal{R}_{opt}) = \alpha \\ \mathbb{P}_{\theta_0}(X \in \mathcal{R}_{opt}) &= \mathbb{P}_{\theta_0}(X \in \mathcal{R}_{opt} \setminus \mathcal{R}_\alpha) + \mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha \cap \mathcal{R}_{opt}) = \alpha\end{aligned}$$

d'où $\mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}) = \mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt})$, soit

$$\int_{x \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}} L(\theta_0, x)\, dx = \int_{x \in \mathcal{R}_{opt} \setminus \mathcal{R}_\alpha} L(\theta_0, x)\, dx$$

On compare maintenant les puissances :

$$\pi(\mathcal{R}_{opt}) = \mathbb{P}_{\theta_1}(X \in \mathcal{R}_{opt}) = \int_{x \in \mathcal{R}_{opt} \setminus \mathcal{R}_\alpha} L(\theta_1, x)\, dx + \int_{x \in \mathcal{R}_{opt} \cap \mathcal{R}_\alpha} L(\theta_1, x)\, dx$$

$$\pi(\mathcal{R}_\alpha) = \mathbb{P}_{\theta_1}(X \in \mathcal{R}_\alpha) = \int_{x \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}} L(\theta_1, x)\, dx + \int_{x \in \mathcal{R}_{opt} \cap \mathcal{R}_\alpha} L(\theta_1, x)\, dx$$

$$\begin{aligned}\int_{x \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}} L(\theta_1, x)\, dx &\leq \int_{x \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}} k_\alpha L(\theta_0, x)\, dx = \int_{x \in \mathcal{R}_{opt} \setminus \mathcal{R}_\alpha} k_\alpha L(\theta_0, x)\, dx \\ &< \int_{x \in \mathcal{R}_{opt} \setminus \mathcal{R}_\alpha} L(\theta_1, x)\, dx \text{ d'où } \pi(\mathcal{R}_\alpha) < \pi(\mathcal{R}_{opt})\end{aligned}$$

NP : Peut-on trouver k_α ?

Cas continu : Théorème des valeurs intermédiaires avec l'application

$$k \to \mathbb{P}_{\theta_0}\Big(L(\theta_1, X) > kL(\theta_0, X)\Big)$$

Cas discret : rarement possible de trouver une région de risque exactement α

- on prend k de façon à définir la région de rejet de risque le plus proche de α tout en étant inférieur à α.
- utilisation d'un test randomisé pour avoir un niveau exact

Exemple

Test de NP de $(H_0) : \theta = \theta_0$ contre $(H_1) : \theta = \theta_1 > \theta_0$ dans le modèle iid $\mathcal{N}(\theta, \sigma^2)$, σ^2 connu.

Il est indépendant de θ_1 et est UPP.

Test optimal de $\theta = \theta_0$ vs $\theta = \theta_1$ au niveau α

Autres propriétés du test de Neyman-Pearson :

- sans biais : $1 - \beta(\theta) = \pi(\theta) > \alpha$ pour tout $\theta \in \Theta_1$
- consistant : $\pi_n(\theta) \to 1$ quand $n \to +\infty$

Si on dispose de plus d'une statistique exhaustive[1] T, la région critique en dépend exclusivement et le test de NP se réduit à une région de rejet de la forme

$$\mathcal{R}_{opt} = \left\{ t | \frac{g(\theta_1; t)}{g(\theta_0; t)} > k_\alpha \right\}; \quad \mathbb{P}(\{T \in \mathcal{R}_{opt}\}) = \alpha$$

1. On dit que $t(X)$ est une statistique exhaustive pour $\theta \in \Theta \subset \mathbb{R}^p$ si la loi de $X = (X_1, \ldots, X_n)$ conditionnellement à $t(X) = t$ ne ne dépend pas du paramètre θ

Hypothèses composites

Quand l'hypothèse alternative est composite :

- la puissance est une fonction de θ : pour $\theta_1 \in \Theta_1$,

$$\pi(\theta_1) = \mathbb{P}_{\theta_1}(T \in \mathcal{R}) = 1 - \beta(\theta_1).$$

Quand l'hypothèse nulle est composite :

- le risque de première espèce est une fonction de θ : pour $\theta \in \Theta_0$,

$$\alpha(\theta) = \mathbb{P}_{\theta}(T \in \mathcal{R})$$

- la taille du test est définie par : $\sup_{\theta \in \Theta_0} \alpha(\theta)$
- Un test est de niveau α si sa taille $\leq \alpha$

Test entre deux hyp. composites au niveau α

Définition

La loi $\mathbb{P}_\theta$ à densité $f_\theta(.)$, $\theta \in \Theta \subset \mathbb{R}$, est dite à rapport de vraisemblance monotone croissant (resp. décroissant), s'il existe une statistique T_n de $\mathcal{X}^n$ dans $\mathbb{R}$ telle que le rapport de vraisemblance d'un n-échantillon $L(\theta_1; x_1, \ldots, x_n)/L(\theta; x_1, \ldots, x_n)$ soit une fonction croissante (resp. décroissante) de T_n pour tout $\theta_1 > \theta$.

Test entre deux hyp. composites au niveau α

Théorème (Lehman : rapport de vraisemblance monotone)

S'il existe une statistique T telle que pour tout couple (θ_1, θ_0) le rapport de vraisemblance $L(\theta_1; x)/L(\theta_0; x)$ soit une fonction monotone de T, alors il existe un test UPP pour les situations d'hypothèses unilatérales :

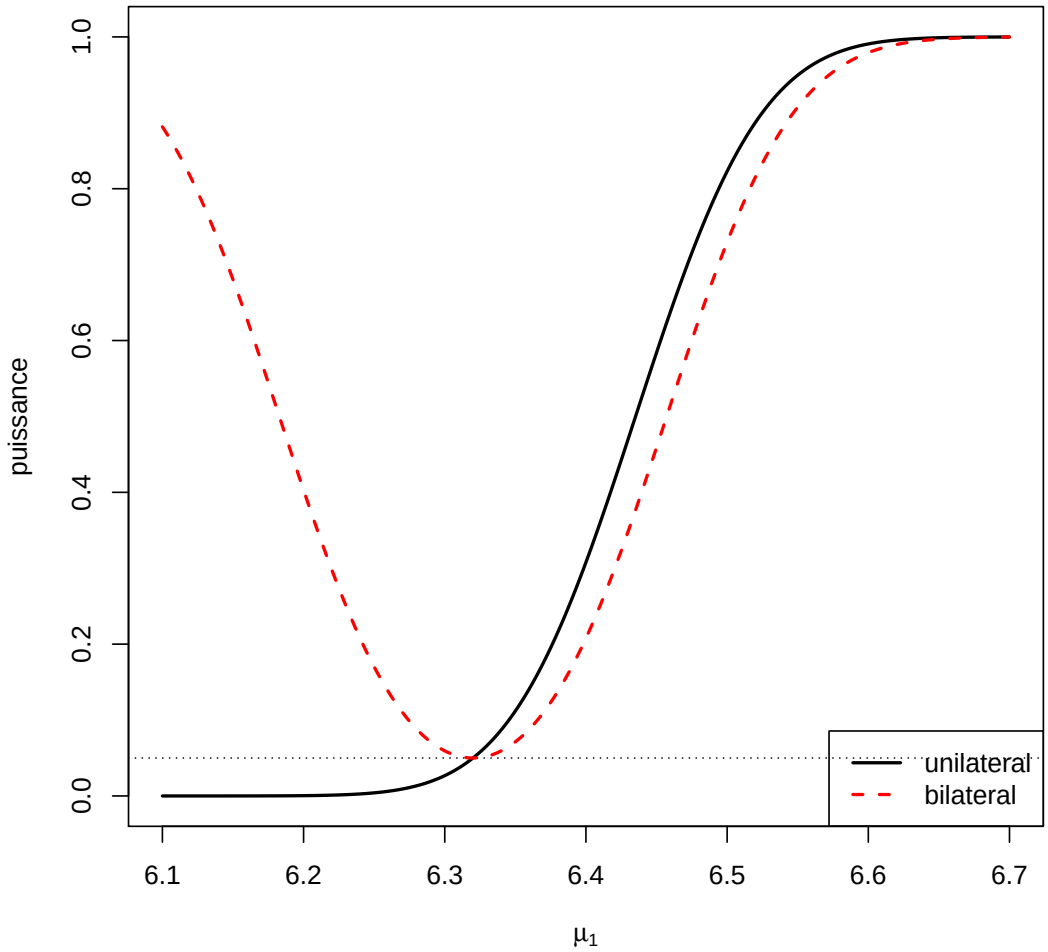
- $(H_0) : \theta \leq \theta_0$ et RV fct $\nearrow$ de T : $\mathcal{R} = \{T > k\}$
- $(H_0) : \theta \geq \theta_0$ et RV fct $\nearrow$ de T : $\mathcal{R} = \{T < k\}$

- $(H_0) : \theta \leq \theta_0$ et RV fct $\searrow$ de T : $\mathcal{R} = \{T < k\}$
- $(H_0) : \theta \geq \theta_0$ et RV fct $\searrow$ de T : $\mathcal{R} = \{T > k\}$

$\hookrightarrow$ cas des statistiques exhaustives des familles exponentielles de loi.

un test pas UPP

Dans le modèle $\mathcal{N}(\theta, 1)$, $\theta = \theta_0$ contre $\theta \neq \theta_0$ de région de rejet $\mathcal{R} = \{|T| > k\}$ n'est pas UPP... mais le meilleur parmi les tests sans biais



Test du rapport des vraisemblances maximales

Cas d'hypothèses plus générales

Définition

Soit une famille paramétrique $\mathbb{P}_\theta, \theta \subset \Theta$ *et les hypothèses* $(H_0) : \theta \in \Theta_0$ *contre* $(H_1) : \theta \in \Theta_1 = \Theta \setminus \Theta_0$. *On appelle rapport de vraisemblance généralisé, la fonction* $RV(X)$ *telle que*

$$RV(X) = \frac{\sup_{\theta \in \Theta_0} L(\theta; X)}{\sup_{\theta \in \Theta} L(\theta; X)}$$

Le test du rapport de vraisemblance (TRV) est le test défini par une région de rejet de la forme

$$\mathcal{R} = \{RV(X) < k_\alpha \leq 1\}.$$

Exemple : Test de $\mu = \mu_0$ contre $\mu \neq \mu_0$ dans un modèle gaussien à variance inconnue.

Propriétés du TRV

Le test TRV n'a pas de propriétés d'optimalité notables,mais on constate dans des situations usuelles qu'il est UPP sans biais.

Théorème (asymptotique du RV)

Soit une famille paramétrique $\mathbb{P}_\theta, \theta \subset \Theta$. *Si* Θ_0 *définit une sous-hypothèse linéaire de* Θ, $dim(\Theta_0) = q$, $dim(\Theta) = p$, *et sous les conditions de régularité de l'EMV, alors, sous* (H_0)

$$-2\log(RV) \xrightarrow{\mathcal{L}} \chi^2(p-q)$$

La région de rejet $\{-2\log(TRV) > q_{\chi^2(p-q)}(1-\alpha)\}$ du test de rapport de vraisemblances maximales est asymptotiquement de niveau α.

Test de Wald

$$(H_0) : A\theta = A\theta_0 \text{ contre } (H_1) : A\theta \neq A\theta_0$$

Soit $\widehat{\theta}$ l'estimateur du maximum de vraisemblance de comportement asymptotiquement normal :

$$\widehat{V}_n^{-1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, Id_p)$$

Si $rang(A) = r$, la loi de la statistique de Wald sous (H_0) est

$$W = [A(\widehat{\theta}_n - \theta_0]'(A\widehat{V}_n A')^{-1} A(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \chi^2(r)$$

La région de rejet du test de Wald est asymptotiquement de niveau α

$$\{W > q_{\chi^2(r)}(1-\alpha)\}$$

cas particulier $A\theta \in \mathbb{R}$

$$(H_0) : A\theta = A\theta_0 \text{ contre } (H_1) : A\theta \neq A\theta_0$$

Soit $\widehat{\theta}$ un emv asymptotiquement normal :

$$\widehat{V}_n^{-1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, Id_p)$$

Sous (H_0),

$$T_n = (A\widehat{V}_n A')^{-1/2} A(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

La région de rejet $\mathcal{R}_\alpha$ est de niveau asympt. α

$$\mathcal{R}_\alpha = \{|T_n| > q^*_{1-\alpha/2}\} = \{W > q^{\chi^2(r)}_{1-\alpha}\}$$

Rem pour l'hypothèse unilatérale $(H_1) : A\theta > A\theta_0$:
$\mathcal{R}_\alpha = \{T_n > q^*_{1-\alpha}\}$
Exemple : Retrouver le test classique de comparaison d'espérances

Test d'une fonction non linéaire de θ

On souhaite tester

$$(H_0) : h(\theta) = h(\theta_0) \text{ contre } (H_1) : h(\theta) \neq h(\theta_0)$$

et on dispose d'un emv $\widehat{\theta}$ asymptotiquement normal

$$\widehat{V}_n^{-1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, Id_p)$$

Soit h une fonction de différentielle $Dh(\theta) = A(\theta)$. Si $rang(A(\theta)) = r$, sous (H_0) :

$$W = [h(\widehat{\theta}_n) - h(\theta_0)]'(A(\theta_0)\widehat{V}_n A(\theta_0)')^{-1}(h(\widehat{\theta}_n) - h(\theta_0)) \xrightarrow{\mathcal{L}} \chi^2(r)$$

La région de rejet du test de Wald est asymptotiquement de niveau α

$$\mathcal{R}_\alpha = \{x | W(x) > q_{1-\alpha}^{\chi^2(r)}\}; \quad \lim_n \mathbb{P}_{(H_0)}(W \in \mathcal{R}_\alpha) \leq \alpha$$

Test d'une fonction non linéaire de θ

Remarque

S'il reste des paramètres non fixés dans l'expression de W, on les estime

$$W = [h(\widehat{\theta}_n) - h(\theta_0)]'[A(\widehat{\theta}_n)\widehat{V}_n A(\widehat{\theta}_n)']^{-1}(h(\widehat{\theta}_n) - h(\theta_0)) \xrightarrow{\mathcal{L}} \chi^2(r)$$

Test de Wald ou RV ?

- Comparent tous les deux des modèles emboîtés $\Theta_0 \subset \Theta_1$, où Θ_0 est défini comme la restriction de Θ_1 aux paramètres vérifiant une contrainte ($h(\theta) = a$ par ex)
- Conçus pour avoir le même niveau asymptotique α
- W est plus simple numériquement :
  - une seule optimisation ; mais dépend de l'estimation (délicate en pratique) de la matrice de covariance
- RV est plus compliqué numériquement :
  - deux optimisations, dont une sous contrainte ; mais souvent meilleur d'un point de vue théorique
- Dans certains cas, ces sont des tests équivalents

En présence de décisions contradictoires sur ces deux tests, on pourra préférer le test du RV

Exemples de tests paramétriques usuels

Test de Student de l'espérance

- Test unilatéral de l'espérance d'une loi $\mathcal{N}(\mu, \sigma^2)$ à variance inconnue $(H_0) : \mu = \mu_0$ contre $(H_1) : \mu > \mu_0$: la variable de décision est la statistique pivotale

$$T_n = \sqrt{n} \frac{\bar{X} - \mu_0}{\sqrt{\sum_i (X_i - \bar{X})^2/(n-1)}} \sim_{(H_0)} \mathcal{T}(n-1)$$

$\mathcal{R} = \{T > qt(n-1, 1-\alpha)\}$, $\mathbb{P}_{(H_0)}(\mathcal{R}) = \alpha$

- Si on ne connait la loi de la statistique de test que de façon asymptotique, les niveau et puissances calculés sont approximativement les niveau et puissance réels

Application

La consommation a-t-elle été sous-estimée au niveau 5% ?

- Modèle : on suppose que les mesures de consommation suivent une loi gaussienne, à variance inconnue
- Hypothèses : (H_0) : $\mu \leq 6.32$ contre (H_1) $\mu > 6.32$
- Statistique de Student : sous (H_0),

$$T = \sqrt{n}\frac{\bar{X} - \mu}{\widehat{\sigma}} \sim_{H_0} \mathcal{T}(n-1)$$

- Règle de décision de niveau 5% : rejet si $T > qt(0.95, n-1)$
- Décision : On observe $\bar{x} = 6.43$, $\widehat{\sigma} = 0.25$ sur un échantillon de $n = 30$ véhicules d'où $t = \frac{\bar{x}-6.32}{0.25/\sqrt{30}} = 2.41 > 1.7$
  On rejette (H_0) au niveau 5%, et on conclut avec un risque de 5% que le constructeur a minoré la consommation

```
> qt(c(0.9, 0.95,0.975,0.99),29)
  [1] 1.31 1.70 2.05  2.46
```

Récréation

Dans le test de l'espérance d'une loi gaussienne $\mu = \mu_0$ contre $\mu \neq \mu_0$, la valeur observée de la statistique de Student sur un échantillon de 12 individus est $t_{obs} = 2$.

- Quelle est la décision au niveau 5% ? Quel est le risque de cette décision ?
- Quelle est la décision au niveau 10% ? Quel est le risque de cette décision ?

```
qt(c(0.9 , 0.95 , 0.975 , 0.99),11)
[1] 1.36   1.80   2.20    2.72
```

Exemples de tests paramétriques usuels

- Test d'une variance
- Test d'une proportion
- Test de la comparaison des moyennes de deux échantillons
- ...

Probabilité critique ou p-value

- C'est le plus petit niveau qui fait rejeter (H_0) au vu des données
- Exemple : test de Student de $\mu = \mu_0$ contre $\mu > \mu_0$,
  - rejet : $\mathcal{R} = \{t; t = T(x) > qt(1-\alpha, n-1)\}$
  - niveau : $\mathbb{P}_{H_0}(T(X) \in \mathcal{R}) = \alpha$
  - valeur observée de la stat de test : $t_{obs} = T(x_{obs})$
  - p-value : $P_c(t_{obs}) = \mathbb{P}_{H_0}(T(X) > t_{obs})$
- Donc,
  - si $P_c(t_{obs}) \leq \alpha$, c'est que t_{obs} est dans la région de rejet de (H_0)
  - si $P_c(t_{obs}) > \alpha$, c'est que t_{obs} est dans la région d'acceptation de (H_0)

$\hookrightarrow$ dessin !

p-value

Définition

Soit la fonction test $\varphi(x;\alpha)$ associée à la région de rejet $\mathcal{R}_\alpha$ de niveau α. La p-value est définie par

$$P_c(t_{obs}) = \inf\{\alpha \in [0,1]; \varphi(x_{obs};\alpha) = 1\}$$

C'est une variable aléatoire.

Dans un test de niveau α, (H_0) est rejetée si $\alpha >$ p-value, conservée si $\alpha <$ p-value :

- si $0.05 >$ p-value > 0.01, le test est significatif,
- si $0.01 >$ p-value > 0.001, le test est très significatif,
- si $0.001 >$ p-value, le test est hautement significatif.

p-value : cas d'une région de rejet bilatère

Exemple test de Student de $\mu = \mu_0$ contre $\mu \neq \mu_0$,

- rejet : $\mathcal{R} = \{x; |T(x)| > qt(1-\alpha/2, n-1)\}$
- niveau : $\mathbb{P}_{H_0}(T(X) \in \mathcal{R}) = \alpha$
- valeur observée de la stat de test : $t_{obs} = T(x_{obs})$
- p-value : $P_c(t_{obs}) = \mathbb{P}_{H_0}(|T(X)| > |t_{obs}|)$
  - si t_{obs} est supérieur à la médiane de T :

$$P_c(t_{obs}) = 2\,\mathbb{P}_{H_0}(T(X) > t_{obs})$$

  - si t_{obs} est inférieur à la médiane de T :

$$P_c(t_{obs}) = 2\,\mathbb{P}_{H_0}(T(X) < t_{obs})$$

Sommaire

Un autre angle de vue

Soit le test $(H_0) : \mu = \mu_0$ contre $(H_0) : \mu \neq \mu_0$ de l'espérance d'une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$ à variance connue. On conserve (H_0) au niveau α si et seulement si

$$|T(X)| = \left| \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} \right| \leq q^*(1 - \alpha/2)$$

soit

$$-q^*_{1-\alpha/2} \leq \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} \leq q^*_{1-\alpha/2}$$

qu'on peut aussi écrire

$$\underbrace{\bar{X} - q^*_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}}_{\widehat{\mu}_{inf}} \leq \mu_0 \leq \underbrace{\bar{X} + q^*_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}}_{\widehat{\mu}_{sup}}$$

et $1 - \alpha = \mathbb{P}\left(|T(X)| \leq q_{1-\alpha/2}\right) = \mathbb{P}\left([\widehat{\mu}_{inf}; \widehat{\mu}_{sup}] \ni \mu_0\right)$

Un autre angle de vue

- Ainsi, pour qu'une valeur hypothétique de μ soit acceptée, il faut et il suffit qu'elle soit dans l'intervalle

$$IC(\mu) = [\widehat{\mu}_{inf}; \widehat{\mu}_{sup}]$$

- Cet intervalle $IC(\mu)$ aux bornes aléatoires est appelé intervalle de confiance de niveau $1 - \alpha$ de l'espérance μ inconnue
- Dans cet exemple, il y a équivalence pour μ entre prendre une valeur acceptée (H_0) dans le test de niveau α et le fait d'être situé dans l'intervalle de confiance de niveau (de confiance) $1 - \alpha$

Fournir un intervalle (fourchette) permet de prendre en compte la fluctuation d'échantillonnage plutôt que de donner une valeur ponctuelle $\widehat{\mu}$

Intervalle de confiance

Définition

Soit $X = (X_1, \ldots, X_n)$ *un n-échantillon de loi* $\mathbb{P}_\theta$, *où* $\theta \in \Theta \subset \mathbb{R}$ *est inconnu. Un intervalle de confiance de niveau* $1 - \alpha$ *pour* θ *est un intervalle* $IC = [\widehat{\theta}_{inf}(X), \widehat{\theta}_{sup}(X)]$ *dont les bornes sont aléatoires, telles que, pour tout* $\theta \in \Theta$

$$P_\theta(IC \ni \theta) \geq 1 - \alpha.$$

où α *est "petit".*
Une réalisation $[\widehat{\theta}_{inf}(x), \widehat{\theta}_{sup}(x)]$ *est obtenue à partir des données* $x = (x_1, \ldots, x_n)$.

Retour sur l'exemple : $X_i \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 connu.

- On choisit $0 \leq \alpha_1, \alpha_2 \leq \alpha$ tq $\alpha_1 + \alpha_2 = \alpha$ et soit q^* la fonction quantile de $\mathcal{N}(0,1)$. Un intervalle de probabilité $1-\alpha$ de $T = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ est $[q^*_{\alpha_1}; q^*_{1-\alpha_2}]$

$$\mathbb{P}(q^*_{\alpha_1} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < q^*_{1-\alpha_2}) = 1-\alpha$$

- d'où un IC de niveau $1-\alpha$ de μ

$$\mathbb{P}\left(\bar{X} + q^*_{1-\alpha_2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + q^*_{\alpha_1}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$IC = \left[\bar{X} - q^*_{1-\alpha_2}\frac{\sigma}{\sqrt{n}}; \bar{X} - q^*_{\alpha_1}\frac{\sigma}{\sqrt{n}}\right]$$

Exemple : $X_i \sim \mathcal{N}(\theta, \sigma^2)$, $\sigma^2 = 1$, n=10.

Une infinité d'ICs de niveau α (4 exemples ci-dessous)

```
> mean(X)
[1] 2.068577
> IC
    alpha1 alpha2      min      max    length
IC1  0.000  0.050 1.904091      Inf       Inf
IC2  0.015  0.035 1.887386 2.285586 0.3982001
IC3  0.025  0.025 1.872580 2.264573 0.3919928
IC4  0.050  0.000     -Inf 2.233062       Inf
```

- IC1 et IC4 sont des intervalles de confiance unilatéraux
- IC2 et IC3 sont bilatéraux
- IC3 est l'intervalle de confiance symétrique, de longueur minimale ici :

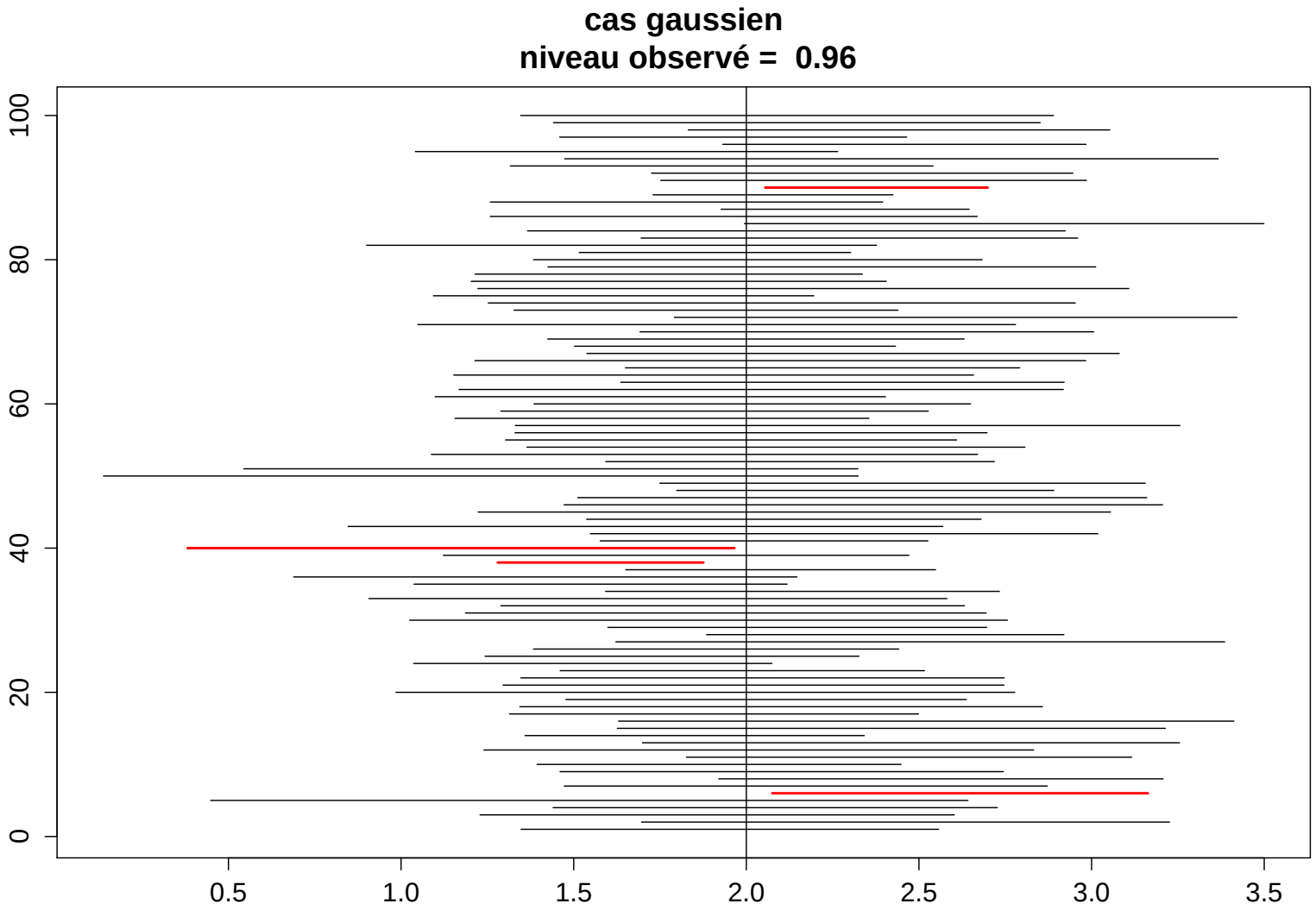
$$IC_{1-\alpha}(\theta) = \left[\bar{X} - q^*_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{X} + q^*_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Interprétation

```
> mean(X)
[1] 2.068577
    alpha1 alpha2      min      max    length
IC3  0.025  0.025 1.872580 2.264573 0.3919928
```

- Interprétation fausse : θ appartient à $[1.872580; 2.264573]$ avec probabilité $1 - \alpha$.
- Interprétation correcte :
  - La vraie valeur de θ (inconnue) appartient ou (exclus.) n'appartient pas à l'intervalle observé `[1.872580 ; 2.264573]`.
  - Si on construit une centaine d'intervalles de confiance à partir d'une centaine de n-échantillons indépendants, en moyenne $100 \times \alpha$ IC observés ne contiendront pas θ
    $\hookrightarrow$ mais on ne sait pas lesquels...

Exemple



Intervalle de confiance : Remarques

- IC est d'autant plus large que α est petit.
  - A l'extrême, l'IC de niveau de confiance 1 contient toutes les valeurs possibles... mais n'est plus informatif !
- IC de l'espérance calculé précédemment est d'autant plus étroit que n est grand
  - Une construction d'IC est convergente si la différence de ses bornes tend en proba vers 0 avec n
- Choisir un estimateur de θ, dont on connaît la loi de probabilité pour tout θ, et le meilleur possible
  - à α et n fixés, l'IC est d'autant meilleur que sa longueur est faible (pour toute réalisation / en moyenne)

Méthodes de construction

Méthode pivotale :

- définir un estimateur $\widehat{\theta}$ de θ
- trouver une statistique pivotale $T_n(\widehat{\theta}, \theta)$ dont la loi ne dépend pas de θ
- exprimer les bornes de l'intervalle de confiance en fonction de T_n et de ses quantiles

Méthodes de construction

IC asymptotique :

- La loi de la statistique n'est pas connue à distance finie, mais tend asymptotiquement vers une loi pivotale.
- On construit l'IC comme si la loi à distance finie était la loi limite.
- le niveau de l'IC construit est approximativement α à distance finie
- l'approximation s'améliore avec n croissant.
- à utiliser par exemple pour l'EMV.

Définition

Une suite d'intervalle de confiance IC_n de $\theta \in \mathbb{R}$ est de niveau asymptotique $1-\alpha$ si, pour tout $\theta \in \Theta$, on a

$$\lim_{n\to\infty} \mathbb{P}(IC_n \ni \theta) = 1-\alpha.$$

Exemple : IC de l'espérance θ d'une loi à variance inconnue

- TLC+Slutsky :

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \xrightarrow{\mathcal{P}} \sigma^2$$

$$T_n = \frac{\bar{X} - \theta}{S_n/\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

d'où, avec q^* le quantile de $\mathcal{N}(0,1)$ d'ordre $1-\alpha/2$:

$$IC = \left[\bar{X} - q^*\frac{S_n}{\sqrt{n}}; \bar{X} + q^*\frac{S_n}{\sqrt{n}}\right] \text{ avec } \mathbb{P}(IC \ni \theta) \simeq 1-\alpha$$

Exemple : IC de l'espérance θ d'une loi à variance inconnue

- Une solution alternative (choisie en général par les logiciels)

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \xrightarrow{\mathcal{P}} \sigma^2$$

$$T_n = \frac{\bar{X} - \theta}{\widehat{\sigma}_n/\sqrt{n}} \simeq \mathcal{T}(n-1) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

d'où, avec $t = qt(1-\alpha/2, n-1)$ le quantile de $\mathcal{T}(n-1)$ d'ordre $1-\alpha/2$

$$IC = \left[\bar{X} - t\frac{\widehat{\sigma}_n}{\sqrt{n}}; \bar{X} + t\frac{\widehat{\sigma}_n}{\sqrt{n}}\right] \text{ avec } \mathbb{P}(IC \ni \theta) \simeq 1-\alpha$$

- Cet IC est exact si la loi de l'échantillon est gaussienne

IC et test de la variance de $\mathcal{N}(\mu, \sigma^2)$

Soit $\widehat{\sigma}_n^2 = \sum_i (X_i - \bar{X})^2/(n-1)$ l'estimateur sans biais de la variance :

$$T_n = \frac{(n-1)\widehat{\sigma}_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

Soit $k_1 = qchisq(\alpha/2, n-1)$ et $k_2 = qchisq(1-\alpha/2, n-1)$

$$\mathbb{P}(k_1 < \frac{(n-1)\widehat{\sigma}_n^2}{\sigma^2} < k_2) = 1 - \alpha$$

- $IC = \left[\frac{(n-1)\widehat{\sigma}_n^2}{k_2}; \frac{(n-1)\widehat{\sigma}_n^2}{k_1}\right]$ avec $\mathbb{P}(IC \ni \theta) = 1 - \alpha$
- Test de $\sigma^2 = \sigma_0^2$ contre $\sigma^2 \neq \sigma_0^2$ est de région d'acceptation $[k_1; k_2]$ pour T_n

Attention, ces constructions sont peu robustes vis à vis de l'hypothèse gaussienne.

IC d'une proportion, pour n assez grand

Modélisation : $Z_i \sim_{i.i.d.} \mathcal{B}(1,\pi)$

$$T_n = \sqrt{n}\frac{\bar{Z}-\pi}{\sqrt{\pi(1-\pi)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ et } \bar{Z}(1-\bar{Z}) \xrightarrow{\mathcal{P}} \pi(1-\pi)$$

d'où, avec $q^* = qnorm(1-\alpha/2)$

$$IC = \left[\bar{Z} - q^*\sqrt{\frac{\bar{Z}(1-\bar{Z})}{n}}; \bar{Z} + q^*\sqrt{\frac{\bar{Z}(1-\bar{Z})}{n}}\right]$$

avec $\mathbb{P}(IC \ni \pi) \simeq 1-\alpha$. Or, $\bar{Z}(1-\bar{Z}) < 1/4$, on peut majorer la précision

$$\Delta\pi = q^*\sqrt{\frac{\bar{Z}(1-\bar{Z})}{n}} \leq \frac{q^*}{2}\frac{1}{\sqrt{n}}$$

soit $n_{max} = q^2/(4\times 0.01^2) \simeq 6765$ pour garantir une précision de $\pm 1\%$

Test d'une proportion, pour n assez grand

- Modélisation : $Z_i \sim_{i.i.d.} \mathcal{B}(1, \pi)$.
- Test de $(H_0) : \pi = \pi_0$ contre $(H_1) : \pi \neq \pi_0$
- sous (H_0)

$$T_n = \sqrt{n} \frac{\bar{Z} - \pi_0}{\sqrt{\pi_0(1-\pi_0)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

- Rejet pour

$$\left| \frac{\bar{Z} - \pi_0}{\sqrt{\pi_0(1-\pi_0)}} \right| > q^*$$

Région de confiance de Bonferroni

Si $\theta = (\theta_1, \theta_2)$, on cherche une région de confiance de la forme $RC(\theta) = IC(\theta_1) \times IC(\theta_2)$

$$\begin{aligned} & \mathbb{P}\left((\theta_1, \theta_2) \notin RC(\theta)\right) \\ = \; & \mathbb{P}(\overline{\theta_1 \in IC(\theta_1) \cap \theta_2 \in IC(\theta_2)}) \\ = \; & \mathbb{P}(\overline{\theta_1 \in IC(\theta_1)} \cup \overline{\theta_2 \in IC(\theta_2)}) \\ \leq \; & \underbrace{\mathbb{P}\left(\theta_1 \notin IC(\theta_1)\right)}_{\leq \alpha/2} + \underbrace{\mathbb{P}\left(\theta_2 \notin IC(\theta_2)\right)}_{\leq \alpha/2} \end{aligned}$$

D'où si $IC(\theta_2)$ et $IC(\theta_2)$ sont de niveau $1 - \alpha/2$, alors $RC(\theta)$ est de niveau simultané $1 - \alpha$:

$$\mathbb{P}\left((\theta_1, \theta_2) \in RC(\theta)\right) \geq 1 - \alpha$$

- l'intersection de K intervalles de confiance de risque α/K forment une région de confiance de risque simultané α.
- procédure en général très conservative

Région de confiance de Wald

Théorème

Soit A une matrice de dimension $r \times p$ et de rang r. Une région de confiance de Wald de niveau asymptotique $1-\alpha$ de $A\theta$ est donnée par

$$RC_\alpha(A\theta) =$$

$$\{A\theta \in \mathbb{R}^r, (A\widehat{\theta} - A\theta)'[A\widehat{V}_n A']^{-1}(A\widehat{\theta} - A\theta) \leq q_{1-\alpha}^{\chi^2_r}\}$$

où $q_{1-\alpha}^{\chi^2_r}$ est le quantile d'ordre $1-\alpha$ de la loi du χ^2_r.

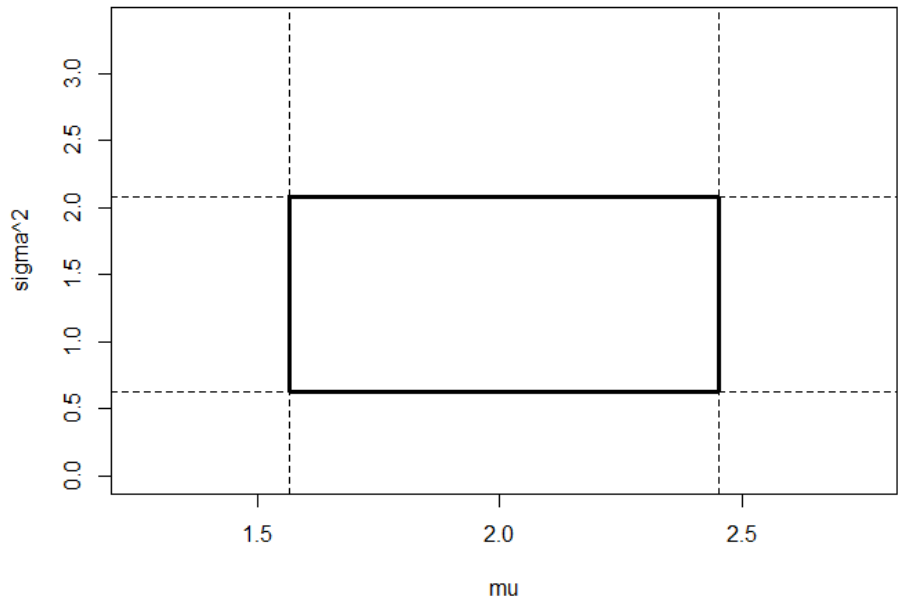
Rem : Possibilité de définir des régions de confiance de niveau exact, en particulier dans le modèle linéaire gaussien.

Région de confiance de Bonferroni

Exemple : région de confiance simultanée de niveau α pour $\theta = (\mu, \sigma^2)$ dans le cas gaussien

$$IC_{1-\alpha/2}(\sigma^2) = \left\{ \sigma^2 | \frac{(n-1)\widehat{\sigma}^2}{q^{\chi^2_{n-1}}_{1-\alpha/4}} \leq \sigma^2 \leq \frac{(n-1)\widehat{\sigma}^2}{q^{\chi^2_{n-1}}_{\alpha/4}} \right\}$$

$$IC_{1-\alpha/2}(\mu) = \left\{ \mu | \sqrt{n} |\bar{X} - \mu| \leq \widehat{\sigma} q^{\mathcal{T}_{n-1}}_{1-\alpha/4} \right\}$$



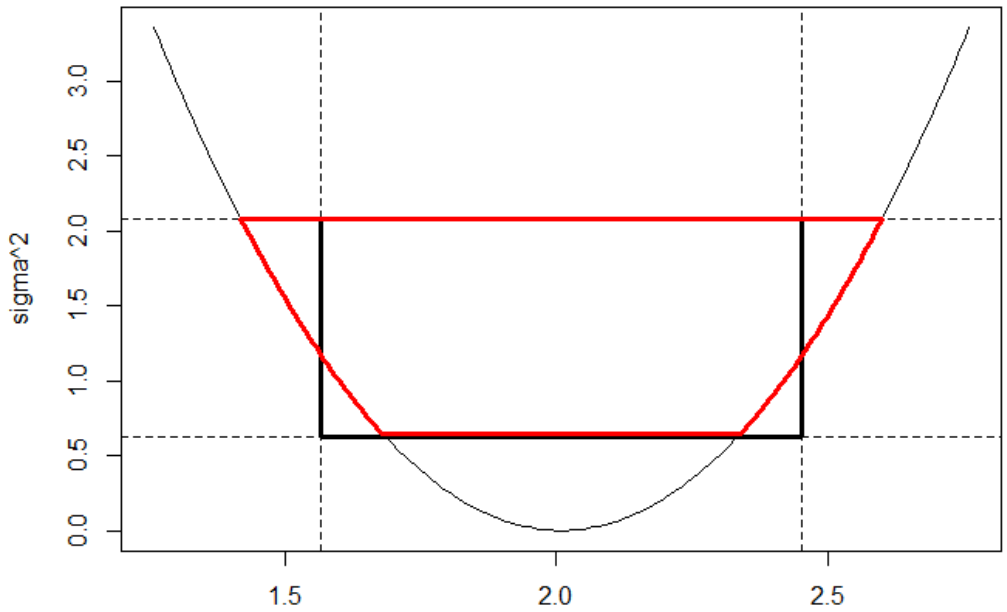
Région de confiance

RC simultanée de niveau α pour $\theta = (\mu, \sigma^2)$, cas gaussien

$$A_1 = \left\{ (\mu, \sigma^2) | \frac{(n-1)\widehat{\sigma}^2}{q^{\chi^2_{n-1}}_{1-\alpha/4}} \leq \sigma^2 \leq \frac{(n-1)\widehat{\sigma}^2}{q^{\chi^2_{n-1}}_{\alpha/4}} \right\}$$

$$\tilde{A}_2 = \left\{ (\mu, \sigma^2) | \sqrt{n}|\bar{X} - \mu| \leq \sigma q^*_{1-\alpha/4} \right\}$$

$$\mathbb{P}(A_1 \cap \tilde{A}_2) = \mathbb{P}(A_1)\mathbb{P}(\tilde{A}_2) = 1 - \alpha + \frac{\alpha^2}{4} > 1 - \alpha$$



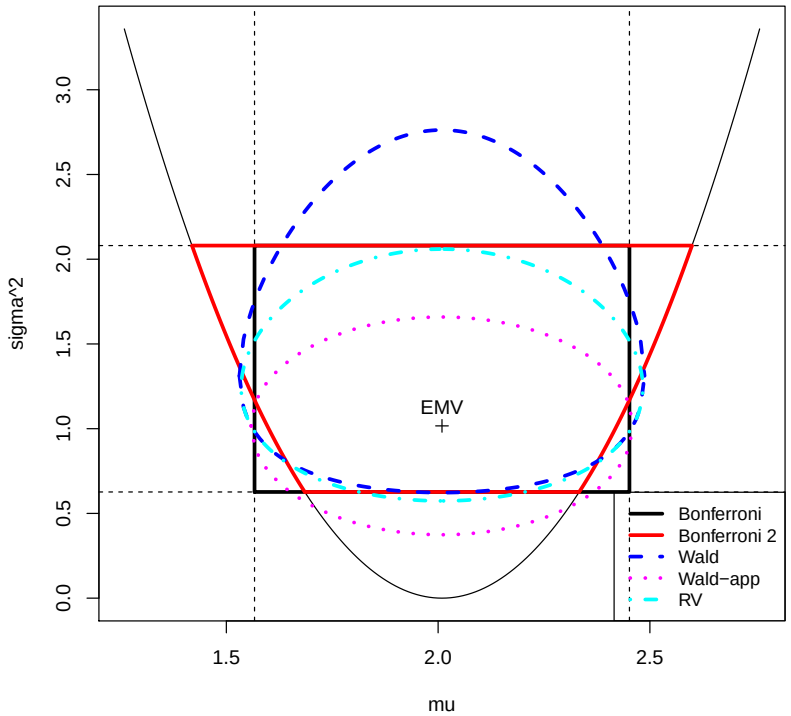
Région de confiance de type Wald

Asymptotique de l'EMV dans le modèle gaussien

$$\frac{n}{\sigma^2}(\bar{X}-\mu)^2+\frac{n}{2\sigma^2}(S_n^2-\sigma^2)^2 \xrightarrow{\mathcal{L}} \chi^2_{(2)}$$

d'où la région de confiance asymptotique de niveau $1-\alpha$

$$RC=\{(\mu,\sigma^2)|\frac{n}{\sigma^2}(\bar{X}-\mu)^2+\frac{n}{2\sigma^2}(S_n^2-\sigma^2)^2 \leq q_{1-\alpha}^{\chi^2_{(2)}}\}$$



RC : influence du nombre d'observations